

TECHNICKÁ UNIVERZITA V KOŠICIACH
Fakulta elektrotechniky a informatiky
Katedra kybernetiky a umelej inteligencie

Ing. Jozef Ivanecký

**Automatická transkripcia a
segmentácia reči**

Dizertačná práca

Školiteľ: **prof. Ing. Dušan Krokavec, CSc.**
Odbor: **38-01-9 Automatizácia a riadenie**

Košice, december 2003

Obsah

Úvod	5
1 Prehľad súčasného stavu	7
1.1 Segmentačné jednotky	7
1.2 Fonetická transkripčia a kódovanie	9
1.2.1 IPA	11
1.2.2 SAMPA	11
1.2.3 Slovenská fonetická transkripčia	18
1.3 Automatická transkripčia	19
1.3.1 Produkčné pravidlá	21
1.3.2 Štatistické metódy	22
1.4 Segmentácia rečového signálu	23
1.4.1 Priame metódy segmentácie	24
1.4.2 Metódy segmentácie využívajúce referenčný systém	26
2 Ciele dizertačnej práce	29
3 Zvolené metódy a vlastný postup riešenia	31
3.1 Zvolené metódy a postupy	31
3.2 Výber segmentačných jednotiek	31
3.3 Fonetická transkripčia a kódovanie	34
3.3.1 Fonematický prepis	34
3.3.2 Fonetický prepis	35
3.4 Automatická transkripčia	42
3.4.1 Samohlásky	43
3.4.2 Dvojhlásky (diftongy)	43
3.4.3 Samohláskové skupiny	46
3.4.4 Tvrdý hlasový začiatok a ráz	49
3.4.5 Samohláska <i>ä</i>	50
3.4.6 Samohlásky <i>ö, ő, ü, ű</i>	50
3.4.7 Príčiny, podmienky a miesta spoluhláskových zmien	51
3.4.8 Znelostná asimilácia	58
3.4.9 Zdvojené spoluhlásky	78
3.4.10 Tvrdé a mäkké spoluhlásky	86

3.4.11	Spoluhlásky [m, F, n, N, $\#$]	100
3.4.12	Iné spoluhláskové skupiny	103
3.4.13	Ostatné pravidlá	107
3.4.14	Výnimky	109
3.4.15	Automatické určovanie morfológických hraníc	110
3.4.16	Zhrnutie	112
3.5	Automatické určovanie hraníc slabík	113
3.5.1	Určovanie hraníc slabík pomocou jednoduchých pravidiel	113
3.5.2	Štatistický prístup pri určovaní hraníc slabík	115
3.6	Segmentácia rečového signálu	119
3.6.1	Proces segmentácie pomocou HMM	119
4	Overenie navrhnutých riešení	125
4.1	Výber segmentačných jednotiek a ich kódovanie	125
4.2	Automatická transkripcia	127
4.3	Automatické určovanie hraníc slabík	129
4.4	Segmentácia rečového signálu	132
5	Výsledky dizertácie	137
5.1	Výsledky dizertácie	137
5.2	Námety pre ďalšie pokračovanie práce	138
	Literatúra	141
	Použité skratky	145
	Zoznam obrázkov	146
	Zoznam tabuliek	147
A	Testy	149
A.1	Zoznam testovacích slov pre automatickú transkripciu a získané výsledky	149
A.2	Zoznam testovacích slov pre slabičnú segmentáciu	151
B	Postupnosť transkripčných pravidiel	153

Úvod

Počítačové rozpoznávanie reči má za sebou už niekoľkodesaťročnú históriu a vývoj sa dnes uberať skôr smerom k vylepšovaniu zabehnutých postupov, než k hľadaniu nových základných princípov. Pri tom je možné v dnešnej dobe v prípade rozpoznávania "menších" jazykov mnohokrát počuť, že daný jazyk je niečím špecifický, že vyžaduje špeciálne postupy a že s takouto problematikou sa nestretávame v žiadnom inom jazyku.

Hlavne posledných 10 rokov v rozpoznávaní reči — ktoré sa nieslo aj v znamení vytvárania viacjazyčných systémov — ukázalo, že základné postupy vypracúvané a overované v posledných desaťročiach sú aplikovateľné na akýkoľvek jazyk. Niekedy spomínaná takzvaná výnimočnosť sa netýka aplikovateľnosti všeobecne zaužívaných postupov, ale skôr nedostatočného rozpracovania konkrétnej problematiky pre daný jazyk, alebo je spojená s nedostatočným oboznámením sa s riešením podobných problémov v iných jazykoch.

Slovenčina nie je v tomto výnimkou. Napriek tomu, že mnohé javy v slovenčine sa nám môžu zdať špecifické a svojím spôsobom výnimočné, z principiálneho hľadiska nie je žiaden rozdiel pri vytváraní systému na rozpoznávanie reči pre slovenčinu a napríklad flámštinu.

Cieľom tejto práce preto nie je vytvoriť niečo principiálne nové, ale ukázať, že slovenčina nie je z pohľadu počítačového rozpoznávania reči výnimočná a že aj keď niektoré problémy doteraz v slovenčine neboli riešené, resp. boli riešené len okrajovo, základné metódy a postupy možno aplikovať aj na ňu.

Predkladaná práca sa venuje problematike jazykovo závislých úloh v procese segmentácie rečového signálu so zameraním na slovenčinu. Bližšie je možné jadro tejto práce označiť ako metódy a postupy na prípravu dát potrebných pre tréning systému na počítačové rozpoznávanie reči (či už pomocou HMM alebo neurónových sietí), zahŕňujúce výber segmentačných jednotiek, návrh ich kódovania a prepis dát do systému segmentačných jednotiek. Súčasťou práce je aj vlastné overenie navrhnutých metód v procese segmentácie.

Kapitola 1

Prehľad súčasného stavu

V tejto kapitole je zhrnutý súčasný stav problematiky so zameraním na oblasti, ktoré sú v tejto dizertačnej práci ďalej rozvinuté. Kapitola je rozdelená do štyroch častí.

Prvá časť je venovaná segmentačným jednotkám. Sú v nej opísané základné segmentačné jednotky používané pre segmentáciu rečového signálu.

Druhá časť je venovaná fonetickej transkripcii z pohľadu kódovania. Sú v nej zhrnuté hlavné fonetické kódovacie systémy, ktoré sa používajú v slovenčine alebo v iných jazykoch, a načrtnutá problematika ich vzájomnej kompatibility, či skôr nekompatibility.

V tretej časti je stručný prehľad problematiky automatického vytvárania ortoepického prepisu textu z jeho ortografickej reprezentácie. Sú v nej opísané dva základné prístupy k riešeniu problematiky a vhodnosť ich použitia.

Štvrtá, záverečná časť, je venovaná segmentácii rečového signálu na zvolené segmentačné jednotky. Tento fundamentálny problém, ktorý je potrebné riešiť v prípade tréningu akéhokoľvek systému na rozpoznávanie reči, je opísaný s poukázaním na dve základné možnosti prístupu k danej problematike.

1.1 Segmentačné jednotky

Dôležitý bod v návrhu celého procesu segmentácie a kódovania je výber segmentačných jednotiek. V publikovaných prácach sa ako jednotky segmentácie objavujú alofóny, fonémy, difóny, slabiky, demislabiky a slová, ale aj iné jednotky. V stručnosti budú spomenuté výhody a nevýhody použitia jednotlivých segmentačných jednotiek podľa [50].

- Alofóny ako pozičné varianty foném môžu poskytnúť informáciu o hranici slova, alebo aj o hranici menších jednotiek, napríklad slabík. Môže to byť veľmi dôležité hlavne pri rozpoznávaní plynulej reči. Segmentácia na alofonickej úrovni pritom eliminuje potrebu aplikovať koartikulačné pravidlá. To vyplýva z faktu, že koartikulačný efekt je obsiahnutý v príslušnej alofóne,

keďže tá zodpovedá konkrétnemu spojeniu určitých foném. Nevýhoda použitia alofón je, že pri fonetickom prepise sa dostane veľký inventár alofón.

- **Fonéma** je fonetická jednotka, ktorá reprezentuje najmenší počet odlišných fonetických tried. Fonéma často zodpovedá našej predstave segmentácie, t. j. vymedzeniu najkratších úsekov v reči odlišujúcich význam dvoch slov. Aj keď je počet foném podstatne menší v porovnaní s počtom alofón, slabík, slov a pod., ich automatické vymedzenie a klasifikácia je stále problematická. Je to spôsobené veľkou variabilitou akustickej realizácie každej fonémy v závislosti na koartikulácii s jej fonetickým okolím. Výhodou foném je ich malý počet (vo väčšine jazykov 30–50), nevýhodou je, že fonéma neobsahuje dôležitú koartikulačnú informáciu.
- **Difóna** je termín, ktorý je používaný pre označenie postupnosti samohláska/spoluhláska tak, že difóne zodpovedajúci úsek reči sa rozširuje zo stredu samohlásky do stredu spoluhlásky alebo naopak. Difóny sú využívané pri segmentácii reči hlavne z dôvodu, že veľká časť akustickej informácie, ktorá je užitočná na identifikáciu spoluhlások, leží v prechodoch medzi spoluhláskami a samohláskami. Výhodou difón je, že obsahujú dôležitú koartikulačnú informáciu, ktorá nebola prítomná pri fonémach, nevýhoda je veľký počet difón (niekoľko tisíc).
- **Slabiky** sú všeobecne považované za fonetické útvary, ktoré obsahujú samohláskové jadro plus začiatkové a koncové spoluhlásky. Slabiky tak obsahujú prechody spoluhláska/samohláska, ako aj prechody samohláska/spoluhláska, vrátane väčšiny koartikulácii a iných fonologických efektov v rámci ich hraníc. V slovenčine môže byť v niektorých prípadoch samohláskové jadro slabiky nahradené slabikotvornou sonórnou spoluhláskou *r*,₁. Príkladmi takýchto slov sú slová "vlk", "vrba" a pod. Dĺžka slabiky v slovenčine sa pritom môže pohybovať od jednej fonémy, ako je napríklad spojka "a", až po päť foném (napríklad jednoslabičné slovo "škrob"). Výhodou segmentácie podľa slabík je, že jadro slabiky je relatívne ľahko lokalizovateľné a identifikovateľné akusticky. Existuje taktiež mnoho dôvodov, ktoré podporujú dôležitú artikulačnú a percepčnú úlohu slabiky, vrátane na slabike založeného prízvuku a rytmu. Existuje však niekoľko nevýhod spojených s využitím slabík ako jednotiek segmentácie. Z fonetického hľadiska neexistuje jasné určenie hraníc slabiky. Hranice slabík sú zreteľné len v postupnostiach, v ktorých sa striedajú spoluhláskové a samohláskové fonémy. Ak je medzi dve susedné jadrá vložená spoluhlásková skupina, slabičné hranice sa problematizujú. Ďalšia nevýhoda je opäť značný počet existujúcich slabík.
- Značná veľkosť inventára slabík môže byť redukovaná použitím **poloslabík**. Každá slabika môže byť pokladaná za kompozíciu začiatkovej poloslabiky, obsahujúcej začiatkový zhuk hlások a príslušného samohláskového jadra, a koncovej poloslabiky, obsahujúcej zostávajúcu časť samohláskového jadra a

koncový zhluk hlások. Použitím poloslabikových jednotiek je možné zredukovať rozsah inventára potrebného na reprezentáciu reči až päťkrát v porovnaní s celými slabikami. Použitie poloslabík však zavádza určitú zložitosť v rozdeľovaní samohláskového jadra. Ak je napríklad hranica medzi začiatočnou a koncovou poloslabikou vytvorená v strede samohláskového jadra, musí sa zvláštna pozornosť venovať procesu porovnávania samohlásky na oboch stranách hranice, hlavne ak sú samohlásky nasledované sonórnymi hláskami (t. j. l, r, j, m, n, ň).

- Ťažkosti vyskytujúce sa pri delení poloslabík podstatne eliminujú **demislabiky**. Od poloslabík sa líšia umiestnením hranice medzi začiatočnou a koncovou demislabikou. Začiatočná demislabika je vymedzená tak, že rozdeľujúca hranica je umiestnená tesne za prechod spoluhláska/samohláska. Umiestnenie hranice týmto spôsobom je efektívne tým, že uvoľňuje obmedzenie na obidvoch stranách hranice a má aj potenciál v redukcii inventára, podobne ako pri poloslabikách.
- Na **slová** ako fonetické útvary je možné pozeráť ako na spojenie konečného počtu foném či slabík. Výhodou priameho použitia slova ako segmentačnej jednotky je eliminácia nutnosti použitia komplikovaných algoritmov segmentácie a identifikácie hierarchicky nižších jednotiek, t. j. alofón, foném a pod. Táto výhoda sa prejaví hlavne v prípade, že slová sú vyslovované izolovane, t. j. sú obklopené dostatočne dlhými pauzami (aspoň 200 až 300 ms). V takomto prípade možno nájsť hranicu medzi slovami relatívne ľahko. V prípade súvislej reči sa problém vymedzenia hraníc slova stáva veľmi obtiažnym.

1.2 Fonetická transkripcia a kódovanie

Pod pojmom *fonetická transkripcia* sa väčšinou rozumie samotná ortoepická reprezentácia nejakej skupiny slov alebo aj proces, ktorým je takáto reprezentácia vytváraná. V procese transkripcie je možné naraziť na ďalšiu problematiku automaticky zahŕňanú pod názov *fonetická transkripcia*. Ide o kódovanie použité pri vytváraní transkripcie, a práve tejto problematike je venovaná táto časť. Na začiatku treba definovať pojmy *transkripcia* a *kódovanie*, ktoré však budú striktne aplikované len v prípade, keď by mohlo dôjsť k ich vzájomnej zámene.

Transkripcia je použitie príslušných znakov abecedy na reprezentáciu zvukov reči. Transkripčný systém môže byť 'ortografický systém', používajúci grafémy konkrétneho jazyka so špeciálnymi diakritickými značkami na odstránenie nejednoznačností, alebo to môže byť 'fonetický (ortoepický) systém' používajúci špeciálne fonetické znaky (IPA, SAMPA alebo iné).

Kódovanie je pridelenie bitového obrazca, pomocou ktorého môžu byť dané znaky (napr. 'a', 'w') reprezentované interne v počítači, za účelom uchovania alebo spracovania. Štandardný kód v počítačoch je ASCII kód, známy aj ako ISO

0	32	SP	48	0	64	@	80	P	96	'	112	p
1	33	!	49	1	65	A	81	Q	97	a	113	q
2	34	"	50	2	66	B	82	R	98	b	114	r
3	35	#	51	3	67	C	83	S	99	c	115	s
4	36	\$	52	4	68	D	84	T	100	d	116	t
5	37	%	53	5	69	E	85	U	101	e	117	u
6	38	&	54	6	70	F	86	V	102	f	118	v
7	39	,	55	7	71	G	87	W	103	g	119	w
8	40	(56	8	72	H	88	W	104	h	120	x
9	41)	57	9	73	I	89	X	105	i	121	y
A	42	*	58	:	74	J	90	Z	106	j	122	z
B	43	+	59	;	75	K	91	[107	k	123	{
C	44	,	60	<	76	L	92	\	108	l	124	
D	45	-	61	=	77	M	93]	109	m	125	}
E	46	.	62	>	78	N	94	^	110	n	126	~
F	47	/	63	?	79	O	95	_	111	o	127	DEL

Tabuľka 1.1: ASCII/ISO 7-bitová znaková sada.

7-bitovo kódovaná sada znakov (ISO 646), ktorá je v Tabuľke 1.1. ISO 646 necháva v základnej znakovkej sade 7 pozícií nepridelených a umožňuje ich rozdielne priradenia podľa národných potrieb. V prípade použitia 8-bitového kódovania už nie je možné hovoriť o štandarde, keďže horných 128 znakov sa v národných abecedách líši. Pri vytváraní týchto štandardov nebola braná do úvahy problematika kódovania špeciálnych fonetických symbolov.

Ak majú byť v počítači použité špeciálne fonetické symboly, je potrebné mapovať každý z nich do ASCII hodnoty. Unicode – napriek tomu, že pokrýva aj IPA symboly – nie je platformovo natolko rozšírený a aplikáciami podporovaný, aby bolo možné jeho širšie nasadenie. Problémom je taktiež ťažšia čitateľnosť IPA symbolov pre menej skúseneho užívateľa, keďže IPA symboly nemajú v mnohých prípadoch žiadnu podobu s im zodpovedajúcimi grafémami¹.

Najrozšírenejším medzinárodným transkripčným systémom je v súčasnosti systém IPA – International Phonetic Association – ku ktorému bola paralelne vyvinutá SAMPA abeceda. Na Slovensku sa prakticky nestretáme s využívaním žiadneho z týchto systémov, keďže v slovenskej fonetike a fonológii sa používa transkripčný systém založený na ortografii slovenčiny. Je pochopiteľné, že tento systém nie je kompatibilný so žiadnym zo spomínaných medzinárodných systémov.

Podrobnejšie je o jednotlivých transkripčných systémoch, ich výhodách a nevýhodách, pojednané v nasledujúcich kapitolách.

¹Tento problém neexistuje pri jazykoch ako čínština, japončina, poprípade nie je taký výrazný v jazykoch ako angličtina a podobne.

1.2.1 IPA

Začiatky IPA transkripcie je možné datovať do roku 1888, keď bola publikovaná prvá verzia IPA transkripčného systému. Jej navrhovatelia neboli zaťaženi žiadnymi obmedzeniami vyplývajúcimi z počítačového spracovania a z toho dôvodu mali pri návrhu IPA symbolov voľnú ruku. Keďže ich snahou bolo taktiež oddelenie ortoepického systému od ortografického, navrhli novú sadu symbolov. Pri niektorých z nich je nápadná podobnosť s ortografiou – a, u, i, l – no na pochopenie mnohých z nich sú potrebné aspoň minimálne znalosti z fonetickej transkripcie – ε, ɔ, ʊ, ʒ, γ. Posledná verzia IPA abecedy bola vydaná v roku 1993 a aktualizovaná v roku 1996 [10] (Tabuľky 1.1, 1.2, 1.3).

IPA vznikla po skúsenostiach s používaním iných transkripcií. Využívajú sa v nej rozličné možnosti rozšírenia počtu znakov. Základný cieľ bol, aby každá fonéma mala osobitný znak bez diakritík, pričom na záznam alofón sa môže použiť aj diakritika, ktorá sa odvodzuje podľa určitých pravidiel. Pôvodne bola určená najmä na vyučovanie výslovnosti cudzích jazykov. Práve z toho vyplýva aj jej fonologická orientácia [35].

Hoci IPA je medzinárodná transkripcia, pri jej vytváraní sa vychádzalo hlavne z poznania západoeurópskych jazykov. Z toho vyplýva, že napriek jej schopnosti opísať relatívne presne fóny z jazykovej skupiny, z ktorej IPA vychádzala, neumožňuje rovnako presný opis ostatných jazykov, slovanské nevynímajúc. V prípade použitia IPA pre slovanské jazyky je nutné sa zmieriť s určitou nepresnosťou.

1.2.2 SAMPA

SAMPA abeceda vznikla za účelom jednoduchého kódovania IPA znakov do 7-bitovej ASCII sady. 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 až dodnes University College v Londýne.

Pri vytváraní tohto štandardného transkripčného systému aplikovateľného na mnohé európske jazyky sa vychádzalo z monojazyčných systémov uvedených v Tabuľke 1.2. Ako vyplýva z tabuľky, problematikou jednoduchého kódovania IPA symbolov sa zaoberalo viacero inštitúcií a v tomto štádiu šlo skôr o formu štandardizácie, ktorá by zaručila vzájomnú kompatibilitu. Rovnako veľký význam ako dosiahnutie vzájomnej kompatibility mala pri jeho vytváraní aj jednoduchosť kódovania pri počítačovom spracovaní.

Z Tabuľky 1.2 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, u – U, e – E, ε – E).

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)

CONSONANTS (PULMONIC)

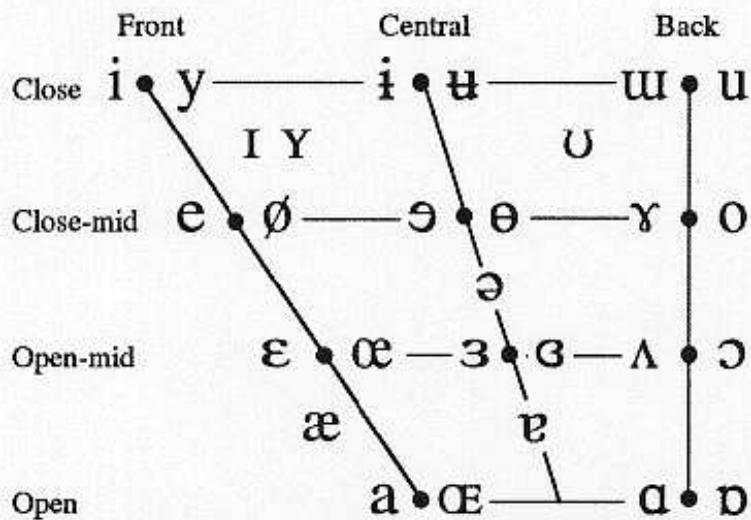
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			ɾ					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
◌ ɔ	ɓ Bilabial	' as in:
Dental	ɗ Dental/alveolar	ɓ' Bilabial
! (Post)alveolar	ɟ Palatal	ɗ' Dental/alveolar
≠ Palatoalveolar	ɠ Velar	ɠ' Velar
Alveolar lateral	ʄ Uvular	ʂ' Alveolar fricative

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

OTHER SYMBOLS

ɱ Voiceless labial-velar fricative	ʎ ʞ Alveolo-palatal fricatives
ɰ Voiced labial-velar approximant	ɺ Alveolar lateral flap
ɸ Voiced labial-palatal approximant	ɧ Simultaneous ʃ and x
ħ Voiceless epiglottal fricative	Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.
ʕ Voiced epiglottal fricative	
ʡ Epiglottal plosive	

k̟p̟ t̟s̟

Obr. 1.2: Schéma IPA z roku 1993.

SUPRASEGMENTALS

'	Primary stress	fou'nə'tɪʃən
ˈ	Secondary stress	
ː	Long	eɪ
ˑ	Half-long	eɪ̯
◌̥	Extra-short	ɛ̥
·	Syllable break	ɪ.ækt
	Minor (foot) group	
	Major (intonation) group	
◌̣	Linking (absence of a break)	

TONES & WORD ACCENTS

LEVEL		CONTOUR	
ē̄ or ˩	Extra high	ē̄ or ˩	Rising
é	High	ē̂	Falling
ē	Mid	ē̄	High rising
è	Low	ē̇	Low rising
ē̇	Extra low	ē̇	Rising-falling etc.
↓	Downstep	↗	Global rise
↑	Upstep	↘	Global fall

DIACRITICS

Diacritics may be placed above a symbol with a descender, e.g. ŋ̣

◌̥	Voiceless	n̥ d̥	◌̤	Breathy voiced	b̤ a̤	◌̪	Dental	t̪ d̪
◌̣	Voiced	ṣ ṭ	◌̰	Creaky voiced	b̰ a̰	◌̺	Apical	t̺ d̺
◌ʰ	Aspirated	tʰ dʰ	◌̱	Linguolabial	ṯ ḏ	◌̻	Laminal	t̻ d̻
◌̙	More rounded	ɔ̙	◌̜	Labialized	tʷ dʷ	◌̃	Nasalized	ẽ
◌̘	Less rounded	ɔ̘	◌̝	Palatalized	tʲ dʲ	◌̚	Nasal release	d̚ⁿ
◌̟	Advanced	u̟	◌̞	Velarized	t˞ d˞	◌̜̚	Lateral release	d̜̚ˡ
◌̠	Retracted	i̠	◌̜̚	Pharyngealized	t̜̚ d̜̚	◌̚	No audible release	d̚
◌̠̠	Centralized	ë	◌̜̚	Velarized or pharyngealized	ɫ			
◌̠̠̠	Mid-centralized	ẽ	◌̚	Raised	e̚ (ɹ̚ = voiced alveolar fricative)			
◌̚	Syllabic	ɹ̚	◌̚	Lowered	e̚ (β̚ = voiced bilabial approximant)			
◌̚	Non-syllabic	e̚	◌̚	Advanced Tongue Root	e̚			
◌̚	Rhoticity	ə̚	◌̚	Retracted Tongue Root	e̚			

Obr. 1.3: Schéma IPA z roku 1993.

	IPA		COST	Esprit	Alvey	VL	BDSON	DIN	ISO	Allen
1	ɑ	!	A	A	A	A	a>	42		a (aa)
2	æ		{ 123	∧	&			44	71	ae
3	ɐ		7	R \	6			41		
4	ɒ		C		Q			43		ao
5	ɔ	!	O	O	O	O	o>	66	a5A	O
6	ç		c	C	C			48		c-
7	ð	!	D	D	D			4A	73	D
8	ɛ	!	E	E	E	E	e>	4D	a37	E
9	ə	!	@ 64	@	@	@	&	4C	a38	6
10	ɜ		3	@/	3			4E		3
11	ɝ	!	G	G	G	G		54	a3C	g-
12	ɪ	!	I	I	I			5A	a3F	I
13	ʎ		l,	L	F			60		lj
14	ɹ		n,	N̄	J		nj	64	a56	nj
15	ŋ	!	N	N	N	N	ng	63	a57	N
16	ø		124	q	{ 123	&	E	67	79	oe
17	œ		\92	Q	9	@	E>	68		oE
18	Œ		&							OE
19	θ	!	T	T	T			73		T
20	ʃ	!	S	S	S	S	S	71	a5F	S
21	ʊ	!	U	U	U			77	a73	U
22	ʌ		V	^	V			79		A
23	ʏ	!	Y	Y	Y			3C		Y
24	ɥ			H	H		Y	57		
25	ʒ	!	Z	Z	Z	Z	Z	3E	a77	Z
26	ʔ	!	?	?	?			3F	a79	?
27	ɹ,R		R	R	K		R	6D,6E		R
28			E~	Ē	E+~		IN	4D22		E,n
29	aɪ		A>I	a/	aɪ	(Y)		405A		ai
30	tʃ		I^S	T/	IS					
31	:	!	:	:	:			2A		:
32	'		(')	'	"			31		"
33	,		(,)	"	%			32		'

Tabuľka 1.2: Prehľad počítačovo reprezentovateľných fonetických transkripcií. Znak '!' v treťom stĺpci poukazuje na dohodu medzi COST 209, Esprit 291 a Alvey tabuľkami. Kódovanie v DIN a ISO stĺpci sú v hexa formáte.

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	ď	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	ĺ	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
			ɥ	H	72
			z	z	122
			ž	Z	90
			'	?	63

Tabuľka 1.3: SAM abeceda (SAMPA) pre fonetickú transkripciu I – 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	Komentár
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 1.4: SAM abeceda (SAMPA) pre fonetickú transkripciu II – 1987.

Aplikovaním týchto pravidiel na Tabuľku č. 1.2 vznikne ASCII reprezentácia pre nasledujúce znaky z fonetickej abecedy IPA:

IPA znaky	ɑ	ɔ	ɔ̃	ɛ	ə	ɜ	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 [58]. Táto fonetická abeceda, predstavujúca prvý – počítačovo jednoducho reprezentovateľný – kódovací systém, dohodnutý medzi predstaviteľmi ôsmich európskych krajín spolupracujúcich na projekte SAM, je v Tabuľke č. 1.3 a č. 1.4. 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 [37].

V posledných piatich rokoch bola SAMPA v rámci niekoľkých projektov značne rozšírená o nové transkripčné znaky a dnes už pokrýva mnohé – nie len indo-európske – jazyky. Vzhľadom na rozsah nie je možné, aby súčasná verzia SAMPA abecedy bola súčasťou tejto práce. Aktuálny stav SAMPA abecedy možno nájsť na [59].

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

1.2.3 Slovenská fonetická transkripcia

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 od štandardizačného návrhu v 60. rokoch rozšírený domáci (s českým koordinovaný) typ systému fonetickej transkripcie [19]. Zohľadňoval sa v ňom jednak dovtedajší úzus v zapisovaní slovenských nárečových textov, jednak dobové technické 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 [19]. O týchto problémoch súvisiacich s

použitím IPA transkripcie pojednáva časť 1.2.1.

Keďže staršie ciele fonetického zápisu (prepisy slovenských najmä nárečových textov boli určené predovšetkým domácim bádateľom, sekundárne aj nelingvistickej verejnosti, z čoho v neposlednom rade vyplývala aj potreba 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 je ho pri riešení otázok fonetickej transkripcie nutné brať do úvahy aj pri pokuse navrhnúť akúkoľvek inú všeobecnejšie zrozumiteľnú fonetickú transkripciu [34].

V Tabuľke č. 1.5 je uvedený zoznam fón spisovnej slovenčiny, kódovaných znakmi používanými v slovenskej fonetickej transkripcii podľa [33]. 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 [12], no v oblasti slovenskej fonetickej transkripcie sa nevelmi využíva a niektoré jeho riešenia možno považovať za problematické.

Problémy, ktoré je potrebné riešiť pri pokuse vytvoriť inventár slovenských fón kódovaných pomocou SAMPA fonetickej abecedy, je možné zhrnúť do dvoch bodov:

- Nedoriešenosť vzťahu slovenskej fonetickej transkripcie s transkripciou IPA. Riešenie tohto problému je 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.
- 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 [24], avšak na vytvorenie fonetickej verzie vhodná nebola. Tento stav bolo možné zovšeobecniť pre väčšinu slovanských jazykov.

1.3 Automatická transkripcia

Medzi najrozšírenejšie techniky pre počítačové rozpoznávanie reči patria v dnešnej dobe skryté Markovove modely (HMM) a neurónové siete (NN). V oboch

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>	päť, mäso	<i>ŕ</i>	vŕba
<i>e</i>	pero	<i>l</i>	skala, vlak
<i>é</i>	dcéra nové	<i>ł</i>	vlk
<i>i</i>	pivo, dym	<i>ł̄</i>	vĺča
<i>í</i>	pískat', nový	<i>ň</i>	vaňa, niest'
<i>o</i>	popol	<i>ľ</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>üa</i>	piatok		v zime, šéflekár
<i>üe</i>	spievať	<i>ű</i>	pravda, otcov
<i>üu</i>	cudziu		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>	košeľa, 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>ĩ</i>	kraj, krajský, zjest'
<i>m</i>	mama	<i>c</i>	cena, pocta
<i>ṁ</i>	amfiteáter, tamvon	<i>ʒ</i>	hádzat', medza
<i>n</i>	rana, nos	<i>č</i>	oči, mačka
<i>ṅ</i>	panský, inžinier	<i>ž</i>	džem, Hodža
<i>ŋ</i>	tenký, cengať		

Tabuľka 1.5: Tabuľka znakov slovenskej fonetickej transkripcie

prípadoch je na vytvorenie systému potrebné väčšie množstvo dát.² Vstupom na tréovanie takýchto systémov je okrem nahratých viet aj ich fonetická transkripcia, od kvality ktorej značne závisí kvalita výsledného systému.

Transkripciu je možné vytvoriť manuálne, čo je v prípade veľkej množiny viet značne pracné a môže viesť k väčšiemu množstvu chýb. V prípade použitia takýchto údajov pri tréovaní sa znižuje úspešnosť rozpoznávania. Ešte väčší vplyv má na druhej strane nepoužitie takýchto údajov v procese samotného rozpoznávania.

Samotný manuálny prepis je možné urobiť dvoma spôsobmi:

- Prepis vstupných viet je urobený bez kontroly toho, čo bolo (mohlo byť) skutočne povedané (vyslovené). V tomto prípade je potrebné generovať všetky možné výslovnosti (v rámci možnosti prepisujúceho). Výhodou takéhoto postupu je skutočnosť, že sa získa väčšie množstvo dát pre použitie v prípadnom rozpoznávaní. Nevýhodou zostáva, že tvorca transkripcie nemusí pokryť výslovnosť použitú hovoriacim.
- Prepis vstupných viet urobený s odposluchom toho, čo bolo skutočne vyslovené. Tento spôsob transkripcie je neoceniteľný hlavne v prípade, že ide o dáta z rôznych nárečových³ oblastí. Netreba podotýkať, že tento spôsob je aj najnáročnejší na realizáciu, avšak v niektorých prípadoch je nevyhnutný.

Druhý prístup je automatická transkripcia. Na rozdiel od prvého je pri nej možné vylúčiť mnohé chyby z nepozornosti. Na druhej strane je vytvorenie systému na automatickú transkripciu značne zložitá záležitosť. V niektorých jazykoch natoľko zložitá, že sa od postupov uvedených v časti 1.3.1 celkom upúšťa. O automatickej transkripcii slovenčiny, avšak za účelom syntézy reči, pojednáva napr. [11]. Možnosťou automatizácie transkripcie v češtine sa zaoberá [50], [8].

1.3.1 Produkčné pravidlá

Najbežnejší spôsob vytvárania fonetického prepisu, pri ktorom je možné začať takpovediac na zelenej lúke, sú produkčné pravidlá. Tento postup je využívaný v prípadoch, keď iné techniky – vyžadujúce tréovaciu množinu – uvedené v časti 1.3.2 nie je možné použiť. Podmienkou pre vytvorenie takéhoto systému je existencia pravidiel – resp. existencia možnosti vytvorenia takýchto pravidiel – pre "správnu" výslovnosť v prípade, že je k dispozícii ortografická reprezentácia textu, ktorý má byť prevedený do ortoepickej formy. Pre slovenčinu je veľmi dobrým zdrojom takýchto pravidiel literatúra [33].

Pri vytváraní systému používajúceho produkčné pravidla sa vychádza z predpokladu, že je definovaný vzťah⁴ medzi postupnosťou znakov v ortografickom

²V závislosti od požadovanej kvality systému môže ísť o niekoľko stovák až niekoľko miliónov tréovacích viet.

³Nemusí sa nevyhnutne jednať o nárečové oblasti.

⁴Otázka zložitosti tohto vzťahu nás na tomto mieste nezaujíma.

prepise a postupnosťou v ortoepickom prepise. Môže ísť o vzťah $1 : 1, 1 : N$ alebo $N : 1$. Vzťahy $N : N$ musí byť možné rozdeliť na pravidlá pozostávajúce zo vzťahov $1 : 1, 1 : N$ alebo $N : 1$. V prípade, že to nie je možné, nie je možné použiť produkčné pravidlá pre generovanie fonetického prepisu pre daný jazyk. Ako príklad pre slovenčinu možno uviesť:

- Vzťah $'1 : 1'$: $'a \rightarrow a'$; graféme a vždy zodpovedá fóna a , bez ohľadu na kontext.
- Vzťah $'1 : N'$: $'i \rightarrow \{ I, L \hat{ } \}'$; graféme i zodpovedá buď fóna I alebo fóna $L \hat{ }$, v závislosti od kontextu⁵.
- Vzťah $'N : 1'$: $'\{ i, y \} \rightarrow I'$; grafémam i, y vždy zodpovedá fóna I :⁶.

Ak je pre daný jazyk možné vytvoriť sadu takýchto pravidiel, ktorá je programovo realizovateľná, je možné týmto spôsobom vytvoriť systém, ktorý na vstupnú postupnosť grafém G bude reagovať generovaním zodpovedajúcej postupnosti foném F . Kvalita výstupnej postupnosti bude závislá od toho, ako citlivo boli vybrané transkripčné pravidlá, t. j. napríklad či bola uvažovaná len spisovná⁷ výslovnosť a podobne.

1.3.2 Štatistické metódy

Metódy uvedené v tejto časti majú spoločné to, že na to, aby dokázali generovať z postupnosti grafém zodpovedajúcu postupnosť foném, potrebujú určité množstvo už prepísaných dát (trénovaciu množinu). Táto trénovacia množina je potrebná na vytvorenie určitých štatistík, ktoré sú neskôr využité na generovanie postupnosti foném pre nové dáta.

Vo väčšine prípadov sa na tento účel používajú rozhodovacie stromy. Základná myšlienka tohto postupu je veľmi jednoduchá.

- Vytvorí sa väčšia trénovacia množina, ktorá bude obsahovať dvojice W, Φ , kde W je slovo v ortografickej forme a Φ je k nemu prislúchajúca fonetická transkripcia. Pre každý takýto pár je vytvorené priradenie medzi grafémami a fonémami. Táto úloha je kľúčová v tomto kroku a prístupuje sa k nej tým spôsobom, že sa najprv vygeneruje priradenie z Φ do W a potom spätne na základe jednoduchých pravidiel priradenie $W \rightarrow \Phi$. Vychádza sa pritom zo skutočnosti, že počet znakov textu nie je nikdy menší než počet fonetických symbolov, alebo inými slovami povedané, každá fonéma (fóna) generuje

⁵Konkrétne v prípade, ak za i nasleduje krátka samohláska, ide buď o samohláskovú skupinu alebo dvojhlásku.

⁶Takéto vzťahy je ale možné zjednodušiť na množinu vzťahov $'1 : 1'$.

⁷Pri automatickom rozpoznávaní autentickej reči vzhľadom na individuálne nárečové a iné neortoepické návyky hovoriacich nevystačíme s výslovnostnými pravidlami spisovnej slovenčiny [25]. Táto skutočnosť je v práci do značnej miery zohľadnená.

aspoň jednu grafému⁸. Na generovanie $\Phi \rightarrow W$ je možné použiť napríklad natrénovaný HMM systém [30].

- Z dvojíc W, Φ vytvorených v predchádzajúcom kroku je vytvorený rozhodovací strom pre každý znak w založený na vlastnostiach postupnosti fón ϕ už generovaných pre predchádzajúce grafémy w a otázky týkajúce sa ortografického kontextu znaku w . Takto vytvorený strom je potom použitý pri prevode z ortografickej do ortoepickej reprezentácie.

Detailnejší popis takéhoto postupu je možné nájsť napríklad v [39], [2], [3]. Výhoda takéhoto riešenia je jeho kompaktnosť a skutočnosť, že v prípade veľkej trénovacej množiny si bude vedieť poradiť aj so slovami, ktoré nie sú súčasťou daného jazyka (produkčné pravidlá by mohli generovať skôr nesprávny ako správny výsledok).

Tieto formy riešenia majú v dnešnej dobe praktické uplatnenie ako súčasť systémov na syntézu reči, kde sa často môžu vyskytnúť nové slová neobsiahnuté v slovníku výslovnosti [55], [9].

Poznámka: Predchádzajúce dve časti boli venované metódam generovania fonetickej transkripcie z – dá sa povedať – opačného konca spektra. Veľmi často je však možné naraziť na systémy, ktoré využívajú určitú formu kombinácie produkčných pravidiel a štatistických metód. Ďalšie práce v tejto oblasti pre slovenčinu a češtinu možno nájsť napríklad v [11] [7], [8] a pre iné jazyky v [40], [60].

1.4 Segmentácia rečového signálu

Pod segmentáciou rečového signálu sa rozumie jeho rozdelenie na úseky zodpovedajúce definovaným jednotkám a označenie každej jednotky symbolom z množiny symbolov segmentačných jednotiek. Ako jednotka segmentácie môže byť uvažovaná veta, slovo, slabika, fonéma, fóna, poprípade ľubovoľne definovaná menšia či väčšia jednotka v závislosti na tom, načo je segmentácia určená. V prípade segmentácie je k dispozícii rečový signál plus informácia o tom, aká postupnosť jednotiek $P = p_1, p_2, \dots, p_n$ segmentácie mu zodpovedá. $p_i \in \Upsilon$, kde Υ je množina všetkých segmentačných jednotiek. Ak je napríklad segmentačná jednotka fóna, tak Υ bude predstavovať fonetický repertoár daného jazyka⁹.

Metódy segmentácie je možné rozdeliť – podobne ako metódy automatickej transkripcie – na dve základné skupiny:

⁸Toto tvrdenie platí pre angličtinu. Pre slovenčinu len v prípade, že pre rás sa nepoužije samostatná fóna. V češtine, kde sa rás vyskytuje oveľa častejšie ako v slovenčine, a býva uvažovaný ako samostatná fóna aj v systéme rozpoznávania reči, už toto tvrdenie neplatí. Obdobná je situácia aj v nemčine.

⁹Samozrejme okrem fón bude obsahovať aj ďalšie symboly na označenie "nerečových" segmentov, ako napríklad ticho, šum a podobne.

- Postupy snažiace sa iteračnými metódami rozdeliť vstupný signál na najpravdepodobnejšie možné segmenty.
- Postupy, ktoré využívajú už existujúci systém a s jeho pomocou segmentujú vstupný signál.

V oboch prípadoch ide o postupy založené na HMM¹⁰ a v oboch prípadoch sa predpokladá, že vstupný signál je už spracovaný a na vstup segmentácie sú privádzané dáta z vektorového kvantizátora¹¹. Z pohľadu segmentácie nezáleží na tom, o aký spôsob výpočtu vektora príznakov ide (kepstrum, LPC, ...), ani aký druh vektorovej kvantizácie je použitý (k-mean, LBG, ...). Podrobnejšie o týchto postupoch pojednáva napríklad [51] alebo [50].

1.4.1 Priame metódy segmentácie

Za priame metódy segmentácie je možné považovať tie, pri ktorých sa nevyužíva referenčný systém, ale počas procesu segmentácie prebieha proces tréningu a následného segmentovania. Tento proces sa opakuje niekoľkokrát, pokiaľ sa nedospeje k uspokojivej segmentácii množiny tréningových dát.

Tento proces, napriek tomu, že sa na prvý pohľad môže zdať komplikovaný, nie je náročný z toho dôvodu, že estimačné algoritmy, akými je Baumov–Welchov (forward-backward) algoritmus, alebo segmentačný k-mean algoritmus, môžu byť inicializované rovnomerným (uniformným) rozdelením (dokonca náhodnou počiatočnou segmentáciou) a napriek tomu veľmi rýchlo konvergujú k optimálnemu odhadu segmentov¹². Väčšinou na uspokojivý odhad postačuje niekoľko iterácií.

Na ilustráciu funkčnosti tohto algoritmu je možné si predstaviť segmentáciu rečového signálu. Ako segmentačná jednotka nech je zvolená napríklad fonéma. K dispozícii je niekoľko stovák tréningových viet s ich textovým prepisom. Okrem toho je k dispozícii slovník, v ktorom je fonemický prepis každého slova obsiahnutého v tréningovej množine. Ďalej sa predpokladá, že na začiatku a konci každej vety je úsek ticha. Samozrejme, ticho sa môže nachádzať aj medzi dvoma slovami. Na základe týchto úvah je možné každú vetu z tréningovej množiny zapísať ako

$$S_W : W_1 W_2 W_3 \dots W_I$$

kde každé $W_i, 1 \leq i \leq I$ je slovo zo slovníka.

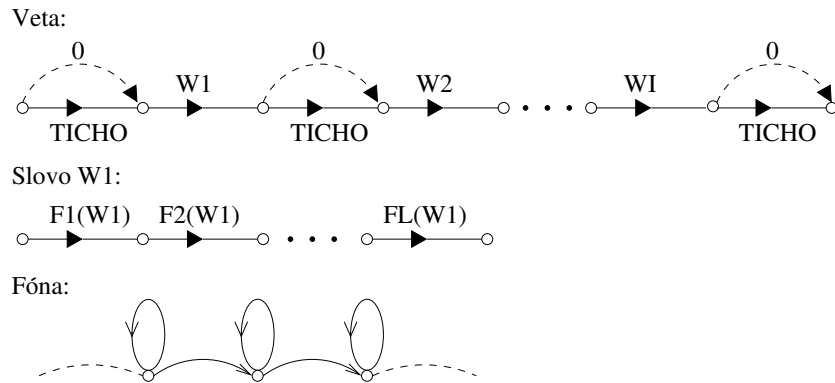
Nech jedna z viet tréningovej množiny je veta *Mama má Emu*. $W_1 = \text{Mama}$, $W_2 = \text{má}$ a $W_3 = \text{Emu}$. Každé slovo je potrebné vyhľadať v slovníku a nájsť jeho fonemický prepis. Na základe neho môže byť veta S zapísaná ako postupnosť foném

$$S_F : F_1(W_1)F_2(W_1) \dots F_{L(W_1)}(W_1) \oplus F_1(W_2)F_2(W_2) \dots F_{L(W_2)}(W_2) \oplus$$

¹⁰V prípade segmentácie pomocou referenčného systému je možné použiť napríklad aj neuronové siete.

¹¹Vektorová kvantizácia nie je nevyhnutnou súčasťou takéhoto systému.

¹²Samozrejme za predpokladu, že je k dispozícii dostatočná tréningová množina.



Obr. 1.4: Spôsob reprezentácie vety, slova a fonémy.

$$F_1(W_1)F_2(W_1) \dots F_{L(W_1)}(W_1) \oplus \dots \oplus F_1(W_I)F_2(W_I) \dots F_{L(W_I)}(W_I)$$

kde $L(W_1)$ je dĺžka slova W_1 v fonémach, $L(W_2)$ dĺžka slova W_2 atď. Nakoniec sa každá fonéma nahradí jej Markovovým modelom (HMM), a HMM pre ticho je vložené aj medzi každé dve slová. Každá fonéma bude modelovaná samostatným trojstavovým HMM. Pre ticho bude použitý HMM s jedným stavom, keďže ticho je väčšinou stacionárne a nemá komplexnejšiu vnútornú štruktúru, tak ako je to v prípade fonémy¹³. V prípade ticha medzi slovami bude umiestnený paralelne s modelom (stavom) pre ticho aj nulový prechod, pre prípad, keď sa medzi dvoma slovami žiadne ticho nenachádza. Obrázok 1.4 zobrazuje takto vytvorený model.

V prípade, že v slovníku je viac než jedna výslovnosť pre niektoré slovo, je potrebné v sieti pridať vo vete, kde je možná dvojaká (trojaká) výslovnosť, paralelnú vetvu pre dané slovo, ktorá bude obsahovať alternatívnu postupnosť HMM.

Pokiaľ je vytvorená takáto sieť pre každú vetu z trénovacej množiny, zostáva odhadnúť parametre modelov foném tak, aby maximalizovali pravdepodobnosť vytvorených modelov pre danú trénovaciu množinu. Tento odhad môže byť urobený pomocou Baumovho–Welchovho (forward-backward) algoritmu, alebo pomocou segmentačného k-mean trénovacieho algoritmu. Spôsob použitia k-mean segmentačného algoritmu na odhad parametrov modelov je nasledujúci:

1. **Inicializácia:** Každá veta z trénovacej množiny sa rozdelí rovnomerne na toľko častí, koľko je vo vete HMM stavov plus jeden segment pre ticho na začiatku a jeden na konci. Ticho medzi slovami sa neuvažuje. Číže v našom prípade, kde každá fonéma je modelovaná trojstavovým HMM, to bude počet foném vo vete krát tri plus dva.
2. **Zhlukovanie:** Všetky vektory príznakov zo všetkých segmentov, prislúchajúcich príslušnému stavu (i) zodpovedajúcemu danej fonéme, sa rozdelia do

¹³Samozrejme, nič nebráni použiť pre ticho taktiež trojstavový HMM, alebo ísť ešte ďalej a definovať rozdielne HMM pre ticho, ktoré je obsiahnuté na začiatku a konci, a pre ticho, ktoré sa nachádza medzi slovami.

M zhlukov pomocou k-mean algoritmu. Tento krok sa vykoná pre všetky fonémy a stavy.

3. **Odhad:** Pre každý zhluk k v stave i je zrátaný vektor stredných hodnôt μ_{ik} , (diagonálna) kovariančná matica \mathbf{U}_{ik} a váhové koeficienty zmesi funkcií hustoty (mixture weights) c_{ik} . Tento krok sa vykoná pre všetky fonémy a stavy.
4. **Segmentácia:** Aktualizovaná a (updatovaná) sada fonematických modelov (ktorá bola pripravená v kroku 3) sa použije na novú segmentáciu každej vetvy z trénovacej množiny na fonémy a stavy (pomocou Viterbiho dekodovacieho algoritmu). V tomto kroku sa prejaví paralelné vetvy vytvorené pre slová s alternatívnou výslovnosťou.
5. **Iterácia:** Kroky 2–4 sa opakujú, pokiaľ proces konverguje.

Pri použití vyššie uvedeného postupu sa už po troch až štyroch iteráciách proces ustáli [51]. Prezentovaný postup demonštruje, ako možno jednoduchým spôsobom z viet, ku ktorým je k dispozícii len ich prepis, odhadnúť parametre potrebné pre natrénovanie HMM, poprípade pre použitie na učenie iného systému, napríklad NN.

Nevýhodou tohto postupu je skutočnosť, že na dosiahnutie požadovanej kvality segmentácie je potrebné vykonať niekoľko iterácií. V prípade že trénovacia množina obsahuje rádovo stovky tisíc viet, je už segmentácia podľa postupov uvedených v tejto časti časovo náročnou záležitosťou, napriek tomu, že sa vykonávajú len tri alebo štyri iterácie¹⁴.

1.4.2 Metódy segmentácie využívajúce referenčný systém

Pri týchto postupoch je k dispozícii už natrénovaný systém na rozpoznávanie reči a pomocou neho je nový vstupný signál, ku ktorému je k dispozícii aj jeho fonetický (fonematický)¹⁵ prepis, segmentovaný na požadované jednotky. Inými slovami, každému jednotkovému úseku signálu je priradený symbol z množiny výstupných symbolov referenčného systému použitého na segmentáciu.

Z predchádzajúcej časti vyplýva, že v tomto prípade vlastne ide o nahradenie iteračného postupu, slúžiaceho k získaniu optimálnych HMM pre segmentáciu, už natrénovaným HMM, takže odpadá tréning a na segmentáciu je postačujúca jedna iterácia¹⁶.

Nech je k dispozícii HMM systém, ktorý na vstupnú postupnosť segmentov $\mathbf{A} = a_1, a_2, \dots, a_k$ reaguje sekvenciou výstupných symbolov y_1, y_2, \dots, y_k , kde

¹⁴V takomto prípade je potrebné zväziť množstvo trénovacích viet, ktoré má byť použité na vytvorenie modelov.

¹⁵Poprípade iný prepis, v závislosti od toho, aký referenčný systém máme k dispozícii.

¹⁶Ako už bolo uvedené, nemusí v tomto prípade ísť o HMM, ale môžu byť použité aj neurónové siete, poprípade iné techniky.

$y_i \in \Psi$ a Ψ je výstupná abeceda¹⁷. Za predpokladu, že ide o fonetickú abecedu, výstupom je postupnosť fón, ktorá by mala zodpovedať vstupnému signálu. V tomto prípade systém pracuje ako rozpoznávač, keďže informácia o tom, aká postupnosť znakov abecedy y_1, y_2, \dots, y_k mala byť vygenerovaná, nebola nikde využitá, rovnako ako nebola obmedzená množina možných výstupov. Možným dôsledkom je, že výstupná množina symbolov y_1, y_2, \dots, y_k nemusí zodpovedať vstupnej postupnosti.

Keďže cieľom je segmentácia, k dispozícii je informácia o tom, aká postupnosť y_1, y_2, \dots, y_k má byť generovaná¹⁸. Čo však je potrebné určiť, je časový úsek, v ktorom sa nachádza y_i . Keďže v systéme je každá fóna modelovaná HMM, je potrebné určiť najpravdepodobnejšiu postupnosť stavov s_1, s_2, \dots, s_k vedúcu ku generovaniu výstupu y_1, y_2, \dots, y_k , čiže

$$P(s_1, s_2, \dots, s_k \mid y_1, y_2, \dots, y_k, s_0) = \frac{P(s_1, s_2, \dots, s_k, y_1, y_2, \dots, y_k \mid s_0)}{P(y_1, y_2, \dots, y_k \mid s_0)}$$

Z rovnice vyplýva, že snahou je nájsť postupnosť s_1, s_2, \dots, s_k , pre ktorú je čitateľ na pravej strane rovnice najväčší. Keďže postupnosť stavov v použítom HMM je Markovov proces, tak pre všetky i platí

$$\begin{aligned} & P(s_1, \dots, s_i, s_{i+1}, \dots, s_k \mid y_1, \dots, y_i, y_{i+1}, \dots, y_k, s_0) \\ &= P(s_1, s_2, \dots, s_i, y_1, y_2, \dots, y_i \mid s_0) P(s_{i+1}, \dots, s_k, y_{i+1}, \dots, y_k \mid s_i) \end{aligned}$$

takže

$$\begin{aligned} & \max_{s_1, s_2, \dots, s_k} P(s_1, \dots, s_i, s_{i+1}, \dots, s_k \mid y_1, \dots, y_i, y_{i+1}, \dots, y_k, s_0) \\ &= \max_{s_i, s_{i+1}, \dots, s_k} \left[P(s_{i+1}, \dots, s_k, y_{i+1}, \dots, y_k \mid s_i) \right. \\ & \times \left. \max_{s_1, s_2, \dots, s_{i-1}} P(s_1, s_2, \dots, s_i, y_1, y_2, \dots, y_i \mid s_0) \right] \end{aligned} \quad (1.1)$$

Ak sa definuje

$$\gamma_i(s_i) \doteq \max_{s_1, s_2, \dots, s_{i-1}} P(s_1, s_2, \dots, s_i, y_1, y_2, \dots, y_i \mid s_0) \quad (1.2)$$

a použije pre substitúciu v (1.1), tak

$$\begin{aligned} & \max_{s_1, s_2, \dots, s_k} P(s_1, \dots, s_i, s_{i+1}, \dots, s_k \mid y_1, \dots, y_i, y_{i+1}, \dots, y_k, s_0) \\ &= \max_s \left\{ \max_{s_{i+1}, \dots, s_k} [P(s_{i+1}, \dots, s_k, y_{i+1}, \dots, y_k \mid s) \gamma_i(s)] \right\} \end{aligned} \quad (1.3)$$

¹⁷V našom zjednodušenom prípade sa neuvažuje HMM s nulovými prechodmi a preto vstup aj výstup sú dĺžky k .

¹⁸Toto platí v ideálnom prípade. V skutočnosti môže byť k dispozícii viacero výstupných postupností Y a v prvom kroku je potrebné určiť, ktorá z nich je správna. Ak je napríklad na vstupe slovo *alebo*, na výstupe môže – v závislosti od výslovnosti – byť postupnosť *alEbO* alebo aj *aLEbO*.

Tento výpočet je možné urobiť rekurzívne. Z definície HMM a rovnice (1.2) vyplýva, že

$$\begin{aligned}
 \gamma_i(s_i) &= \max_{s_1, \dots, s_{i-1}} P(s_1, s_2, \dots, s_i, y_1, y_2, \dots, y_i \mid s_0) \\
 &= \max_{s_{i-1}} p(y_i, s_i \mid s_{i-1}) \max_{s_1, \dots, s_{i-2}} P(s_1, \dots, s_{i-1}, y_1, \dots, y_{i-1} \mid s_0) \quad (1.4) \\
 &= \max_{s_{i-1}} p(y_i, s_i \mid s_{i-1}) \gamma_{i-1}(s_{i-1}) \quad (1.5)
 \end{aligned}$$

Rovnica (1.4) vedie priamo k Viterbiho algoritmu [57]. Tento postup predstavuje štandardný prístup pre segmentáciu v prípade, že je k dispozícii natrénovaný systém. Za predpokladu, že segmentované dáta sa nelíšia veľmi od dát použitých na samotné natrénovanie systému, výsledky segmentácie sú veľmi dobré. Touto metódou možno uspokojivo segmentovať jeden jazyk modelom druhého¹⁹. V prípade, že segmentované dáta a tréningové dáta sú dosť odlišné, výsledky nemusia byť uspokojujúce²⁰.

Ďalšie obmedzenie vyplývajúce z tohto druhu segmentácie spočíva v tom, že signál nie je možné segmentovať na menšie jednotky, ako sú výstupy generované referenčným systémom. Ak sú napríklad výstupmi HMM referenčného systému symboly z množiny všetkých foném pre daný jazyk, nie je možné týmto postupom získať fonetickú, poprípade akúkoľvek inú "nižšiu" segmentáciu.

¹⁹Tu samozrejme vyvstáva otázka správneho mapovania fonetiky jedného jazyka do druhého jazyka, rovnako ako aj miera odlišnosti týchto dvoch fonetik.

²⁰To nastáva napríklad v tom prípade, že na tréning systému boli použité čisté (nezašumené) dáta a takýmto systémom sa pokúšame segmentovať dáta značne zašumené.

Kapitola 2

Ciele dizertačnej práce

V posledných desaťročiach v rozpoznávaní reči prevládajú dve techniky – skryté Markovove modely (HMM) a neurónové siete (NN). Systémy založené na HMM sú výpočtovo menej náročné a jednoduchšie implementovateľné, preto sa presadili v dnešných komerčných systémoch. Z hľadiska kvality rozpoznávania sú obidva postupy porovnateľné. V oboch prípadoch je na vytvorenie systému potrebná množina tréningových dát a v oboch prípadoch ide o segmentované dáta¹.

Tréningová množina obsahuje okrem samotných nahrávok tréningových viet aj ich prepis do formátu použitých segmentačných jednotiek. Tento proces – až na niektoré kroky – je väčšinou robený automaticky, hlavne v prípade, že nejde o prvotný tréning nového jazyka.

V opačnom prípade je nutné vyriešiť výber segmentačných jednotiek, ich kódovanie a podobne. Kroky, ktoré je potrebné urobiť pred samotnou segmentáciou rečového signálu nového jazyka, je možné zhrnúť do nasledujúcich bodov:

- Definícia segmentačných jednotiek.
- Definícia ich kódovania.
- Prepis ortografického opisu tréningových viet do systému segmentačných jednotiek.

Tieto kroky sú jazykovo závislé. Po ich úspešnom zvládnutí sú v ďalšom kroku aplikovateľné všeobecne zaužívané a jazykovo nezávislé postupy na samotnú segmentáciu rečového signálu.

Z poznania súčasného stavu v rozpoznávaní reči pre slovenčinu vyplynulo, že doteraz neboli pre slovenčinu vyriešené základné jazykovo závislé kroky potrebné pre segmentáciu. Okruh problémov sa ešte rozšíri z pohľadu kompatibility s okolitým svetom. Na základe tohto poznania a vyššie uvedených faktov boli sformulované nasledujúce ciele dizertácie:

¹V prípade HMM môže byť segmentácia priamo súčasťou tréningu, avšak v zložitejších systémoch je proces segmentácie oddelený.

1. Výber vhodných segmentačných jednotiek a návrh ich kódovania kompatibilného s inými jazykmi a vhodného nielen pre účely počítačového rozpoznávania reči, ale akceptovateľného aj v jazykovednej sfére.
2. Analýza možností automatickej transkripcie v slovenčine s následným návrhom a realizáciou takéhoto systému.
3. Analýza možností automatického slabičného a morfematického delenia slov v slovenčine s následným návrhom a realizáciou takéhoto systému.
4. Segmentácia rečového signálu slovenčiny pomocou HMM s overením navrhnutých riešení.

Kapitola 3

Zvolené metódy a vlastný postup riešenia

3.1 Zvolené metódy a postupy

Táto práca sa zaoberá procesom segmentácie rečového signálu ako základným krokom k vybudovaniu akéhokoľvek zložitejšieho systému na rozpoznávanie reči. Proces samotnej segmentácie vyžaduje prípravu dát, z ktorej niektoré kroky sú jazykovo závislé, a preto je potrebné riešiť túto problematiku pre každý jazyk.

Pre proces segmentácie je potrebný predspracovaný rečový signál, čo je možné považovať za jazykovo nezávislú úlohu, a popis spracovaného signálu v systéme požadovaných segmentačných jednotiek. Táto úloha je už jazykovo závislá a obsahuje niekoľko krokov a potenciálnych nedoriešených problémov.

3.2 Výber segmentačných jednotiek

Výber segmentačných jednotiek je otázka, na ktorú je potrebné zodpovedať ako na jednu z prvých. Pre proces samostatnej segmentácie totiž nepostačuje len samotný rečový signál, ale aj jemu zodpovedajúci text, ktorý musí byť "segmentovaný" na rovnaké segmentačné jednotky, na aké má byť segmentovaný samotný rečový signál.

Ako segmentačné jednotky môžu byť použité slová, slabiky, fonémy, fóny, poprípade ľubovoľné iné definované jednotky v závislosti od toho, za akým účelom je segmentácia vykonávaná.

Slová

Segmentácia na slová je vhodná v tom prípade, ak cieľom je natrénovať systém, ktorý bude modelovať každé slovo ako samostatný HMM¹. Typický príklad pre takúto aplikáciu je rozpoznávanie čísloviek. Pre každú číslovku od nula do deväť je

¹Samozrejme, tie isté segmentované úseky môžeme použiť na tréning neurónovej siete.

vytvorený HMM. Okrem toho je potrebné natrénovať modely pre ticho, resp. šum. Nevýhodou takéhoto systému je skutočnosť, že okrem čísloviek – vo všeobecnosti okrem slov, ktoré boli použité na tréning – nebude vedieť rozpoznávať nič iné. Na druhej strane takéto systémy môžu byť veľmi spoľahlivé.

Slabiky

Slabika ako segmentačná jednotka je vhodná pre také typy jazykov ako napríklad japončina, kde sú slabiky presne definované a ich počet nie je príliš veľký. Japončina obsahuje okolo 50 slabík a slabika je v nej základnou stavebnou jednotkou reči.

Pre slovenčinu slabika ako segmentačná jednotka nie je príliš vhodná. Jednak z dôvodu, že už samotná definícia slabiky v slovenčine je značne vágna, a jednak preto, že celkový počet slabík v slovenčine je príliš vysoký. Ako vyplýva z výsledkov dosiahnutých v časti 4.3, ich celkový počet je cez 4000. Skutočný počet sa v dostupnej literatúre nepodarilo zistiť a podľa autora práce vytváraniu nových slabík nič nestojí v ceste.

Fonémy

Fonéma sa môže javiť ako najprirodzenejšia segmentačná jednotka v slovenčine. Vyplýva to hlavne z pocitu, že do značnej miery prirodzene korešponduje s grafémami. Slovenčina obsahuje nasledujúce fonémy²:

- **Samohlásky a dvojhásky:** *a, e, i, o, u, ä, á, é, í, ó, ú, ĩa, ĩe, ĩu, ô*
- **Spoluhlásky:** *p, b, t, d, ě, ě, k, g, m, n, r, ř, l, ľ, ň, ě, f, v, s, z, š, ž, x, h, j, c, 3, ě, ž*

Ako vyplýva zo zoznamu, počtom aj podobnosťou sa tento repertoár podobá repertoáru grafém v slovenčine. Fonéma ako segmentačná jednotka v slovenčine má okrem lepšej čitateľnosti aj tú výhodu, že pre slová použité v tréningovej množine je jednoduchšie generovať fonematický prepis, než prepis fonetický. V prípade manuálneho vytvárania prepisu nie sú potrebné fonetické znalosti na takej úrovni, ako je to v prípade vytvárania prepisu na úrovni fón.

Fóny

Napriek vyššie uvedeným výhodám segmentácie na fonémy bola ako najvýhodnejšia segmentačná jednotka pre slovenčinu zvolená fóna. Z definície fóny vyplýva, že fóna je najmenšia akustická jednotka jazyka³ [33], [52]. Ďalšia výhoda použitia fón namiesto foném pre rozpoznávanie reči spočíva v tom, že niektoré fóny

²Tu je použité kódovanie používané v slovenskej fonológii aby sa zvýraznila podobnosť s grafémami.

³Na rozdiel od fonémy, ktorú môžeme považovať za najmenšiu význam rozlišujúcu jednotku.

sa vyskytujú len v určitom – presne definovanom – kontexte. Túto skutočnosť je možné využiť pri rozpoznávaní reči na zlepšenie kvality systému⁴.

Nevýhody fón spočívajú v horšej čitateľnosti a už spomínanej komplikovanejšej transkripcie. Zoznam slovenských fón – v kódovaní zvyčajne používanom v slovenskej fonetike a fonológii je nasledujúci:

- **Samohlásky dvojhásky:** *a, e, i, o, u, ä, á, é, í, ó, ú, ĩa, ĩe, ĩu, ůo*
- **Spoluhlásky:** *p, b, t, d, t̃, d̃, k, g, m, m̃, n, ñ, ŋ, r, r̥, r̥̃, l, l̥, l̥̃, ň, l̃, f, v, w, ů, s, z, š, ž, x, h, γ, j, ĵ, c, ʒ, č, ʒ*

Zoznam slovenských fón, aj s ich kódovaním IPA, SAMPA a kódovaním používaným v slovenskej fonetike, je v tabuľke 3.6.

Zhrnutie

- Vo všeobecnosti platí, že segmentačné jednotky vyššej úrovne je možné odvodiť zo segmentačných jednotiek nižšej úrovne. Teda ak je urobená segmentácia na úrovni fón, je možné z nej jednoducho prejsť na fonematickú úroveň prepísaním alofón na fonémy. Pre slovenčinu to znamená aplikovať nasledujúce substitúcie:

Fonéma	Alofóny
m	m, m̃
n	n, ñ, ŋ
r	r, r̥
l	l, l̥
h	h, γ
v	v, w, ů
j	j, ĵ

Týmto spôsobom je možné pokračovať ďalej. Samozrejme, dolná hranica nie sú fóny. Ak je napríklad definovaný pre každú fónu trojstavový HMM, tak ako je to v navrhnutom riešení, každý stav je možné považovať za segment – segmentom v skutočnosti aj je – a uvažovať ho ako začiatok fóny, jej centrálnu časť a koniec fóny.

- Nič nebráni v použití ľubovoľných segmentačných jednotiek. V [1] bola úspešne použitá fonematická abeceda, ktorá neobsahovala dlhé samohlásky. Bolo na to niekoľko praktických dôvodov vrátane obmedzení, o ktorých bližšie pojednáva časť 1.2.2. Tá istá abeceda bola použitá na definovanie segmentačných jednotiek. Detailnejšiemu popisu segmentačných jednotiek sa autor venuje v [21].

⁴Toto platí samozrejme aj pre niektoré fonémy, avšak tu je myslený fonetický kontext ako napríklad v prípade výslovnosti spoluhlásky *n*.

3.3 Fonetická transkripcia a kódovanie

V tejto časti sú opísané prístupy k problematike vytvárania SAMPA abecedy pre slovenčinu. Keďže tomuto problému sa autor venoval niekoľko rokov, budú tu opísané dva prístupy k riešeniu problému. Prvá verzia SAMPA abecedy pre slovenčinu bola vytváraná v čase, keď samotná SAMPA abeceda bola v začiatkových štádiách vytvárania. Nedostatok potrebných znakov viedol k tomu, že bola vytvorená len fonemická verzia.

Druhá verzia, vytvorená s niekoľkoročným odstupom, sa snaží popísať kompletný repertoár slovenských fón abecedou SAMPA a tým pádom aj IPA⁵.

Pri vytváraní IPA verzie slovenskej transkripcie boli okrem neúplnosti zdrojov zaoberajúcich sa možnosťami aplikácie IPA systému na slovenčinu problémom aj isté malé rozdiely v [33] a [52] v opise slovenského transkripčného systému. Ako východiskový zdroj bol použitý [33].

3.3.1 Fonematický prepis

Prvá verzia SAMPA abecedy pre slovenčinu bola vzhľadom na vyššie spomínané problémy navrhnutá na prepis na fonematickej úrovni. To znamená, že alofóny boli prepísané jedným znakom. To čiastočne umožnilo vysporiadať sa s problémom absencie znakov v SAMPA abecede. Išlo o alofóny foném *m*, *n*, *r*, *l*, *h*, *v*, *j*⁶.

Fonéma	Alofóny
m	m, <i>m̥</i>
n	n, <i>n̥</i> , ŋ
r	r, <i>r̥</i>
l	l, <i>l̥</i>
h	h, <i>ɣ</i>
v	v, w, <i>ʋ</i>
j	j, <i>j̥</i>

Týmto bol zredukovaný počet potrebných znakov na vytvorenia SAMPA abecedy o 7.

Pri vytváraní SAMPA transkripcie boli vzhľadom k absencii IPA transkripcie pre slovenčinu – v tom čase SAMPA ani nebola popisovaná pomocou IPA znakov – pre správne priradenie použité opisy jednotlivých hlások podľa [58] v prípade SAMPA znakov a [33] v prípade slovenskej transkripcie. Takto bol pre slovenské samohlásky *a*, *ä*, *e*, *i*, *o*, *u*, *á*, *é*, *í*, *ó*, *ú* získaný SAMPA prepis *a*, *#*, *E*, *i*, *o*, *u*, *a:*, *E:*, *i:*, *o:*, *u:*. V prípade dvojhlások *ia*, *ie*, *iu*, *ô* bol použitý prepis *ia*, *iE*, *iu*, *uo*. Ako je uvedené v kapitole 3.3.2, v tomto prepise sa vyskytuje niekoľko chýb,

⁵To vyplýva zo skutočnosti, že proces vývoja SAMPA abecedy dospel do štádia kompatibility s IPA abecedou a na popis SAMPA znakov sa začali používať aj znaky IPA. V počiatkoch vývoja SAMPA abecedy boli jednotlivé znaky opisované len ich fonetickými vlastnosťami.

⁶Použitie kódovania zodpovedá kódovaniu v slovenskej fonetickej transkripcii.

ST	SAMPA	ST	SAMPA	ST	SAMPA	ST	SAMPA
<i>a</i>	a	<i>ǐe</i>	iE	<i>m</i>	m	<i>z</i>	z
<i>á</i>	a:	<i>ǐu</i>	iu	<i>n</i>	n	<i>š</i>	S
<i>ä</i>	#	<i>ô</i>	uo	<i>r</i>	r	<i>ž</i>	Z
<i>e</i>	E	<i>p</i>	p	<i>ř</i>	r:	<i>x</i>	x
<i>é</i>	E:	<i>b</i>	b	<i>l</i>	l	<i>h</i>	h
<i>í</i>	i:	<i>t</i>	t	<i>ĺ</i>	l:	<i>j</i>	j
<i>o</i>	o	<i>d</i>	d	<i>ň</i>	J	<i>c</i>	ts
<i>ó</i>	o:	<i>ť</i>	7	<i>ľ</i>	L	<i>ʒ</i>	dz
<i>u</i>	u:	<i>ď</i>	8	<i>f</i>	f	<i>č</i>	tS
<i>ú</i>	u:	<i>k</i>	k	<i>v</i>	v	<i>ž</i>	dZ
<i>ǎ</i>	ia	<i>g</i>	g	<i>s</i>	s		

Tabuľka 3.1: Fonematická verzia SAMPA abecedy pre slovenčinu.

ktoré vyplývali hlavne zo štádia, v ktorom sa vývoj abecedy SAMPA v tom čase nachádzal.

V prípade prepisu spoluhlások vznikol v niekoľkých prípadoch problém určenia správneho ekvivalentu k slovenským spoluhláskam. V dvoch prípadoch nebolo možné priradiť slovenským hláskam SAMPA znaky z dôvodu neexistencie SAMPA ekvivalentov pre slovenské hlásky. Vyplýva to z dôvodov uvedených v časti 1.2.2.

Problém v prípade spoluhlások boli palatály *t* a *d*. Keďže žiaden zo západoeurópskych jazykov, na základe ktorých bola navrhnutá prvá verzia SAMPA abecedy, neobsahuje palatály *t* a *d*, bolo nutné priradiť im vlastné symboly. Za týmto účelom boli použité dovtedy nevyužitá znaky 7 (*t*) a 8 (*d*). Táto verzia bola napriek obmedzeniam vyplývajúcim z absencie opisu alofón, ako aj neúplného popisu spoluhlások, úspešne použitá v systéme na rozpoznávanie reči v rámci projektu SQEL [1]. Kompletná fonematická verzia SAMPA abecedy pre slovenčinu navrhnutá v roku 1995 je v tabuľke 3.1.

3.3.2 Fonetický prepis

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 doterajšej snahy je transkripčný systém SAMPA pre slovenčinu. Problematika vypracovaného návrhu bude ďalej rozobratá 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 [56], je však niekoľko nepresností.

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 ranných verziách SAMPA.

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

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

Týmto znakom bolo možné 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, ʒ̣.

Týmto bol repertoár slovenských spoluhlások, prepísateľný pomocou prvej verzie SAMPA abecedy, vyčerpaný.

Palatálam *ṭ*, *ḍ*, pre ktoré v abecede SAMPA v čase vzniku prvej verzie pre slovenčinu ešte neexistovali znaky, boli priradené vlastné symboly 7 a 8, 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. Znak bol zavedený z dôvodu nedostatku voľných symbolov pre ďalšie rozširovanie SAMPA abecedy.
- *Rozšírenie o gréčtinu*, ktoré obsahuje znaky pre už spomínané palatály *ṭ – c* a *ḍ – J̣*, pre ktoré boli v prvej verzii pre slovenčinu použité vlastné symboly.

Z týchto rozšírení SAMPA prepis pre slovenčinu získal symboly pre *ṭ – c*, *ḍ – J̣* a slovenské znelé *h – ḥ*.

Na tomto mieste je vhodné poznamenať, že znak pre palatalizáciu, ktorý bol zavedený skôr ako rozšírenie pre gréčtinu, je pre slovenčinu nevyhovujúci, napriek tomu, že sa používa v slovenskom aj českom transkripčnom systéme. Slovenčina rovnako ako aj čeština má len palatálne konsonanty a nemá palatalizované konsonanty, no znak pre palatalizáciu je v našich jazykoch používaný napriek možnej zavádzajúcej a mylnej interpretácii pri konfrontácii so skutočnými palatalizovanými konsonantmi, aké má napr. ruština. Navrhovatelia slovenskej a českej transkripcie si boli tejto skutočnosti vedomí, no zostali pri tomto nejednoznačnom

prepise (upozorňujúc zase na iné možné problémy s využívaním symbolov IPA pre jednotlivé palatály a argumentujúc rozšírenosťou vedomia o tom, v ktorých jazykoch má svoje miesto palatálnosť a v ktorých palatalizovanosť), aj napriek rozporu s požiadavkou po jednoznačnosti symbolov v medzinárodných podmienkach [19].

V procese tvorby abecedy SAMPA pre slovenčinu sa vyskytol jeden vážnejší problém, ktorý sa doteraz nepodarilo celkom uspokojivo vyriešiť. Je to prepis nazály $\#$, pre ktorú v IPA ani v SAMPA systéme nie je symbol. Vychádzajúc z predpokladu, že SAMPA rovnako ako ani IPA nepredpokladá pre tento prípad také podrobné rozlíšenie, prepisuje sa ako n (napokon aj v doterajšej slovenskej transkripcii existujú v zachytení alofón n popri kvalitatívnych aj kvantitatívne rozdiely – napr. [33] a [52]).

Samohlásky

Hlavný problém pri vytváraní SAMPA abecedy pre slovenčinu tvoria samohlásky. V tomto prípade nejde v súčasnosti 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 zatvorenými vokálmi v systéme IPA a každý prepis do IPA tak predstavuje vlastne len určité priblíženie k optimálnemu riešeniu na základe niektorých charakteristík samohlások.

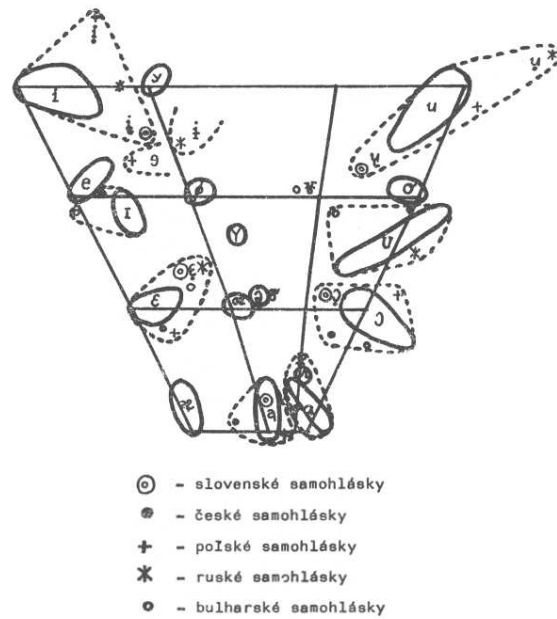
Tento problém sa netýka len slovenčiny, ale takisto aj ďalších slovanských jazykov. Na obrázku 3.1 je ukážka rozloženia samohlások pre slovenčinu, češtinu, poľštinu, ruštinu a bulharčinu. Z obrázku vyplýva, že pre žiaden z uvedených jazykov nie je možný jednoznačný prepis samohlások do IPA – tým pádom aj SAMPA – systému.

V prvej verzii abecedy SAMPA pre slovenčinu boli slovenským samohláskam a , e , i , o , u priradené IPA⁷ znaky a , ε , i , o , u [24]. Táto verzia bola aktuálna až do roku 2000.

V ďalšej verzii SAMPA pre slovenčinu [25] bola v prepise samohlások uskutočnená zmena, ktorá sa týkala zápisu samohlások e , i , o , u . Pri tomto kroku sa vychádzalo z novšieho uvedenia IPA transkripcie pre češtinu, publikovaného v medzinárodnom zborníku IPA z roku 1999 [10]. Vychádzalo sa pritom z poznatkov spomínaných vo fonetickej literatúre, že "slovenské spoluhlásky sa kryjú skoro úplne s českými, a to ako z hľadiska artikulácie, tak aj akustiky [16], [17], citované aj v [13], ktoré však, ako ukazujú novšie výskumy, nie sú celkom presné.

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 spomínaného v súvislosti s abecedou IPA už vyššie,

⁷Podľa súčasného prepisu medzi IPA a SAMPA.



Obr. 3.1: Rozloženie samohlások v niektorých slovanských jazykoch.

Samohláska	F1 (Hz)	F2 (Hz)	F3 (Hz)
a	682	1315	2293
e	428	1718	2365
i	285	1916	2656
o	481	1084	2194
u	326	967	2059
ä	700	1510	2300

Tabuľka 3.2: Formantové frekvencie pre slovenské samohlásky.

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 3.2 a 3.3 je uvedená vokalická schéma podľa rozpracovaného grafu IPA z r. 1996 [18] a uloženie slovenských vokálov, vychádzajúce z ich formantových charakteristík, podľa slovenskej fonetickej literatúry [20], [32], [35].

Formantové frekvencie pre slovenské samohlásky sú uvedené v tabuľke 3.2. V tabuľke 3.3 sú uvedené formantové frekvencie pre IPA samohlásky v systéme kardinálnych vokálov. Pre tretí formant sú v dostupnej literatúre uvádzané značne rozdielne hodnoty, takže údaje v tabuľke sú len orientačné. Pre niektoré fóny sa frekvencie pre tretí formant nepodarilo získať.

Z uloženia slovenských krátkych samohlások v schéme vyplýva, že využitie

Samohláska	F1 (Hz)	F2 (Hz)	F3 (Hz)
a	740	1300	2300
ɑ	740	1125	1900
e	360	2400	2630
ɛ	510	1950	2280
i	250	2800	3400
ɪ	280	2340	2590
o	360	800	
ɔ	510	950	
u	250	720	
ʊ			
æ	740	1650	

Tabuľka 3.3: Formantové frekvencie pre kardinálne vokály v systéme IPA.

IPA symbolov *a*, *e*, *i*, *o*, *u* na ich prepis, ktoré sa nachádza v niektorých prepisoch slovenských hlások do IPA aj SAMPA [12], [56], posúva predstavu o ich vlastnostiach. Realite bližší je prepis pomocou znakov *a*, *ɛ*, *ɪ*, *ɔ* a *ʊ*, preto boli tieto znaky použité ako podklad pre zápis v SAMPA.

Na overenie tejto skutočnosti bola určená vzdialenosť slovenských samohlások k jednotlivým adeptom na prepis. Na určenie vzdialenosti medzi dvoma samohláskami $S_A = (F_{1A}, F_{2A}, F_{3A})$ a $S_B = (F_{1B}, F_{2B}, F_{3B})$, ktoré sú reprezentované ich formantovými frekvenciami F_1, F_2 a F_3 , bola ako základ použitá euklidovská vzdialenosť

$$D(S_A, S_B) = \sqrt{\sum_{i=1}^n (S_{Ai} - S_{Bi})^2}. \quad (3.1)$$

Vzdialenosti medzi jednotlivými dvojicami boli zrátané podľa vzťahu (3.1) pre $n = 2$ a $n = 3$. Získané hodnoty sú uvedené v tabuľke 3.4 a 3.5. Pre presné meranie vzdialenosti by bolo vhodnejšie použitie Mahanobisovej vzdialenosti, poprípade inej vzdialenosti uvažujúcej nielen strednú hodnotu, ale takisto aj rozptyl, avšak z dôvodu absencie presného štatistického popisu formantových frekvencií pre slovenské samohlásky bola použitá rovnica 3.1. Vzdialenosť pre samohlásku *a* bola vynechaná, keďže v tomto prípade s prepisom do systému IPA nie sú problémy. Z výsledkov uvedených v tabuľkách 3.4 a 3.5 vyplýva, že prepis *ɛ*, *ɪ*, *ɔ*, *ʊ* je aj na základe vypočítaných vzdialeností formantových frekvencií vhodnejší pre slovenské samohlásky než prepis *e*, *i*, *o*, *u*.

Á. Král v analýze venovanej zápisu slovenských samohlások v znakoch IPA [32] pripomína väčšiu zatvorenosť slovenského *e* aj *o*, než je reprezentované týmto zápisom, čo naznačuje bodkou pod znakom, teda *ɛ̣*, *ɔ̣*, keďže však pravidlá pre prepis kódovania IPA do SAMPA diakritiku tohto typu neobsahujú, v prepise boli použité ako východisko IPA znaky *ɛ* a *ɔ*.

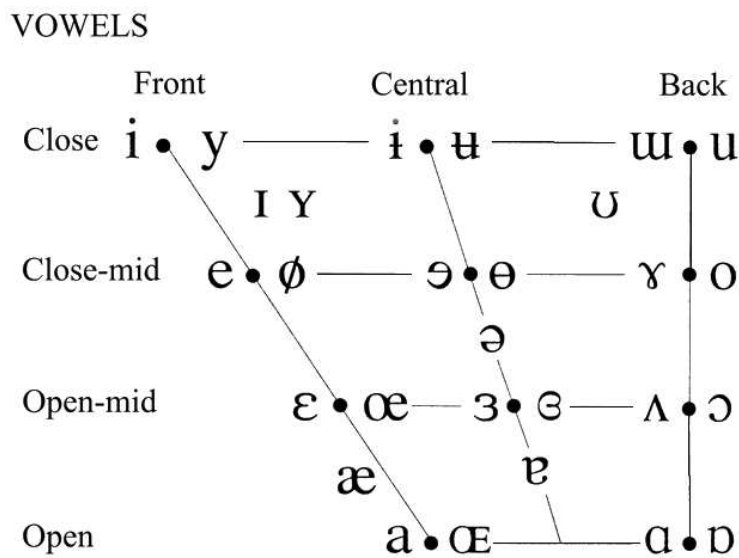
Pri prepise samohlásky *ä* bol použitý IPA znak *æ* (Á. Král v cit. článku vzhľa-

e → e	455,10	e → ε	246,10
i → i	884,69	i → ɪ	424,02
o → o	308,70	o → ɔ	137,10
u → u	267,04	u → ʊ	

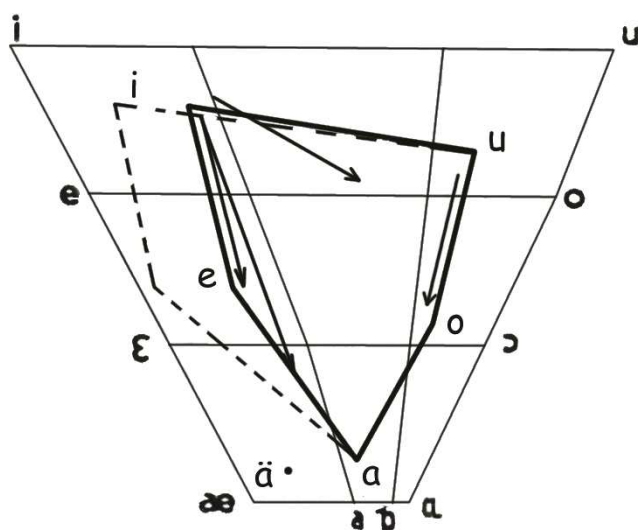
Tabuľka 3.4: Vzdialenosti slovenských samohlások a samohlások v schéme CV pre F_1 a F_2 .

e → e	526,63	e → ε	260,36
i → i	1155,94	i → ɪ	425,39

Tabuľka 3.5: Vzdialenosti slovenských samohlások a samohlások v schéme CV pre F_1 , F_2 a F_3 .



Obr. 3.2: Vokalická schéma IPA z roku 1996.



Obr. 3.3: Slovenské samohlásky a dvojhlásky v systéme kardinálnych vokálov. Čiarkovane je vyznačená poloha *e*, *i* po prepočte $(F2 + F3)/2$.

dom 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).

Pri prepise dlhých samohlások boli na základe ich artikulačnej aj akustickej analógie s krátkymi samohláskami bez signifikantných rozdielov [32], [35] využité tie isté znaky ako pri krátkych, s označením dĺžky: *a:*, *ɛ:*, *i:*, *ɔ:*, *ɔ:*.

Pri prepise slovenských dvojhlások 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: $\downarrow a$, $\downarrow \varepsilon$, $\downarrow i$, $\downarrow \varepsilon$, resp. ($I \hat{a}$, $I \hat{E}$, $I \hat{U}$, $U \hat{O}$). Diskusia a alternatívne riešenia prepisu sa tu môžu odvíjať od realizačných špecifik slovenských dvojhlások. 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 dvojhlásky *iu* v porovnaní so samohláskou *u* a druhého segmentu v *ia* v porovnaní so samohláskou *a*.

V prípade *iu* bol vzhľadom na centrálnejšie postavenie druhého segmentu využitý symbol, ktorý na naznačenie centrálneho postavenia hlásky inak prepisovanej pomocou \downarrow ponúka SAMPA: $U \setminus$. Zodpovedajúci znak v IPA ʊ je uvedený aj na mieste IPA symbolu v porovnávacíj tabuľke, je však v zátvorkách vzhľadom na to, že vo vokalických schémach IPA, ktoré boli k dispozícii, sa tento znak nevyužíva, a napokon sa na jeho predbežnú absenciu v IPA upozorňuje aj v prekóduvacích pravidlách IPA–SAMPA [59].

Alternatívny zápis dvojhlásky *ia* ponúka [32], keď pre jej druhý segment, ktorý sa postavením blíži k postaveniu samohlásky *ä*, uvádza aj možnosť prepisu pomo-

ST	IPA	SAMPA	ST	IPA	SAMPA	ST	IPA	SAMPA
<i>a</i>	a	a	<i>p</i>	p	p	<i>ň</i>	ɲ	J
<i>á</i>	a:	a:	<i>b</i>	b	b	<i>ĺ</i>	ʎ	L
<i>ä</i>	æ	{	<i>t</i>	t	t	<i>f</i>	f	f
<i>e</i>	ɛ	E	<i>d</i>	d	d	<i>v</i>	v	v
<i>é</i>	ɛ:	E:	<i>č</i>	c	c	<i>w</i>	ʋ	f_v
<i>i</i>	ɪ	I	<i>ď</i>	ɟ	J\	<i>ů</i>	ʊ	U_^
<i>í</i>	ɪ:	I:	<i>k</i>	k	k	<i>s</i>	s	s
<i>o</i>	ɔ	O	<i>g</i>	g	g	<i>z</i>	z	z
<i>ó</i>	ɔ:	O:	<i>m</i>	m	m	<i>š</i>	ʃ	S
<i>u</i>	ʊ	U	<i>rn</i>	ɱ	F	<i>ž</i>	ʒ	Z
<i>ú</i>	ʊ:	U:	<i>n</i>	n	n	<i>x</i>	x	x
<i>ia</i>	ɪa	I_^a	<i>ñ</i>	-	n	<i>h</i>	ɦ	h\
<i>ie</i>	ɪɛ	I_^E	<i>ŋ</i>	ŋ	N	<i>γ</i>	ɣ	G
<i>iu</i>	ɪ(ʊ)	I_^U\	<i>r</i>	r	r	<i>j</i>	j	j
<i>uo</i>	ʊɔ	U_^O	<i>ř</i>	ʃ	r=	<i>ĩ</i>	ɪ	I_^
			<i>ṛ̌</i>	ʃ:	r=:	<i>c</i>	ts	ts
			<i>l</i>	l	l	<i>ž</i>	dz	dz
			<i>ḷ</i>	ɫ	l=	<i>č</i>	tʃ	tS
			<i>ḷ̣</i>	ɫ:	l=:	<i>ž</i>	dʒ	dZ

Tabuľka 3.6: Porovnanie slovenskej transkripcie, IPA a SAMPA systému.

cou znakov æ (SAMPA znak {). Keďže však na základe artikulačno-percepčných vlastností má tento segment skôr a-ový charakter, z daných alternatív prepisu dvojhĺasky *ia* bola uprednostnená možnosť ɪa (SAMPA I_^a).

Tabuľka 3.6 obsahuje konečnú verziu SAMPA pre slovenčinu, ktorá bola publikovaná v [27]. Okrem SAMPA znakov a znakov slovenskej transkripcie sú v nej uvedené aj príslušné IPA znaky.

3.4 Automatická transkripcia

Pred vytvorením systému na automatické generovanie fonetickej transkripcie pre slovenčinu bolo potrebné rozhodnúť, či použiť niektorú zo štatistických metód, alebo sa pokúsiť navrhnuť pravidlá na produkciu správnej výslovnosti. Vzhľadom k tomu, že pre prvý prístup nie je k dispozícii dostatočné množstvo dát, bola analyzovaná možnosť definovať množinu pravidiel na generovanie možnej výslovnosti a výsledky analýzy implementované.

V nasledujúcej časti sú uvedené transkripčné pravidlá, ktoré boli navrhnuté pre prepis z ortografickej do ortoepickej reprezentácie. Keďže navrhnuté transkripčné pravidlá vychádzajú z pravidiel slovenskej ortoepie, je pred každým transkripčným pravidlom opis ortoepických pravidiel, z ktorých sa vychádzalo. Pravidlá

priamo prebrané z [33] sú uvedené menším písmom a označované ako §. Na kódovanie fonetických znakov je použitá SAMPA abeceda, opísaná v predchádzajúcej časti.

3.4.1 Samohlásky

V slovenčine je medzi zápisom samohlások a ich výslovnosťou jednoznačný vzťah, a preto je transkripcia samohlások pre spisovnú výslovnosť bez ťažkostí. Diskutabilný je prepis znaku *ä*, preto je mu venovaná samostatná časť 3.4.5. Na prepis samohlások sú použité nasledujúce pravidlá:

Krátke samohlásky:

$$\mathbf{a} \rightarrow \mathbf{a} \quad (p - a - pier) \quad (3.2)$$

$$\mathbf{e} \rightarrow \mathbf{E} \quad (p - e - ro) \quad (3.3)$$

$$\mathbf{i} \rightarrow \mathbf{I} \quad (p - i - vo) \quad (3.4)$$

$$\mathbf{y} \rightarrow \mathbf{I} \quad (d - y - m) \quad (3.5)$$

$$\mathbf{o} \rightarrow \mathbf{O} \quad (p - o - p - o - l) \quad (3.6)$$

$$\mathbf{u} \rightarrow \mathbf{U} \quad (p - u - to) \quad (3.7)$$

Dlhé samohlásky:

$$\mathbf{\acute{a}} \rightarrow \mathbf{a} : \quad (p - \acute{a} - s) \quad (3.8)$$

$$\mathbf{\acute{e}} \rightarrow \mathbf{E} : \quad (dc - \acute{e} - ra) \quad (3.9)$$

$$\mathbf{\acute{i}} \rightarrow \mathbf{I} : \quad (p - \acute{i} - skat\acute{e}) \quad (3.10)$$

$$\mathbf{\acute{y}} \rightarrow \mathbf{I} : \quad (nov - \acute{y}) \quad (3.11)$$

$$\mathbf{\acute{o}} \rightarrow \mathbf{O} : \quad (p - \acute{o} - l) \quad (3.12)$$

$$\mathbf{\acute{u}} \rightarrow \mathbf{U} : \quad (p - \acute{u} - pava) \quad (3.13)$$

Tieto pravidlá sú na samohlásku aplikované vtedy, ak sa samohláska nachádza medzi dvoma spoluhláskami alebo je na hranici slova.

3.4.2 Dvojhlásky (diftongy)

V spisovnej slovenčine sú štyri dvojhlásky: *ia*, *ie*, *iu*, *ô*. V prípade, že po samohláske *i* bezprostredne nasleduje ďalšia samohláska, nemusí ísť o dvojhlásku. Príkladom sú predponové slová (priučíť, vyučiť) a prevzaté slová (Ázia, duo).

Kedže určenie príslušnosti slova k skupine cudzích resp. prevzatých slov nie je možné bez rozsiahlej databázy opisujúcej možné prevzaté, resp. cudzie slová v slovenčine, bola uvažovaná len transkripcia slovenských slov. V prípade, že rozšírenie pravidiel o cudzie slová nebolo komplikované, boli zavedené aj tieto pravidlá. Dôvodom pre toto riešenie je skutočnosť, že v prípade, ak sa na vstupe objavajú aj cudzie, resp. prevzaté slová, sú urobené možné kroky redukovajúce množstvo manuálnej transkripcie.

Dvojhĺáska \hat{o}

Transkripcia dvojhĺásky \hat{o} je jednoznačná a bolo pre ňu definované nasledujúce pravidlo:

$$\hat{o} \rightarrow \mathbf{U_O} (v - \hat{o} - \mathit{la}) \quad (3.14)$$

Rovnaká jednoznačnosť platí aj pre dvojicu samohlások uo . V slovenčine sa dvojhĺáska \hat{o} nikdy nezapíše ako uo a z toho vyplýva, že dvojica samohlások uo sa nachádza v prevzatom alebo cudzom slove a je na ňu aplikovaná výslovnosť ako na samohláskovú skupinu.

$$\mathbf{uo} \rightarrow \mathbf{UO} \quad (3.15)$$

Predpona $vy-$

Tretie pravidlo, ktoré je možné jednoznačne definovať, využíva dvojaký znak pre fónu [i] v slovenskej ortografii. V prípade, že sa slovo začína dvojicou znakov $vy-$ a za nimi nasleduje jedna zo samohlások a, e, i, o, u , táto samohlásková kombinácia tvorí skupinu dvoch hlások, a nie dvojhĺásku.

Toto pravidlo, obdobne ako predchádzajúce, pokrýva aj množinu cudzích slov. Z toho dôvodu je uvedená kombinácia predpony vy so všetkými piatimi slovenskými samohláskami napriek tomu, že slovenčina obsahuje len tri dvojhĺásky začínajúce na i .

$$\hat{v}\mathbf{ya} \rightarrow \mathbf{vIa} \quad (vy - \mathit{asfaltovať}) \quad (3.16)$$

$$\hat{v}\mathbf{ye} \rightarrow \mathbf{vIE} \quad (vy - \mathit{expedovať}) \quad (3.17)$$

$$\hat{v}\mathbf{yi} \rightarrow \mathbf{vII} \quad (vy - \mathit{inkasovať}) \quad (3.18)$$

$$\hat{v}\mathbf{yo} \rightarrow \mathbf{vIO} \quad (vy - \mathit{obliekať}) \quad (3.19)$$

$$\hat{v}\mathbf{yu} \rightarrow \mathbf{vIU} \quad (vy - \mathit{učtovať}) \quad (3.20)$$

V ostatných prípadoch nie je možné len na základe jednoduchých pravidiel určiť, či dvojice samohlások ia, ie, iu tvoria samohláskovú postupnosť, alebo dvojhĺásku.

Predpona $pri-$

Ďalšia predpona, ktorá obsahuje i , je predpona $pri-$. Keďže táto predpona obsahuje znak i a nie y , v kombinácii so slovom začínajúcim na samohlásku je potrebné určiť, či ide skutočne o predponové slovo – teda slovo vytvorené z predpony pri a slovného základu ($priučít$) – alebo ide o slovo s dvojhĺáskou ($prienik$). Problém je rozobrať pre každú dvojicu samostatne.

Vzhľadom na to, že určiť príslušnosť slova k danému jazyku nie je možné bez znalosti jazyka, v zjednodušenom prípade to značí bez databázy pripisujúcej príslušnosť každého slova k domácim alebo prebraným, resp. cudzím slovám,

bolo potrebné aj v tomto prípade zvoliť mechanizmus klasifikácie prepisovaných slov. Keďže cieľom nebolo vytvoriť systém, ktorý bude so 100% správnosťou prepisovať ľubovoľné vstupné dáta do ortoepickej reprezentácie, ale systém, ktorý dokáže s vysokou spoľahlivosťou prepisovať bežný text, v ktorom sa môžu vyskytovať, a teda sa aj vyskytujú aj prebrané, resp. cudzie slova. Z tohto dôvodu bol ako zdroj informácií o jazyku zvolený korpus slovenského jazyka vytváraný v oddelení Slovenského národného korpusu v Jazykovednom ústave Ľudovíta Štúra SAV. Korpus v čase vytvárania tohto systému obsahoval vyše 3 000 000 slov v textoch rôznych žánrov. Všetky konštatovania týkajúce sa štatistických výskytov ortoepických javov v slovenčine boli odvodené z tohto korpusu (pokiaľ v texte nebude uvedené inak⁸).

Hlásková postupnosť *pria-* Ak sa na začiatku slova nachádza postupnosť znakov *pria-*, na určenie, či ide o dvojhlásku alebo samohláskovú skupinu, bola využitá skutočnosť, že slovenčina má len 23 slov [33], v ktorých základnom tvare sa na začiatku vyskytuje postupnosť znakov *pria* a znaky *ia* tvoria dvojhlásku. V prípade, že slovo patrí do tejto skupiny slov, je použité pravidlo

$$\hat{p}ria \rightarrow pr\mathbf{I}\hat{a}(pria - dka) \quad (3.21)$$

v opačnom prípade

$$\hat{p}ria \rightarrow pr\mathbf{Ia}(pri - agresívne) \quad (3.22)$$

Zoznam slov bez ohýbacej prípony bol vytvorený podľa korpusu a literatúry [33] a začlenený do prehľadávacej databázy.

V korpuse bola taktiež sledovaná početnosť opačného prípadu, t.j. keď *pria-* tvorí samohláskovú skupinu. Výskyt takýchto slov je menší než v prípade slov s dvojhláskou, avšak vytváraniu takýchto nových slov nič nebráni.

Hlásková postupnosť *prie-* V prípade postupnosti znakov *prie-* bol zvolený opačný prístup, keďže slov, v ktorých postupnosť *pri* netvorí predponu ako v predchádzajúcom prípade, je podstatne viac. V sledovanom korpuse bolo nájdených 593 rôznych slov začínajúcich sa na *prie-*, no vo všetkých prípadoch šlo o dvojhlásku a nie o samohláskovú postupnosť. Ani v literatúre [33] sa neuvádzajú takéto slová, avšak také slová môžu vzniknúť spojením prevzatého, resp. cudzieho slova začínajúceho sa na *e* a predpony *pri-*. Ako skutočný príklad z korpusu slovenčiny je možné uviesť slovo *priexotický*. Pre tento prípad je za predpokladu, že dané slovo sa nachádza v pomocnom slovníku, použité nasledujúce pravidlo:

$$\hat{p}rie \rightarrow pr\mathbf{IE}(pri - exotický) \quad (3.23)$$

v opačnom prípade

$$\hat{p}rie \rightarrow pr\mathbf{I}\hat{E}(prie - nik) \quad (3.24)$$

⁸Slovenský národný korpus v jeho aktuálnej verejne dostupnej verzii možno nájsť na internetovej adrese <http://korpus.juls.savba.sk>.

Hlásková postupnosť *priu-* Posledná z tretice je postupnosť *iu*. Keďže slovenčina nemá bezpredponové slovo začínajúce sa na *priu-*, je možné na každé slovo aplikovať nasledujúce pravidlo

$$\hat{p}riu \rightarrow pr\mathbf{IU}(pri - učiť) \quad (3.25)$$

V sledovanom korpuse bolo jediným slovom začínajúcim sa na *priu-* sloveso *priučiť* a jeho tvary. Celkový výskyt dvojhlásky *iu* je v slovenčine veľmi nízky, podľa [41] je to len okolo 0.016 %.

Po ošetrení predponových eventualít sú na zvyšok slova aplikované nasledujúce pravidlá

$$ia \rightarrow \mathbf{I}\hat{\mathbf{a}} \quad (p - ia - tok) \quad (3.26)$$

$$ie \rightarrow \mathbf{I}\hat{\mathbf{E}} \quad (sp - ie - vať) \quad (3.27)$$

$$iu \rightarrow \mathbf{I}\hat{\mathbf{U}} \setminus (cudz - iu) \quad (3.28)$$

ktoré zabezpečia správnu transkripciu dvojhlások v zvyšku slova za predpokladu, že slovo nepatrí medzi prevzaté, resp. cudzie slová.

Alternatívne ošetrenie predpony *pri-*

Vyššie opísané riešenie vyžaduje, aby bol zavedený slovník výnimiek. Napriek tomu, že bez takéhoto slovníka sa nie je možné — ako bude ukázané neskôr — zaoberať, po realizácii a overení postupov navrhnutých v časti 3.5 sa ukázalo ešte jedno riešenie vyššie spomínaného problému.

Z definície slabiky a dvojhlásky vyplýva, že slabičná hranica nemôže rozdeľovať dvojhlásku. Ak bude toto tvrdenie aplikované na vyššie uvedenú časť o predpone *pri-*, problém je možné definovať nasledujúcim spôsobom: *Ak slovo začína postupnosťou "pri", za "i" nasleduje samohláska a medzi "i" a samohláskou je slabičná hranica, potom ide o samohláskovú skupinu. V opačnom prípade ide dvojhlásku.*

V prípade, že bude použitá táto metóda, odpadá problém slovníka výnimiek pre určenie predpony. Na zaručenie spoľahlivosti tohto postupu stačí pridať slová, ktoré tvorili slovník výnimiek, do trénovacej množiny pre slabičné delenie.

3.4.3 Samohláskové skupiny

Pri prepise samohláskových skupín sa vychádzalo z nasledujúcej skutočnosti:

§1 Pretože samohlásky sú nositeľmi slabičnosti, v slove je toľko slabík, koľko je v ňom samohlások. Ak sa bezprostredne vedľa seba vyskytnú dve samohlásky, patria do rozličných slabík a slabičná hranica prebieha medzi nimi. Ak by obidve samohlásky mali zostať v tej istej slabike, jedna z nich by musela stratiť slabičnosť napríklad zmenšením zvučnosti, skrátením trvania pod hranicu krátkych samohlások, alebo inou akustickou zmenou.

- §2 V koreni domácich slov sa v slovenčine samohláskové skupiny nevyskytujú. Môžu byť iba v predponových slovách (*naučiť, poistiť, preinačiť, vyučiť, zaistiť*), v zložených slovách (*sebaistý, sivoooký, veľkoobchod*) a v prevzatých slovách. Samohlásková skupina v domácom slove signalizuje jeho zloženosť (predponou, skladaním).
- §3 V procese zdomácnovania prevzatých slov sa v mnohých prípadoch prejavilo odstránenie samohláskovej skupiny. Oslabením vokalickej kvality prvej samohlásky môže vzniknúť zvukový komplex podobajúci sa na dvojhlásku. Tak vznikli dvojhlásky v slovách *fašiangy, hierarchia, hygiena, pacient, pieta, socializmus* a iné.
- §4 V niektorých slovách má zdomácnenie ešte väčší rozsah, lebo dvojhláska mäkkí aj predchádzajúce spoluhlásky *t, d, n, l*, ktoré sa inak v prevzatých slovách nemäkkia. Tak je to v slovách *Daniel, diabol, gavalier, golier, halier, inžinier, pionier* a iné.

Bližšie sa týmto problémom zaoberá časť 3.4.10.

Výslovnosť samohláskových skupín

- §5 Vzorom pre výslovnosť samohláskových skupín v cudzích a prevzatých slovách sú domáce slová ako *dozaista, naučiť, okoloídúci, poistka, poobliekať, poučiť, protiídúci, veľkoobchod* a podobne. V slovenčine vznikajú samohláskové skupiny aj na hranici slov, ktoré sa vyslovujú bez prestávky, ako napríklad v spojeniach *mama a otec, Jano i Eva, má u Adama*, ale i *o ústave* a podobne. V slovenčine sa v týchto spojeniach samohlásky nemenia tak, žeby strácali svoju slabikotvornú hodnotu, a medzi ne sa nevšúva nijaký pomocný zvuk. Samohlásky si zachovávajú plnú samohláskovú kvalitu a vyslovujú sa v rozličných slabikách. Tým sa líšia od výslovnosti dvojhlások.
- §6 V slovenskej spisovnej výslovnosti je prechod medzi samohláskami hladký a plynulý a znenie hlasu sa neprerušuje. V spisovnej slovenčine sa v samohláskových skupinách nevyslovuje ani ráz. Táto výslovnosť je charakterizovaná ako splyvavá slovenská výslovnosť.

Na základe vyššie uvedeného boli zavedené nasledujúce transkripčné pravidlá:

$$\mathbf{ai} \rightarrow \mathbf{aI} \quad (\text{doz} - \text{ai} - \text{sta}) \quad (3.29)$$

$$\mathbf{au} \rightarrow \mathbf{aU} \quad (n - \text{au} - \text{čiť}) \quad (3.30)$$

$$\mathbf{oi} \rightarrow \mathbf{OI} \quad (p - \text{oi} - \text{stka}) \quad (3.31)$$

$$\mathbf{oo} \rightarrow \mathbf{OO} \quad (p - \text{oo} - \text{bliekať}) \quad (3.32)$$

$$\mathbf{ou} \rightarrow \mathbf{OU} \quad (p - \text{ou} - \text{čiť}) \quad (3.33)$$

$$\mathbf{ii} \rightarrow \mathbf{II} \quad (\text{prot} - \text{ii} - \text{dúci}) \quad (3.34)$$

$$\mathbf{yu} \rightarrow \mathbf{IU} \quad (v - \text{yu} - \text{čiť}) \quad (3.35)$$

Rovnako bolo možné definovať ďalšie pravidlá pre samohláskové skupiny vyskytujúce sa v prevzatých slovách, ktoré neboli ošetrené doterajšími pravidlami. Skupiny dvoch samohlások:

$$\mathbf{ao} \rightarrow \mathbf{AO} \quad (\text{ch} - \text{ao} - \text{s}) \quad (3.36)$$

$$\mathbf{ea} \rightarrow \mathbf{Ea} \quad (\text{id} - \text{ea}) \quad (3.37)$$

$$\mathbf{eá} \rightarrow \mathbf{Ea} : (\text{oc} - \text{eá} - n) \quad (3.38)$$

$$\mathbf{éo} \rightarrow \mathbf{E : O} (\text{kliš} - \text{éo} - \text{vity}) \quad (3.39)$$

$$\mathbf{ié} \rightarrow \mathbf{IE} : (\text{prem} - \text{ié} - r) \quad (3.40)$$

$$\mathbf{io} \rightarrow \mathbf{IO} (\text{id} - \text{io} - t) \quad (3.41)$$

$$\mathbf{ió} \rightarrow \mathbf{IO} : (\text{mil} - \text{ió} - n) \quad (3.42)$$

$$\mathbf{oe} \rightarrow \mathbf{OE} (\text{kan} - \text{oe}) \quad (3.43)$$

$$\mathbf{oi} \rightarrow \mathbf{OI} (\text{paran} - \text{oi} - k) \quad (3.44)$$

$$\mathbf{oo} \rightarrow \mathbf{OO} (\text{z} - \text{oo} - \text{lógia}) \quad (3.45)$$

$$\mathbf{uu} \rightarrow \mathbf{UU} (\text{kontin} - \text{uu} - m) \quad (3.46)$$

Skupiny troch samohlások:

$$\mathbf{eae} \rightarrow \mathbf{EaE} (r - \text{eae} - \text{rácia}) \quad (3.47)$$

$$\mathbf{eeu} \rightarrow \mathbf{EEU} (n - \text{eeu} - \text{klidovský}) \quad (3.48)$$

$$\mathbf{eoa} \rightarrow \mathbf{EOa} (\text{ost} - \text{eoa} - \text{rtróza}) \quad (3.49)$$

$$\mathbf{eoi} \rightarrow \mathbf{EOI} (\text{ost} - \text{eoi} - t) \quad (3.50)$$

$$\mathbf{oio} \rightarrow \mathbf{OIO} (\text{fot} - \text{oio} - \text{nizácia}) \quad (3.51)$$

Samohláskové skupiny na konci slova

§7 V prevzatých slovách sa skupina samohlások na konci ohybného slova nemení na dvojhlásku. Príkladom je prevzaté slovo *hierarchia*, ktoré ma v koreni dvojhlásku *ie* no koncová samohlásková skupina sa nediftongizuje. Spisovná výslovnosť je [hI[˘]ErarxIa]. Nediftongická výslovnosť sa zachováva aj v slovách *beštia*, *ceremónia*, *lalia*, *mizéria*, *pasia*, *patália*, *porcia* a pod.

V prípade, že slovo patrí do skupiny s nediftongickou výslovnosťou na konci, aplikujú sa naň pravidlá:

$$\mathbf{ia} \parallel \rightarrow \mathbf{Ia} (\text{bešt} - \text{ia}) \quad (3.52)$$

$$\mathbf{ie} \parallel \rightarrow \mathbf{IE} (\text{bešt} - \text{ie}) \quad (3.53)$$

$$\mathbf{iu} \parallel \rightarrow \mathbf{IU} (\text{bešt} - \text{iu}) \quad (3.54)$$

v opačnom prípade

$$\mathbf{ia} \parallel \rightarrow \mathbf{I}^{\wedge} \mathbf{a} (\text{cudz} - \text{ia}) \quad (3.55)$$

$$\mathbf{ie} \parallel \rightarrow \mathbf{I}^{\wedge} \mathbf{E} (\text{cudz} - \text{ie}) \quad (3.56)$$

$$\mathbf{iu} \parallel \rightarrow \mathbf{I}^{\wedge} \mathbf{U} (\text{cudz} - \text{iu}) \quad (3.57)$$

Znakom \parallel sú označované hranice slov.

Na určenie diftongickej a nediftongickej výslovnosti sú k dispozícii dva postupy. Buď definovať slovník na určenie toho, kedy ide o samohláskovú postupnosť a kedy o dvojhlásku, alebo podobne, ako bolo uvedené na konci časti 3.4.2, použiť metódu slabičnej hranice.

Skupiny [aU_{-}^{\wedge} , OU_{-}^{\wedge} , EU_{-}^{\wedge} , IU_{-}^{\wedge} , aI_{-}^{\wedge} , EI_{-}^{\wedge} , OI_{-}^{\wedge} , UI_{-}^{\wedge}]

- §8 Samohláskové skupiny so samohláskou u , i na druhom mieste sa vo výslovnosti častejšie menia na skupiny *samohláska* + U_{-}^{\wedge} , I_{-}^{\wedge} , pričom sa tieto skupiny skôr menia v prvej a druhej slabike slova. Napríklad slová *auto*, *Kaukaz*, *oidipovský*, *oidium*, *pozauna* vyslovujeme ako [aU₋[∧]tO, kaU₋[∧]kas, OI₋[∧]dIpOU₋[∧]skI:, OI₋[∧]dIU_m, pOzaU₋[∧]na].
- §9 S hláskou U_{-}^{\wedge} , resp. I_{-}^{\wedge} sa vyslovujú aj mnohé cudzie a prevzaté slová, v ktorých sa po samohláske píše w , v , j : *clown*, *cowboy*, *fajn*, *ofsajd*, *pilav*, *policajt*. Vyslovujú sa ako [klaU₋[∧]n, kaU₋[∧]bOI₋[∧]/kOU₋[∧]bOI₋[∧], ofsaI₋[∧]t, faI₋[∧]n, pIlaU₋[∧], pOIItsaI₋[∧]t].
- §10 Pred zvučnými spoluhláskami l , r najmä však pred l sa vyskytuje dvojaká výslovnosť: detail [detail/detaI₋[∧]], email [Email/EmaI₋[∧]], faul [faUI/faU₋[∧]l], joule [dZaUI/dZaU₋[∧]l].
Obidva spôsoby výslovnosti sú považované v slovenčine za rovnocenné, no prvá výslovnosť sa vyskytuje zriedkavejšie.
- §11 V odvodených slovách, v ktorých spoluhlásku l oddeľuje od pôvodnej samohláskovej skupiny slabičná hranica, ustálila sa výslovnosť s hláskami [I_{-}^{\wedge} , U_{-}^{\wedge}] [EmaI₋[∧]IOvac, faU₋[∧]IOvac].

3.4.4 Tvrдый hlasový začiatok a ráz

- §12 V slovenčine sa tvrdý hlasový začiatok a ráz vyskytujú pomerne zriedka. Tvrдый hlasový začiatok a ráz sa vôbec nevyskytujú vnútri slov a na hranici morfému:
[naUtSIc, prEInatSIc, zaOrac],
vnútri rečového taktu na hraniciach slov:
[prEdOktO:brOm, naU:stax, UzUs, zalhlOU₋[∧]],
na hraniciach slovných základov zložených slov:
[ElEktrOInZIJI₋[∧]Er, mnOhOaspEktOvI:, pravOUhII:, sIvOOkI:, vElkOOpxOt].
- §13 Tvrдый hlasový začiatok a ráz môžu byť na začiatku fonácie po prestávke, resp. na začiatku vety, keď sa prvé slovo začína samohláskou:
[?IstI: bOhatI: kUpEc; ?a znOva prISla jEsEJ],
no správna je aj výslovnosť [IstI: bOhatI: kUpEc; a znOva prISla jEsEJ].
- §14 Tvrдый hlasový začiatok a ráz sa častejšie vyskytujú v citovo zafarbených výpovediach, v citovo zafarbených slovách, v citoslovciach, vo zvolaniach a zvukomalebných slovách:
[?OtstU:p s tsEsti; ?EI, al'E sI vIvI₋[∧]EdOl; ?EI₋[∧], ?ax, ?aLE; ?I?a:, mE?E?E:].
Ak sa samohláska vysloví s rázom a predchádzajúce slovo sa končí znelou párovou spoluhláskou (pričom medzi slovami nie je prestávka), asimiluje sa znelá hláska na neznelú. Tu sa tvrdý hlasový začiatok a ráz správajú ako neznelé spoluhlásky. Vtedy sa napríklad slovné spojenie *vlak už odišiel* chybne vysloví [vlak ?US ?OdISI₋[∧]El]. V splývajúcej výslovnosti sa však vyslovuje [vlag UZ OdISI₋[∧]El].
- §15 Pri vyjadrení ostrého nesúhlasu a údivu sa môže ráz vyskytnúť aj na konci slova pred prestávkou:
[JI₋[∧]E?]!, [a:nO?]?, [ba?]!.

Vzhľadom na to, že pre výslovnosť rázu neexistujú presné pravidlá a jeho výskyt je závislý na hovoriacom, ako vyplýva z vyššie uvedeného, neboli preň vytvorené žiadne transkripčné pravidlá.

Ďalším dôvodom na nevytvorenie transkripčných pravidiel je skutočnosť, že nie je použitý v systéme navrhnutých segmentačných jednotiek, resp. v použitom fonetickom repertoári slovenčiny. V systémoch na rozpoznávanie slovenčiny sa ráz nezvykne používať. Inak je to napríklad v nemčine či češtine.

3.4.5 Samohláska *ä*

- §16 V súvislosti so štýlovou diferenciáciou spisovnej slovenčiny sa výslovnosť *ä* (päta, mäso) začína pociťovať ako príznaková – ako znak vyššieho štýlu výslovnosti. V neutrálnom štýle spisovnej výslovnosti sa dnes pokladajú za základné tvary [mEkI:, mEsO, naL_ˆmE] atď.
- §17 Slová *bábä, dúpä, holúbä, chlápä, púpä, sôvä, žriebä* a od nich odvodené slová ako *žriebäcí, žriebätko* a podobne sa v neutrálnom štýle vyslovujú so samohláskou *a* alebo *E*. [ba:ba / ba:bE, dU:pa / dU:pE, hOIU:ba / hOIU:bE, xla:pa / xla:pE, pU:pa / pU:pE, sU_ˆOva / sU_ˆOvE, ZrL_ˆEba / ZrL_ˆEbE].

Z vyššie uvedeného vyplýva takmer úplná jednoznačnosť prepisu samohlásky *ä*. V siedmich slovách, v ktorých je možná dvojaká výslovnosť, bola kvôli zjednodušeniu zvolená výslovnosť *s e*.

$$\ddot{a} \rightarrow \mathbf{E}(m - \ddot{a} - so) \quad (3.58)$$

Z pohľadu počítačového rozpoznávania reči sa môže vyššie uvedené zjednodušenie javiť ako nutnosť v prípade, že výskyt fóny *ä* bude v trénovacej množine len v malom počte prípadov.

V sade pravidiel pre alternatívnu výslovnosť je obsiahnutý aj prepis *ä* vo vyššom štýle výslovnosti:

$$\ddot{a} \rightarrow \{(m - \ddot{a} - so)\} \quad (3.59)$$

3.4.6 Samohlásky *ö, ø, ü, ů*

- §18 Samohlásky *ö, ø, ü, ů* sa v slovenčine vyskytujú iba v prevzatých slovách a vyslovujú sa aj v cudzích slovách. Vyslovujú sa napríklad v slovách: *angström, budžet, etuda, fleuret, föhn, jury, menu, nuansa, oersted, revue, resumé, sujet, surrealismus*.
- §19 Samohlásky *ö, ø, ü, ů* sa vyslovujú aj v cudzích menách a názvoch, napr. vo francúzskych menách: *Bruzelles, Duclos, Dumas, Hugo, Montreux*, v nemeckých a švédskych a iných menách a názvoch: *Böhm, Dürer, Göteborg, Kitzbühel, Münster, Utrecht, Zürich*.
- §20 Zdomácnovanie cudzích slov so samohláskami *ö, ø, ü, ů* sa prejavuje aj tak, že sa nahrádzajú najbližšími príbuznými slovenskými samohláskami. Môžu to byť samohlásky *i, u, e, o*. Napríklad slovo *parfum* sa v neutrálnom a vyššom štýle spisovnej výslovnosti vyslovuje [parfYm, parf2:m], v neutrálnom štýle sa pripúšťa aj výslovnosť [parfUm, parfEm].

Pre cudzie samohlásky boli v prípade, že sú v spisovnej slovenčine nahradené najbližšou príbuznou slovenskou samohláskou, prirodzene použité rovnaké prepisovacie pravidlá ako pre slovenské samohlásky. V prípade, že sa v cudzích slovách vyskytujú originálne samohlásky, je im priradený príslušný znak z fonetickej abecedy napriek tomu, že nepatrí do repertoára slovenských fón.

$$\ddot{o} \rightarrow \mathbf{2} \quad (G - \ddot{o} - teborg) \quad (3.60)$$

$$\ddot{u} \rightarrow \mathbf{Y} \quad (Z - \ddot{u} - rich) \quad (3.61)$$

$$\mathbf{\check{O}} \rightarrow \mathbf{2} : \quad (3.62)$$

$$\mathbf{\check{U}} \rightarrow \mathbf{Y} : \quad (3.63)$$

Tieto pravidlá však nie su spoľahlivé, pretože v rôznych jazykoch prislúchajú rovnakým grafémam rôzne fóny. Konkrétne v tomto prípade pre grafému *ö* je v švédskej SAMPA abecede iný prepis, než pre tú istú grafému v maďarčine. Preto sú pravidlá (3.60) až (3.63) uvedené skôr ako príklad. V takomto prípade je vhodnejšie použiť pre prepis najbližšiu slovenskú fónu:

$$\ddot{o} \rightarrow \mathbf{E} \quad (G - \ddot{o} - teborg) \quad (3.64)$$

$$\ddot{u} \rightarrow \mathbf{U} \quad (Z - \ddot{u} - rich) \quad (3.65)$$

3.4.7 Príčiny, podmienky a miesta spoluhláskových zmien

- §21 Vo výslovnosti spoluhlások a spoluhláskových skupín je pri transkripcii slovenčiny omnoho viac problémov, než pri samohláskach a samohláskových skupinách. Súvisí to s väčším počtom spoluhlások, no najmä s hláskovou stavbou rytmických a významových jednotiek reči v spojitosti s fonologickou štruktúrou slov, morfém a slabík.
- §22 Najmenšia jednotka, ktorú vieme vysloviť samostatne, je slabika. Ako signálová jednotka je slabika určená princípom sonórnosti, ako funkčná jednotka princípom kontrastu.

Slabika vzniká podľa princípu kontrastu medzi samohláskou a spoluhláskou. Je to kontrast súboru vlastností samohlásky oproti súboru vlastností spoluhlásky. Zabezpečuje sa ním ľahké a rýchle sluchové rozlišovanie a poznávanie samohláskovej a spoluhláskovej zložky slabiky. Základom slabiky je samohláska, lebo samohláska je najzvučnejšia (najsonórnejšia), má najväčšiu zvukovú plnosť.

Pretože samohlásky sú sonórnejšie ako spoluhlásky a kontrast vzniká prítomnosťou samohlásky a spoluhlásky, jadro slabiky býva obklopené menej sonórnymi zvukmi, takže v slove je toľko slabík, koľko je v ňom vrcholov sonórnosti. Preto sonórny svah nemôže byť v hraniciach jednej slabiky prerušený iným vrcholom sonórnosti. V slabike ako sonórnej jednotke môže byť [v] + [a] [vaJa], [tS] + [a] [tSaSa], [v] + [E] [vEnO], [tS] + [E] [tSElO] atď., no nemôže byť [vtSEra, vpI:sac, vzcax, vscax], ale iba [ftSEra, fpi:sac, fscax], lebo *č*, *p*, *z*, *s*, *t* sú hlásky menej zvučné ako [v] (resp. vôbec nie sú zvučné). Sonórny princíp stavby slabiky umožňuje aj výslovnosť [kraI_ˆ, faI_ˆnovI:, staU_ˆba, aU_ˆto] a vynútil si aj výslovnosť [maI_ˆzena, mEdaI_ˆla, EmaI_ˆlOvI:, EmaII, kOzmOnaUt].

Slabika je v spisovnej slovenčine jedno z miest úpravy a zmien (modifikácie) hlások v súvislej reči.

- §23 Druhým a ešte dôležitejším miestom modifikácie hlások je v slovenčine hranica morfémy, čiže aj hranica slov. Tu sa pri odvodzovaní, ohýbaní a pri skladaní a spájaní slov môžu stretnúť aj také hlásky (resp. fonémy), ktoré sa nevyslovujú vnútri slabiky a morfémy (tak je to v slovách *vzťah, prosba, prosme, prudší, egyptský, drozd príletel*), a môže vzniknúť aj taká skupina spoluhlások, ktorá nám spôsobuje artikulačné ťažkosti, ako je napríklad v slovách *odstúpiť, ľudskosť, mestský*. Preto je morfematická hranica miestom mnohých spoluhláskových zmien, najmä takzvaných asimilačných zmien, čiže asimilácií.
- §24 Vyskytujú sa aj také spoluhláskové skupiny, ktoré z dnešného hľadiska patria do tej istej slabiky a do tej istej morfémy, neprotirečia princípom ich stavby a spájania, a napriek tomu v nich vznikajú rozličné úpravy hlások. Takéto skupiny vznikli v procese historického vývinu jazyka zánikom, zámenou alebo zmenou niektorých hlások, preberaním a prispôbovaním cudzích slov a prejavujú sa v nich niektoré všeobecné zákonitosti artikulácie. V nich sa spoluhlásky oproti iným postaveniam rozličným spôsobom artikulačne zjednodušujú alebo upravujú. Tak sa napríklad modifikuje artikulácia *d* v slovách *dno, padla*, spoluhláska *l* v slovách *poltón, holdovať*, spoluhláska *m* v slove *lampa*.

Ďalej budú podrobne rozobraté jednotlivé ortoepické javy a budú pre ne definované skupiny pravidiel.

Hlásky [j, I[^]]

- §25 V slovenčine sa znakom *j* zaznačujú dva zvuky: tvrdopodnebná úžinová trená spoluhláska [j], ako v slovách *dvaja, jeden, žijú, prijať, spojiť*, a tvrdopodnebná úžinová nešumová hláska [I[^]], ako v slovách *kraj, dajte*. O ich výskyte v spisovnej slovenčine platia tieto pravidlá:

Hláska [j]

- §26 Na začiatku slabiky pred samohláskou, kde píšeme *j*, vyslovuje sa šumová spoluhláska [j] bez ohľadu na to, či stojí na absolútnom začiatku slova, alebo nie. V spisovnej slovenčine sa vyslovuje [j] v takých pozíciách, ako je v slovách *jama, jasný, jeden, juh; hocijako, hocijaký; bojisko, hokejista, najímať, pomyje, prijateľný, prijímať; moji, tvoji, svoji* atď. Vyslovuje sa

[jama, jasnI:, jEdEn, jUx; hOtsIjakO, hOtsIjakI:; bOjIskO, hOkEjIsta, naji:mac, pOmIjE, prIjacELnI:, prIjI:mac; mOjI, tvOjI, svOjI] atď.

- §27 Takto sa vyslovuje *j* aj vtedy, keď pred ním predchádza spoluhláska a medzi ním a predchádzajúcou spoluhláskou je slabičná hranica (*j* je vtedy prvou hláskou slabiky). Vyslovuje sa teda aj v slovách *adjektívum, benjamín, injekcia, injikovať, konjugácia, nadjazd, objasniť, objať, objav, objednať, objem, odjakživa, odjedat', podjazd, predjariet, rozjačať sa, rozjasniť sa, rozjímať, subjekt*. Vyslovuje sa

[adjEktI:vUm, bEnjamI:n, InjEktsIa, InjIkOvac, kOnjUga:tsIa, nadjast, ObjasJIc, Objac, Objau[^], ObjEdnac, ObjEm, OdjagZIva, OdjEdac, pOdjast, prEdjarI[^]E, rOzjatSacs, rOzjasJIcsa, rOzjI:mac, UbjEkt].

Toto pravidlo platí aj v medzislovnej fonetike, ako napríklad v spojeniach *veľká jama*, *poď jesť*, *on je doma*, *dosť jasný* atď. Vyslovuje sa

[vELka:jama, pOJ\jEsc, OnjEdOma, dOzJ\jasnI:].

Hlásoka [I[˘]]

§28 Na konci slabiky po samohláske sa na mieste pravopisného znaku *j* vyslovuje sonórna (nešumová) hlásoka [I[˘]] bez ohľadu na to, či nasleduje spoluhlásoka, samohlásoka alebo *j*. Slová ako *dvojaktovka*, *dvojazyčný*, *kraj*, *krajší*, *kroj*, *vylej*, *vylejte* vyslovujú sa

[dvOI[˘]aktOU[˘]ka, dvOI[˘]jazItSnI:, kraI[˘], kraI[˘]SI:, krOI[˘], vILEI[˘], vILEI[˘]cE].

Toto pravidlo platí aj v medzislovnej fonetike bez ohľadu na to, či je *j* na konci slova pred prestávkou, alebo či po ňom nie je prestávka, a bez ohľadu na to, či sa nasledujúce slovo začína samohláskou, alebo spoluhláskou. Výrazy a spojenia ako *daj mi*, *daj jej*, *roj ôs*, *tvoj Ivan*, *Stoj! To je pekný kraj* vyslovujú sa

[daI[˘]mI, daI[˘]jEI[˘], rOI[˘]U[˘]Os, tvOI[˘]Ivan, stOI[˘], tOjEpEknI:kraI[˘]].

§29 Výslovnosť [j] aj [I[˘]] sa pripúšťa v dvoch prípadoch:

- Keď znak *j* stojí v písme na konci slabiky a pred ním prichádza znak *i*, *y*, vyslovuje sa v spisovnej slovenčine [j] alebo [I[˘]]. V slovách *Ázijčan*, *kolokvijný*, *kyj*, *ľbyjský*, *medvedobijca*, *patricij*, *režijný* atď. vyslovuje sa [j] alebo [I[˘]]:

[a:zIjtSan, kOIOkvIjnI:, kIj, II:bIjskI:, mEdvEJ\ObIjtsa, patrItsIj, rEZIjnI:]

alebo

[a:zII[˘]tSan, kOIOkvII[˘]nI:, kII[˘], II:bII[˘]skI:, mEdvEJ\ObII[˘]tSa, patrItsII[˘], rEZII[˘]nI:].

Výslovnosť so spoluhláskou [j] je častejšia vo vyššom štýle spisovnej výslovnosti. Slová *kyjačikovitý*, *kyjanica* sa spisovne vyslovujú [kIjatSIkovItI:, kIjanItsja], lebo tu stojí *j* na začiatku slabiky. To isté platí pre slová *prijímať*, *prijímací*, *prijímač*, *prijímanie*.

- Keď znak *j* stojí v začiatkovej časti slabiky, ale ešte v tej istej slabike pred ním je iná spoluhlásoka, vyslovuje sa [j] alebo [I[˘]]. Slová *zjazd*, *zjazdový*, *zjav*, *zjaviť*, *zjednať*, *zjesť*, *zjedol*, *zjedla* atď. vyslovujú sa

[zjast, zjazdOvI:, zjaU[˘], zjavIc, zjEdnac, zjEsc, zjEdol, zjEdla]

alebo

[zI[˘]ast, zI[˘]azdOvI:, zI[˘]aU[˘], zI[˘]avIc, zI[˘]Ednac, zI[˘]Esc, zI[˘]Edol, zI[˘]Edla].

Tak sa vyslovujú aj cudzie slová *fjEld*, *fjOrd*. Vyslovuje sa

[fjElt, fjOrt] alebo [fI[˘]Elt, fI[˘]Ort].

Na základe vyššie uvedeného boli definované nasledujúce transkripčné pravidlá:

- Pre výslovnosť "j":

$$| \mathbf{j}([aEIOUa : E : I : O : U :]) \rightarrow \mathbf{j} \setminus 1(j - ama) \quad (3.66)$$

$$\hat{\mathbf{j}}([aEIOUa : E : I : O : U :]) \rightarrow \mathbf{j} \setminus 1(dostj - asný) \quad (3.67)$$

- Pre výslovnosť "I[^]":

$$([aEOUa : E : O : U :])\mathbf{j} | \rightarrow \setminus 1\mathbf{I}^{\wedge}(dvo - j - aktovka) \quad (3.68)$$

$$([aEOUa : E : O : U :])\mathbf{j} || \rightarrow \setminus 1\mathbf{I}^{\wedge}(da - jje - j) \quad (3.69)$$

- Pre výslovnosť "j" aj "I[^]":

$$([I])\mathbf{j} | \rightarrow \setminus 1\mathbf{j} \quad (medvedobi - j - ca) \quad (3.70)$$

$$\rightarrow \setminus 1\mathbf{I}^{\wedge} \quad (medvedobi - j - ca) \quad (3.71)$$

$$([I])\mathbf{j} || \rightarrow \setminus 1\mathbf{j} \quad (ky - j) \quad (3.72)$$

$$\rightarrow \setminus 1\mathbf{I}^{\wedge} \quad (ky - j) \quad (3.73)$$

$$| ([SP])\mathbf{j}([aEIOU]) \rightarrow \setminus 1\mathbf{j} \setminus 2 \quad (predz - j - azd) \quad (3.74)$$

$$\rightarrow \setminus 1\mathbf{I}^{\wedge} \setminus 2(predz - j - azd) \quad (3.75)$$

$$|| ([SP])\mathbf{j}([aEIOU]) \rightarrow \setminus 1\mathbf{j} \setminus 2 \quad (z - j - azd) \quad (3.76)$$

$$\rightarrow \setminus 1\mathbf{I}^{\wedge} \setminus 2(z - j - azd) \quad (3.77)$$

kde "|" označuje hranicu slabiky a "SP" spoluhlásku. Delením slov na slabiky sa podrobnejšie zaoberá časť 3.5. Z dôvodu väčšej jasnosti a čitateľnosti zápisu sú na zápis transkripčných pravidiel použité regulárne výrazy. Ďalej sa predpokladá, že slabičná hranica je pri aplikovaní navrhnutých pravidiel určená podľa postupov opísaných v časti 3.5.

Pravidlá pre výslovnosť "j" aj "I[^]" je možné obísť a pre všetky prípady, ktoré nepatria do prvej ani do druhej skupiny, bude aplikovaná obojaká výslovnosť. Tento spôsob prepisu sa ukazuje užitočný v prípadoch, kde voľba správnej výslovnosti vyžaduje hlbšie jazykové znalosti.

§30 Ako už bolo spomenuté v časti 3.4.3 §3 o výslovnosti samohláskových skupín, hláska [I[^]] sa vyslovuje aj v cudzích a prevzatých slovách *eis*, *maizena*, *medaila*, *oidipovský* apod. Vyslovuje sa aj v slovách *detail*, *email*, *koktail*.

§31 Slová *anjel*, *evanjeliár*, *evanjelický*, *evanjelik*, *evanjelium* sa vyslovujú

$$[a\mathbf{J}\mathbf{I}^{\wedge}\mathbf{E}\mathbf{l}, \text{Eva}\mathbf{J}\mathbf{I}^{\wedge}\mathbf{E}\mathbf{l}\mathbf{I}\mathbf{a}:r, \text{Eva}\mathbf{J}\mathbf{I}^{\wedge}\mathbf{E}\mathbf{l}\mathbf{I}\mathbf{c}\mathbf{k}\mathbf{l}:, \text{Eva}\mathbf{J}\mathbf{I}^{\wedge}\mathbf{E}\mathbf{l}\mathbf{I}\mathbf{k}, \text{Eva}\mathbf{J}\mathbf{I}^{\wedge}\mathbf{E}\mathbf{l}\mathbf{I}\mathbf{U}\mathbf{m}].$$

Tu sa skupina *je* zmenila na dvojhlásku, pred ktorou sa mäkkí aj predchádzajúca spoluhláska. Tieto slová boli zaradené medzi výnimky, ktoré sa testujú pred aplikovaním ostatných transkripčných pravidiel. Spôsobom ošetrovania výnimiek sa podrobnejšie venuje časť 3.4.14.

Hlásky [v, U[^]]

Hláska [v]

§32 Sonórne sa vyslovuje na začiatku slabiky (pred vrcholom slabiky), kde sa píše *v*, a to pred samohláskou, pred slabičným [r=, l=, r=:, l=:] a pred [r, l, L, j].

Slová a spojenia ako *adverbium*, *bezvetrie*, *dvere*, *obúvať sa*, *slovo*, *svadba*, *sviatok*, *uvažovať*, *v okamihu*, *v oku*, *veje vetrík*, *voda*, *vústiť*, *zväzok*, *zver*; *vrch*, *vrba*, *vlk*, *vĺča*; *vlak*, *vliecť*, *vlek*, *vrana*, *v jame*, *v lete*, *v lodi*, *v roku* vyslovujeme

[advErbiUm, bEzvEtrI[^]E, dvErE, ObU:vacsA, slOvO, svadba, svL[^]atOk, UvaZOvac, vOkamIhU, vOkU, vEjEvEtrI:k, vOda, vU:scIc, zvEzOk, zvEr; vr=x, vr=:ba, vl=k, vl=:tSa; vlak, vLL[^]Etsc, vLEk, vrana, vjamE, vLEcE, vIOJ\I, vrOkU].

Hláska [U[^]]

§33 Sonórne [U[^]] sa vyslovuje na konci slabiky po jej vrchole, kde sa píše *v*, po slabičnej hláske (po samohláske a po slabičnom [r=, l=, r=:, l=:]), ako v slovách *bratov*, *dav*, *dievča*, *Janov klobúk*, *kov*, *lev*, *pravda*, *zastav sa*, *zo stromov krv*, *obrv*, *prv* a pod. Vyslovuje sa

[bratOU[^], daU[^], J\I[^]EU[^]tSa, janOU[^]klobU:k, kOU[^], LEU[^], praU[^]da, zastaU[^]sa, zOstrOmOU[^]; kr=U[^], Obr=:U[^], pr=U[^]].

Na základe vyššie uvedeného boli definované nasledujúce transkripčné pravidlá:

- Pre výslovnosť "v":

$$| \mathbf{v}([SM]) \rightarrow \mathbf{v}\backslash 1 \quad (slo - v - o) \quad (3.78)$$

$$| \mathbf{v}([r = l = r =: l =: rlLj]) \rightarrow \mathbf{v}\backslash 1 \quad (v - \acute{r}ba) \quad (3.79)$$

$$| ([SP])\mathbf{v}([SM]) \rightarrow \backslash 1\mathbf{v}\backslash 2 (bez - v - etrie) \quad (3.80)$$

$$| ([SP])\mathbf{v}([r = l = r =: l =: rlLj]) \rightarrow \backslash 1\mathbf{v}\backslash 2 (pos - v - rbi) \quad (3.81)$$

$$\| \mathbf{v}([SM]) \rightarrow \mathbf{v}\backslash 1 \quad (v - oda) \quad (3.82)$$

$$\| \mathbf{v}([r = l = r =: l =: rlLj]) \rightarrow \mathbf{v}\backslash 1 \quad (v - \acute{I}ča) \quad (3.83)$$

$$\| ([SP])\mathbf{v}([SM]) \rightarrow \backslash 1\mathbf{v}\backslash 2 (d - v - ere) \quad (3.84)$$

$$\| ([SP])\mathbf{v}([r = l = r =: l =: rlLj]) \rightarrow \backslash 1\mathbf{v}\backslash 2 (s - v - rbi) \quad (3.85)$$

- Pre výslovnosť "U[^]":

$$([SM])\mathbf{v} | \rightarrow \backslash 1\mathbf{U}^{\wedge} (die - v - \acute{c}a) \quad (3.86)$$

$$([r = l = r =: l =:])\mathbf{v} | \rightarrow \backslash 1\mathbf{U}^{\wedge} (kr - v - n\acute{e}) \quad (3.87)$$

$$([SM])\mathbf{v} \| \rightarrow \backslash 1\mathbf{U}^{\wedge} (chlie - v) \quad (3.88)$$

$$([r = l = r =: l =:])\mathbf{v} \| \rightarrow \backslash 1\mathbf{U}^{\wedge} (pr - v) \quad (3.89)$$

SM označuje kvôli prehľadnosti samohlásku.

§34 Hláska [U[^]] sa po slabičnej hláske vyslovuje aj v niektorých prevzatých slovách v samohláskových skupinách, kde sa píše *u*, ako v slovách *auto*, *automat*, *automatický*, *kaučuk*, *Kaukaz* atď. Vyslovuje sa

[aU[^]tO, aU[^]tOmatItskI:, kaU[^]tSuk, kaU[^]kas] atď.

§35 Zriedkavo sa [U[^]] vyslovuje na mieste, kde je v písme *u*, *w* aj v iných prevzatých slovách. Napríklad pri jednoslabičnej výslovnosti je v slovách *crawl*, *faul*, *joul*. Vtedy sa vyslovuje

[kraU[^]l, faU[^]l, dZaU[^]l].

Slová *auto*, *crawl*, *faul*, *joul* boli pridané medzi výnimky. Slovo *auto* v prípade vhodne navrhnutého algoritmu pre ošetrovanie výnimiek správne prepisuje aj sledovanú postupnosť grafém v slovách *automat*, *automatický*, poprípade v ďalších slovách odvodených od týchto slov. Podrobnosti sú uvedené v časti 3.4.14.

§36 V domácich slovách sa [U[^]] vyslovuje v siedmom páde jednotného čísla podstatných a prídavných mien ženského rodu, kde sa píše *u* (ženou, slávnou, novou ulicou), v slovách *náukobeh*, *náhodou* a v slovách *mraučať/mravčať*, *mrauk/mravk*. Vyslovuje sa

[ZEnOU[^], sla:vnOU[^]/sla:U[^]nOU[^], nOvOU[^]ULItsOU[^]; na:U[^]kObEx, na:hO-dOU[^], mraU[^]tSac, mraU[^]k].

Drvivá väčšina slov, ktoré sa v slovenčine končia na *ou*, sú podstatné a prídavné mená ženského rodu. Okrem nich sa končia na *ou* niektoré i príslovky, ako napríklad *väčšinou*, avšak z ortoepického hľadiska ide o tú istú výslovnosť ako pri podstatnom mene *väčšina* v siedmom páde jednotného čísla. Z tohto dôvodu bolo pre výslovnosť hlásky [U[^]] definované aj nasledujúce transkripčné pravidlo:

$$OU \parallel \rightarrow OU[^] (\textit{ženo} - u) \quad (3.90)$$

Slová *náukobeh*, *náhodou*, *mraučať/mravčať*, *mrauk/mravk* boli pridané medzi výnimky. Jediné nájdené slovo s *ou* na konci, ktoré nepatrí do vyššie spomínanej skupiny, je podľa [42] cudzie slovo *manitou*, ktoré sa vyslovuje ako [manItU:] a ktoré bolo zaradené medzi výnimky.

§37 Ak však *u* stojí na začiatku slabiky, nevyslovuje sa [U[^]], lež [u]. Slová *náuka*, *naučiť*, *ponaučenie*, *vyučiť*, *zaucho* vyslovujú sa

[na:Uka, naUtSIc, pOnaUtSEJL[^]E, vIUtSIc, zaUxO].

Keďže je transkripčné pravidlo (3.90) aplikované až v tejto časti, je na ľavej strane pravidla už SAMPA kódovanie, keďže pri prepise samohlások boli na *o* a *u* aplikované samohláskové pravidlá.

§38 Výslovnosť [v] aj [U[^]] sa pripúšťa:

- ak v tej istej slabike pred *v* predchádza *u*, ako v slovách *cúvnuť*, *čuv*, *náčuv*, *nemluviňa*, *obuv*, *obuvník*, *počúvnuť*, *posuv*, *posuvka*, *prezwka*, *Vezuv*;

- na konci slabiky v koreni slova a pred príponou, ak nasleduje *n, ň, l, ľ, r*, ako v slovách *krvný, pavlač, rovnako, rovnák, slávny, slovník, šanovlivý, škovrán, zarovnať, zívla*;
- na konci slabiky a slova, keď predchádza neslabičné *r, l*, ako v slovách *červ, konzerv, nerv, sálv*.

Vyslovuje sa

- [tsU:U^ˆnUc, tSUU^ˆ, na:tsUU^ˆ, JEmlUU^ˆJa, ObUU^ˆ, ObUU^ˆJl:k, pOtSU:U^ˆnUc, pOsUU^ˆ, pOsUU^ˆka, prEzUU^ˆka, vEzUU^ˆ];
- [kr=U^ˆnI:, paU^ˆlatS, rOU^ˆnakO, rOU^ˆJl:k, sla:U^ˆnI, slOU^ˆJl:k, SanOU^ˆLlIvI:, SkOU^ˆra:n, zarOU^ˆnac, zI:U^ˆla];
- [tSErU^ˆ, kOnzErU^ˆ, nErU^ˆ, sa:lU^ˆ];

alebo

- [tsU:vnUc, tSUv, na:tsUv, JEmlUvJa, ObUv, ObUvJl:k, pOtSU:vnUc, pOsUv, pOsUvka, prEzUvka, vEzUv];
- [kr=vnI:, pavlatS, rOvnakO, rOvJl:k, sla:vnI, slOvJl:k, SanOvLlIvI:, SkOvra:n, zarOvnac, zI:vla];
- [tSErv, kOnzErv, nErv, sa:lv].

Ak medzi *u* a *v* prechádza slabičná hranica, takže *v* je na začiatku druhej slabiky, vyslovuje sa aj v tejto skupine iba [v]:

[natSU:vac, nadU:vac, ObU:vac, pOtSU:vac, sU:vIsI^ˆEc, UvaZOvac, UvEdnU:c, UvE-rIc, UvIJ\I^ˆEc] atd.

Pre tento prípad nebolo potrebné definovať žiadne transkripčné pravidlo, keďže táto skutočnosť vyplýva z (3.80) a (3.81).

§39 Keď pred *v* predchádza samohláska a po *v* nasleduje znelá (nie však zvučná) spoluhláska a slabikovanie slova nie je ustálené, vyslovuje sa alebo [U^ˆ] alebo [f_v]. Slovo *závdavok* sa teda vyslovuje podľa spôsobu slabikovania buď [za:U^ˆdavOk] (slabičná hranica je pred *d*), alebo [za:f_vdavOk] (slabičná hranica je po *á* pred *v*).

Na začiatku slabiky pred spoluhláskou, kde sa píše *v*, môže sa v určitých situáciách vyslovovať aj šumové [f_v].

Slová, v ktorých existuje možnosť výslovnosti [f_v], boli pridané medzi výnimky a boli k nim uvedené obidva spôsoby výslovnosti.

§40 Na začiatku časti slabiky (v tzv. presonantickom postavení) pred samohláskou *o* v spisovnej slovenčine nikdy nevyslovujeme v domácich slovách [U^ˆ]. Vtedy by sa totiž skupina *vo* zmenila na dvojhlásku [U^ˆO]. Slová ako *dvojka, dvojmo, dvor, dvořit, kvočka, kvokať, obvod, potvora, stvora, stvorit, svoj, svojho, svokor, svokra, tvoj, tvojho, tvoji, tvojich* atď. vyslovujeme

[dvOI^ˆka, dvOI^ˆmO, dvOr, dvOrIc, kvOtSka, kvOkac, ObvOt, pOtvOra, stvOra, svOrIc, svOI^ˆ, svOI^ˆhO, svOkOr, svOkra, tvOI^ˆ, tvOI^ˆhO, tvOjI, tvOjIch].

Tento problém bol vyriešený tým spôsobom, že v slovách začínajúcich na *dvo-*, *kvo-*, *obvo-*, *potvo-*, *stvo-*, *svo-*, *tvo-* sa v prvej slabike *v* prepisuje ako [v].

§41 Spoluhláska *v* sa vyslovuje aj v slovách *distingvovaný*, *Guinea*, *guinejský*, *sangvinik*. Vyslovuje sa

[dlstINgvOvanI:, gvI:nEa, gvI:nEI_ˆskI:, saNgvInIk].

Tieto slová boli pridané medzi výnimky.

Pre výslovnosť "v" aj "U_ˆ" bola z dôvodu absencie znalostí, potrebných k správnej interpretácii pravidiel pre výslovnosť, využitá skutočnosť, že ak nebolo aplikované žiadne pravidlo pre výslovnosť "v" rovnako ako aj pre výslovnosť "U_ˆ", je možné aplikovať obojakú výslovnosť. Pre ostatné prípady teda platí:

$$\mathbf{v} \rightarrow \mathbf{v} \quad (\text{obu} - v) \quad (3.91)$$

$$\rightarrow \mathbf{U}_-^{\wedge}(\text{obu} - v) \quad (3.92)$$

Po dôkladnej analýze týchto pravidiel bolo potrebné definovať ďalšie pomocné pravidlo. Z pravidiel pre dvojakú výslovnosť totiž vyplýva, že ak sa aplikuje pravidlo (3.86) alebo (3.88), nebude už aplikované pravidlo o dvojakej výslovnosti, keďže v spomínaných pravidlách sa uvažujú všetky samohlásky a teda aj *u*. Preto bolo potrebné pred tieto prípady zaradiť nasledujúce pravidlo:

$$U\mathbf{v} \mid \rightarrow UU_-^{\wedge}(\text{obu} - v - \text{ník}) \quad (3.93)$$

$$\rightarrow U\mathbf{v} \quad (\text{obu} - v - \text{ník}) \quad (3.94)$$

3.4.8 Znelostná asimilácia

a) Všeobecne o asimilácii

§42 Asimilácia (spodobovanie) je proces, v ktorom sa jedna alebo viacej hlások mení tak, že sa pripodobňuje k inej (susednej) hláske. V tomto procese hláska (alebo skupina hlások) stráca niektorú vlastnosť alebo nadobúda niektorú inú vlastnosť. Stráca vlastnosť, ktorou sa odlišovala od susednej hlásky, alebo priberá takú novú vlastnosť, ktorou sa artikulačne a akusticky približuje k susednej hláske. Z ortoepického hľadiska sú dôležité aj také asimilácie, ktoré nemajú platnosť fonologických neutralizácií. Patria k nim napríklad zmeny, o ktorých sa hovorí v súvislosti s výslovnosťou spoluhláskových skupín *tl*, *dl*, *tn*, *dn*, *mp*, *bp* a pod.

Rozlišujeme viacej druhov asimilácií – asimilácie miesta a spôsobu artikulácie, asimilácie znelosti; podľa smeru asimilačného pôsobenia hlások sa rozlišuje spätná (regresívna) a postupujúca (progresívna) asimilácia.

§43 Z ortoepického hľadiska je v slovenčine najdôležitejšia znelostná asimilácia. Znelostná asimilácia je v slovenčine spätná. Patrí k typickým črtám slovenčiny a je jedným z činiteľov spôsobujúcich efekt splyvavej slovenskej výslovnosti. Chyby v znelostnej asimilácii sú nápadné a hodnotia sa ako hrubé ortoepické chyby.

Znelostná asimilácia spočíva v tom, že znelé párové spoluhlásky v istej situácii strácajú znelosť, takže sa vyslovujú ako ich neznelé páry, a neznelé spoluhlásky zasa v istej situácii znelosť získavajú.

b) Spoluhlásky podliehajúce znelostnej asimilácii.

§44 Znelostná asimilácia sa prejavuje v tzv. znelostných pároch. V slovenčine sú tieto páry neznelých a znelých spoluhlások:

neznelé: [p, f, t, s, ts, S, tS, c, k, x]
znelé: [b, f_v, d, z, dz, Z, dZ, J\, g, G, h]

Ostatné spoluhlásky sú nepárové znelé. Znelostné páry (p–b, f–f_v atď.) sa od seba líšia iba znelosťou. Znelostnej asimilácii podliehajú iba párové spoluhlásky.

Párovú symetriu znelých a neznelých spoluhlások porušuje protiklad neznelého [x] oproti znelým [G, h], lebo v určitých situáciách sa aj [G] aj [h] správajú ako znelý pár neznelého [x].

Spoluhláska [v] nie je v zozname párových znelých spoluhlások, lebo je to nepárová (sonórna, zvučná) a nešumová hláska, ktorá sa v inom type asimilácie strieda s nepárovou zvučnou a nešumovou hláskou [U_ ^]. Hláska v nemôže stratiť vlastnosť znelosti. Znelým párom k neznelému [f] je spoluhláska [f_v], ktorá sa od [f] líši iba znelosťou.

c) Podmienky a miesta znelostnej asimilácie

§45 Znelostná asimilácia sa v slovenčine uskutočňuje na morfe-matických hraniciach, t. j. na týchto miestach:

- na hranici slov,
- na hranici slovných základov v zložených slovách,
- na hranici predpony a koreňa alebo základu slova,
- na hranici koreňa alebo základu slova a odvodzovacej (výnimočne aj ohýbacej) prípony.

d) Všeobecné pravidlá znelostnej asimilácie

§46 Znelostná asimilácia sa v spisovnej slovenčine uskutočňuje vtedy, keď sa na morfe-matickej hranici stretne neznelá spoluhláska so znelou (so znelou párovou, nepárovou alebo so samohláskou), alebo keď sa na tomto mieste stretne znelá párová spoluhláska s neznelou, resp. keď po znelej párovej spoluhláske nasleduje prestávka.

§47 Prestávka má v znelostnej asimilácii rovnaký účinok ako neznelá spoluhláska. Takto sa správa aj tvrdý hlasový začiatok a ráz. Ako už bolo uvedené v 3.4.4, tvrdý hlasový začiatok a ráz v našich transkripčných pravidlách neuvažujeme.

Zo všeobecného pravidla o znelostnej asimilácii existujú viaceré výnimky.

§48 Na hranici koreňa (základu) slova a odvodzovacej a ohýbacej prípony sa znelostná asimilácia pri stretnutí neznelej spoluhlásky so zvučnou uskutočňuje iba v niektorých prípadoch a vôbec sa neuskutočňuje pri stretnutí neznelej spoluhlásky so samohláskou. Z hľadiska podmienok, rozsahu a výsledku sa znelostná asimilácia v spisovnej slovenčine uskutočňuje rovnako na hranici slov, na hranici slovných základov v zložených slovách a na hranici predpony a koreňa alebo základu slova.

Asimilácia znelej spoluhlásky na neznelú

§49 Keď sa na hranici dvoch samostatných slov, medzi ktorými nie je prestávka, na hranici slovných základov v zložených slovách (kompozitách) a na hranici predpony a koreňa alebo základu slova stretne znelá párová šumová spoluhláska (alebo skupina takýchto spoluhlások) s neznelou, asimiluje sa znelá spoluhláska na neznelú.

Táto asimilácia sa uskutočňuje aj na konci slova pred prestávkou.

§50 Znelostná asimilácia je na hranici slov, medzi ktorými nie je prestávka a nasledujúce slovo sa začína neznelou spoluhláskou:

Spojenia *dážď prší, drozd spieva, hvizd pišťalky, dub spadol, keď sa dozvedel, loď pláva, buď tichšie, mozog pracuje, plod práce, zväz priateľov* vyslovujeme

[da:Scpr=Si:, drostspL^eva, hvistpi:Scalki, dupspadol, kecsadozveJ\el, locpla:va, buc-cixSL^e, mozokpratsuje, plotpra:tse, zvesprL^aceLoU^].

Transkripčné pravidlá:

$$\mathbf{ZN} \parallel \mathbf{NZ} \rightarrow \mathbf{NZa} \parallel \mathbf{NZ} (\textit{droz} - \textit{dspieva}) \quad (3.95)$$

$$\mathbf{ZN} \vdash \mathbf{NZ} \rightarrow \mathbf{NZa} \mathbf{NZ} \quad (\textit{be} - \textit{zp} - \textit{ečný}) \quad (3.96)$$

$$\mathbf{ZN} \ddagger \mathbf{NZ} \rightarrow \mathbf{NZa} \mathbf{NZ} \quad (\textit{be} - \textit{zt} - \textit{oho}) \quad (3.97)$$

$$\mathbf{ZN} \mathbf{NZ}\$ \rightarrow \mathbf{NZa} \mathbf{NZ} \quad (\textit{bo} - \textit{zk}) \quad (3.98)$$

kde \mathbf{ZN} označuje znelú spoluhlásku, \mathbf{NZ} neznelú a \mathbf{NZa} neznelú vzniknutejšiu asimiláciou znelej. Znak \parallel označuje hranicu základu slova a predpony, a znak \ddagger hranicu slov v zložených slovách.

Pravidlo (3.98) je definované pre asimiláciu na konci slova pred prestávkou. Aplikácia tohto pravidla je komplikovaná, keďže hovoriaci môže urobiť v reči prestávku kdekoľvek a z textu to nemusí vyplývať. Tento druh asimilácie je v implementácii uvažovaný na konci vety a pred čiarkou, bodkočiarkou alebo dvoj-bodkou.

Znelostná asimilácia v predložkových spojeniach:

§51 Predložkové spojenia *nad plotom, nad tebou, od teba, od včera, od priateľa, pod kolesom, pod parou, popred šatňu, popod cestu, nad všetko, pred stromom, pred tebou, v chlebe, v Prahe, z cesta, z čista-jasna, z Prešova* vyslovujeme

[natplotom, natcebou, otceba, otftSera, otrL^aceLa, potkoLesom, potparoU^, popretSatJu, popotcestu, natfSetko, pretstromom, pretcebou, fxLebe, fpPrahe, stsesta, stSistaJasna, spreSova].

Predložkové spojenia sú ošetrené rovnako ako v predchádzajúcom prípade transkripčným pravidlom (3.95).

§52 Znelostná asimilácia na konci slov pred prestávkou:

Pred prestávkou slová ako *dážď, dub, glg, hádž, hlad, hvizd, jedz, mosadz, mráz, ps-truh, rád, raž, zväz* vyslovujeme

[da:Sc, dup, gl=k, ha:tS, hlat, hvist, jets, mosats, mra:s, pstrux, ra:t, raS, zves].

Tak sa vyslovujú aj slová ako *bozk*, *hrýzť*, kde znelá párová spoluhláska predchádza pred neznelou: [bOsk, hrI:sc].

Tieto slová boli pridané medzi výnimky. Pre slová ako *bozk*, *hrýzť*, kde asimilácia vzniká na konci slova končiaceho sa dvoma spoluhláskami, bolo definované transkripčné pravidlo (3.98).

Znelostná asimilácia v zložených slovách:

- §53 V slovenčine je málo zložených slov, v ktorých sa prvý slovný základ končí znelou párovou spoluhláskou. Takéto slová sú *beztak*, *beztoho*, *kníhtlačiareň*, *nadto*. Vyslovujeme ich

[bEstak, bEstOhO, kJI:xtlatSI_ˆarEJ, nattO].

Tieto slová boli zaradené medzi výnimky, napriek tomu, že sú ošetrené pravidlom (3.97). Je to z toho dôvodu, že detekcia zložených slov je problematická.

Znelostná asimilácia na hranici predpony a koreňa alebo základu slova:

- §54 Slová ako *bezcieľny*, *bezcharakterný*, *bezpečný*, *bezprostredný*, *nadchodiť*, *občadiť*, *občas*, *občerstviť*, *občiahnuť*, *občina*, *obchod*, *obchodník*, *obsadiť*, *obsah*, *obsevok*, *obsluha*, *obširny*, *obštrukcia*, *obtlčť*, *podfarbiť*, *podfuk*, *podkova*, *podpis*, *predpis*, *rázcestie*, *rázpin*, *rážtoka*, *rozkaz*

a slová ako *vehod*, *vhklad*, *vpád*, *vstať*, *vtierať*, *vzchádzať*, *vzchopiť sa*, *vzkriesenie*, *vzpažiť*, *vzpažmo*, *vzpera*, *vzpierač*, *vzpínadlo*, *vzpor*, *vzpriamiť sa*, *vzpruha*, *vzpružiť*, *vzťah*, *vzťahovať sa*, *vztiahnuť*, *vztlak*, *vztýčiť*, *vztyk*, atď. vyslovujeme

[bestSI_ˆeLni, besxarakterni:, bespetSni:, besprostredni:, natxoJ\ic, optSaJ\ic, optSas, optSerstvic, obtSI_ˆahnuc, optSina, opxot, opxodJI:k, opsaJ\ic, opsax, opsevok, opsluha, opSi:rni, opStruktsia, optl=tsc, potfarbic, potfuk, potkova, potpis, pretpis, ra:stsescI_ˆe, ra:spin, ra:stoka, roskas;

fxot, fklat, fpa:t, fstac, fcI_ˆerac, fsxa:dzac, fsxopicsa, fskrI_ˆeseJI_ˆe, fspaZic, fs-paZmo, fspeda, fspI_ˆeratS, fspi:nadlo, fspor, fsprI_ˆamicsa, fspruha, fspruZic, fscax, fscahovacs, fscI_ˆahnuc, fstlak, fsti:tSic, fstik] atď.

- §55 Takto sa vyslovujú aj slová, v ktorých sa výnimočne píše *v*, akoby *v* bola predpona, a slová, ktoré kedysi boli predponové, no dnes sú to iba pravopisné výnimky a niektoré iné slová.

Také slová sú *avšak*, *včela*, *včera*, *všade*, *(zo)všadiaľ*, *však*, *všakhej*, *všakže*, *všetko*, *vták*, *vtedy*, *(dô)vtip*; *nebezpečenstvo*, *(ne)bezpečný*, *rozkoš*, *rozkošnícky*, *roztomilý*, *roztopaš*, *roztopašník*, *roztopašný*, *zabezpečenie*. Vyslovujeme ich

[afSak, ftSEla, ftSEra, fSaJ\E, (zo)fSaJ\I_ˆaL, fSak, fSaghEI_ˆ, fSagZE, fSEtkO, fta:k, ftEdI, (dU_ˆo)fcIp; JEbEspEtSEnstvO, (Je)bEspEtSnI:, rOskOS, rOskOSJI:ts-kI, rOstOmIII:, rOstOpaS, rOstOpaSJI:k, rOstOpaSnI:, zabEspEtSEJI_ˆE].

Slová *nádcha*, *nadchnúť*, *nadšenie* v spisovnej slovenčine vyslovujeme: [na:txa, natxnU:c, natStSEJL_ˆE].

Všetky tieto slová boli zaradené medzi výnimky.

§56 Ak sa od takýchto slov odvodzujú nové slová predponou, ktorá sa končí znelou spoluhláskou, náležite sa tu uskutočňuje znelostná asimilácia predponovej znej spoluhlásky na neznelú. Tak je to v slovách *odvšade*, *odvšadiaľ*, *odvtedy*, *predvčerajší*, *predvčerajšok*, *predvčerom*. Vyslovujeme ich

[otfSaJ\ə, otfSaJ\I_ˆaL, otftedi, pretftSeraI_ˆSi:, pretftSeraI_ˆSok, pretftSerom].

V slove *voš* sa pri skloňovaní strieda v koreni *v/f*. Vyslovuje sa [vOS], ale [fSI] a znelostnou asimiláciou aj [otfSivavI_ˆEc].

Slova *voš*, *vši*, *odvšivavieť* boli zaradené medzi výnimky.

Asimilácia neznej spoluhlásky na znelú

§57 Keď sa na hranici dvoch samostatných slov, medzi ktorými nie je prestávka, na hranici slovných základov v zložených slovách (kompozitách) a na hranici predpony a koreňa alebo základu slova stretne neznelá spoluhláska (alebo skupina takýchto spoluhlások) so znelou hláskou, t. j. so znelou párovou, so zvučnou (sonórnou) alebo so samohláskou, asimiluje sa neznelá spoluhláska na znelú.

Na hranici predpony a koreňa (základu) slova je v slovenčine táto asimilácia celkom výnimočná (predpony sa končia znelou hláskou).

Asimilácia na hranici slov, medzi ktorými nie je prestávka, a prvé slovo sa končí neznelou spoluhláskou:

Slovné spojenia ako *chlap zavolal*, *hoc oklamal*, *hrst dukátov*, *vlak hučí*, *vlak mešká*, *les vonia*, *most opravili*, *pec hreje*, *pes jastrí*, *prst ho bolí*, *prst i ruka*, *taft nemáme*, *šéf odišiel*, *tak isto*, *viac-menej*, *viac ráz*, *vliecť opitého*, *znelosť hlások* vyslovujeme

[xlabzavolal, hOdzOklamal, hr=J\dUka:tOU_ˆ, vlaghUtSI:, vlagmESka:, LEzvOJL_ˆa, mOzdOpravILI, pEdzhrEjE, pEzjastrI:, pr=zdhObOLI:, pr=zdlrUkU, taf_vdJEmat:mE, SE:f_vOdISI_ˆEl, tagIstO, vI_ˆadzmeJEj, vI_ˆadzra:s, vLI_ˆEdzJ\OpItE:hO, zJE-IozJ\hla:sOk].

Tento druh asimilácie bol ošetrený pravidlami:

$$\mathbf{NZ} \parallel \mathbf{ZN} \rightarrow \mathbf{ZNa} \parallel \mathbf{ZN} \text{ (} le - svonia \text{)} \quad (3.99)$$

$$\mathbf{NZ} \ddagger \mathbf{ZN} \rightarrow \mathbf{ZNaZN} \text{ (} tan - cm - ajster \text{)} \quad (3.100)$$

Pre asimiláciu na hranici predpony a základu slova nebolo z vyššie uvedených dôvodov definované pravidlo a takéto prípadné slová sú zaradené medzi výnimky.

Asimilácia na hranici slovných základov v zložených slovách:

§58 Zložené slová ako *akiste*, *bársaký*, *čerthovie*, *desaťuholník*, *hocako*, *hocaký*, *jakby*, *jak-živ*, *lesmajster*, *liftboj*, *mostmajster*, *naničhodník*, *päťväzkový*, *postoperatívny*, *rotmajster*, *spakručie*, *šéfdirigent*, *šéflekár*, *šéfredaktor*, *štvrtrok*, *takhľa*, *takisto*, *takzvaný*, *takže*, *tancmajster*, *tľhuba*, *včasráno*, *vrtnajster*, *všakáno*, *všakver*, *watthodina* vyslovujeme

[agIscE, ba:rzakI:, tSErdhOvI_ˆE, J\EsaJ\UhOlJI:k, hOdzakO, hOdzakI:, jagbI, jag-ZIU_ˆ, LEzmaI_ˆstEr, lIf_vdbOI_ˆ, mOzdmaI_ˆstEr, naJIdZhOdJI:k, pEJ\zvEskOvI:, pOzdOpEratI:vnI/pOzdOpEratI:U_ˆnI, rOdmaI_ˆstEr, spragrUtSI_ˆE, SE:f_vdIrIgEnt, SE:f_vLEka:r, SE:f_vrEdaktOr, Stvr=J\rOk, taghLa, tagIstO, tagzvanI:, tagZE, tandzmaI_ˆstEr, tl=dZhUba, ftSazra:nO, vr=dmaI_ˆstEr, fSaga:nO, fSagvEr, vadhOJ\Ina] atď.

V slovách *hocičo*, *hocijaký*, *vrtohlav* sa znelostná asimilácia neuskutočňuje, lebo morfematická hranica prebieha takto: *hoci/čo*, *hoci/jaký*, *vrt/o/hlav*, resp. *vrto/hlav*.

Tieto slová boli zaradené medzi výnimky.

Asimilácia na hranici predpony a koreňa slova:

§59 Táto asimilácia je iba v slovách *smena*, *Smena*, ktoré sú pravopisnou výnimkou. Vyslovujeme [zmena].

Slovo *smena*, podobne ako aj nižšie uvedené odvodeniny, boli pridané medzi výnimky. Vzhľadom k tomu, že nespisovná výslovnosť [smena] je tiež často badaateľná, pre výslovnosť bola pridaná aj táto eventualitu.

§60 Tak sa vyslovujú aj odvodené slová ako *smenmajster*, *smennosť*, *smenový*. Vyslovujeme

[zmenmaI_ˆster, zmennosc, zmenovi:].

Slová *Smena*, *smena* i *zmena* sa vyslovujú rovnako aj v spojeniach ako *denník Smena*, *pracovná smena*, *programová zmena*. Spisovná výslovnosť je

[J\eJJI:gzmena, pratsovna:/prasoU_ˆna:zmena, programova:zmena].

Predchádzajúce dva paragrafy už pre súčasnosť nie sú aktuálne, keďže od roku 1989 sa podľa kodifikovaného pravopisu píše *zmena*, *zmenmajster* a pod. Z dôvodu compatibility so staršími textami boli tieto slová pridané medzi výnimky.

§61 Slová *nesmierny*, *sloh*, *sloha*, *směr*, *směrovat*, *směrovka*, *svah* nie sú z dnešného hľadiska predponové. Vyslovujú sa

[JESmI_ˆErnI, slOx, slOha, smEr, smErOvac, smErOU_ˆka, svax].

Vyslovuje sa však

[zmI^ˆEr, zvaZOvac],

no tu sa aj píše *zmier*, *zvažovať (sa)*. Rozlišujú sa slová *sloh* (písomná práca) a *zloh* (orať do zlohu, na seba).

Všetky tieto slová boli pridané medzi výnimky.

Slová *kde*, *sme* a *-kdy*

§62 Z hľadiska súčasnej slovenčiny sú pravopisnými výnimkami slová *kde*, *sme* a slová so zložkou *-kdy*. Vyslovujú sa

[gJ\ē, zme, -gdi].

Preto sa pred nimi uskutočňuje znelostná asimilácia ako pred inými slovami (pred základmi slov), ktoré sa začínajú znelou hláskou, čiže slová *bárskde*, *dakde*, *hockde*, *niekde*, *niekdajší*, *nikde*, *nikdy* vyslovujú sa

[ba:rzgJ\ē, dagJ\ē, hOdzgJ\ē, JI^ˆEgJ\ē, HI^ˆEgdaI^ˆSI:, JIgdJ\ē, JIgdI].

§63 Vyslovuje sa však

[ktO, JIktO, JI^ˆEktO, daktO, hOtsktO], nie [gdO, JIgdO, JI^ˆEgdO, dagdO, hOdzgdO].

§64 Slovo *takmer* vyslovujeme [tagmEr].

Slová *sme*, *takmer*, *niekdajší* boli pridané k výnimkám a pre ostatné boli definované nasledujúce transkripčné pravidlá:

$$\mathbf{NZkde} \parallel \rightarrow \mathbf{ZNagJ\ē}(bárs - kde) \quad (3.101)$$

$$\mathbf{NZkdy} \parallel \rightarrow \mathbf{ZNagdI} \quad () \quad (3.102)$$

$$([SMrlí])kde \parallel \rightarrow \backslash 1gJ\ē \quad (nie - kde) \quad (3.103)$$

$$([SMrlí])kdy \parallel \rightarrow \backslash 1gdI \quad (ni - kdy) \quad (3.104)$$

Ak za pravidlom nie je uvedený príklad, znamená to, že sa príklad nepodarilo nájsť. Pravidlo však bolo definované v prípade, že je explicitne opísané v pravidlách slovenskej výslovnosti.

Predložky *s*, *k*

§65 Podľa pravidla o znelostnej asimilácii na hranici slov sa vyslovujú aj neslabičné predložky *s*, *k*: ak nasleduje neznelá spoluhláska, vyslovujú sa [s, k], ak nasleduje znelá hláska (znelá párová, zvučná alebo samohláska), vyslovujú sa [z, g].

Predložka *s* :

§66 Predložkové spojenia *s čistým, s tetkou* vyslovujeme [stSisti:m, scetkoU[^]], lebo po *s* nasleduje neznelá spoluhláska. Predložkové spojenia *s babkou, s veľkou bábou, s humorom, s deťmi, s milými deťmi, s Olgou, s jemnou výčitkou, s vaším priateľom, s ich právomocou, s rokmi, s vďačnosťou, s veľmi dobrou náladou* vyslovujeme

[zbapkOU[^], zvELkOU[^]ba:bOU[^], zhUmOrOm, zJ\EcmI, zmIII:mIJ\EcmI, zOLgOU[^], zjEmnOU[^]vI:tSItkOU[^], zvaSI:mprI[^]acELOm, zIxpva:vOmOcOU[^], zrOkmI, zf.vJ\atSnOscOU[^], zvELmIdOvrOU[^]na:ladOU[^]].

(Tu sa slovo nasledujúce po predložke *s* začína znelou hláskou.)

Predložka *k* :

§67 Predložkové spojenia *k plotu, k stolu* vyslovujeme [kplOtU, kstOIU], lebo po *k* nasleduje neznelá spoluhláska.

Predložkové spojenia *k hore, k domu, k bielej stene, k murárom, k naklonenej rovine, k jasliam, k vode, k vdove, k Andrejovi, k ich záveru, k môjmu postoju, k nášmu stanovisku, k vaším rečiam* vyslovujeme

[ghOrE, gdOmU, gbL[^]EIEI[^]scEJE, gmUra:rOm, gnaklOJEnEI[^]rOvIJE, gjasLI[^]am, gvOdE, gf_vdOvE, g_andrEjOvI, gIG/gIhza:vErU, gmU[^]OjmUpOstOjU, gna:ZmUstanOvIskU, gvaSI:mrEtSI[^]am]. (Tu sa slovo nasledujúce po predložke *k* začína znelou hláskou.)

Na základe toho bolo možné definovať nasledujúce pravidlá:

$$\mathbf{s} \parallel (NZ) \rightarrow \mathbf{s} \parallel \setminus 1 (s - tetkou) \quad (3.105)$$

$$\mathbf{k} \parallel (NZ) \rightarrow \mathbf{k} \parallel \setminus 1 (k - stolu) \quad (3.106)$$

$$\mathbf{s} \parallel (ZNH) \rightarrow \mathbf{z} \parallel \setminus 1 (s - humorom) \quad (3.107)$$

$$\mathbf{k} \parallel (ZNH) \rightarrow \mathbf{g} \parallel \setminus 1 (k - vode) \quad (3.108)$$

ZNH označuje znelú hlásku. Pravidlá (3.107) a (3.108) sú komplementárne k pravidlám (3.105) a (3.106) a súčasne sú podmnožinou všeobecných asimilačných pravidiel na hranici slov (3.95) a (3.99).

§68 Znelostná asimilácia neslabičných predložiek *s, k* sa výnimočne neuskutočňuje pred niektorými osobnými zámenami. Vyslovujeme

[sJI:m, sJOU[^], snamI, svamI, sJImI; kJEmU, kJEI[^], kna:m, kva:m, kJIm].

Výslovnosť [zJI:m, zJImI, gJEmU, gJIm] atď. nie je spisovná.

Informácia o nespisovnej výslovnosti je z nášho pohľadu informácia o možnosti výskytu aj tejto alternatívnej výslovnosti. Z tohto dôvodu boli medzi výnimky pridané obidva spôsoby výslovnosti.

§69 Aj vokalizované predložky *so*, *ku* sa v spojení s osobnými zámenami *mnou*, *mne*, t. j. *so mnou*, *ku mne*, vyslovujú [sO, kU]. Vyslovujeme

[sOmnOU[^], kUmJE]. Všade inde sa vokalizované predložky *so*, *ku* vyslovujú [zO, gU]. Výslovnosť pre *so mnou*, *ku mne* sa pridala medzi výnimky.

Znelostná asimilácia na hranici základu slova a prípony

Asimilácia znelej spoluhlásky na neznelú

§70 Keď sa na hranici základu slova a prípony (t.j. v koncovej časti slova) stretne znelá šumová spoluhláska s neznelou, asimiluje sa znelá spoluhláska na neznelú.

Slová *arabčina*, *batôžtek*, *blízko*, *bodkovať*, *bozkať*, *dovozca*, *drábsky*, *drobček*, *drobčiť*, *dubkom*, *hrudkovitý*, *choďte*, *chrbtica*, *jazdkyňa*, *Jožko*, *klbčiť sa*, *klbko*, *ľahko*, *ľahšie*, *mužstvo*, *násobok* – *násobku*, *násobkom*, *násobky*, *nebožtík*, *nízky*, *občan*, *oblahčiť*, *odiedzka*, *odpadkový*, *otázka*, *pochabstvo*, *rehtať*, *sladkastý*, *svedok* – *svedka*, *sudkyňa*, *súdržka* – *súdržky*, *so súdržkou*, *ťažkosť*, *viezť*, *vozte*, *vlhčina*, *volžský*, *záhradkár*, *zblízka*, *zobkať* vyslovujeme

[araptSIna, batU[^]OScEk, bLI:skO, bOtkOvac, bOskac, dOvOtsa, dra:pskI, drOptSEk, drOptSIc, dUpkOm, hrUtkOvItI:, xOccE, xr=pcItsa, jastkIJa, jOSkO, kl=p-tSIcsa, kl=pkO, LaxkO, LaxSI[^]E, mUSstvO, na:sOpkU, na:sOpkOm, na:sOpkI, JE-bOScI:k, JI:skI, OptSan, ObLaxtSIc, OJ\I[^]Ecka, OtpatkOv:, Ota:ska, pOxapstO, rExtac, slatkastI:, svEtka, sUtkIJa, sU:drUSka, sU:drUSkI, zOsU:drUSkOU[^], caSkOsc, vI[^]Esc, vOscE, vl=xtSIna, vOlSskI:, za:hratka:r, zbLI:ska, zOpkac] atď.

V tomto prípade je potrebné vyriešiť problém správneho morfeomatického delenia slova. Tento problém je podrobnejšie rozpracovaný v časti 3.4.15. Ďalej sa predpokladá, že morfeomatická hranica bola určená. Na základe toho bolo definované nasledujúce transkripčné pravidlo:

$$\mathbf{ZNs} \vdash \mathbf{NZ} \rightarrow \mathbf{NZsaNZ}(\mathit{blí} - z - \mathit{ko}) \quad (3.109)$$

kde \vdash označuje hranicu základu slova a prípony. Obdobne je pre hranicu základu slova a predponu použitý znak \dashv . Obidva znaky môžu byť nahradené znakom \wr označujúcim morfeomatickú hranicu. Znaky \vdash a \dashv sú používané len vtedy, keď je potrebné zvýrazniť buď predponu alebo príponu. *ZNs* označuje znelú šumovú spoluhlásku a *ZNsa* znelú šumovú asimilovanú spoluhlásku.

§71 V druhom páde množného čísla postatných mien typu *ceruzka*, *klbko*, *obchôdzka*, *súdržka* sa pred odvodzovací príponu vsúva vkladná samohláska (resp. dvojhláska). Pred ňou sa však znelostná asimilácia neuskutočňuje. Vyslovuje sa

[tsErUska, kl=pkO, OpXU[^]Otska, sU:drUSka],

ale

[tsErUZI[^]Ek, kl=bI[^]Ek, OpXU[^]OdZOk, sU:drUZI[^]Ek] atď.,

lebo základ slova sa končí na znelú spoluhlásku: *ceruz-*, *klb-*, *obchôdz-*, *súdrž-*.

Pre tento prípad nie je potrebné vytvárať žiadne pravidlo, keďže predchádzajúce pravidlo (3.109) aplikované nebude. To iste platí aj o skutočnostiach uvedených v nasledujúcom odstavci.

§72 Podobne aj v slovách *kliatba – kliatob, kosba – kosieb, liečba – liečob, mlatba – mlatieb, platba – platieb, prosba – prosieb, spisba – spisieb, veštbá – veštieb, výučba – výučieb* atď. sa pred pohyblivou a vkladnou samohláskou (dvojháskou) znelostná asimilácia neuskutočňuje. V slovách s pohyblivou alebo vkladnou samohláskou je ukazovateľom zakončenia koreňa slova práve ten tvar, v ktorom po koncovej koreňovej spoluhláske nasleduje samohláska. Vyslovuje sa

[kLI^ˆatOp, kOsI^ˆEp, lI^ˆEtSOp, mlacI^ˆEp, placI^ˆEp, prOsI^ˆEp, spIsI^ˆEp, vES-cI^ˆEp, vI:UtSI^ˆEp].

§73 V slovách *bozk, (po/vy)bozkávať, dubkom, mrzký, redkovka* sa dnes nepocituje morfemická hranica. Vyslovujeme ich však

[bOsk, (pO/vI)bOska:vac, dUpkOm, mr=skI:, rEckOU^ˆka].

§74 Pravopisné dvojtvary sú *udlabčiť – udlapčiť, udlabkať – udlapkať, zaehkať – zaechkať, zarehtať – zarechtať*. Vyslovujeme

[udlaptSIc, udlapkac, zaExkac, zarExtac].

Tieto slová boli zaradené medzi výnimky.

Asimilácia neznej spoluhlásky na znelú

§75 Keď sa na hranici základu slova a prípony stretne neznelá spoluhláska so znelou, znelostná asimilácia nastáva v plnom rozsahu iba vtedy, ak je znelou hláskou šumová spoluhláska. Slová ako *kliatba, kresba, liečba, mlatba, prosba, veštbá, výučba*, podobne aj slová *nášho, vášho* vyslovujeme

[kLI^ˆadba, krEzba, lI^ˆEdZba, mladba, prOzba, vEZdba, vI:UdZba,; na:ZhO, va:Z-hO] atď.

§76 Znelostná asimilácia sa neuskutočňuje, ak sa v podstatných menách a prídavných menách na hranici základu a prípony neznelá spoluhláska základu slova stretne so samohláskou alebo zvučnou spoluhláskou ohýbacej alebo slovotvornej prípony.

V slovách a tvaroch ako *chlap, chlapa, chlapmi, platnosť, sobášny šťastlivý, šťastný, taktný, taktná, taktné, taktne, väčšmi* atď. vyslovujeme

[xlap, xlapa, xlapmI, platnOsc, sOba:SnI, ScastLIvI:, ScastnI:, taktnI:, taktna:, taktnE:, taktJE, vEtSmI] atď.

$$\mathbf{NZ} \vdash \mathbf{ZNs} \rightarrow \mathbf{ZNaZNs}(kre - s - ba) \quad (3.110)$$

§77 Znelostná asimilácia nastáva v slovesnom tvare 1. osoby množného čísla rozkazovacieho spôsobu pred príponou *-me*, ak predchádza neznelá spoluhláska, ako v slovách *kosme, kúpme, lešme, nosme, plačme, plaťme, prosme, pusťme, trafme* atď.

Vzhľadom k skutočnosti, že slová končiace na *[SP]me* môžu byť napríklad aj podstatné mená a nie len slová podľa paragrafu 77 *krčme, norme, siedme*, tento druh asimilácie je aplikovaný len vtedy, ak veta končí znakom ”!”.

§78 Znelostná asimilácia nastáva aj v tvaroch zámen *náš, váš – nášmu, vášmu*, podobne ako *nášho, vášho*. Vyslovujeme

[kOzmE, kU:bmE, LEZJ\mE, nOzmE, plaJ\mE, prOzmE, pUzJ\mE, traf_vmE; na:ZmU, va:ZmU, na:ZhO, va:ZhO].

Tieto slová boli zaradené medzi výnimky.

§79 V príslovkových tvaroch zo športovej terminológie s príponou *-mo* (*obkročmo, pripätmo, rozkročmo, skočmo, skrčmo, súručmo, vznesmo*) a okrem toho aj v slove *skusmo* sa pripúšťa výslovnosť

[OpkrOtSmO, prIpEtmO, rOskrOtSmO, skOtSmO, skr=tSmO, sU:rUtSmO, vzJE-smO; skUsmO],

ale aj

[OpkrOdZmO, prIpEdmO, rOskrOdZmO, skOdZmO, skr=dZmO, sU:rUdZmO, vzJE-zmO; skUzmO].

Tieto slová boli zaradené medzi výnimky.

Znelostná asimilácia [f, f_v]

Všeobecne o [f, f_v]

§80 V slovenskom pravopise sa znakom *v* označujú štyri hlásky: [v, U_ˆ, f_v, f]. Z nich [v] je pernozubná nešumová sonórna spoluhláska, [U_ˆ] je pernoperná nešumová sonórna spoluhláska, [f_v] je pernozubná šumová znelá a [f] je pernozubná šumová neznelá spoluhláska. Sonórna spoluhláska [v] nemôže (ako nepárová znelá) stratiť znelosť – nemôže sa asimilovať na neznelú hlásku, čiže nemôže sa v spisovnej slovenčine v znelostnej asimilácii strieďať so spoluhláskou [f]. Výslovnosť [obuf, J\I_ˆeftSa, kr=f, pr=f] sa hodnotí v spisovnej slovenčine ako hrubá a veľmi nápadná ortoepická chyba. V znelostnej asimilácii vystupuje iba pár [f, f_v].

§81 Zo zákonitosti výskytu hlások [v-U_ˆ] vyplýva, že [f, f_v] sa v znelostnej asimilácii môžu strieďať v slabike po slabičnom vrchole iba vtedy, ak sa na tomto mieste podľa pravidiel slovenského pravopisu píše písmeno *f* (napr. v spojení *Jozef Gregor*). Na mieste písmena *v* sa hlásky [f - f_v] v znelostnej asimilácii môžu zasa vyskytnúť iba v časti slabiky, ktorá je pred vrcholom sonórnosti (a iba na tých miestach, kde sa podľa zásad slovenského pravopisu píše písmeno *v*). Hláska [f_v] nie je v slovenčine osobitná fonéma a nemá ani osobitné pravopisné označenie.

§82 Vznik a výskyt hlásky [f_v] súvisí s princípom stavby slabiky v slovenčine. V spisovnej slovenčine nemôže v tej istej slabike sonórne [v] predchádzať pred šumovými spoluhláskami *z, s, d, t, ě, ř* atď. Tu sa v slovenčine vyslovuje tá istá hláska ako v slovách a spojeniach *šéfredaktor, šéf opery, huf ľudí, Jozef Gregor Tajovský*. Vyslovujeme

[SE:f_vrEdaktOr, SE:f_vOpErI, hU:f_vLUJ\I:, jOzEf_vgrEgOrtahOU_ˆskI:].

Je to tá šumová spoluhláska, ktorá sa vyslovuje aj v tvaroch rozkazovacieho spôsobu, kde podľa zákonitosti znelostnej asimilácie máme

[kOslc - kOzmlE, kU:pIc - kU:bmE, nOslc - nOzmlE, vOJ\Ic - vOJ\mlE], a tak aj [trafIc - traf_vmlE].

Podobne aj v slovách a spojeniach ako *vbiť, vdova, vďaka, vhodiť, vzájomný, navzájom, vzbĺknúť, vzdor, vzduch, vzor, vždy, vždycky, vžiť sa, v Bratislave, v dome, v humne, v zime, v Žiline, v žite* vyslovujeme

[f_vbIc, f_vdOva, f_vJ\ aka, f_vhOJ\Ic, f_vza:jOmnI:, naf_vza:jOm, f_vzbl=:knUc, f_vzdOr, f_vzdUx, f_vzOr, f_vZdI, f_vZdItskI, f_vZlcsa, f_vbracIslavE, f_vdOmE, f_vhUmJE, f_vzImE, f_vZILIJE, f_vZlce].

Pravidlá výslovnosti [f – f_v]

Hláska [f]

§83 Vyslovuje sa na začiatku a vnútri slova pred samohláskou, kde sa píše *f*, ako v slovách *dúfať, farba, Fero, fialka, kofola, trafiť, úfať, zúfať*, atď. Vyslovujeme

[dU:fac, farba, fErO, flalka, kOfOla, trafIc, U:fac, zU:fac];

na konci slova (v onomatopojach a citoslovciach, v cudzích a prevzatých slovách), kde sa píše *f*, ak nasleduje prestávka, alebo ak sa nasledujúce slovo začína neznelou spoluhláskou, ako v slovách *baf, golf, graf, gróf, húf, luft, pif-paf, reliéf, šéf, šerif, štift, štráf, uf*. Vyslovuje sa

[ba:f, gOlf, graf, grO:f, hU:f, lUft, pIfpaf, rElIE:f, SE:f, SErIf, StIf, Stra:f, Uf];

na začiatku slabiky, kde sa píše *v*, ak nasleduje neznelá spoluhláska, ako v slovách *včela, včera, vpiť, vpustiť, vši, všiť, vták, vtekať*,

ako súčasť predpony *vz-*, ak nasleduje neznelá spoluhláska, ako v slovách *vzchádzať, vzhodiť, vzhopiť sa, vzťah, vztýčiť, vztyk* atď.

§84 Okrem slov *včela, včera, vták, vši* ide o postavenie *v* v predpone. Tak sa vyslovuje aj predložka *v*, ak nasleduje neznelá spoluhláska, ako v spojeniach *v čistote, v práci, v továrni, v Poprade, v tomto mestečku* atď. Vyslovujeme

[ftSEra, ftSEla, fta:k, fcEkac, fpUscIc, fSIc, fpIc; fsxOpIcsa, fscax, fstIk, fstI:tSIc, fsxOJ\Ic, fsxa:dzac; ftSIstOcE, fpra:tsI, ftOva:rJ\I, fpOpraJ\E, ftOmtOmEscEtSkU].

Na základe vyššie uvedeného bolo možné definovať nasledujúce transkripčné pravidlá:

$$\mathbf{f}[aEIOUa : E : I : O : U :] \rightarrow \mathbf{f}\backslash 1 \quad (f - arba) \quad (3.111)$$

$$\mathbf{f}, \rightarrow \mathbf{f}, \quad (gol - f) \quad (3.112)$$

$$\mathbf{f} \parallel (NZ) \rightarrow \mathbf{f} \parallel \backslash 1(pif - paf) \quad (3.113)$$

$$| \mathbf{v}(NZ) \rightarrow \mathbf{f} \backslash 1 (v - čela) \quad (3.114)$$

$$\parallel \mathbf{v}z(NZ) \rightarrow \parallel \mathbf{f}z \backslash 1(v - zťah) \quad (3.115)$$

$$\mathbf{v} \parallel (NZ) \rightarrow \mathbf{f} \parallel \backslash 1(v - práci) \quad (3.116)$$

§85 V slove *návšteva* sa podľa spôsobu slabikovania pripúšťa výslovnosť [na:fScEva] i [na:U_ ScEva].

Toto slovo bolo zaradené medzi výnimky.

Hláska [f_v]

§86 Vyslovuje sa na začiatku slabiky, kde sa píše *v*, ak bezprostredne nenasleduje samohláska alebo [r, l, r=, l=, r=:, l=:, L, j], ako v slovách *vbiť*, *vdova*, *vhodiť*, *vhodný*, *vžiť sa*, pričom môže byť aj súčasťou predpony *vz-*, ako v slovách *vzbúriť sa*, *vzrast*, *vzbĺknúť*, alebo môže byť aj predložkou, ako v spojeniach *v hore*, *v dome*, *v meste*, *v Banskej Bystrici*, *v Galante* atď. Tak sa vyslovuje aj v tvare 1. osoby množného čísla rozkazovacieho spôsobu slovesa *trafiť* - *trafme*, kde sa píše *f*. Vyslovuje sa

[f_vbIc, f_vdOva, f_vhOJ\Ic, f_vhOdnI:, f_vZIcsa; f_vzbU:rIcsa, f_vzrast, f_vzbl=:knUc, f_vhOrE, f_vdOmE, f_vmEstE, f_vbanskEI_ bIstrItsI, f_vgalancE; traf_vmE] atď.

Na konci slova a v zložených slovách na konci prvého slovného základu, kde je v písmo *f*, ak sa nasledujúce slovo (alebo slovný základ druhej časti kompozita) začína samohláskou, znelou alebo zvučnou spoluhláskou a medzi týmito slovami nie je prestávka, ako v slovách a spojeniach *Jozef Gregor Tajovský*, *šéf odišiel*, *šéf zavolať*, *graf znázorňuje*, *šéfredaktor*, *liftboj*, *húf ľudí*, *halfvolej*. Vyslovuje sa

[jOzEf_vgrEgOrtajOU_ skI:, SE:f_vOdISI_ El, SE:f_vzavOlal, graf_vzna:zOrJUjE, hU:f_v-LUJ\I:, SE:f_vrEdaktOr, llf_vdbOI_ , half_vvOIEI_] atď.

Vyslovuje sa aj v slovách *Afgan*, *Afganistan*, *Afgár*:

[af_vgan, af_vganIstan, af_vga:r].

Slová *trafme*, *Afgan*, *Afganistan*, *Afgár* boli zaradené medzi výnimky a pre ostatne prípady boli definované nasledujúce pravidlá:

$$| \mathbf{v} [\hat{SM}rlr = l = r =: l =: Lj] \rightarrow \mathbf{f} \backslash 1 (v - biť) \quad (3.117)$$

$$\parallel \mathbf{v}z [\hat{SM}rlr = l = r =: l =: Lj] \rightarrow \mathbf{f} \backslash 1 (vz - búra) \quad (3.118)$$

$$\mathbf{v} \parallel [\hat{SM}rlr = l = r =: l =: Lj] \rightarrow \mathbf{f} \parallel \backslash 1 (v - hore) \quad (3.119)$$

$$\mathbf{f} \parallel (ZNH) \rightarrow \mathbf{f} \parallel \backslash 1(\text{šéf} - \text{odišiel}) \quad (3.120)$$

SM označuje z dôvodu lepšej čitateľnosti samohlásku a ZNH znelú hlásku. Takto definované pravidlá (3.117), (3.118) a (3.119) sa prekrývajú s pravidlami (3.114),

(3.115) a (3.116). Z toho vyplýva, že by prípadoch, kde platia obidva, mala byť možná dvojaká výslovnosť. V [33] sa však nič také neuvádza.

Keďže detekcia zdvojených slov je zložitá a prípadov, kde sa prvá časť zloženého slova končí na *f* a druhá začína samohláskou, znelou alebo zvučnou spoluhláskou, nie je veľa (*šéflekár*, *šéfredaktor*), nebolo pre tento prípad definované pravidlo, ale tieto slová boli pridané medzi výnimky.

§87 Dvojaká výslovnosť sa pripúšťa

- keď predchádza samohláska a nasleduje znelá (nie zvučná) spoluhláska a slabikovanie slova, v ktorom sa píše, nie je ustálené. Podľa slabikovania slova sa vyslovuje buď [za:U_ˆdavOk], alebo [za:wdavOk];
- na začiatku slabiky, kde sa píše *v*, ak nasleduje spoluhláska *n*, *ň* a slabičná hranica je pred *v*. Slová spojenia ako *vnem*, *vníknúť*, *vníčka*, *vnútri*, *v nemocnici* vyslovujeme
[vJEm, vJIknU:c, vnUtSka, vnU:trI, vJEmOtsnItsI]
alebo
[f_vJEm, f_vJIknU:c, f_vnUtSka, f_vnU:trI, f_vJEmOtsnItsI].

Slovo *závdavok* bolo zaradené medzi výnimky a pre ostatné prípady boli definované nasledujúce pravidlá:

$$| \mathbf{v}[n] \rightarrow \mathbf{v} \setminus 1 \quad (v - nem) \quad (3.121)$$

$$\rightarrow \mathbf{f}_v \setminus 1 \quad (v - nem) \quad (3.122)$$

$$\mathbf{v} \parallel [n] \rightarrow \mathbf{v} \parallel \setminus 1 \quad (v - nemocnici) \quad (3.123)$$

$$\rightarrow \mathbf{f}_v \parallel \setminus 1 (v - nemocnici) \quad (3.124)$$

Znelostná asimilácia *ch*, *h*

Spoluhlásky [x, G, h\]

§88 Znelostná asimilácia *ch*, *h* sa uskutočňuje dôsledne podľa pravidiel o znelostnej asimilácii v spisovnej slovenčine. Osobitná situácia v znelostnej asimilácii *ch*, *h* vzniká iba tým, že oproti neznelému [x] stoja v slovenskej výslovnosti dve znelé hlásky – [G, h\]. Rozdiel medzi nimi je v spôsobe a mieste artikulácie. Spoluhláska [G] sa od [x] líši iba znelosťou. V spisovnej slovenčine sa nevyslovuje hláska, ktorá by bola hrtanová a neznelá.

V spisovnej slovenčine sa hlásky [x, G, h\] vyslovujú takto:

§89 Ak sa podľa zákonitostí znelostnej asimilácie ocitne na asimilačnom mieste pred znelou spoluhláskou hláska *h*, ako v slovách a spojeniach *juh*, *lúh*, *juh* či *sever*, *lúh páchne*, *roh stola*, *prah prekročil*, vyslovujeme

[jUx, IU:x, jUxtSIsEvEr, IU:xpa:xJE, rOxstOla, praxprEkrOtSI] atď.

§90 Ak sa však podľa zákonitostí znelostnej asimilácie ocitne na asimilačnom mieste pred znelou hláskou (pred znelou párovou, pred zvučnou, alebo samohláskou) spoluhláska *ch*, ako v slovách a spojeniach *achjaj*, *cechmajster*, *nechže*, *vzduch veľkomesta*, *nech*

ide, kalich a fľaša, z hnedých očí a pod., vyslovuje sa

bud'

[ah\jaL^, cEh\maL^stEr, JEh\ZE, f_vzdUh\vELkOmEsta, JEh\IJ\E, kaLIh\afLaSa, zh\JEdI:h\OtSI:],

alebo

[aGjaL^, cEGmaL^stEr, JEGZE, f_vzdUGvELkOmEsta, JEGIJ\E, kaLIGafLaSa, zhJEdI:GOTSI:] atď.

- §91 Ak sa ocitne *h* v asimilačnom postavení pred znelou hláskou (pred znelou párovou, pred zvučnou alebo pred samohláskou, no nie pred *h*), môže sa vysloviť buď [h\], alebo [G]. V slovách a spojeniach ako *bohdá, bohdaĵ, Bohdan, bohuchovaj, bohvie, onehdajší, onehdy, juh i sever, prah domu, roh ulice, Váh i Hron, vrh guľou, druh vína* vyslovujeme

bud'

[bOh\da:, bOh\daL^, bOh\UxOvaL^, bOh\vL^E, OJEh\daL^SI:, OJEh\dl, jUh\I-sEvEr, prah\dOmU, rOh\ULItsE, va:h\Ih\rOn, vr=h\gULOUL^, drUh\vI:na],

alebo

[bOGda:, bOGdaL^, bOGUxOvaL^, bOGvL^E, OJEGdaL^SI:, OJEGdl, jUGIsE-vEr, praGdOmU, rOGULItsE, va:GIhrOn, vr=GgULOUL^, drUGvI:na] atď.

Pre výslovnosť *ch* a *h* boli definované nasledujúce transkripčné pravidlá:

$$\mathbf{h}, \rightarrow \mathbf{x}, \quad (ju - h) \quad (3.125)$$

$$\mathbf{h} \parallel [NZ] \rightarrow \mathbf{x} \parallel \setminus 1 \quad (ro - h \text{ stola}) \quad (3.126)$$

$$\mathbf{h} \wr [NZ] \rightarrow \mathbf{x} \setminus 1 \quad (dl - h - ší) \quad (3.127)$$

$$\mathbf{ch} \parallel [ZNH] \rightarrow \mathbf{h} \parallel \setminus 1 (ne - ch \text{ ide}) \quad (3.128)$$

$$\rightarrow \mathbf{G} \parallel \setminus 1 (ne - ch \text{ ide}) \quad (3.129)$$

$$\mathbf{ch} \wr [ZNH] \rightarrow \mathbf{h} \setminus \setminus 1 \quad (zmrtvy - ch - vstanie) \quad (3.130)$$

$$\rightarrow \mathbf{G} \setminus 1 \quad (zmrtvy - ch - vstanie) \quad (3.131)$$

$$\mathbf{h} \parallel [ZNH] \rightarrow \mathbf{h} \parallel \setminus 1 (dru - h \text{ vína}) \quad (3.132)$$

$$\rightarrow \mathbf{G} \parallel \setminus 1 (dru - h \text{ vína}) \quad (3.133)$$

$$\mathbf{h} \wr [ZNH] \rightarrow \mathbf{h} \setminus \setminus 1 \quad (bo - h - viečo) \quad (3.134)$$

$$\rightarrow \mathbf{G} \setminus 1 \quad (bo - h - viečo) \quad (3.135)$$

- §92 Ak sa *ch* alebo *h* ocitne v postavení, kde sa v slovenčine uskutočňuje znelostná asimilácia, a bezprostredne po *ch* alebo *h* nasleduje *h*, vyslovuje sa obyčajne [Gh\]. V spojeniach ako *z vysokých hôr, pach hniloby, z umelých hmôt, strach ho donútil, duch Helsínk, Váh hučí, na roh hodil, lúh hryzie kožu, lieh horí, prah hnije* a pod. vyslovujeme obyčajne

[zvIsOkI:Gh\U^Or, paGh\JIIObI, zUmELI:Gh\mU^Ot, straGh\OdOnU:cIl, dUG-h\ElsI:Nk, va:Gh\UtSI:, narOGh\OJ\Il, IU:Gh\rIzI^EkOZU, LI^EGh\OrI:, praG-h\JIjE] atď.

Takéto postavenie sa v slovenčine vyskytuje len na hranici slov, takže iné druhy asimilačných hraníc nebolo potrebné uvažovať a bolo možné definovať nasledujúce transkripčné pravidlo:

$$[\text{chh}] \parallel \mathbf{h} \rightarrow \mathbf{G} \parallel \mathbf{h} \backslash (pa - ch \ h - \text{niloby}) \quad (3.136)$$

V korpuse slovenčiny bolo nájdených niekoľko cudzích slov, v ktorých sa vyskytuje postupnosť *chh*, resp. *hh*. Za zmienku stoja slová *Muhhamad*, *Buchholz*.

- §93 Výslovnosť [h\h\] nie je chybná, no vyžaduje veľa výdychového prúdu, je menej zreteľná (výrazná) ako výslovnosť [Gh\], ba namiesto programového [h\h\] sa často vyslovuje iba [h\].

Znelostná asimilácia v prevzatých slovách

- §94 Pri používaní, preberaní a zdomácnovaní cudzích slov v slovenčine sa zákonitosti znelostej asimilácie uplatňujú podobne ako v domácich slovách. Znelostná asimilácia je jeden zo spôsobov, ako sa prevzaté slová prispôbujú slovenčine.

Všeobecné pravidlo je takéto:

Ak sa na morfematickej hranici stretne neznelá spoluhláska so znelou hláskou alebo znelá spoluhláska s neznelou, prispôbuje sa predchádzajúca hláska nesledujúcej v znelosti tak, že obidve sú buď znelé, buď neznelé (to je princíp tzv. regresívnej asimilácie). Znelostne rovnaké hlásky sa nemenia.

Z tohto pravidla sú v prevzatých slovách rozličné výnimky. Zmena znelosti alebo neznelosti spoluhlások v prevzatých slovách oproti ich pravopisu sa nedá vždy vysvetliť zákonitosťami znelostnej asimilácie tak, ako sa uvádza v predchádzajúcich častiach.

Pravidlá znelostnej asimilácie v prevzatých slovách

- §95 V prevzatých slovách sa znelostná asimilácia uskutočňuje takto:

- Keď sa na hranici dvoch samostatných slov, medzi ktorými nie je prestávka, na hranici slovných základov v zložených slovách a na hranici predpony a koreňa alebo základu slova stretne znelá párová spoluhláska (alebo skupina takýchto spoluhlások) s neznelou, asimiluje sa znelá spoluhláska na neznelú. Asimilácia sa uskutočňuje aj na konci slova pred prestávkou.
- Keď sa na tomto mieste stretne neznelá spoluhláska (alebo skupina takýchto spoluhlások) so znelou hláskou, t. j. so znelou párovou, so zvučnou alebo so samohláskou, asimiluje sa neznelá samohláska na znelú.
- Znelostná asimilácia sa uskutočňuje aj vo viacslovnom cudzom výraze, v spojení prevzatého (resp. cudzieho) slova s domácim slovom a v zloženom slove aj na hranici slovných základov, ktoré pochádzajú z rôznych jazykov.
- Na hranici koreňa (základu) slova a odvodzovacej prípony zriedkakedy vznikajú podmienky na znelostnú asimiláciu.

- §96 Znelostná asimilácia na hranici samostatných slov a v zložených slovách

Príklady: *ad personam*, *deus ex mechina*, *ex abrupto*, *ex lex*, *ex libris*, *ex offa*, *vis-a-vis*;
achát a tópas, *blamáž to bola*, *džes počúvam*, *džús vypiť*, *fjord sa tiahne*, *flox rastie*,

kontakt so všetkými, kontext je neznámy, kurz potrvá, plus-mínus, prax a teória, reflex na svetlo;

avantgarda, baletmajster, basbarytón, basketbal, futbal, fyzikultúra, jurisdikcia, kvin-takord, luzmeter, maršruta, politbyro, rádiusvektor, septakord, sexbomba, transakcia, transalpský, wathhodina, wattmeter. Vyslovuje sa

[atpEso:nam, dEUzEgzmaxI:na, EgzabrUptO, EgzIeks, EgzIibrIs, EgzOfO, vIzavI; axa:tatOpa:s, blama:StObOla, dZEspOtSU:vam, dZU:zvIpIc, fl^OrtsacL^ah\JE, flOgz-rascL^E, kOntagdzOfSEtkI:mI, kOntEgzdjEJEzna:mI, kUrspOtr=va:, plUzmI:nUs, pragzatEO:rIa, rEflEgznasvEtO;

avandgarda, balEdmaL^stEr, bazbarItO:n, baskEdbal, fUdbal, flskUltU:ra, jUrIz-dIktsIa, kvIndakOrt, IUgzmEtEr, marZrUta, pOIIdbIrO, ra:diUzvEktOr, sEbdakOrt, sEgzbOmba, tranzaktsIa, tranzalpskI:, vadh\OJ\Ina, vadmEtEr] atď.

§97 Znelostná asimilácia na hranici predpony a základu slova

Príklady: *abstrahovať, abstrakcia, absurdnosť, exhaustor, disharmónia, disjunkcia, dislokácia, dysbázia, dysgenéza, exhalácia, obscénny, observetórium, postdatovať, postembryonálny, postglaciál, postgraduálny, postrealistický, retranslácia, retranslátör, subfebrilný* atď. Vyslovuje sa

[apstrah\Ovac, apstraktsIa, apsUrduOsc, Egzh\au^stOr, dIzh\armO:nIa, dIzjUNkt-sIa, dIzlOka:tsIa, dIzba:zIa, dIzgEnE:za, Egzh\ala:tsIa, OpsE:nnI, OpsErvat=OrIUm, pOzdatOvac, pOzdembrIOna:lnI, pOzdgIatsIa:l, pOzdradUa:lnI, pOzdrEallstItsKI:, rEtranzla:tsIa, rEtranzla:tOr, sUpfEbrInI:] atď.

Vzhľadom k vyššie uvedeným problémom pre tieto asimilačné pravidlá neboli definované žiadne transkripčné pravidlá a uvedené slová boli pridané medzi výnimky.

Znelostná asimilácia v skupine dvoch a viacej spoluhlások vnútri slova

§98 V prevzatých slovách sa znelostná asimilácia uskutočňuje aj vtedy, keď sa vnútri slova stretnú dve alebo tri spoluhlásky, ktoré v domácich slovách nemôžu stáť vedľa seba. Tak je to v slovách *angström, augsburský, Aztékovia, Desdemona, ekzém, fosgén, gangster, Habsburg, krepdešín, lesbický, Lesbos, plebs, presbyter, presbyterián, rotbartka, röntgen, synekdocha, wartburg, Wolfgang* a pod. Vyslovujeme

[aNkstrOm, aU^gzbUrskI:/ aUgzbUrskI:, astE:kOvL^a, dEzdEmOna/dEzdEmO:na, EgzE:m, fOzGE:n, gaNkstEr, h\abzbUrK, krEbdESI:n, lEzbItsKI:, lEzbOs, plEps, prE-zbItEr, prEzbItErla:n, rOdbartka, rOndgEn, sInEgdOxa, vardbUrK, vOlf_vgaNk] atď.

§99 Prevzaté slová so spoluhláskou s vnútri slova a niektoré iné prevzaté slová

Namiesto neznelého [s] sa vyslovuje znelé [z] v niektorých prevzatých slovách v koreni medzi dvoma samohláskami, napr. v slovách *hosana*, resp. *hosanna, mesalína, resorbovať, resorpcia, resumé, visáž*. Namiesto s sa v prevzatých slovách vyslovuje z aj v iných pozíciách, napr. v slovách *maestoso, spiritiuoso*. Vyslovujeme

[h\Ozana/h\Ozanna, mEzall:na, rEzOrbOvac, rEzOrptsIa, rEzUmE:, vIza:S, ma-EstO:zO, spIrltUO:zO].

Keďže v prípade výslovnosti slov *hosana*, *hossana* sa v bežnej reči vyslovuje hlavne [h\Osana/h\Osanna], čiže so [s], bola táto forma výslovnosti pridaná medzi výnimky.

- §100 V prevzatých slovách, v ktorých sa pôvodne písalo *ss*, sa vo väčšine prípadov zachovala výslovnosť so *s*, aj keď sa takéto slová píšu dnes v slovenčine iba s jedným *s*. Sú to napríklad slová *asimilácia*, *asistent*, *cesia*, *esencia*, *fisúra*, *komisia*, *komposesor*, *koncesionár*, *konfesiónálny*, *oposum*. Vyslovujeme

[asImIla:tsIa, asIstEnt, tsEsIa, EsEntsIa, flsU:ra, kOmIsIa, kOmpOsEsOr, kOntsEsIOna:r, kOnfEsIOna:lnI, OpOsUm].

V slovách, v ktorých podľa dnešnej pravopisnej normy slovenčiny píšeme *ss*, vyslovujeme [s]. Píšeme napríklad *glissando*, *pianissimo*, vyslovujeme

[gIIsandO, pIanIsImO].

- §101 V mnohých slovách sa za pôvodné *s* (zriedkavejšie aj *ss*) dnes už píše *z* v súlade so spisovnou výslovnosťou. Tak je to v slovách *burza*, *cenzor*, *dezertér*, *deziderátum*, *ezižnovať* *dizertácia*, *dorzálny*, *exkurzia*, *kauzálny*, *klauzula*, *konverzácia*, *konzul*, *kozmos*, *kríza*, *múza*, *múzeum*, *muzikant*, *muzikológia*, *musulman* (pripúšťa sa ešte aj tvar *musulman*), *nazalizácia*, *nazálny*, *perzekúcia*, *senzitívny*, *senzualizmus*, *senzus* atď.

Tieto slová neboli pridané medzi výnimky, keďže transkripčné pravidlá pre domáce slová by mali generovať správnu výslovnosť aj pre túto skupinu slov.

- §102 V prevzatých slovách sa namiesto písaného *s* vyslovuje [z] v mnohých slovách, v ktorých sa *s* nachádza po zvučnej spoluhláske. Tak je to v slovách *personálny*, *garsoniéra*, *konsonancia*, *konsonant*, *konsonovať*, *in extenso*, *nelson*, *persifláž*, *persóna*, *personál*, *personalista*, *personalizmus*, *personifikovať*, *skansen* (pripúšťa sa už aj písanie *skanzen*), *šanson*, *šansoniér*, *šansoniérka*, *tyrsus*, *versus*, *vice versa* atď. Vyslovujeme

[pErzOna:lnI, garzOJL^Era, kOnzOnantsIa, kOnzOnant, kOnzOnOvac, InEkstEnzO, nElzOn, pErzÍfla:S, pErzO:na, pErzOna:l, pErzOnalIsta, pErzOnalIzmUs, pErzOnIfIkOvac, skanzEn, SanzO:n, SanzOJL^Er, SanzOJL^Erka, tIrzUs, vErzUs, vItsEvErza] atď.

V niektorých slovách sa pripúšťa aj výslovnosť so [s]: [perso:na, persona:l, persona:lni, personalista, personalizmus, Sanso:n].

- §103 V základnom tvare slovo *blankvers* vyslovujeme [s], v ostatných tvaroch (o blankverse, blankversom atď.) vyslovujeme [z]. Vyslovujeme

[blaNkvErs],

ale

[blaNkvErzU, OblaNkvErzE, blaNkvErzOm] atď.

§104 Vo všetkých tvaroch slova *dres* (dres, dresu, dresom atď.) vyslovujeme však iba [s]. Vyslovujeme

[drEs, drEsU, drEsOm] atď.

V slovách *jansenizmus*, *Komsomol* sa tiež vyslovuje [s]:

[jansEnIzmUs, kOmsOmOl].

§105 V slovách s opačným poradím spoluhlásky *s* a zvučnej spoluhlásky, ako napr. v slovách *islam*, *Island*, *osmóza*, *osram*, sa vyslovuje [s]. Vyslovujeme

[Islam, Islant, OsmO:za, Osram].

V slove *kvizling* sa však ustálila výslovnosť [kvizliNk].

§106 V zložených slovách sa na hranici slovných základov, keď neznelá spoluhláska predchádza pred zvučnou spoluhláskou, ako v slovách *finišman*, *pivotman*, *pivotmanka*, *športsmanský*, sa znelostná asimilácia neuskutočňuje. Vyslovuje sa

[fInISmEn, pIvOtman, pIvOtmaNka, SpOrtsmEskI:].

Slová *bundeswehr* (*Bundeswehr*), *businessman* sa vyslovujú [bUndEsvE:r, bIznIsmEn] alebo [bUndEzvE:r, bIznIzmEn].

V slove *kongresman* je spisovná výslovnosť [kOngrEzman] i [kOngrEzmEn].

§107 V slovách ako *makrosómia*, *nanosómia*, *pansofia*, *persulfát* sa znelostná asimilácia neuskutočňuje, lebo tu morfeimaická hranica prechádza medzi samohláskou a neznelou spoluhláskou, resp. medzi zvučnou spoluhláskou a *s* (makro-sómia, pan-sofia). Vyslovuje sa

[makrOsO:mIa, nanOsOmIa, pansOfIa, pErsUlfa:t].

Slovo *unisono* sa však vyslovuje [unIzO:nO].

§108 Treba rozlišovať latinskú predponu *di* a latinskú resp. grécku predponu *dis-*, *dys-*. V slovách s predponou *di-* sa regresívna znelostná asimilácia obyčajne neuskutočňuje. Slová *dismilácia*, *disimilovať*, *disolúcia*, *disonancia* vyslovujeme [disimilAtsia, disimilovac, disolu:tsia, disonantsia].

Predchádzajúce prípady boli zaradené medzi výnimky.

§109 V slovách s predponou *dis-*, *dys-* sa znelostná asimilácia uskutočňuje, ak nasleduje znelá hláska. Slová *disharmónia*, *disjunkcia*, *dislokácia*, *dysbázia*, *dysgenéza*, *dyslália* atď. vyslovujeme

[dizharmo:nia, dizjuNktsia, dizloka:tsia, dizba:zia, dizgene:za, dizla:lia] atď.

Keďže slovenčina neobsahuje domáce slová začínajúce na *dis-*, *dys-*, bolo možné pre tento prípad definovať transkripčné pravidlo:

$$\| \mathbf{d[iy]s}(ZNH) \rightarrow \mathbf{dIz}\backslash 1(di - s - junkcia) \quad (3.137)$$

Slov s predponou *dys-* sa v slovenčine vyskytuje málo, avšak zastúpenie predpony *dis-* je časté.

§110 V mnohých slovách takéhoto pôvodu a zloženia píšeme už v slovenčine predponu zhodne s výslovnosťou. Píšeme *dysartria*, *dysopia*, *dyszozmia*, *dyszúria*.

Prevzaté slová, v ktorých sa píše *x*

§111 V cudzích a prevzatých slovách sa písmeno *x* číta (vyslovuje) [gz] v spojení *ex-*, ak *ex-* stojí na začiatku slova a ak nasleduje samohláska, znelá alebo zvučná spoluhláska. Tak sa *x* číta (vyslovuje) aj na konci slova, ak nenasleduje prestávka a ak sa nasledujúce slovo začína znelou alebo zvučnou spoluhláskou alebo samohláskou. Slová a spojenia ako *exámen*, *exemplár*, *exhalát*, *exil*, *existovať*, *exminister*, *exogénny*, *exoparazit*, *exoterický*, *exotický*, *exulant*, *ex lex* a pod., tak aj *box* a *lyžovanie*, *prax* má vyslovujeme

[agza:mEn, EgzEmpIa:r, Egzh\ala:t, EgzIl, EgzIstOvac, EgzmInIstEr, EgzOgE:nnI, EgzOparazIt, EgzOtErItskI:, EgzOtItskI:, EgzUlant, EgzIEks; bOgzalIZOvanL^E, pragzma:] atď.

Na základe vyššie uvedeného boli definované nasledujúce transkripčné pravidlá:

$$\| \mathbf{ex}(SZNS) \rightarrow \mathbf{Egz}\backslash 1(ex - \textit{emplár}) \quad (3.138)$$

$$\mathbf{x} \| (SZNS) \rightarrow \mathbf{gz}\backslash 1 (\textit{pra} - \textit{xmá}) \quad (3.139)$$

SZNS označuje samohlásku, znelú alebo zvučnú (sonórnu) spoluhlásku.

§112 Písmeno *x* čítame (vyslovujeme) ako [ks] v ostatných prípadoch:

- pred neznelou spoluhláskou v slovách ako *dextrín*, *expanzia*, *explózia*, *exponovaný*, *juxtapozícia*, *text*, *textil*;
- na konci slova pred prestávkou (*bórax*, *box*, *prax*);
- na konci slova, ak nasledujúce slovo začína na neznelú spoluhlásku (*box trvá*, *prax v škole*);
- na absolútnom začiatku slova (*xantipa*, *Xaver*, *xenón*, *xerografia*, *xylofón*);
- vnútri slova, ak po *x* nie je morfeatická hranica a ak *x* nestojí po začiatočnom *e* (*epoxid*, *lexéma*, *oxidácia*, *oxidovať*, *plexisklo*);
- ak *x* stojí pred slovotvornou alebo ohýbacou príponou, ako v slovách *bóraxový*, *boxer*, *boxovať*, *epizeuxa*, *fixatív*, *fixný*, *reflex*, *taxa*, *taxe*, *taxou* atď.

Vyslovuje sa

[dEkstrI:n, EkspanzIa, EkspLO:zIa, EkspOnOvanI:, jUkstapOzI:tsIa, tEkst, tEkstII;
bO:raks, bOks, praks;
bOkstr=va:, praksfSkOLE;
ksantIpa, ksavEr, ksEnO:n, ksErOgrafIa, ksIIOfO:n;
EpOksIt, lEkSE:ma, OksIda:tsIa, OksIdOvac, plEksIsklO;
bO:raksOvI:, bOksEr, bOksOvac, EpIzEUksa, flksatI:U_ˆ, flksnI:, rEfIkEs, taksa,
taksE, taksOU_ˆ] atď.

Pre ostatné prípady bolo možné definovať:

$$\mathbf{x} \rightarrow \mathbf{ks}(bo - x - er) \quad (3.140)$$

Samozrejme, pre správny prepis je potrebné dodržať správne poradie aplikovania transkripčných pravidiel (3.138), (3.139) a (3.140).

Prevzaté slová a tvary, v ktorých sa znelostná asimilácia neuskutočňuje

§113 Znelostná asimilácia sa neuskutočňuje:

- v prevzatých slovách, v ktorých *s*, *š* (resp. aj iné neznelé spoluhlásky) stoja vnútri slovotvornej prípony alebo bezprostredne pred slovotvornou príponou (*flautista*, *kongresista*, *parnasizmus*);
- pred pádovou príponou a pred koncovou samohláskou podstatného mena (*praxe*, *praxi*, *praxou*, *adresa*, *baronesa*, *melasa*, *noblesa*, *pirueta*, *rasa*, *rímša*). Pre výslovnosť slov s predchádzajúcou zvučnou spoluhláskou môže byť rozhodujúca tendencia vysloviť celú skupinu znelo;
- v nesklonných slovách končiacich sa na *é* (*matlasé*, *plisé*);
- v niektorých slovách končiacich sa na *-us*, *-os*, *-as*, *-es* (*lapsus*, *Pegasus*, *Xerxes*);
- v prevzatých slovách končiacich sa na *-ér* (*necesér*, *režisér*).

Vyslovujeme

[flaUtIsta/flaU_ˆtIsta, kONgrEsIsta, parnasIzmUs;
praksE, praksI, praksOU_ˆ;
adrEsa, barOnEsa, mElasa, nOblEsa, pIrUEta, rasa, rI:msa;
matlasE:, plIsE:;
lapsUs, pEgasUs, ksErksEs;
nEsEsE:r, rEZIsE:r].

Tieto slová boli pridané medzi výnimky.

3.4.9 Zdvojené spoluhlásky

§114 Zdvojené spoluhlásky sa v spisovnej slovenčine vyslovujú na morfematických hraniciach, a to na hraniciach slov, medzi ktorými nie je prestávka, na hranici slovných základov v zložených slovách, na hranici predpony a koreňa (základu) slova, na hranici koreňa (základu) slova a prípony. Vnútri slova sa zdvojené spoluhlásky vyskytujú

pri stretaní dvoch spoluhlások. V skupinách troch a viacej spolhlások zdvojené spoluhlásky vznikajú zriedkavo. Tu nastávajú rozličné zjednodušenia spoluhláskových skupín.

Jestvujú dva typy (stupne) splývania hlások do zdvojených spoluhlások – tzv. voľné a tesné spojenie. Voľné spojenie sa vyskytuje najmä na hranici predpony a základu slova, na hranici slov a osobitne vo vyššom štýle výslovnosti.

Zdvojené spoluhlásky namiesto dvoch spoluhlások

§115 V spisovnej slovenčine sa zdvojené spoluhlásky vyslovujú vtedy, keď sa vnútri slova na hranici morféme medzi dvomi samohláskami, ale aj na hranici samostatných slov stretnú dve rovnaké spoluhlásky (*nn, tt, tṭ, šš* atď.), alebo spoluhlásky, ktoré sa odlišujú iba znelosťou (*bp, dt, dṭ, zs, žš* atď.), alebo aj mäkkosťou (to sú dvojice *tṭ, tḍ, dṭ, dḍ, tṭ, tḍ, dṭ, dḍ*). Slová a spojenia ako *Anne, dennodenne, chyťte, mäkký, oddych, päťtisíc, pollitrový, rozzúriť, vyšší, vyššie, zákonný, zákonne, zlomme; bezsemenný, nižší, nižšie, oddelenie, oddiel, odtade, päťtonový, podte, rozsúdiť, subpolárny; podťať, podtiahnuť, odtekať, odtiaľ; váš žiak, voz stojí, pár rokov, nad tunelom, mám málo, pred tebou* a pod. vyslovujeme

[aJJE, J\EnnOJ\EJJE, xIccE, mEkkI:, OddIx, pEccIsI:ts, pOllItrOvI:, rOzzU:rIc, vISSI:, vISSI_ˆE, za:kOnnI:, za:kOJJE, zLOmmE; bEssEmEnnI:, JISSI:, JISSI_ˆE, OJ\J\ELEJI_ˆE, OJ\J\I_ˆEl, OttaJ\E, pE[tt/ct]OnOvI:, pOccE, rOssU:J\Ic, sUppOla:rnI; pOccac, pOcI_ˆah\nUc, OccEkac, OccI_ˆaL; va:ZZI_ˆak, vOstOjI:, pa:rrOkOU_ˆ, nattUnElOm, ma:mma:lO, prEccEbOU_ˆ] atď.

Zdvojené spoluhlásky vyslovujeme aj v cudzích a prevzatých slovách ako *citissimo, commodo, Edda, larghetto, Mekka, motto, netto* a pod. Vyslovujeme

[tSIItIssImO, kOmmO:dO, Edda, largEttO, mEkka, mOttO, nEttO] atď.

V slovách *erráta, espresso* vyslovujeme jednoduché spoluhlásky. Vyslovujeme [Era:ta, EsprEsO].

Slová *erráta, espresso* boli pridané medzi výnimky a pre ostatné prípady bolo definované:

$$\begin{aligned} (SM)(\mathbf{SP}) \wr \mathbf{2}(SM) &\rightarrow \setminus 1[\mathbf{2}][\mathbf{2}]\setminus 3(ro - zz - \acute{u}riť) & (3.141) \\ (SM)\mathbf{NZ} \wr (\mathbf{ZNp})(SM) &\rightarrow \setminus 1[\mathbf{2}][\mathbf{2}]\setminus 3(ho - td - og) & (3.142) \\ (SM)\mathbf{ZN} \wr (\mathbf{NZp})(SM) &\rightarrow \setminus 1[\mathbf{2}][\mathbf{2}]\setminus 3(o - dt - ial) & (3.143) \\ (SM)[\mathbf{td}] \wr \mathbf{t}(SM) &\rightarrow \setminus 1\mathbf{cc}\setminus 2 \quad (po - dt - at) & (3.144) \\ (SM)[\mathbf{td}] \wr \mathbf{d}(SM) &\rightarrow \setminus 1\mathbf{J}\setminus \mathbf{J}\setminus 2 \quad (o - dd - iel) & (3.145) \\ (SM)[\mathbf{td}^*] \wr \mathbf{t}(SM) &\rightarrow \setminus 1\mathbf{tt}\setminus 2 \quad () & (3.146) \\ (SM)[\mathbf{td}^*] \wr \mathbf{d}(SM) &\rightarrow \setminus 1\mathbf{dd}\setminus 2 \quad () & (3.147) \end{aligned}$$

kde znak \wr označuje morfematickú hranicu, ZNp – znelú párovú a NZp – neznelú párovú. Hranaté zátvorky na pravej strane znamenajú fonetický prepis prísluš-

nej spoluhlásky. Časť z vyššie definovaných pravidiel je redundantná, keďže ich pokrývajú všeobecné pravidlá pre asimiláciu.

§116 Zdvojené *j* sa v spisovnej slovenčine nevyskytuje. V slovách a spojeniach ako *dvojje-diný, dvojazyčný, dať jej jeden* sa prvé *j* vyslovuje ako hláska [I^ˆ]. Vyslovujeme

[dvOI^ˆjEJ\InI:, dvOI^ˆjazItSnI:, daI^ˆjEI^ˆjEdEn] atď.

$$(SM)j \wr j(SM) \rightarrow \backslash 1I_{-}^{\wedge} \wr j \backslash 2 (dvo - jj - azyčny) \quad (3.148)$$

$$(SM)j \parallel j(SM) \rightarrow \backslash 1I_{-}^{\wedge} \parallel j \backslash 2 (da - jj - e - jj - eden) \quad (3.149)$$

Aj v tomto prípade sa pravidlá prekrývajú. Pravidlo (3.149) totiž vyplýva z všeobecných pravidiel o výslovnosti *j* a I^ˆ, a preto ho nie je nutné implementovať.

§117 Zriedka sa vyskytuje aj zdvojené *h*.

§118 Ak je dvojica rovnakých spoluhlások alebo spoluhlások, ktoré sa odlišujú znelosťou alebo mäkkosťou, v susedstve ďalšej spoluhlásky, nastávajú rozličné zjednodušenia.

§119 Na hranici samostatných slov, najmä vo vyššom štýle, sa môžu vysloviť aj dve úplné spoluhlásky.

V zložených slovách na hranici slovných základov sa tiež pripúšťa aj výslovnosť [pEctOnOvI:, pEctaktOvI:, pEJ\dEkOvEts, J\EvEJ\dObI:] atď., t. j. výslovnosť so zachovaním skupín *tt, dd*.

Tieto slova boli zaradené medzi výnimky.

§120 Zdvojené spoluhlásky vznikajú v skupine dvoch spoluhlások na hranici koreňa (základu) slova a odvodzovacej prípony, ak sa tu stretne niektorá zo spoluhlások *t, d, ĺ, đ* so sykavou úžinovou alebo záverovoužinovou spoluhláskou. Ak je na druhom mieste *s, z, c, dz*, vznikne zdvojené *c* alebo *dz*, ak je na druhom mieste *š, ž, č, dž*, vznikne zdvojené *č* alebo *dž*. Znelosť alebo neznelosť zdvojenej spoluhlásky závisí od zákonitosti znelostnej asimilácie. Slová ako *otec - otca, štetec - štetca, zajatec - zajatca, vedec - vedca* (podobne *vedcovi, vedcom, vedci* atď.), *rozhodca, sudca, porotca, platca, dobytčí, radšej, všetci*, ale aj slová, v ktorých je už morfológická hranica zastretá, ako sú *dvadsať, tridsať, štyridsať, predsa, Bytča, Čadca*, vyslovujeme

[Ottsa, ScEttsa, zajatsta, vEttsa, vEtstsOvI, vEtstsOm, vEtstsI, rOzh\Ottsa, sUttsa, pOrOttsa, platsta, dObItStSI:, ratStSEI^ˆ, fSEtstsI, dvatstsac, trItstsac, StIrItstsac, prEttsa, bItStSa, tSatsta] atď.

$$[tdđd] \sim [szcdz] \rightarrow ts \sim ts (zaja - tc - a) \quad (3.150)$$

$$[tdđd] \sim [szcdz] \rightarrow dz \sim dz (o - dz - voň) \quad (3.151)$$

$$[tdđd] \sim [šžčdž] \rightarrow tS \sim tS (By - tč - a) \quad (3.152)$$

$$[tdđd] \sim [šžčdž] \rightarrow dZ \sim dZ (štvr - tž - blnk) \quad (3.153)$$

kde '˘' označuje hranicu odvodzovacej prípony alebo predpony.

§121 V číslovkách (dvadsať, tridsať atď.) sa v neutrálnom štýle pripúšťa aj zjednodušená výslovnosť, t.j.

[dvatsac, trItsac, StIrItsac, pEJ\EsIˆat, J\EvEJ\EsIˆat, pEcIsI:ts, J\EsacIsI:ts].

Tento spôsob výslovnosti bol pridaný medzi výnimky rovnako ako aj nasledujúce slová

Hodža, srdce.

§122 V priezvisku *Hodža* a v slove *srdce* (*srdcový, osrdcovník*) sa tiež vyslovujú zdvojené spoluhlásky:

[hOdZdZa, sr=tstsE, sr=tstsOvi:, Osr=tstsOvJi:k/Osr=tstsOUˆJI:k].

Pripúšťa sa aj výslovnosť [hOdZa].

§123 Zdvojené spoluhlásky vznikajú na hranici predpony a koreňa (základu) slova a na hranici slovných základov v zložených slovách, ak sa tu medzi dvoma samohláskami vyskytne niektorá zo spoluhlások *t, d, ě, d'* pred záverovoužinovou spoluhláskou *c, dz, č, dž*. Slová ako *desaťciferný, nadčelný, odcestovať, odcudziť, odčerpať, odčiniť, odčítať, päťciferný, päťčiarkový, podceňovať, podčiarknúť, predčítavať* a pod. vyslovujeme

[J\EsatstsIfErnI:, natStSElnI:, OtstsEstOvac, OtstsUdzIc, OtStSErpac, OtStSIJic, OtStSI:tac, pEtstsIfErnI:, pEtStSIˆarkOvI:, pOtstsEJOvac, pOtStSIˆarknUc, prEtStSI:tavac] atď.

Pre tento prípad nebolo potrebné definovať nové transkripčné pravidla. Stačilo rozšíriť pravidlá (3.150) až (3.153) o morfeomatickú hranicu slovných základov v zložených slovách.

§124 Pre stretaní sa takýchto spoluhlások na hranici samostatných slov sa tiež môže (ale nemusí) vysloviť zdvojená spoluhláska.

[**tdfd'**] || [szcdz] → **tsts** (*o – d s – vojich*) (3.154)

→ **ts** (*o – d s – vojich*) (3.155)

→ **dzdz** (*naháňa – ě s – trach*) (3.156)

→ **dz** (*naháňa – ě s – trach*) (3.157)

[**tdfd'**] || [šžčdž] → **tStS** (*pre – d č – asom*) (3.158)

→ **tS** (*pre – d č – asom*) (3.159)

→ **dZdZ** (*pre – d š – kandáлом*) (3.160)

→ **dZ** (*pre – d š – kandáлом*) (3.161)

Programová implementácia nevyslovenia zdvojenej samohlásky je problematická, keďže je potrebné riešiť otázku, odkiaľ má byť v prípade generovania výslovnosti pre jednotlivé slová danej vety zdvojená spoluhláska odtrhnutá. V implementácii bola teda uvažovaná len zdvojená výslovnosť.

§125 V zložených slovách na hranici slovných základov sa pripúšťa aj výslovnosť [pEctSI^ˆErnI:, pEctSI^ˆarkovI:], no vyslovuje sa [SEstSI^ˆErnI:, SEstSI^ˆEnnI:].

§126 Ak sa na hranici predpony a koreňa (základu) slova stretne spoluhláska *d* (resp. v znelostnej asimilácii *t*) s úžinovou sykavou spoluhláskou *s*, *dz*, *č*, *dž*, alebo sa uskutoční iba znelostná asimilácia a zdvojená spoluhláska sa nevyslovuje.

Slová ako *nadšiť*, *odsúdiť*, *odsunúť*, *odšumiť*, *odzubadiť*, *odžať*, *pod sada*, *podsekretár*, *podsiať*, *podsunúť*, *podšiť*, *podšívka*, *podzáмок*, *podzemnica*, *podžierať*, *predsieň*, *predsudok*, *predzápas* a pod. vyslovujú sa v neutrálnom štýle

bud' [natStSIc, OtstsU:J\Ic, OtstsUnU:c, OtStSUMI^ˆEc, OdZdZUbadLIc, OdZdZac, pOtstsada, pOtstsEkrEta:r, pOttsaI^ˆac, pOtstsUnU:c, pOtStSIc, pOtStSI:U^ˆka, pOdZdza:mOk, pOdZdZEmJI^ˆtsa, pOdZdZI^ˆErac, prEtstsI^ˆEJ, prEtstsUdOk, prEdZdza:pas] atď.,

bud' [natSIc, Ot_sU:J\Ic, Ot_sUnU:c, Ot_sUMI^ˆEc, Od_zUbadLIc, Od_Zac, pOt_sada, pOt_sEkrEta:r, pOt_saI^ˆac, pOt_sUnU:c, pOt_SIc, pOt_SI:U^ˆka, pOd_za:mOk, pOd_zEmJI^ˆtsa, pOd_ZI^ˆErac, prEt_sI^ˆEJ, prEt_sUdOk, prEd_za:pas] atď.

Tieto slová boli pridané medzi výnimky rovnako ako aj slovo *dcéra*.

§127 Slovo *dcéra* vyslovujeme [tse:ra].

§128 V niektorých prípadoch sa zdvojená spoluhláska vyslovuje vtedy, keď sa na morfeomatickej hranici slova stretnú záverovoúžinová a úžinová spoluhláska, ktoré sa neodlišujú miestom artikulácie (napr. *čš*), ako je to v slovách *väčší*, *väčšina*, *väčšinou*. Vyslovujeme

[vEtStSI:, vEtStSI^ˆna, vEtStSI^ˆnoU^ˆ].

Tieto slová sme pridalí medzi výnimky.

Výslovnosť so zdvojenou spoluhláskou sa pripúšťa aj v zložených slovách ako *viacsemenný*, *viacsýtny*. Može sa vyslovovať [vI^ˆatstsmEnnI:, vI^ˆatstSI:tnI].

Častejšia je však výslovnosť so zachovaním obidvoch spoluhlások, t. j. [vI^ˆatssEmEnnI:, vI^ˆatssI:tnI].

Obidva druhy výslovnosti boli zaradené medzi výnimky.

§129 Ak vedľa spoluhlások *čš* stojí ďalšia spoluhláska (t. j. v skupine troch spoluhlások), zdvojená spoluhláska nevzniká, lež vypadáva *š*. Tak je to v slovách *väčšmi*, *najväčšmi*.

Vyslovujeme [vEtSmI, naI^ˆvEtSmI].

Tieto slová boli pridané medzi výnimky.

Zdvojené spoluhlásky na hranici samostatných slov

§130 Na hranici samostatných slov sa zdvojené spoluhlásky môžu vyslovovať iba vtedy, ak sa tu stretnú dve rovnaké spoluhlásky, spoluhlásky, ktoré sa odlišujú iba znelosťou alebo aj mäkkosťou, alebo keď sa tu stretne záverová spoluhláska so záverovoúžinovou. V iných prípadoch sa na hranici samostatných slov zdvojené spoluhlásky nevyslovujú.

§131 Na hranici samostatných slov sa zdvojené spoluhlásky nevyslovujú, ak sa tu stretne záverová spoluhláska s úžinovou. Spojenia ako *pod sem, rád som, hrad Šášov, kúpať sa, spievať si* atď. vyslovujeme

[pOcsEm, ra:tsOm, hratSa:SOU₂[^], kU:pacsa, spL₂[^]EvacsI] atď.

$$(SM)(SP) \parallel \backslash 2(SM) \rightarrow \backslash 1[\backslash 2] \parallel [\backslash 2]\backslash 3(má - mm - álo) \quad (3.162)$$

$$(SM)NZ \parallel (ZNp)(SM) \rightarrow \backslash 1[\backslash 2] \parallel [\backslash 2]\backslash 3(ku - sz - eme) \quad (3.163)$$

$$(SM)ZN \parallel (NZp)(SM) \rightarrow \backslash 1[\backslash 2] \parallel [\backslash 2]\backslash 3(vo - zs - toji) \quad (3.164)$$

$$(SM)[td] \parallel t(SM) \rightarrow \backslash 1c \parallel c\backslash 2 \quad (o - d t - archy) \quad (3.165)$$

$$(SM)[td] \parallel d(SM) \rightarrow \backslash 1J \parallel J\backslash 2 \quad (živo - t ďalej) \quad (3.166)$$

$$(SM)[td] \parallel t(SM) \rightarrow \backslash 1t \parallel t\backslash 2 \quad (ve - d' t - ento) \quad (3.167)$$

$$(SM)[td] \parallel d(SM) \rightarrow \backslash 1d \parallel d\backslash 2 \quad (vypáli - t d - om) \quad (3.168)$$

Ani v tomto prípade nebolo potrebné implementovať všetky pravidlá, keďže niektoré vyplývajú zo skôr definovaných pravidiel.

§132 Zriedka sa vyslovuje zdvojené *h* [hh]. Ak by takýto prípad mal nastať, namiesto [hh] sa obyčajne vyslovuje skupina [Gh]. Spojenie ako *Váh hučí, z umelých hmôt, duch Helsínk, lieh horí* vyslovujeme

[va:GhUtSI:, zUmEI:GhmU₂[^]Ot, dUGhElsINk, LI₂[^]EGhOrI:].

$$(SM)h \parallel h(SM) \rightarrow \backslash 1Gh\backslash 2(Vá - hh - učí) \quad (3.169)$$

Zdvojené spoluhlásky v zložených slovách

§133 Na hranici slovných základov v zložených slovách sa zdvojené spoluhlásky záväzne vyslovujú iba vtedy, ak sa stretnú dve rovnaké spoluhlásky alebo spoluhlásky líšiace sa iba znelosťou.

Ak sa stretnú spoluhlásky líšiace sa mäkkosťou (*päťtonový, deväťdobý*), alebo ak sa stretne záverová spoluhláska so záverovouúžinovou (*päťciferný, päťčiarkový*), resp. záverovouúžinová s úžinovou (*viacsemenný, viacskýtny*), môže sa vysloviť zdvojená spoluhláska alebo dve samostatné spoluhlásky.

Zdvojené spoluhlásky v skupine troch a viacej spoluhlások

§134 Zdvojené spoluhlásky vznikajú vnútri slova na hranici morfému aj v skupine viac ako dvoch spoluhlások, ak treťou spoluhláskou v skupine je [r, l, r=, l=, v] alebo aj [m, n, J, L]. Pritom sa musia spĺňať ostatné podmienky výslovnosti zdvojených spoluhlások ako v skupine dvoch spoluhlások. Dôležité je tiež, či ide o predponu a základ slova, alebo o základy slova v kompozite.

Zdvojené spoluhlásky sa teda vyslovujú aj v slovách *bezznáškový, desaťčlenný, hrdší, hrdšie, miliardtina, obplieť, odcloniť, odčleniť, predtlač, predtiedny, rozslabikovať,*

*rozslziť, zorsmútiť, rozsnívať, rozsnovať, rozsrdiť, rozsvietiť, rozzvoniť, rozzvučať, tvr-
dší, tvrdšie, urdčina, žltší* a pod. Vyslovujeme

[bEzzna:SkOvI:, J\EsatStSLEnnI:, h\r=tStSI:, hr=tStSI^E, mIIIarccIna,
OppLI^Esc, OtstslOJic, OtStSLEJic, prEttlatS, prEttrI^EdnI, rOsslabIkOvac, rOs-
sl=zIc, rOssmU:cIc, rOssJI:vac, rOssnOvac, rOssr=J\Ic, rOssvI^EcIc, rOzzvOJic,
rOzzvUtSac, tvr=tStSI:, tvr=tStSI^E, UrtStSIa, Zl=tStSI:] atď.

§135 Ak je v zložených slovách v skupine troch spoluhlások, ktoré inak vyhovujú pod-
mienkam vzniku zdvojených spoluhlások, treťou spoluhláskou *r, l, v, m, n, ň*, a ak
nepredchádza *c, dz, č, dž* pred inou znelou spoluhláskou (príklad: *viacznačný*), môže
sa vyslovovať zdvojená spoluhláska. Preto sa môže vyslovovať aj

[pEtstslabItSnI:, pEtstslOvnI:, J\EvEJ\J\JOvI:].

V slove *markgróf*, kde je morfeomatická hranica medzi *k - g* a *a* túto spoluhláskovú dvo-
jicu predchádza a po nej nasleduje *r*, vyslovuje sa zdvojená spoluhláska: [marggrO:f].

Tieto slová boli pridané medzi výnimky.

Pozície, v ktorých nevznikajú zdvojené spoluhlásky

§136 Zdvojené spoluhlásky nevznikajú na hranici slov, ak sa tu nestretnú dve rovnaké
spoluhlásky, spoluhlásky ktoré sa líšia iba znelosťou (resp. mäkkosťou), alebo spolu-
hlásky *t, d, t', d'* so spoluhláskami *c, dz, č, dž*. Spojenia ako *umývať sa, predstaviť si,
šesť stupňov, hoc čakal, veď sme* atď. vyslovujeme

[umI:vacsa, prEtstavIcsI, SEscstUpJoU^, hOtstSakal, vEJ\zmE] atď.

Táto skutočnosť je už ošetrená pravidlami (3.162), (3.163) a (3.164).

§137 Zdvojené spoluhlásky nevznikajú v skupine troch a viacej spoluhlások, ak treťou spo-
luhláskou nie je [r, l, r=, l=, v, m, n, J, L], ako napríklad v slovách *francúzky, odpustte,
podstata, podstatný, predspev, predstava, predstavenie, predstavený, predstaviteľ, pred-
stierať, predstih, predškolsý, pusťte, rozskočiť, rozstreknuť, rozstrieť, rozstrihnúť, vi-
ťazstvo* a pod. Vyslovujeme

[frantsU:skI, OtpUscE, pOtstata, pOtstatnI:, prEtspeU^, prEtstava,
prEtstavEJI^E, prEtstavEnI:, prEtstavIcEL, prEtscl^Erac, prEtsclx, prEtSkOlslI:,
pUscE, rOskOtSIc, rOstrEknU:c, rOstrI^Ekac, rOstrIh\nU:c,
vI:castvO] atď.

Táto skutočnosť taktiež vyplýva z už definovaných pravidiel o výslovnosti zdvo-
jených spoluhlások.

§138 Zdvojené spoluhlásky nevznikajú ani v slovách, v ktorých je už morfeomatická hranica
celkom zastretá, alebo v ktorých sa morfeomatická hranica nepociťuje, ako napríklad
v slovách *odznak, odzemok*. Vyslovujeme

[Odznak, odzEmOk].

Tieto slová boli pridané medzi výnimky.

- §139 V skupine troch a viacej spoluhlások sa spoluhlásková skupina všeobecne zjednodušuje najmä v slovách, ktorých morfológické zloženie sa zastiera (*predstavenie, podstata*).
- §140 Zdvojené spoluhlásky v spisovnej slovenčine nevznikajú, ak sa na morfológickej hranici stretne svetlá sykavá spoluhláska (s, z, c, dz) s temnou sykavou spoluhláskou (š, ž, č, dž), alebo ak sa stretne takáto úžinová spoluhláska (s, z, š, ž) so záverovouúžinovou (c, dz, č, dž). Slová a spojenia ako *rozšíriť, rozžiarit, rozžialit, rozžeravený, raz žil, les šumel, pes šiel, dnes žijú; rozčesať, sčítať, sčesať, sčervenieť, sčervivieť, voz čerešien, hlas čriedy; hocčo, hoc čakal, noc černie, vec čarodejná*, a pod. vyslovujeme

[rOsSI:rIc, rOzZIL^rIc, rOzZIL^aLIc, rOzZEravEnI:, razZIL, LEsSUMEl,
pEsSI^El, dJEzZIJU:; rOstSEsac, stSI:tac, sTsEsac,
stSErvEJL^Ec, stSErvIvL^Ec, vOstSErESI^EJ, h\lastSrI^EdI; hOtstSO, hOtstSa-
kal, nOtstSErJL^E, vEtstSarOJ\EI^na:] atď.

Aj v tomto prípade, podobne ako aj v nasledujúcich, by vytváranie pravidiel viedlo k vytváraniu redundantných komplementárnych pravidiel.

- §141 V spoluhláskových skupinách *stc, stčl, stš, sts, stz, stst, ststr, stst, stsl, stďň, std, stď* sa v neutrálnom štýle pripúšťa výslovnosť s vynechaním spoluhlásky *ť*, resp. *t*.

Ak po *ť, t* nasleduje *s* alebo *z*, vynechaním *ť, t* vzniknú podmienky na výslovnosť zdvojenej spoluhlásky. Vyslovuje sa

[bEssElEr, SEssEdadlOvI:, SEsstUpJOvI:, SEsstranOvI:, SEzza:prax].

Číslovka *šeststo* sa však vyslovuje [Sesto].

Ak v takejto skupine po prvej spoluhláske *s* nasledujú dve záverové spoluhlásky alebo dvojica záverová + záverovouúžinová spoluhláska, po vynechaní *ť, t* nevzniká zdvojená spoluhláska. Vyslovuje sa

[SEzJ\EnnI:, SEzJ\I^ElnI, SEstsIfErnI:, SEzJ\EsI^at, SEzJ\EsI^atka,
SEzdJ\OvI:].

Tieto slová boli pridané medzi výnimky.

- §142 V skupine troch a viac spoluhlások sa v zložených slovách na hranici slovných základov spoluhlásková skupina nemusí zjednodušiť a nemusí sa vysloviť ani zdvojená spoluhláska, a to ani vtedy, keď treťou spoluhláskou je [r, l, r=, l=, v, m, n, J, L]. Je tu podobná situácia ako na hranici dvoch samostatných slov. Preto slová ako *viacznáčny, päťslabičný, päťslovný, desaťsten, šesťdadlový, šesťslabičný, šesťsten, šesťstenný, šesťstranový, šesťstranný* sa môžu vyslovovať

[vI^adzdnatSnI:, pEcslabItSnI:, pEcslOvnI:, J\EsacscEn, SEscstranOvI:, SEscstrannI:]
a pod.

Vyslovujeme však iba

[SEzJ\esI^at, SEzJ\esI^acina, SEzJ\esI^atka, SEzJ\esI^atJik,
SEzJ\esI^adrotSni:, SEzJ\esI^ati] atď.

Na hranici samostatných slov sa však nezjednodušuje ani spoluhlásková skupina [dzzn]. Napríklad spojenie *viac znamená* vyslovujeme [vLˆadzznamena:].

V spisovnej slovenčine sa obyčajne nevyslovuje ani zdvojené h [hh].

3.4.10 Tvrdé a mäkké spoluhlásky

Spoluhlásky *t, d, n, l* – *ť, ď, ň, ľ*

§143 V slovenčine sa z pravopisného hľadiska rozlišujú tvrdé, mäkké a obojaké spoluhlásky. K tzv. tvrdým spoluhláskam sa okrem iných zaraďujú *t, d, n, l* a k mäkkým zasa spoluhlásky *ť, ď, ň, ľ*.

Vo výslovnosti spoluhlások *t, d, n, l* – *ť, ď, ň, ľ* sa v spisovnej slovenčine vyskytuje viacej ortofonických a ortoepických chýb. Prejavujú sa ako nedostatočná mäkkosť spoluhlásky *ľ* a ako chybné používanie tvrdých spoluhlások *t, d, n, l* namiesto mäkkých *ť, ď, ň, ľ* a naopak.

Niektoré z týchto ortoepických chýb súvisia aj so slovenským pravopisom. Platí v ňom totiž pravidlo, že výslovnosť *ť, ď, ň, ľ* sa neoznačuje mäkkčeňom, ak nasleduje *i, í, e, ia, ie, iu*.

Lenže pravidlo o výslovnosti *ť, ď, ň, ľ* pred *i, í, e, ia, ie, iu* neplatí bezvýnimočne. V mnohých domácich slovách vyslovujeme tvrdé *t, d, n, l* aj pred samohláskovými zvukmi *i, í, e, ia, ie, iu* a prevzaté slová sa vo výslovnosti odlišujú od domácich markantne práve tým, že pravidlo o mäkkčení *t, d, n, l* v nich neplatí. A aj tu sú ešte rozličné obmedzujúce pravidlá.

Spoluhlásky *t, d, n, l* – *ť, ď, ň, ľ* v domácich slovách

§144 O výslovnosti *t, d, n, l* – *ť, ď, ň, ľ* v domácich slovách platia tieto pravidlá:

- Na konci slov pred samohláskami *a, o, u, á, ó, ú, é*, pred dvojhláskou *ô* a pred spoluhláskou sa *t, d, n, l* – *ť, ď, ň, ľ* vyslovuje v zhode s pravopisom. V týchto pozíciách sa spoluhlásky *ť, ď, ň, ľ* vyslovujú iba vtedy, ak sa tzv. mäkká výslovnosť označuje v pravopise mäkkčeňom.
- Tam, kde v písme po *t, d, n, l* stojí ypsilon, vyslovuje sa dôsledne *t, d, n, l*, hoci písmeno *y* označuje tú istú hlásku ako písmeno *i*.

§145 Spoluhlásky *ť, ď, ň, ľ* sa v domácich slovách vyslovujú aj pred samohláskovými zvukmi *i, í, e, ia, ie, iu*, hoci tu sa na *ť, ď, ň, ľ* mäkkčeň nepíše.

V slovách *baňa, ďakovať, ňuchať, páľ, poľovať, posuň, ľava*, ale aj v slovách ako *delo, dielo, dívať sa, leto, liek, lievať, list, naproti, nikto, paniu, polia, telo* a pod. vyslovujeme *ť, ď, ň, ľ*.

[baJa, J\akOvac, JUxac, pa:L, pOLOvac, pOsUJ, cava; J\EIO, J\IˆEIO, J\I:vacs, LEtO, LIˆEk, LIˆEvac, LIst, naprOcI, JIktO, PaJIˆU, pOLIˆa, cElO] atď.

Na základe vyššie povedaného bolo možné definovať:

$$(\mathbf{tdnl}\check{\mathbf{d}}\check{\mathbf{n}}\check{\mathbf{l}})([aou\hat{o}SP]) \parallel \rightarrow [\mathbf{1}]\mathbf{2} \quad (ma - l - \acute{e}) \quad (3.170)$$

$$(\mathbf{tdnl})y \rightarrow [\mathbf{1}]I \quad (p - l - yn) \quad (3.171)$$

$$(\mathbf{tdnl}) \rightarrow [\backslash \mathbf{1}]I : (pek - n - ý) \quad (3.172)$$

$$\mathbf{t}([i\acute{e}iaieiu]) \rightarrow \mathbf{c} \backslash \mathbf{1} \quad (napro - t - i) \quad (3.173)$$

$$\mathbf{d}([i\acute{e}iaieiu]) \rightarrow \mathbf{J} \backslash \backslash \mathbf{1} \quad (d - ielo) \quad (3.174)$$

$$\mathbf{n}([i\acute{e}iaieiu]) \rightarrow \mathbf{J} \backslash \mathbf{1} \quad (n - ikto) \quad (3.175)$$

$$\mathbf{l}([i\acute{e}iaieiu]) \rightarrow \mathbf{L} \backslash \mathbf{1} \quad (l - ist) \quad (3.176)$$

$$\rightarrow \mathbf{l} \backslash \mathbf{1} \quad (l - ist) \quad (3.177)$$

V prípade výslovnosti [l/L] bolo okrem pravidla (3.176) definované aj pravidlo (3.177) keďže v nektorých oblastiach Slovenska je mäkká výslovnosť spoluhlásky *l* skôr výnimkou⁹. Pravidlá (3.171), (3.172) sú v implementácii umiestnené pred pravidlom (3.5). Pravidlá (3.173) až (3.177) sú všeobecné pravidlá, z ktorých existuje mnoho výnimiek, opísaných v nasledujúcich častiach. Tieto pravidlá boli v implementácii umiestnené až na koniec za všetky ostatné pravidlá ošetrojúce výslovnosť mäkkých a tvrdých spoluhlások.

Výslovnosť *t, d, n, l* pred *i, í, e, ie* v domácich slovách

§146 Pred *i, í, e, ie*, sa v domácich slovách výnimočne vyslovuje *t, d, n, l* v týchto prípadoch: v tvaroch prídavných mien vzorov *pekný* a *matkin*, v tvaroch niektorých čísloviek a trpných prídavných, ako v slovách a spojeniach *vážený prítomní, milí, stáli, matkini, príbuzní, zaviati, matkinej; malej, stáleho, stálemu, stálej, zaviateho, zaviatemu, zaviatej; jedni, jednej, hodni, vinní, vinnej, žiadni, žiadne, žiadnej, žiadneho, žiadnemu*. Vyslovujeme

[va:ZEnI:prI:tOmnI:, mIII:, sta:II, matkInIprI:bUznI:, zavI.^atI, matkInEI.^;
maIEI.^, sta:lEh\O, sta:lEmU, sta:lEI.^, zavI.^atEh\O, zavI.^atEmU, zavI.^atEI.^;
jEdnI, jEdnEI.^, h\OdnI, vInnI:, vInnEI.^, ZI.^adnI, ZI.^adnE, ZI.^adnEI.^, ZI.^a-
dnEh\O, ZI.^adnEmU].

Tieto slová boli pridané medzi výnimky, zohľadňujúc pri tom možné tvary. Viac o spracovaní výnimiek pojednáva časť 3.4.14.

§147 Spoluhlásky *d, n* vyslovujeme aj v slovách *žaden, hoden, vinen*. Vyslovujeme

[ZI.^adEn, h\OdEn, vInEn].

§148 Spoluhlásky *t, d, n, l* sa pred *i, í, e, ie* vyslovujú aj v tvaroch osobných zámen *ten, tento, onen*, t. j. v slovách *ten, tej, tí, tie, tento, tejto, títo, tieto, onen, onej, oní, oné*. Vyslovujeme

[tEn, tEI.^, tI:, tI.^E, tEntO, tEI.^tO, tI:tO, tI.^EtO, OnEn, OnEI.^, OnI:, OnE:].

Nevyslovujeme však [OnIkac], lež [OJIkac]. Zámeno *oni* vyslovujeme [OJI]. Zámeno *ony* vyslovujeme [OnI]. Chybná je výslovnosť [cI.^E, cI.^EtO].

⁹Rovnaké problémy sú aj s výslovnosťou *de, te, ne*, no keďže ide o menšie odchýlky než v prípade *l*, rieši sa to výnimkami.

Tieto slová boli pridané medzi výnimky. Medzi výnimky bola pridaná takisto aj "chybná" výslovnosť, keďže takáto výslovnosť je v prípade zámen *tie* a *tieto* bežná.

- §149 Spoluhlásky *t, d, n, l* sa vyslovujú v predponách, keď po predponovej spoluhláske nasleduje spoluhláska *i, í, e*, ako v slovách *odinakiaľ, odísť, podísť, predísť, nadísť, odídeme, odišiel, odíď, odíďme, odíďte, odídúc, odistiť, odistil, predizba*, v slovách *nadeň, podeň, predeň* a v zložených slovách typu *polizba*. Vyslovujeme

[OdInak!aL, OdI:sc, pOdI:sc, prEdI:sc, nadI:sc, OdI:J\Em, OdISL^El, OdI:c, OdI:J\mE, OdI:ccE, OdI:dUts, OdIscIc, OdIscIl, prEdIzba; nadEJ, pOdEJ, prEdEJ; pOIzba] atď.

Slová *nadeň, podeň, predeň, polizba* boli pridané medzi výnimky a pre ostatné sú definované nasledujúce pravidlá:

$$(\mathbf{tdnl}) \dashv ([ie]) \rightarrow \backslash \mathbf{1} \backslash 2(o - d - i\text{šiel}) \quad (3.178)$$

- §150 Spoluhlásky *t, d, n, l* sa pred *i, í, e* vyslovujú vo zvukomalebných slovách a citosloviaciach a v slovesách s týmto základom: *cingi-lingi, čvrli-čvrli, čvrlikať, dínom-dánom, fidli-fidli, fidlikať, ne, né, pidlikať, pidlikovať, štidiri, tidlikovať, tikať, tikot, tik-tak, tudlikať, tuli, tulikať, tutuli* a pod. Vyslovujeme ich

[tsINgIIINgI, tSvr=ItSvr=II, tSvr=IIkac, dI:nOmda:nOm, fdlIfIdII, fdlIkac, nE, nE:, pIdIkac, pIdIkOvac, StIdIrI, tIdIkac, tIkac, tIkOt, tIktak, tUdIkac, tUII, tUIIkac, tUtUII] atď.

- §151 Spoluhlásky *t, d, n, l* sa pred *i, í, e* vyslovujú aj v slovách *jeden, jedenásť, okienečko, pijatika, putika, štek, štep, zaštepovať* (vo význame "zošívvať", iné je *štepíť, zaštepíť, štep*), *štyri, teda, temer, terajší, teraz, terč, vínečko, všetečnica, všetečný, vtedajší, vtedy, jún - v júni, júl - v júli*. Vyslovujeme ich

[jEdEn, jEdEna:sc, OkI^EnEtSkO, pIjatIka, pUtIka, StEk, StEp, zaStEpOvac, StIrI, tEda, tEmEr, tEraL^SI:, tEras, tErtS, vI:nEtSkO, fSEtEtSJItsa, fSEtEtSnI:, ftEdaL^SI:, ftEdI, vJU:nI, vJU:II].

Tieto slová boli pridané medzi výnimky. Pre sloveso *zaštepovať* bola uvedená výslovnosť [zaStEpOvac], ale aj [zaScEpOvac], keďže bez kontextu nevieme určiť, o ktorý význam ide.

- §152 V slovách *mestečko, miestečko*, dávame prednosť výslovnosti [mEscEtSkO, mI^EscEtSkO], v slove *poveternostný* a v odvodených slovách a tvaroch pripúšťame výslovnosť [pOvEtErOstnI:] i [pOvEcErnOstnI:], no slová *veternica, veterník, veterno, veterný* vyslovujeme [vEcErJIItsa, vEcErJI:k, vEcErnO, vEcErnI:].

Tieto slová boli zaradené medzi výnimky. Pre slová *mestečko, poveternostný* boli zaradené obidva spôsoby výslovnosti.

Slová a tvary, v ktorých je vo výslovnosti *t, d, n, l – ě, ě, ň, ě* kolísanie

§153 V spisovných prejavoch sa niekedy zjavuje neistota a kolísanie vo výslovnosti slov a tvarov, v ktorých sa striedajú hlásky *t, d, n, l – ě, ě, ň, ě* i v niektorých iných slovách. Sú to najmä tieto skupiny slov: druhové prívlastňovacie prídavné mená, príslovky, slovesá so spoluhláskou *l* pred neurčitkovou kmeňotvornou príponou a tvary druhého a tretieho stupňa prídavných mien. Okrem toho sa neistota objavuje vo výslovnosti niektorých pomnožných podstatných mien s koncovou samohláskou *e* a vo výslovnosti niektorých iných slov. V spisovnej slovenčine sa ustálila táto výslovnosť:

§154 Druhové prívlastňovacie prídavné mená (odvodené od zvieracích podstatných mien) vyslovujeme s mäkkou spoluhláskou. Slová ako *čmelí, danielí, hadí, krokodíli, orlí, oslí, sloní, sobolí, včelí* a pod. vyslovujeme

[tSmELI:, daJI_ˆELI:, h\JI:, krOkOdI:LI:, OrLI:, OsLI:, sLOJI:, sObOLI:, ftSELI:] atď.

Príslušnosť k danej množine slov nie je možné určiť bez databázy zvieracích podstatných mien. Uvedené slová boli pridané do databázy výnimiek. Krajové rozdiely v mäkkej a tvrdej výslovnosti *l* sú ošetrené už definovanými pravidlami 3.176 a 3.177.

§155 V príslovkách sa pred koncovým *e* vždy vyslovuje *ě, ě, ň, ě*. Slová ako *minule, náhle, nezávisle, ojedinelé, okamžite, popravde, roztomile, šťastne, zle, znamenite* a pod. vyslovujeme

[minuLe, na:hLe, Jeza:visLe, ojeJ\iJeLe, okamZice, popraU_ˆJ\e, rostomiLe, ScastJe, zLe, znameJice] atď.

Rozlišujeme napríklad prídavné mená [kra:sne, sla:vne], resp. [sla:U_ˆne] a príslovky [kra:sJe, sla:vJe], resp. [sa:U_ˆJe]. Aj tu platí dvojité výslovnosť pre *l* podľa 3.176 a 3.177.

§156 Spoluhlásky *l/ě* sa striedajú v slovách a tvaroch ako *poslať – pošlem, pošli, poslal, pošlú; postlať – posteliem, posteľ, postelú; kašať – kašlem, kašli, kašlal; zakašlal; piplať – piplem, pipli, piplal; klať – kolem, kolú, klal; zaklať – zakolem, zakoľ, zaklal; mlieť – mlel, meľ, meľme, melú*.

V slovách s príponou *-ka* odvodených od sloviess na *-liť*, napr. v slovách *prestreliť – prestrelka, zapáliť – zápalka*, sa vyslovuje *l*.

Toto pravidlo nepatrí medzi pravidlá o výslovnosti tvrdých a mäkkých spoluhlások, keďže v tomto prípade za *l* nasleduje spoluhláska.

Ak je neurčitková kmeňotvorná prípona *i, ia, ie*, vyslovuje sa *ě* vo všetkých slovných tvaroch. Tak je to v slovách a tvaroch *kliat – klajem, klajme; plieť – plejem, plej, plel; páliť – pálim, páliť, páľ, páľme* atď. Vyslovujeme

[pOslac - pOSLEm, pOSLI, pOslal, pOSIU:; pOstlac - pOscELI_ˆEm, pOscEL, pOscELU:; kaSlac - kaSLEm, kaSLI, kaSlal; zakaSlal; pIplac - pIpLEm, pIpLI, pIpIal; klac, kOLEm, kOIU:, klal; zaklac - zakOLEm, zakOL, zaklal; mLI_ˆEc - mLEl, mEL, mELmE, mEIU:;]

prEstrELlc - prEstrElka, zapa:Llc - za:palka;
 kLL[^]ac - kLajEm, kLL[^]al, kLaL[^]mE; pLL[^]Ec - pLEjEm, pLEL[^], pLEl;
 pa:Llc - pa:LIm, pa:LIl, pa:L, pa:LmE] atď.

Vytváranie transkripčných pravidiel pre výslovnosť [L] komplikuje skutočnosť, že v niektorých oblastiach slovenská sa hláska [L] už takmer nevyskytuje. Z toho dôvodu bola pre tento prípad definovaná obojaká výslovnosť a zjednodušené pravidlo (3.179) bolo zaradené na koniec pravidiel o výslovnosti obojakých spoluhlások.

$$(I)([ieiaie]) \rightarrow \mathbf{L} \setminus 1(k - l - iať) \quad (3.179)$$

$$\rightarrow \mathbf{l} \setminus 1(k - l - iať) \quad (3.180)$$

§157 Slovesá, ktoré majú pred koncovou spoluhláskou prítomníkového kmeňa, t. j. pred *t*, *d*, *n*, *l*, samohlásku, končia sa v tvaroch rozkazovacieho spôsobu spoluhláskami *ť*, *ď*, *ň*, *ľ*. Vyslovujeme

[pLEtU: - pLEc, pLEJ\mE, pLEccE; (Od)vEdU: - (Od)vEc, (Od)vEJ\mE, (Od)vEccE;
 (JE)mInU: - (JE)mIJ, (JE)mIJmE, (JE)mIJcE; (zO)mElU: - (zO)mEL, (zO)mELmE,
 (zO)mELcE].

Táto výslovnosť však vyplýva aj z ortografie a, preto pre ňu nebolo nutné definovať žiadne špeciálne transkripčné pravidlo.

§158 Spoluhlásky *t*, *d*, *n*, *l* - *ť*, *ď*, *ň*, *ľ* sa striedajú aj pri stupňovaní prídavných mien: *zaťatý*, *zaťatejší*, *najzaťatejší*; *rovnorodý*, *rovnorodejší*, *najrovnorodejší*; *rýchly*, *rýchlejší*, *najrýchlejší*; *stály*, *stálejší*, *najstálejší* atď. Vyslovujeme

[zacatI:, zacacEL[^]SI:, naL[^]zacacEL[^]sI:; rOvnOrOdI:, rOvnOrOJ\EL[^]SI:,
 naL[^]rOvnOrOJ\EL[^]SI:/ rOU[^]nO-;
 mL[^]ErnI, mL[^]ErJEL[^]SI:, naL[^]mL[^]ErJEL[^]SI:; rI:xII, rI:xLEL[^]SI:,
 naL[^]rI:xLEL[^]SI:; sta:II, sta:LEI[^]SI:, naL[^]sta:LEI[^]SI:] atď.

§159 Mäkké *ľ* sa vyslovuje vo všetkých tvaroch slov typu *kúpele*, t. j. v slovách a tvaroch *kúpele*, *kúpeľov*, *kúpeľom* atď.; *machle*, *machľov*; *zubále*, *zubáľov*; *korále*, *koráľov*; *Vráble*, *Vrábeľ*; *behúle*, *behúľov*; *Tále*, *Táľov*; *Štále*, *Štáľov* a pod. Vyslovujeme

[kU:pELE, kU:pELOU[^], kU:pELOm; maxLE, maxLOU[^], maxLOm; zUba:LE, zUba:-
 LOU[^]; kOra:LE, kOra:LOU[^]; vra:bLE, vra:bEL; bEh\LE, bEh\LOU[^]; ta:LE,
 ta:LOU[^]; Sta:LE, Sta:LOU[^]] atď.

V tomto prípade platí to, čo bolo o výslovnosti *l* uvedené vyššie.

§160 Vyslovujeme však [sanda:Le], ale [sanda:loU[^]], lebo je aj jednotné číslo sandál so spoluhláskou [l].

§161 Výslovnosť *ť*, *ď*, *ň*, *ľ* sa ustálila alebo ustaľuje v slovách *anenský*, *cmíter*, *čudesný*, *deptať*, *František*, *františkán*, *galiba*, *geleta*, *glezg*, *kalika*, *kolimaha*, *lekno*, *leknovec*, *lepok*, *lengat sa*, *lept*, *lichobežník*, *lichý*, *limba*, *linajka*, *lis*, *lišaj*, *lišajník*, *lišta*, *mater-nica*, *molestovať*, *neter*, *odedza*, *oflinkovať*, *opatera*, *panenčiť*, *panenka*, *paneneský*,

panenstvo, paníc, partička, paznecht, peleš, poniklec, Potisie, pudivietor, redikať sa, sesternica, sesterský, slečna, sled, šmelína, Talian, taliančina, taliansky, tuliga, tepať, terigať sa, test, testiná, testovci, tetrov, tis, Tisa, tisina, Tisovec, trpaslík, vyligotanj, ťinenka atď. Spoluhláska *ľ* sa vyslovuje aj v slovách *malichernosť, neplecha, pohoršlivý, polený, pomaličky, pomalinky, pospolitosť, premilený, silený, striezlivý*. Vyslovujeme

[aJEnskI:, tsmIcEr, tSUJ\EsnI:, J\Eptac, francISEk, francISka:n, gaLIba, gELEta, gLEsk, kaLIka, kOLImah\, LEknO, LEknOvEts, LEpOk, LENgacsa, LEpt, LIxObEZJI:k, LIxI:, LImba, LIaI_ka, LIs, LISaI_ , LISaI_ JI:k, LISta, macErJItsa, mOLEstOvac, JEcEr, OJ\Edza, OfLINKOvac, OpacEra, paJEntSIc, paJENka, paJEnskI:, paJEnstvO, paJIts, parcItSka, pazJExt, pELES, pOJIKLEts, pOcIsI_ E, pUJ\IvI_ EtOr, rEJ\Ikacsa, sEscErJItsa, sEscErskI:, sLEtSna, sLEc, SmELIna, taLI_ an, taLI_ antSIna, taLI_ anskI, taLIga, cEpac, cErIgcacsa, cEsc, cEscIna:, cEscOU_ tsI, cEtrOU_ , cIs, cIsa, cIsIna, cIsOvEts, tr=pasLI:k, vILIgOtanI:, ZIJENka; maLIxErnOsc, JEpLExa, pOh\rSLIvI:, pOLEnI:, pOmaLItSkI, pOmaLInkI, pOspOLItOsc, prEmILEnI:, sILEnI:, strI_ EzLIvI:].

Tieto slová boli pridané medzi výnimky. K niektorým z nich však boli priradené aj alternatívne výslovnosti. Napríklad [tr=pasLI:k/tr=pasII:k; vILIgOtanI:/vIIIgOtanI:] atď.

Spoluhlásky *t, d, n, l, – ť, ď, ň, ľ* v prevzatých slovách

§162 Jedným zo znakov "cudzosti" prevzatých slov v spisovnej slovenčine je aj výslovnosť *t, d, n, l* pred *i, í, e*. Tie pôvodne cudzie a prevzaté slová, ktoré celkom zdomácnili, prispôsobili sa zákonitostiam výslovnosti domácich slov – v nich sa namiesto *t, d, n, l* pred *i, í e* vyslovuje *ť, ď, ň, ľ*. V prevzatých slovách sa aj pred ohýbacími odvodzovacími príponami a vnútri týchto prípon pred samohláskovými zvukmi *i, í, e, ia, ie, iu* vyslovujú namiesto *t, d, n, l* spoluhlásky *ť, ď, ň, ľ*.

O výslovnosti spoluhlások *t, d, n, l – ť, ď, ň, ľ* v cudzích a prevzatých slovách platia tieto základné pravidlá:

§163 Pôvodne cudzie slová ktoré, celkom zdomácnili, sa vyslovujú ako domáce slová. Také slová sú napr. *diabol, košela, tehla, truhla* a i. Vyslovujeme

[J\I_ abOl, kOSELa, cEh\ a, trUh\ a] atď.

§164 V cudzích a prevzatých slovách, ktorých pôvod sa uvedomuje, vyslovujú sa aj pred *i, í, e* spoluhlásky *t, d, n, l*. Takto sa *t, d, n, l* vyslovuje v koreni (základe) slova a pred cudzími slovtvornými príponami.

Spoluhlásky *t, d, n, l* v základe prevzatých slov

§165 Spoluhlásky *t, d, n, l* sa vyslovujú v základe slova pred *i, í, e* napr. v týchto cudzích a prevzatých slovách (resp. vlastných menách): *Adela, Dalmatínek, decht, dekan, denár, disciplína, diván, elidovať, etnický, etnikum, fenek, Feničan, fetiš, Filip, generál, Chetit, chuligán, idea, Júlia, káder, klika, klíma, kolega, kolibrík, Kolín, komplikácia, kredenc, krokodíl, kulisa, lečo, licitovať, lido, líkér, lilipután, línia, liter (deci-, mili-atď), Martin, matematika, mestic, meter, miliarda, milión, model, molino, moslim,*

muslim, Napoleon, napolitánka, negácia, negližé, nestor, nikel, nikotín, nistej, obelisk, olivín, Palestína, paleta, palisáda, panoptikum, paralela, partiový, partizán, pelendrek, pelerína, pelikán, Peter, pikník, planírovať, plebs, plika, podesta, polenta, poliklinika, politika, politruk, prestíž, problém, proletariát, Pyreneje, roleta, ruleta, rutina, sangvinik, sanítka, satelit, segedínsky, signalizácia, sveter, šledrián, špiritizovať, štepona, tableta, talizman, technik, tektonika, teleskop, terč, termín, tetovať, tér, textilka, tiger, titovka, triplicita, univerzálie, univerzita, univerzum, zenit a i. Vyslovujeme ich

[adEla, dalmatIJEts, dExt, dEkan, dEna:r, dIstsIplI:na, dIva:n, EIIdOvac, EtnItskI:, EtnIkUm, fEnEk, fEnItSan, fEtIS, fIlIp, gEnEra:l, xEtIt, xUlIga:n, IdEa, jU:IIa, ka:dEr, kIIka, kII:ma, kOIeGa, kOIIBrI:k, kOI:n, kOmplIka:tsIa, krEdEnts, krOkOdI:l, kUlIsa, IEtSO, IItsItOvac, IIdO, IlkE:r, IIIpUta:n, II:nIa, ItEr (dEtsI-, mIIIatd), martIn, matEmatIka, mEstIts, mEtEr, mIIIarda, mIIIO:n, mOdEl, mOIInO, mOslIm, mUsIm, napOIeOn, napOIIta:Nka, nEga:tsIa, nEglIZE:, nEstOr, nIkEl, nIkOtI:n, nIstEl_^, ObElIsk, OIIvI:n, palEstI:na, palEta, palIa:da, panOptIkUm, paralela, partIOvI:, partIza:n, pElEndrEk, pElErI:na, pElIka:n, pEtEr, plknIk, planI:rOvac, plEps, plIka, pOdESta, pOIEnta, pOIklInIka, pOIItIka, pOIItItrUk, prEstI:Z, prObLE:m, prOIetarIa:t, pIrEnEjE, rOIeta, rUIeta, rUtIna, saNgvInIk, sanItka, satElIt, sEgEdI:nskI, sIgnalIza:tsIa, svEtEr, SIedrIa:n, SpIrItIzOvac, StEpOvac, tableta, talIzman, tExnIk, tEktOnIka, tElEskOp, tErtS, tErmI:n, tEtOvac, tE:r, tEks-tIIka, tIgEr, tItOU_^ka, trIplItsIta, UnIvErza:IIe, UnIvErzIta, UnIvErzUm, zEnIt] a iné.

§166 Takto sa vyslovujú aj slová *kaleráb, krinolína, mandolína, plastelína, rizling, vatelín, vellín, zeler*. Vyslovujeme ich

[kalEra:p, krInOI:na, mandOI:na, plastElI:na, rIzIInk, vatElI:n, vEltlI:n, zElEr].

§167 Spoluhlásky *t, d, n, l* sa vyslovujú pred *i, e* aj v niektorých prevzatých slovách rozšírených aj v nárečiach: *ceremónia, šindel, štemovať, štimovať, tanistra* a pod. Vyslovuje sa

[tsErEmO:nIa, SIndEL, StEmOvac, StImOvac, tanIstra].

Všetky tieto slová boli pridané medzi výnimky.

Spoluhlásky *t, d, n, l* pred cudzími príponami

§168 Spoluhlásky *t, d, n, l* sa vyslovujú v prevzatých slovách končiacich sa na *ik, ík*, ak nejde o zdrobnené podstatné mená (deminutíva). Tak sa *t, d, n, l* vyslovuje napríklad v slovách *alkoholik, astenik, Atlantik, baladik, botanik, dogmatik, dramatik, cynik, energetik, erotik, evanjelik, gramatik, hedonik, hygienik, ironik, kanonik, katolik, matematik, medik, machanik, melancholik, narkotik, neurotik, olympionik, patalytik, platonik, politik, praktik, romantik, skeptik, sklerotik, technik* a v mnohých iných slovách. Tak sa vyslovuje aj meno *Beňadik*. Vyslovujeme

[alkOh\IIk, astEnIk, atlantIk, baladIk, bOtanIk, dOgmatIk, dramatik, tsInIk, EnEr-gEtIk, ErOtIk, EvaJI_^ElIk, gramatik, h\EdOnIk, h\Igl_^EnIk, IrOnIk, kanOnIk,

katOII:k, matEmatIk, mEdIk, maxanIk, mElanxOIIk, narkOtIk, nEUrOtIk, OlympI-OnIk, patalytIk, platOnIk, pOIIItIk, praktIk, rOmantIk, skEptIk, sklErOtIk, tExnIk; bEJadIk].

Keďže nevieme automaticky určiť, kedy ide o cudzie slovo a kedy o zdobnenu, všetky vyššie uvedené slová boli pridané medzi výnimky.

§169 Spoluhlásky *t, d, n, l* sa vyslovujú pred príponou *-ický (-ícky)* v slovách ako *alkoholický, biblický, drakonický, hymnický, idylický, kryštálický, lakonický, melancholický, melodický, neurastenecký, platonický, symbolický* a v mnohých iných slovách. Vyslovujeme

[alkOh\OIItskI:, bIblItsI:, drakOnItsI:, h\ImmItsI:, IdIIIItsI:, krIStalItsI:, lakOnItsI:, mElanxOIIItsI:, mElOdItsI:, nEUrastEnItsI:, platOnItsI:, sImbOIIItsI:] atď.

Na základe toho boli definované nasledujúce pravidla:

$$(\mathbf{tdnl})ick \parallel \rightarrow \backslash \mathbf{1}ItsI : (alkoho - l - ický) \quad (3.181)$$

$$(\mathbf{tdnl})cky \parallel \rightarrow \backslash \mathbf{1} \quad (kato - l - ícky) \quad (3.182)$$

Pravidlo (3.182) nakoniec nebolo zaradené medzi pravidlá, ktoré boli implementované, keďže sa vzťahuje len na prevzaté slova, no slovenčina má mnoho slov končiacich sa na *-ícky*, avšak spoluhlásky *d t n l* sa pred nimi vyslovujú mäkko. Napríklad slová ako *banícky, kadernícky, pivovarnícky, mladícky* a podobne. Tieto slová museli byť ošetrené výnimkami.

§170 Spoluhlásky *t, d, n, l* sa vyslovujú pred príponou *-ickosť* v slovách ako *baladickosť, démonickosť, dramatickosť, mechanickosť, melancholickosť, mikrofonicnosť, sklerotickosť* a v mnohých iných slovách. Vyslovujeme

[baladItsOsc, dE:mOnItsOsc, dramatiItsOsc, mExanItsOsc, mElanxIIItsOsc, mIkroFonItsOsc, sklErOtItsOsc] atď.

$$(\mathbf{tdnl})ickos \parallel \rightarrow \backslash \mathbf{1}ItsOsc(sklero - t - ickosť) \quad (3.183)$$

To, čo platí pre pravidlo (3.182), platí aj pre pravidlo (3.183). Z toho dôvodu nebolo nakoniec ani toto pravidlo implementované.

§171 Spoluhlásky *t, d, n, l* sa vyslovujú pred príponou *-izovať* v slovách ako *aktualizovať, amortizovať, automatizovať, centralizovať, civilizovať, dramtizovať, federalizovať, hospitalizovať, idealizovať, idylizovať, industrializovať, kolonizovať, kompletizovať, kritizovať, krištálizovať, mechanizovať, mobilizovať, modernizovať, normalizovať, penalizovať, racionalizovať, radikalizovať, sfanatizovať, signalizovať, stabilizovať, symbolizovať, sympatizovať, synchronizovať, syntetizovať, škandalizovať, štandardizovať, štylizovať, taktizovať, teoretizovať, tyranizovať* a v mnohých iných slovách. Vyslovujeme

[aktUaIzOvac, amOrtIzOvac, aU_ˆtOmatIzOvac, tsEntralIzOvac, tsIvIIIzOvac, dramIzOvac, fEdEraIzOvac, h\spItaIzOvac, IdEaIzOvac, IdIIIzOvac, IndUstrIaIzOvac, kOIOnIzOvac, kOmplEtIzOvac, krItIzOvac, krIStalIzOvac, mExanIzOvac, mObIIIzOvac, mOdErnIzOvac, nOrmaIzOvac, pEnaIzOvac, ratsIOnaIzOvac, radIkaIzOvac, sfanatIzOvac, sIgnalIzOvac, stabIIIzOvac, sImbOIzOvac, sImpatIzOvac, sInxrOnIzOvac, sIntEtIzOvac, SkandaIzOvac, StandardIzOvac, StIIIzOvac, taktIzOvac, tEOrEtIzOvac, tIranIzOvat] atď.

$$(\mathbf{tdnl})izova \parallel \rightarrow \setminus \mathbf{1}IzOvac(tyra - n - izovat) \quad (3.184)$$

§172 V slove *bagatelizovat* pripúšťame výslovnosť s *l* aj *l*, lebo základný tvar je *bagateľ*. Vyslovujeme

[bagacEL, bagacELnI:, bagacELIzOvac], ale aj [bagacElIzOvac].

Slovo *bagatelizovat* bolo pridané s obidvoma spôsobmi výslovnosti medzi výnimky.

§173 Spouhlásky *t*, *d*, *n*, *l* sa vyslovujú pred príponou *-ismus* v slovách ako *defetizmus*, *diabolizmus*, *dogmatizmus*, *fanatizmus*, *formalizmus*, *germanizmus*, *chvostizmus*, *idealizmus*, *imperializmus*, *metabolizmus*, *nominalizmus*, *patriotizmus*, *polonizmus*, *radikalizmus*, *sionizmus*, *realizmus*, *romantizmus*, *socializmus*, *symbolizmus*, *univerzalizmus*, *verbalizmus* a v mnohých iných slovách. Vyslovujeme

[dEfEtIzmUs, J\I_ˆabOIzmUs, dOgmatIzmUs, fanatIzmUs, fOrmaIzmUs, gErma-nIzmUs, xvOstIzmUs, IdEaIzmUs, ImpErIaIzmUs, mEtabOIzmUs, nOmInaIzmUs, patriOIzmUs, pOIOnIzmUs, radIkaIzmUs, sIOIzmUs, rEaIzmUs, rOmantIzmUs, sOtsI_ˆaIzmUs, symbOIzmUs, UnIvErzaIzmUs, vErbaIzmUs] atď.

$$(\mathbf{tdnl})ismus \parallel \rightarrow \setminus \mathbf{1}IzmUs(fana - t - izmus) \quad (3.185)$$

§174 Spouhlásky *t*, *d*, *n*, *l* sa vyslovujú pred príponou *-ista* v slovách ako *artista*, *bicyklista*, *cimbalista*, *čembalista*, *činelista*, *defetista*, *externista*, *filatelista*, *finalista*, *formalista*, *futbalista*, *gardista*, *germanista*, *interbrigadista*, *kapitalista*, *kolonialista*, *komparatista*, *komunista*, *latifundista*, *latinista*, *materialista*, *motocyklista*, *nudista*, *organista*, *pianista*, *pólista*, *propagandista*, *radikalista*, *realista*, *sadista*, *separatista*, *sionista*, *socialista*, *špecialista*, *štatista*, *štipendista*, *štilista*, *telefonista*, *tenista*, *volejbalista*, *žurnalista* a v mnohých iných slovách. Vyslovujeme

[artIsta, bItsIklIsta, tsImbaIsta, tSEmbaIsta, tSInElIsta, dEfEtIsta, ExtErnIsta, fIlatElIsta, fInaIsta, fOrmaIsta, fUtbalIsta, gardIsta, gErmaIsta, IntErbrIgadIsta, kapItalIsta, kOIOnIalIsta, kOmparatIsta, kOmUnIsta, latIfUndIsta, latInIsta, matErIalIsta, mOtOtsIklIsta, nUdIsta, OrganIsta, pIanIsta, pO:Ista, prOpagandIsta, radIkaIsta, rEaIsta, sadIsta, sEparatIsta, sIOIsta, sOtsI_ˆaIsta, SpEtSialIsta, StatIsta, StIpEndIsta, StIIIsta, tElEfOnIsta, tEnIsta, vOIEl_ˆbaIsta, ZUrnaIsta] atď.

$$(\mathbf{tdnl})ista \parallel \rightarrow \setminus \mathbf{1}Ista(exter - n - ista) \quad (3.186)$$

§175 V slove *huslista* pripúšťame výslovnosť *l* aj *l*, lebo v základe slova je *l*: *husle, huslový*. Vyslovujeme

[h\UsLE, h\UsLOvI:, h\UsLIsta], ale aj [h\UsIIsta].

Slovo *huslista* bolo pridané s obidvoma spôsobmi výslovnosti medzi výnimky.

§176 Spoluhlásky *t, d, n, l* sa vyslovujú pred príponou *-ita* v slovách ako *admiralita, afinita, aktualita, alkalita, banalita, bandita, beštialita, bonita, brutalita, debilita, egalita, elita, eventualita, finalita, formalita, frigidita, ilegalita, imunita, individualita, invalidita, Izraelita, kolegialita, kozmopolita, kriminalita, kvalita, kvantita, labilita, legalita, lokalita, mentalita, mobilita, modalita, neutralita, oficialita, originalita, permeabilita, prediktabilita, racionalita, realita, rentabilita, rivalita, sanita, sentimentalita, senzibilita, sexualita, stabilita, sterilita, stupidita, suverenita, špecialita, tonalita, totalita, variabilita, vitalita* a v mnohých iných slovách. Tak sa vyslovujú aj mená *Afrodita, Edita, Judita* a iné. Vyslovujeme

[admIralIta, affnIta, aktUalIta, alkalIta, banalIta, bandIta, bEStIalIta, bOnIta, brUtalIta, dEbIIIIta, EgalIta, ElIta, EvEntUalIta, fInalIta, fOrmalIta, frIgIdIta, IIEgalIta, ImUnIta, IndIvIdUalIta, InvalIdIta, IzraElIta, kOIeGIalIta, kOzmOpOIIta, krImInalIta, kvalIta, kvantIta, labIIIIta, lEgalIta, lOkalIta, mEntalIta, mObIIIIta, mOdalIta, nEUtralIta, OffcIalIta, OrIgInalIta, pErmEabIIIIta, prEdIktabIIIIta, racIONalIta, rEalIta, rEntabIIIIta, rIvalIta, sanIta, sEntImEntalIta, sEnzIbIIIIta, sEksUalIta, stabIIIIta, stErIIIIta, stUpIdIta, sUvErEnIta, SpEcIalIta, tOnalIta, tOtalIta, varIabIIIIta, vItaIIIta; afrOdIta, EdIta, jUdIta].

V tomto prípade už nie je možné vytvoriť jednoduché transkripčné pravidlo pre domáce slová ako napríklad *záštita*. Tieto slová boli pridané medzi výnimky.

§177 Spoluhlásky *t, d, n, l* sa vyslovujú pred príponou *-ika* v slovách ako *akustika, anglicka, antika, aritmetika, atletika, bazilika, cyklistika, cyrilika, dietika, dogmatika, encyklika, enklitika, erotika, etika, exotika, fonetika, genetika, gotika, gramatika, harmonika, hlaholika, hydraulika, klinika, kozmetika, kozmonautika, kriminalistika, kritika, kybernetika, lingvistika, matematika, metodika, neotika, optika, panika, pneumatika, politika, problematika, proklitika, replika, republika, symbolika, systematika, statika, štatistika, štylistika, taktika, tematika, žurnalistika* a v mnohých iných slovách. Tak sa vyslovujú aj slová *píjatika, putika, rusistika, slavistika*. Vyslovujeme

[akUstIka, aNgIIstIka, antIka, arItmEtIka, atIEtIka, bazIIka, cIkIIstIka, cIrIIIka, dIEtIka, dOgmatIka, EncIkIIka, ENklItIka, ErOtIka, EtIka, EgzOtIka, fOnEtIka, gEnEtIka, gOtIka, gramatIka, harmOnIka, hlahOIka, hIdraUIka, kIIInIka, kOzmEtIka, kOzmOnaUtIka, krImInalIstIka, krItIka, kIbErnEtIka, IIIngvIstIka, matEmatIka, mEtOdIka, nEOtIka, OptIka, panIka, pnEUmatIka, pOIItIka, prObIEmatIka, prOkIIItIka, rEplIka, rEpUblIka, sImbOIka, sIstEmatIka, statIka, StatIstIka, StIIIs-tIka, taktIka, tEmatIka, ZUrnalIstIka; píjatika, putika, rusistika, slavistika].

Ani v tomto prípade nebolo možné definovať jednoduché transkripčné pravidlo kôli tvarom domácich slov ako *živôtika, žeriavnika* a podobne. Tieto slová boli pridané medzi výnimky.

§178 Spoluhlásky *t, d, n, l* sa vyslovujú pred príponou *-id, -it* v slovách ako *amonit, anionit, antimonit, argentit, bakelit, bentonit, brizolit, eternit, galenit, granit, hematit, heraklit, igelit, jodid, kationit, kyanid, kredit, xylofit, magnetit, megalit, molybdenit, monolit, neutronit, paleolit, perlit, petit, saponit, satelit, selenit, sololit, stroncianit, teodolit, teralit, titanit, uralit, uranit, vulkanit* a v mnohých iných slovách. Vyslovujeme

[amOnIt, anIOOnIt, antImOnIt, argEntIt, bakElIt, bEntOnIt, brIzOlIt, EtErnIt, galEnIt, granIt, hEmatIt, hEraklIt, IgElIt, jOdIt, katIOOnIt, kIanIt, krEdIt, ksIlOlIt, magnEtIt, mEgalIt, mOlIbdEnIt, mOnOlIt, nEUtrOnIt, palEOlIt, pErIt, pEtIt, sapOnIt, satElIt, sElEnIt, sOlOlIt, strOnsIanIt, tEOdOlIt, tErallIt, tItanIt, UralIt, UranIt, vUlkanIt] atď.

Tieto slová boli pridané medzi výnimky.

§179 Spoluhlásky *t, d, n, l* sa vyslovujú pred príponami *-ín, -ínka, (-ínka)* v slovách ako *anilín, bernardín, blondín, diftín, formalín, globulín, chitín, inzulín, kaolín, lignín, luetín, manganín, melanín, mezanín, mušelin, naftalín, nikotín, pektín, penicilín, popelín, saponín, smaltín, tanín, terpentín, tetralín, travertín, turmalín, uranín, vatelín, vazelín*. Tak sa vyslovujú aj slová *perlínka, pralínka, uršulínka, názvy Berlín, Kolín* a mená *Augustín, Konštantín*. Vyslovujeme

[anIlI:n, bErnardI:n, blOndI:n, dIfItI:n, fOrmalI:n, glObUlI:n, xItI:n, InzUlI:n, kaOlI:n, lIgnI:n, lUEtI:n, maNganI:n, mElanI:n, mEzanI:n, mUSElI:n, naftalI:n, nIkOtI:n, pEktI:n, pEnItsIlI:n, pOpElI:n, sapOnI:n, smaltI:n, tanI:n, tErpEntI:n, tEt-rallI:n, travErtI:n, tUrmalI:n, UranI:n, vatElI:n, vazElI:n;
pErIlINka, pralINka, UrSUI:Nka;
bErIlI:n, kOlI:n;
aU_`gUstI:n, kOnStantI:n].

V názvoch *Budín, Hodonín* vyslovuje sa *ď, ň*. Vyslovujeme [bUJ\I:n, h\OdOJI:n].

Tieto slová boli pridané medzi výnimky.

§180 Spoluhlásky *t, d, n, l* sa vyslovujú pred príponou *-ent* v slovách ako *abonent, abstinent, asistent, dekadent, deponent, descendent, disponent, exponent, incident, independent, komponent, kontinent, korešpondent, oponent, prezident, proponent, remitent, repetent, revident, rezident, študent* a v mnohých iných slovách. Vyslovujeme

[abOnEnt, apstInEnt, asIstEnt, dEkadEnt, dEpOnEnt, dEstsEndEnt, dIspOnEnt, EkspOnEnt, IntsIdEnt, IndEpEndEnt, kOmpOnEnt, kOntInEnt, kOrESpOndEnt, OpOnEnt, prEzIdEnt, prOpOnEnt, rEmItEnt, rEpEtEnt, rEvIdEnt, rEzIdEnt, StU-dEnt] atď.

Keďže neboli nájdené domáce slová končiacie na príponu *-ent*, bolo možné definovať nasledujúce transkripčné pravidlo:

$$(\mathbf{tdnl})ent \parallel \rightarrow \backslash \mathbf{1Ent}(absti - n - ent) \quad (3.187)$$

§181 Spoluhlásky *t, d, n, l* sa vyslovujú pred príponou *-er, -ér* v slovách ako *káder, operatér, outsider, parter, partner, púder, reportér, sabotér, salamander, žonglér* a v iných slovách. Vyslovujeme

[ka:dEr, OpEratE:r, aU_ˆtsaI_ˆdEr/ aU_ˆt_saI_ˆdEr, partEr, partnEr, pU:dEr, rEpOrtE:r, sabOtE:r, salamandEr, ZONglE:r] atď.

Pre prípony *-er, -ér* platí to isté, čo bolo uvedené pre príponu *-ent*. Z toho dôvodu bolo možné definovať nasledujúce transkripčné pravidlá:

$$(\mathbf{tdnl})er \parallel \rightarrow \backslash \mathbf{1Er} \quad (dí - l - er) \quad (3.188)$$

$$(\mathbf{tdnl})r \parallel \rightarrow \backslash \mathbf{1E} : r(ama - t - ér) \quad (3.189)$$

§182 Spoluhlásky *t, d, n, l* sa vyslovujú pred zakončením *-es*, ako v slovách *majáles, notes*. Vyslovujeme [maja:lEs, nOtEs].

Bolo zistené, že prevzatých slov končiacich sa na *-es* je oveľa viac než slov domácich. V domácich slovách vo všetkých prípadoch šlo o zakončenie *-nes* a to konkrétne v slovách *dnes, dodnes, doneš, nadnes, nes, oddnes, odnes, podnes, povznes, prednes, prenes, prines, vnes, vznes, zanes*. Tieto slová boli zaradené medzi výnimky a bolo definované nasledujúce transkripčné pravidlo:

$$(\mathbf{tdnl})es \parallel \rightarrow \backslash \mathbf{1Es}(no - t - es) \quad (3.190)$$

§183 Spoluhlásky *t, d, n, l* sa vyslovujú pred koncovým *-e*, ako v slovách *finále, nacionále, parte, pastorále, penále*, a v iných slovách. Vyslovujeme

[ffna:lE, natsIOna:lE, partE, pasOra:lE, pEna:lE] atď.

Tieto slová boli pridané medzi výnimky.

§184 Spoluhlásky *t, d, n, l* sa vyslovujú v prevzatých slovách pred koncovou samohláskovou skupinou, v ktorej prvou samohláskou je *i, e*, t. j. v koncových samohláskových skupinách *ia, ie, iu*, a pod., ako v slovách *aktínium, alumínium, berylium, beštia* (podobne *beštie, beštiu, beštiu, beštiou, beštiu, beštiám, beštiách, beštiami*), *ceremónia, domínium, evanjelium, galimatiáš, gálium, gaudium, harmónium, hélium, hostia, idea, Judea, kleptománia, konzilium, lapálie, ľalia, mánia, médium, nobelium, ódium, partia, patália, pivónia, pódium, prelúdium, prezídium, rádio, saturnálie, skandium, štípendium, štúdium, vanádium* a v mnohých iných slovách. Vyslovujeme

[aktI:nIUm, alUmI:nIUm, bErI:IUUm, bESTIa, bESTIE, bESTII, bESTIU, bESTIOU_ˆ, bESTII:, bESTIa:m, bESTIa:x, bESTIamI, tsErEmO:nIa, dOmI:nIUm, EvaJI_ˆEIUUm, galImatIa:S, ga:IUUm, gaU_ˆdIUUm, h\rmO:nIUUm, h\IUUm, h\stIa, IdEa, jUdEa, klEp_tOma:nIa, kOnzI:IUUm, lapa:IIIE, LaIIa, ma:nIa, mE:dIUUm, nObEIUUm, O:dIUUm, partIa, pata:IIa, pIvO:nIa, pO:dIUUm, prElIU:dIUUm, prEzI:dIUUm, ra:dIO, satUrna:IIIE, skandIUm, StIpEndIUUm, StU:dIUUm, vana:dIUUm] atď.

V časti 3.4.1, kde sa pojednáva aj o nediftongickej výslovnosti samohláskových skupín, boli definované pravidlá (3.52), (3.53) a (3.54). V prípade, že sa *t, d, n, l*

nachádza pred koncovou samohláskovou skupinou, na ktorú bolo aplikované jedno z pravidiel (3.52), (3.53) a (3.54), *t*, *d*, *n*, *l* sa prepisuje tvrdo, teda ako [t, d, n, l].

Spoluhlásky *t*, *d*, *n*, *l* pred domácimi príponami

§185 V prevzatých slovách pred domácimi slovotvornými (odvodzovacími) a ohýbacími príponami začínajúcimi sa samohláskovými zvukmi *i*, *í*, *e*, *ia*, *ie*, *iu* (alebo pred príponami, ktoré sa hodnotia ako domáce) vyslovuje sa *t*, *d*, *n*, *l*.

§186 Spoluhlásky *t*, *d*, *n*, *l* sa vyslovujú pred príponou *-ik* (*-ík*), ktorou sa tvoria zdrobenené podstatné mená (deminutíva), ako v slovách *anjelik*, *advokátik*, *banánik*, *bicyklík*, *ci-gánik*, *detailík*, *figlík*, *hotelík*, *kanálik*, *kanónik*, *kastrólik*, *miliónik*, *mementík*, *mopslík*, *paplónik*, *pedálik*, *pikulík*, *pinčlík*, *praclík*, *puclík*, *referátik*, *románik*, *šálik*, *škandálik*, *téglík* a v mnohých iných slovách. Vyslovujeme

[aJL^ELIk, advOk:a:cIk, bana:JIk, bItskLI:k, tsIga:JIk, dEtaL^LI:k, fl:gLIk, h\OtE-LI:k, kana:LIk, kanO:JIk, kastrO:LIk, mIIIO:JIk, mEmEncI:k, mOpsLI:k, paplO:JIk, pEda:LIk, plkULI:k, pIntSLI:k, pratsLI:k, pUdLI:k, rEfEra:cIk, rOma:JIk, Sa:LIk, Skanda:LIk, tE:gLIk].

$$\mathbf{t} \vdash ik \parallel \rightarrow \mathbf{cIk} \quad (\text{advoká} - t - ik) \quad (3.191)$$

$$\mathbf{d} \vdash ik \parallel \rightarrow \mathbf{J} \backslash Ik \quad (\text{ria} - d - ik) \quad (3.192)$$

$$\mathbf{n} \vdash ik \parallel \rightarrow \mathbf{JIk} \quad (\text{baló} - n - ik) \quad (3.193)$$

$$\mathbf{l} \vdash ik \parallel \rightarrow \mathbf{LIk} \quad (\text{pedá} - l - ik) \quad (3.194)$$

$$\rightarrow \mathbf{IIk} \quad (\text{pedá} - l - ik) \quad (3.195)$$

$$\mathbf{t} \vdash ík \parallel \rightarrow \mathbf{cI} : k \quad (\text{memen} - t - ík) \quad (3.196)$$

$$\mathbf{d} \vdash ík \parallel \rightarrow \mathbf{J} \backslash I : k \quad (\text{bu} - d - ík) \quad (3.197)$$

$$\mathbf{n} \vdash ík \parallel \rightarrow \mathbf{JI} : k \quad (\text{obuv} - n - ík) \quad (3.198)$$

$$\mathbf{l} \vdash ík \parallel \rightarrow \mathbf{LI} : k \quad (\text{detai} - l - ík) \quad (3.199)$$

$$\rightarrow \mathbf{II} : k \quad (\text{detai} - l - ík) \quad (3.200)$$

Pre spoluhlásku *l* boli definované pravidlá pre mäkkú aj tvrdú výslovnosť ktoré su však zbytočné keďže sú pokryté pravidlami 3.176 a 3.177.

§187 Spoluhláska *l* sa vyslovuje aj v slovách *špendlík*, *štamperlík*, *trpaslík*, *žolík*, v názvoch prvkov *draslík*, *hliník* (podobne *hliníkáraň*, *hliníkový*), *kyslík*, *uhlík*. Vyslovujeme

[SpEndLI:k, StampErLI:k, tr=pasLI:k, ZOLI:k;
drasLI:k, h\LIJI:k, h\LIJIka:rEJ, h\LIJI:kOvI:, kIsLI:k, Uh\LI:k].

Keďže bežnejšia je tvrdá výslovnosť, nie je potrebné definovať žiadne pravidlá, keďže správnu výslovnosť pre *l* generujú pravidlá 3.176 a 3.177.

§188 Spoluhlásky *t, d, ň, l* sa vyslovujú v prevzatých slovách s príponou *-nik (-ník)*, ako v slovách *agrárnik, beatnik, ciferník, finančník, chinínovník, kompromisník, kriminálnik, sanitník, sírnik, škandálnik, textilník, tradičník*, a v iných slovách. Vyslovujeme

[agra:rJIk, bI:tJIk, tsIfErJI:k, fInantSJI:k, xInI:nOvJII/xInI:nOU_^JI:k, kOmprOmIs-JI:k, krImIna:lJIk, sanItJI:k, sI:rJIk, Skanda:lJIk, tEkstIlJI:k, tradItSJI:k] atď.

Vyššie uvedené pravidlo je značne vágne. V tomto prípade sú problémom slová, ktoré sú pôvodom cudzie, koreň slov končí na *n* a za ním nasleduje prípona *-ik*. Sú to napríklad slová *elektronik, botanik, arzenik*. Tieto slová končia síce na *-nik*, avšak vyslovujú sa tvrdo. Pojednáva o tom časť venujúca sa spoluhláskam *t, d, n, l* pred cudzími príponami. Vzhľadom k tomuto problému, neboli pre spomínané prípady definované pravidlá, ale slová boli pridané medzi výnimky.

§189 Spoluhlásky *t, d, ň, l* sa vyslovujú v prevzatých slovách končiacich sa na *-iak, -iar, -ier*, kde *ia, ie* je dvojhĺaska, ako v slovách *fortieľ, gavalier, golier, granatier, halier, hotelier, inžinier, kanonier, krauliari, kupliari, maniak, monastier, mušketer, palier, pilier, pionier, pranier, rentier, rytier, šarnier, špalier, škapuliar, švindliari, žoldnier* atď. Vyslovujeme

[fOrcL^EL, gavaLI^Er, gOLL^Er, granacL^Er, h^aLI^Er, h^OtELI^Er, InZIJL^Er, kanOJL^Er, kraUL^LI^ar, kUpLI^ar, maJL^ak, mOnascL^Er, mUSkEcL^Er, paLL^Er, pILL^Er, pIOJL^Er, praJL^Er, rEncL^Er, rIcL^Er, SarJL^Er, SpaLI^Er, SkapULI^ar, SvIndLI^ar, ZOldJL^Er].

Slová *dieža, evanjelický, evanjelik, evanjelium, pliaga, predsálie, Sliezske, Talianske* vyslovujeme

[J^I^Eza, EvaJL^EItskI:, EvaJL^EIk, EvaJL^EIUm, pLI^aga, prEtstsa:LIE / prEt_sa:LI^E, sLI^EskO, taLI^anskO].

Slovo *dieža* bolo pridané medzi výnimky, pre ostatné platí to isté čo pre §187.

§190 Spoluhlásky *t, d, ň, l* sa vyslovujú v prevzatých slovách pred príponou *-ec*, ako v slovách *Abesínek, Argentínek, Bengálek, Dalmatínek, kontinentálek, neanderlák, Portugálek, Sarínek* atď. Vyslovujeme

[abEsIJEtS, argEntI:JEts, bENga:LEts, dalmatI:JEts, kOntInEnta:LEts, nEandErta:LEts, pOrtUgaLEts, sardi:JEts] atď.

§191 Spoluhlásky *t, d, ň, l* sa vyslovujú v prevzatých slovách pred príponou *-ica*, ako v slovách *mangalica, terkelica*. Vyslovujeme [maNgaLItsa, tErIELItsa].

§192 Spoluhlásky *t, d, ň, l* sa vyslovujú v prevzatých slovách pred príponou *-ena*, ako napríklad v slove *fiľfena*. Vyslovujeme [fIfLEna].

§193 Spoluhlásky *t, d, ň, l* sa vyslovujú v prevzatých slovách pred príponou *-ina*, ako v slovách *kremelína, tónina*. Vyslovujeme [krEmELIna, tO:JIna].

Slová v ktorých ide o výslovnosť *d, t, n* boli pridané medzi výnimky. Pre výslovnosť *l* platia pravidlá 3.176 a 3.177.

§194 Spoluhlásky *t, d, ň, l* sa vyslovujú v prevzatých slovách pred ohýbacími príponami, ak bezprostredne nasleduje samohláska *i* alebo *e*, ako v slovách *bál - na bále, bicykel - na bicykli, dubleta - dubliet, futbal - na futbale, o futbale, hominid - hominidi, hotel - v hoteli, Chetit - Chetiti, komunista - komunisti, konzul - konzuli, kotleta - kotliet, lokalita - v lokalite, paleta - paliet, paralelka - paraleliek, pilulka - piluliek, terén - v teréne, turbína - pri turbíne* a pod. Vyslovujeme

[ba:l - naba:LE] a tak aj

[nabItsIkLI, dUblI^Et, nafUdbaLE, OfUdbaLE, h\OmInIJ\I, f_vh\OtELI, xEtIcI, kOmUnIscI, kOnzULI, kOtLI^Et, vlOkalIcE, paLI^Et, paralELI^Ek, pIIULI^Ek, ftErE:JE, prItUrbI:JE] atď.

§195 Spoluhlásky *t, d, ň, l* sa vyslovujú aj v slovách *koleda, lajblík, latinčina, latinčinár, latinka, latinský, limonáda, linajka, linka*. Vyslovujeme

[kOLEda, laL^bLI:k, lacIntSIa, lacIntSIa:r, lacINka, lacInskI:, LIImOna:da, LIInaL^ka, LINka].

Slová v ktorých ide o výslovnosť *d, t, n* boli pridané medzi výnimky. Pre výslovnosť *l* platia pravidlá 3.176 a 3.177.

3.4.11 Spoluhlásky [m, F, n, N, ɳ]

§196 Na mieste, kde sa v slovenskom pravopise píše písmeno *n*, vyslovuje sa v spisovnej slovenčine okrem predodasnovej záverovej nosovej zvučnej hlásky [n] aj mäkkopodnebná záverová nosová zvučná hláska [N] a predodasnová úžinová nosová spoluhláska [ɳ]. V niektorých slovách sa namiesto [n] vyslovuje aj [m]. Výskyt hlások [N],[ɳ],[F] závisí od spoluhláskového okolia.

Podobne pred spoluhláskami *v, f* sa vyslovuje namiesto pernoperného [m] pernozubné [F].

Znak *ɳ* patrí do repertoára znakov slovenskej fonetickej transkripcie 1.2.3. V transkripcii IPA ani SAMPA preň neexistuje znak, a preto hláska [ɳ] je stále prepisovaná rovnako ako hláska [n], tak ako je to uvedené v časti 3.3.2.

Spoluhláska [N]

§197 Spoluhláska [N] sa v spisovnej slovenčine vyslovuje vtedy, keď po *n* bezprostredne nasleduje *k, g*, ako v slovách *banka, cengať, dorastenka, doručenka, dožinky, dribling, drienka, dutinka, funkcia, hrozienka, inkasovať, inklinovať, mangán, mienka, panghart, pingpong, radovánky, sankcia, solingenský, strunga, synklinála, synkopa, tenký* atď. Vyslovujeme

[baNka, tsENgac, dOrascENka, dOrUtSENka, dOZINKI, drIblINK, drL^ENka, dUcINKa, fUNktsIa, h\rOzL^ENka, INkasOvac, INklInOvac, maNga:n, mL^ENka, paNg-h\art, pINKpONk, radOva:NkI, saNktsIa, sOIINgEnskI:, strUNga, sINKlIna:la, sINkOpa, cENkI:] atď.

Pre výslovnosť [N] bolo možné definovať jednoduché transkripčné pravidlo:

$$\mathbf{nk} \rightarrow \mathbf{Nk}(\text{doži} - n - \text{ky}) \quad (3.201)$$

$$\mathbf{ng} \rightarrow \mathbf{Ng}(\text{ce} - n - \text{gať}) \quad (3.202)$$

§198 Spoluhláska [N] sa môže striedať s [n], ale aj [J], ako v slovách a tvaroch *doručenka* - *doručeníek*, *dožinky* - *dožíniek*, *drienka* - *drienok*, *dutinka* - *dutíniek*, *hrozienka* - *hrozienok*, *mienka* - *mienok*, *radovánky* - *radovánok*, *skriňa* - *skrinka*, *sviňa* - *svinka*, *účinko* - *účinku*, *zinok* - *zinku* atď. Vyslovujeme

[dOrUtSEnka - dOrUtSEJl_ˆEk, dOZINky - dOZIJl_ˆEk, drI_ˆENka - drI_ˆEnOk, dUcINka - dUcIJl_ˆEk, h\rOzI_ˆENka - h\rOzI_ˆEnOk, mI_ˆENka - mI_ˆEnOk, radOva:NkI - radOva:nOk, skrIJa - skrINka, svIJa - svINka, U:tSInOk - U:tSINkU, zInOk - zINkU] atď.

§199 Spoluhláska [N] sa v spisovnej slovenčine vyslovuje v prevzatých a cudzích slovách aj pred spoluhláskovou skupinou, ktorú v ortografii zapisujeme znakom *x*, ako v slove *anxiozita*. Vyslovujeme [aNksIOzIta].

Túto skutočnosť bolo možné ošetriť pravidlom:

$$\mathbf{nx} \rightarrow \mathbf{Nks}(a - n - \text{xiozita}) \quad (3.203)$$

Okrem slova *anxiozita* bol zaznamenaný relatívne vysoký výskyt tejto spoluhláskovej postupnosti zastúpenej v slovách *Bronx* a *Lynx*.

§200 Na hranici slov sa v priemernom a pomalšom rečovom tempe žiada v spisovnej slovenčine aj pred *k*, *g* vyslovovať [n]. Slovné spojenia ako *on kráča*, *len keby*, *ten gunár*, *pán kráľ* vyslovujeme

[onkra:tSa, Lenkebi, tenguna:r, pa:nkra:L].

Keďže pravidlo (3.203) sa neaplikuje na hranici slov, nebolo nutné vytvárať žiadné ďalšie pravidlo.

§201 Výslovnosť [n] pred *k*, *g* vnútri slova sa v spisovnej slovenčine hodnotí ako nápadná, umelá a strojená, resp. hyperkorektná výslovnosť. Je to tzv. písmenková výslovnosť.

Spoluhláska [ɲ]

§202 Spoluhláska [ɲ] sa vyskytuje na mieste, kde sa v slovenskom pravopise píše *n*, ak nasleduje ďasnová úžinová spoluhláska *s*, *z*, *š*, *ž*. Toto úžinové [ɲ] vzniká asimiláciou spôsobu artikulácie [n], t. j. prispôbením sa [n] k nasledujúcej úžinovej spoluhláske. Vyslytuje sa v slovách ako *banský*, *inšie*, *inžinier*, *konský*, *konžský*, *lanský*, *lanštiak*, *penzia*, *Slovensko*, *Tanzánia* a pod. Úžinové [ɲ] sa však vyslovuje aj pred inými úžinovými spoluhláskami.

§203 Úžinové [ɲ] sa vyskytuje aj pred spoluhláskou [x], ako v slove *bronchitída*, *melanchólia*.

§204 Záverový dotyk jazyka o podnebnú klenbu sa oslabuje aj v iných pozíciách, napríklad pred spoluhláskou *l*, ako v slovách *nedbanlivec*, *nedbanlivý*, *príčinlivý*, *trvanlivý* a pod.

Ako už bolo uvedené vyššie, pre znak $\#$ neexistuje ekvivalent ani v IPA ani SAMPA systéme. Preto do pravidiel bol pre $\#$ zahrnutý prepis ako pre obyčajné n .

$$\mathbf{ns} \rightarrow \mathbf{ns} (ba - n - ský) \quad (3.204)$$

$$\mathbf{nz} \rightarrow \mathbf{nz} (pe - n - zia) \quad (3.205)$$

$$\mathbf{n} \rightarrow \mathbf{nS} (i - n - šie) \quad (3.206)$$

$$\mathbf{n} \rightarrow \mathbf{nZ} (i - n - žinier) \quad (3.207)$$

$$\mathbf{nch} \rightarrow \mathbf{nx} (bro - n - chitída) \quad (3.208)$$

$$\mathbf{n} \rightarrow \mathbf{nL} (nedba - n - livec) \quad (3.209)$$

Výslovnosť [m] namiesto [n]

§205 Vnútri slova sa v spisovnej slovenčine pred pernopernou spoluhláskou namiesto [n] vyslovuje spoluhláska [m]. Tak je to v slovách *bonbón*, *Donbas*, *hanba*, *honba*, *klenba*, *ženba*. Vyslovuje sa

[bOmbO:n, dOmbas, h\amba, h\Onba, kLEmba, ZEmba].

$$\mathbf{nb} \rightarrow \mathbf{mb} (ha - n - ba) \quad (3.210)$$

§206 V prevzatých slovách sa však aj v tejto pozícii (v kombinácii $n + p, b$) najmä vo vyššom štýle vyslovuje [n]. Napríklad slová ako *medicinbal*, *nonpareille* sa vyslovujú [meditsinbal, nonpareil[˘]].

Tieto slová boli pridané medzi výnimky.

§207 Pred p, b sa spoluhláska [n] vyslovuje aj na hranici slov. Slovné spojenia ako *on počul*, *zvon bije* vyslovujeme [onpotSul, zvonbije].

Výslovnosť [F] namiesto [m, n, J]

§208 Spoluhláska [F] sa v spisovnej slovenčine vyslovuje vtedy, keď po m vnútri slova bezprostredne nasleduje v, f , ako v slovách *domvedúca*, *emfáza*, *komfort*, *lymfa*, *Memfis*, *memfiska*, *symfónia*, *tamvontriumf*, *tromf*, a v slove *franforce*. Vyslovuje sa

[dOFvEdU:tSa, EFfa:za, kOFfOrt, IIFfa, mEFfIs, mEFfIska, sIFfO:nIa, taFvOn, triUFf, trOFF, fraFfOrtsE] atď.

$$\mathbf{mv} \rightarrow \mathbf{Fv} (do - m - vedúca) \quad (3.211)$$

$$\mathbf{mf} \rightarrow \mathbf{Ff} (tro - m - f) \quad (3.212)$$

§209 V neutrálnom a vyššom štýle výslovnosti sa v slovenčine nepripúšťa výslovnosť pernozubnej spoluhlásky [F] namiesto [n, J]. V slovách ako *dezinfekcia*, *fanfáry*, *infekcia*, *informácia*, *kanva*, *konfekcia*, *konferencia*, *konflikt*, *konvergencia*, *konvoj*, *panvica*,

panvička, skonfiškovať a v slovných spojeniach ako *ten váš, len volaj, jeleň vyskočil* vyslovujeme

[dEzInfEktsIa, fanfa:rI, InfEktsIa, InfOrma:tsIa, kanva, kOnfEktsIa, kOnfErEntsIa, kOnflkt, kOnvErgEntsIa, kOnvOI˘, panvItsa, panvItSka, skOnfISkOvac; tEnva:S, LEnvOlaL˘, jELEJvIskOtSI] atď.

Pre ostatné prípady bolo pre grafému *m* definované pravidlo:

$$\mathbf{m} \rightarrow \mathbf{m}(m - a\acute{t}) \quad (3.213)$$

3.4.12 Iné spoluhláskové skupiny

Spoluhláskové skupiny *t, d + n, ň, l, ľ*

- §210 V spoluhláskových skupinách *tn, tň, dn, dň, tl, dl, tľ, dľ* sa ústny záver prvej spoluhlásky spája so záverom druhej spoluhlásky. Konček jazyka zotrva v záverovom postavení na ďasne od začiatku artikulácie *t, d* do konca artikulácie nasledujúcej spoluhlásky *n, ň, l, ľ*.
- §211 V skupinách *tn, tň, dn, dň* sa na prechode k artikulácii druhej spoluhlásky uvoľní iba podnebnohltanový záver, lebo druhá spoluhláska je nosová. Ústnu explóziu *t, d* nahrádza v tejto kupine explózia nosová. Výslovnosť s ústnou explóziou *t, d* v skupinách *tn, dn, tň, dň* sa hodnotí ako prepíata.

Tieto spoluhláskové skupiny sa vyskytujú v slovách ako *sadni si, stretne, stretnúť sa, zadný* a pod.

V skupinách *tl, dl, tľ, dľ* nasleduje po *t, d* tzv. boková spoluhláska. Na prechode od *t, d* k *l, ľ* sa nezruší dotyk končeka jazyka na ďasne, lež uvoľní sa priechod na boku jazyka. Ďasnová explózia *t, d* sa teda nahradí bokovou explóziou. Výslovnosť s končekom explóziou sa v týchto skupinách tiež hodnotí ako prepíata. Tieto spoluhláskové skupiny sa vyskytujú v slovách ako *dláto, podľa, tlama, tlieť* a pod.

Spoluhláskové skupiny *n, ň + t, d, ě, ď; l, ľ + t, d, ě, ď, n, ň*

- §212 Záverový dotyk končeka jazyka na ďasne pretrváva počas artikulácie dvoch susedných spoluhlások aj v skupinách *nt, nd, nť, ndť, ňt, ňd, ňť, ňď*. Tu sa na prechode od artikulácie prvej spoluhlásky k druhej zruší podnebnohltanový otvor (priechod), ktorý vyžaduje výslovnosť spoluhlások *n, ň*.
- §213 V skupinách *lt, ld, ľt, ľď, ln, ľň, ľt, ľď, ľn, ľň* sa na prechode od artikulácie prvej spoluhlásky k druhej zruší bočný (bokový) otvor a vznikne úplný ústny záver. Priehrada medzi končekom (resp. bokom) jazyka, ktorá vzniká pri artikulácii *l*, resp. *ľ*, sa bez explózie spája so záverom nasledujúcej spoluhlásky.

Takto sa vyslovujú uvedené spoluhláskové skupiny v slovách ako *inde, tento, žandár, zostaňte, študentík; poldeň, Poltár, príľne, príľnúť, školník, vlna, zvolte*, v spojených spojeniach ako *kôň drobčí, peň topoľa, pochváľ druhého*, ak medzi slovami nie je prestávka.

Spoluhláskové skupiny *t, d + n, ň, l, ľ* po ďasnovej úžinovej alebo záverovej spoluhláske – skupiny *stn, zdn* atď.

§214 Spoluhláskové skupiny *tn, tň, tl, tľ, dn, dň, dl, dľ* sa vyslovujú s jedným záverom a s nosovou alebo s bokovou ústnou explóziou aj vtedy, keď pred nimi predchádza spoluhláska *s, z, š, ž, c*.

§215 Skupiny *stn, stň, zdn, zdň, štň, ždn, ždň, stl, zdl, stľ, ctň* sa teda nezjednodušujú. Slová ako *čestný, dáždňik, ľúbostný, miestny, námestník, naprázdno, nectný, postlať, prázdniny, prázdny, rastlina, starostlivý, stlačiť, sústrastný, šťastný, týždne, vlastne, vlastniť, zdláviť, zvláštny* a pod. vyslovujeme

[tSEstnI:, da:ZdJIk, LU:bOstnI:, mL^EstnI, na:mEstJI:k, napra:zdnO, nEtstnI:, pOstlac, pra:zdJIny, pra:zdnI, rastLIna, starOstLIVI:, stlatSIc, sU:strastnI:, ScastnI:, tI:ZdJE, vlastnE, vlastJIc, zdla:vIc, zvla:StnI] atď.

Chybná je výslovnosť [ScasnI:, mL^EsnI, tSEsnI:, pra:znI, tI:ZJE, starOsLIVI:] a pod. Výnimočne sa však vyslovuje [SEsna:sc, SEsna:stI, SEsna:scI, tsnOta]. Nezjednodušuje sa ani spoluhlásková skupina *str*. Slová ako *prostredný, prostriedok, stred, streda* a pod. vyslovujeme

[prOstrEdnI:, prOstrI^EdOk, strEt, strEda] atď.

Spoluhláskové skupiny *ntn, ntň*

§216 V spoluhláskových skupinách *ntn, ntň* pretrvávajú počas výslovnosti celej skupiny jediný ústny záver, zatvára sa a otvára sa iba podnebnohltnanový priechod (priechod do nosa). Tak je to napríklad v slovách *bažantnica, latentne, latentný, permanentne, permanentný* a pod.

Spoluhláskové skupiny *šs, žs, sš, zš, zž, sč, čs, zč*

§217 Ak sa na morfematickej hranici stretne tmavá sykavá spoluhláska so svetlou, resp. svetlá s tmavou, táto dvojica spoluhlások sa nezjednodušuje a neuskutočňuje sa asimilácia svetlosti – temnosti. Podľa zákonitostí znelostnej asimilácie sa uskutočňuje iba znelostná asimilácia. Slová ako *černošský, kováčsky, mužský, rozčarovať, rozšíriť, rozžiariť, sčesať, zžiť* sa atď. vyslovujeme

[tSErnOSskI:, kOva:tSskI, mUSskI:, rOstSarOvac, rOsSI:rIc, rOzZI^aLIc, rOzZI^arIc, stSEsac, zZIc] atď.

Výslovnosť [tSErnOSkI:, kOva:tskI, mUskI:, rOstSarOvac, StSEsac] nie je spisovná. Nespisovná je aj výslovnosť [rOSSl:rIc, rOZZI^arIc, rOZZI^aLIc] a pOd.

$$\check{s} \setminus s \rightarrow \mathbf{Ss} \text{ (černo – šs – ký)} \quad (3.214)$$

$$\check{z} \setminus s \rightarrow \mathbf{Zs} \text{ (mu – žs – ký)} \quad (3.215)$$

$$s \setminus \check{s} \rightarrow \mathbf{sS} \text{ (fitne – sš – túdia)} \quad (3.216)$$

$$z \text{ } \check{s} \rightarrow \mathbf{zS} \text{ (} ro - z\check{s} - \acute{i}ri\check{t} \text{)} \quad (3.217)$$

$$z \text{ } \check{z} \rightarrow \mathbf{zZ} \text{ (} ro - z\check{z} - \acute{i}ari\check{t} \text{)} \quad (3.218)$$

$$s \text{ } \check{c} \rightarrow \mathbf{stS} \text{ (} s\check{c} - \acute{e}sa\check{t} \text{)} \quad (3.219)$$

$$\check{c} \text{ } s \rightarrow \mathbf{tSs} \text{ (} kov\acute{a} - \check{c}s - ky \text{)} \quad (3.220)$$

$$z \text{ } \check{c} \rightarrow \mathbf{ztS} \text{ (} ro - z\check{c} - \acute{a}rova\check{t} \text{)} \quad (3.221)$$

Slová s vyššie uvedenou "nespisovnou" výslovnosťou boli spolu so "spisovnou" formou výslovnosti pridané medzi výnimky. Slovo *fitnessštúdia* nie je rýdzo slovenské slovo, avšak bol to jediný príklad, ktorý sa v korpuse slovenčiny so skupinou *sš* podarilo nájsť.

Spoluhláskové skupiny *zsk*, *zšť*, *tsk*, *dsk*, *dzsk*; *dt*, *tstv*, *dšť*, *tšť*, *ptč*, *dštv*, *ndsk*

§218 Ak sa spoluhlásky *ts*, *ds*, *zs*, *tš*, *dš*, *tč*, *dč* stretnú na morfematickej hranici a ak sa nachádzajú medzi samohláskami, vznikajú z nich zdvojené spoluhlásky. V susedstve ďalšej spoluhlásky vznikajú zdvojené spoluhlásky iba zriedkavo.

V spoluhláskových skupinách *tsk*, *dsk*, *dst*, *zsk*, *dzsk*, *tstv*, *dštv*, *ndsk* nevznikajú zdvojené spoluhlásky. Tu sa spoluhláskové skupiny alebo zjednodušujú, alebo v asimilačnom procese vznikajú nové spoluhlásky.

§219 Spoluhláskové skupiny *ts*, *ds*, *dš*, *dzs* v spoluhláskových skupinách *tsk*, *dsk*, *dst*, *dzsk*, *tstv*, *dštv*, *ndsk* sa asimilujú na *c*. Slová ako *bohatstvo*, *hradská*, *huncútsky*, *kamarátsky*, *lehotský*, *ľudský*, *opodstatnenie*, *podstata*, *prievidzský*, *semitský*, *svetský*, *šatstvo*, *Švedsko*, *švédsky*, *loďstvo*, *Burgundsko* vyslovujeme

[bOh\atstvO, h\ratska:. h\UntsU:tskI, kamara:tskI, LEh\OtskI:. LUtskI:, OpOtstatJEJI_^E, pOtstata, prI_^EvItskI:, sEmItskI:, svEtskI:, SatstvO, SvE:tskO, SvE:tskI, lOtstvO, bUrgUntskO] atď.

$$\mathbf{tsk} \rightarrow \mathbf{tsk} \text{ (} hunc\acute{u} - tsk - y \text{)} \quad (3.222)$$

$$\mathbf{dsk} \rightarrow \mathbf{tsk} \text{ (} hra - dsk - a \text{)} \quad (3.223)$$

$$\mathbf{dst} \rightarrow \mathbf{tst} \text{ (} opo - dst - atnenie \text{)} \quad (3.224)$$

$$\mathbf{dzsk} \rightarrow \mathbf{tsk} \text{ (} prievi - dzsk - \acute{y} \text{)} \quad (3.225)$$

$$\mathbf{tstv} \rightarrow \mathbf{tstv} \text{ (} boha - tstv - o \text{)} \quad (3.226)$$

$$\mathbf{dštv} \rightarrow \mathbf{tstv} \text{ (} lo - dštv - o \text{)} \quad (3.227)$$

$$\mathbf{ndsk} \rightarrow \mathbf{ntsk} \text{ (} Burgu - ndsk - o \text{)} \quad (3.228)$$

§220 Spoluhláskové skupiny *zsk*, *zšť* sa zjednodušujú odpadnutím poslednej spoluhlásky základu slova, t. j. spoluhlásky *z*, pred príponou *-ský*, *-sko*, *-ština*. Tak je to v slovách *Francúzsko*, *francúzsky*, *francúzština*, *indonézsky*, *perzský*, *perzština*, *rhodézsky*, *Sliezsko*, *sliezsky*, *tunguzský* a pod. Vyslovujeme

[frantsU:skO, frantsU:skI, frantsU:ScIna, IndOnE:skI, pErskI:, pErScIna, rOdE:skI, sLI_^EskO, sLI_^EskI, tUNgUskI:] atď.

$$\mathbf{zsk} \parallel \rightarrow \mathbf{skI} : (\textit{tunggu} - \textit{zský}) \quad (3.229)$$

$$\mathbf{zsky} \parallel \rightarrow \mathbf{skI} \quad (\textit{indoné} - \textit{zsk} - \textit{y}) \quad (3.230)$$

$$\mathbf{zsko} \parallel \rightarrow \mathbf{skO} \quad (\textit{Slie} - \textit{zsk} - \textit{o}) \quad (3.231)$$

$$\mathbf{zština} \parallel \rightarrow \mathbf{ScIna}(\textit{per} - \textit{zšt} - \textit{ina}) \quad (3.232)$$

§221 Spoluhláskové skupiny *dš*, *tš*, *tč* sa vyslovujú ako *č*, ak sa pri nich nachádza ďalšia spoluhláska (t. j. v skupinách *dšť*, *tšť*, *ptč*, ako v slovách *egyptčina*, *odštiepok* (iné je *oštiepok*), *posvetštit*, *zľudštiet*, *zľudštit* a pod. Vyslovujeme

[OtScL^EpOk, EgIptSIIna, pOsvEtScIc, zLUtScL^Ec, zLUtScIc] atď.

$$\mathbf{dšť} \rightarrow \mathbf{tSc}(o - \textit{dš} - \textit{tiepok}) \quad (3.233)$$

$$\mathbf{tšť} \rightarrow \mathbf{tSc}(\textit{posve} - \textit{tš} - \textit{tit}) \quad (3.234)$$

$$\mathbf{ptč} \rightarrow \mathbf{ptS}(\textit{egyp} - \textit{tč} - \textit{ina}) \quad (3.235)$$

Spoluhláskové skupiny *ntsk*, *ndsk*

§222 Po spoluhláske *n* sa spoluhláskové skupiny *tsk*, *dsk* v neutrálnom a vyššom štýle zjednodušujú asimiláciou *ts*, *ds* na *c*, ako sa *ts*, *ds* zjednodušuje v susedstve iných spoluhlások.

Slová ako *asistentský*, *cirkusantský*, *dekadentský*, *dirigentský*, *emigrantský*, *holandský*, *prezidentský*, *študentský* v neutrálnom a vyššom štýle vyslovujeme

[asIstEntskI:, tsIrkUsantskI:, dEkadEntskI:, dIrigEntskI:, EmIgrantskI:, hOlantskI:, prEzIdEntskI:, StUdEntskI:].

V nižšom štýle sa môžu tieto skupiny vyslovovať s vynechaním spoluhlásky *t*, *d*, t. j. [asIstEnskI:, EmIgranskI:, prEzIdEnskI:].

$$\mathbf{ntsk} \rightarrow \mathbf{ntsk}(\textit{cirkusan} - \textit{tsk} - \textit{ý}) \quad (3.236)$$

$$\rightarrow \mathbf{nsk} (\textit{cirkusan} - \textit{tsk} - \textit{ý}) \quad (3.237)$$

$$\mathbf{ndsk} \rightarrow \mathbf{ntsk}(\textit{holan} - \textit{dsk} - \textit{ý}) \quad (3.238)$$

$$\rightarrow \mathbf{nsk} (\textit{holan} - \textit{dsk} - \textit{ý}) \quad (3.239)$$

Tieto pravidla museli byť v implementácii umiestnené pred pravidlo (3.222), inak by neboli nikdy aplikované. Pravidlo (3.228) bolo zrušené ¹⁰.

¹⁰Na tomto mieste treba poznamenať, že všetky pravidlá o spoluhláskových skupinách boli v implementácii umiestnené ešte pred asimilačnými pravidlami, čiže hneď za výslovnosť *j* a *v*, aby sa zabránilo nežiadúcemu aplikovaniu niektorého z týchto pravidiel na spoluhláskové skupiny.

Spoluhláskové skupiny *ptsk*, *stsk*, *ststv*, *sťst*

§223 Tieto spoluhláskové skupiny sa zjednodušujú vypadnutím spoluhlásky *t* (resp. spoluhlások *st*, *st*). Slová ako *egyptský*, *koptský*, *mestský*, *modistsvo*, *šesťsto* vyslovujeme

[EgIpskI:, kOpskI:, mEskI:, mOdIstvO, SEstO] atď.

$$\mathbf{ptsk} \rightarrow \mathbf{psk} \text{ (egy - pťský)} \quad (3.240)$$

$$\mathbf{stsk} \rightarrow \mathbf{ssk} \text{ (me - stsk - ý)} \quad (3.241)$$

$$\mathbf{ststv} \rightarrow \mathbf{sstv} \text{ (modi - ststv - o)} \quad (3.242)$$

$$\mathbf{sťst} \rightarrow \mathbf{sst} \text{ (še - sťst - o)} \quad (3.243)$$

Spoluhláskové skupiny *stk*, *štk*; *nstv*, *mstv*; *tkn*, *tkň*, *ktn*, *ktň*; *pt*, *ptk*

§224 Tieto spoluhláskové skupiny sa v spisovnej výslovnosti nezjednodušujú. Slová ako *čiastka*, *dotkne*, *dotknúť sa*, *dvanásťka*, *efektne*, *efektný*, *Egypt*, *egyptka*, *jedenásťka*, *klamstvo*, *milosrdenstvo*, *striktne*, *striktný*, *stvoríť*, *vytkne*, *vytknúť* atď. vyslovujeme

[tSl_`astka, dOtkJnE, dOtknU:csa, dvana:stka, EfEktJE, EfEktnI:, EgIpt, EgIptka, jEdEna:stka, klamstvO, mlOsr=J\EnstvO, striktJE, striktnI:, stvOrIc, vItkJE, vItknU:c] atď.

3.4.13 Ostatné pravidlá

Pravidlá uvedené v predchádzajúcich častiach vychádzali z pravidiel slovenskej výslovnosti, v ktorej sú však uvedené len pravidla pre výslovnosť v prípadoch, kde by mohli vzniknúť nejasnosti, resp. neboli tam uvedené pravidlá pre prípady, kde niet pochýb o výslovnosti. Napríklad pravidlá o výslovnosti spoluhlások sa týkali buď asimilácií, zdvojených spoluhlások, spoluhlások na hranici slov, mäkkých a tvrdých spoluhlások, avšak nikde nebola uvedená napríklad výslovnosť znelych a neznelých spoluhlások, ak sa takáto spoluhláska nachádza medzi dvoma samohláskami. Automaticky sa predpokladá, že výslovnosť v takýchto prípadoch je bezproblémová a nie je nutné sa o nej zmieňovať.

Výnimkou je v tomto prípade výslovnosť slabičného a neslabičného *r* a *l*, pre ktoré neboli v [33] popísané žiadne pravidla. Keďže pravidlá pre výslovnosť slabičného a neslabičného *r* a *l* nie sú úplne samozrejmé, boli použité pravidlá z [47], ktoré hovoria:

§225 Slabičnú funkciu majú fonémy *r*, *l* v týchto podmienkach:

- Musí ich predchádzať šumový konsonant alebo konsonant *m*, *v*. Môže ich predchádzať aj skupina spoluhlások obvyklých na začiatku slabiky.
- Spoluhláska, ktorá nasleduje za slabičným *r* alebo *l* je šumová spoluhláska alebo niektorá zo spoluhlások *m* *v* *n*. Táto spoluhláska môže uzatvárať slabiku, alebo sa ňou môže začínať nasledujúca slabika, v ktorej je nositeľom slabičnosti vokál (krátky, dlhý, diftong). Zasa tu môže byť aj skupina spoluhlások.

- Po slabičnom r môže nasledovať ako konsonant aj sonóra l alebo $ḷ$, ale iba ako konsonant, ktorým sa začína nasledujúca slabika, v ktorej je nositeľom slabičnosti vokál (krátky, dlhý, diftong).

Aj výber spoluhlások, v okolí ktorých sa vyskytujú slabičné r , l na niečo ukazuje. Slabičné r a l môže stáť medzi všetkými šumovými spoluhláskami. Zo sonórnych spoluhlások môžu byť len spoluhlásky m , v vo všetkých tých pozíciách, v ktorých sú šumové spoluhlásky. Spoluhlásky n , $ŋ$ môžu stáť len v pozícii za slabičným r alebo l a môžu uzatvárať slabiku. Spoluhlásky l , $ḷ$ môžu stáť tiež len za slabičným r , ale musí sa nimi začínať nová slabika. Spoluhláska r nemôže tvoriť konsonantické okolie slabičnej sonóry. Je zo sonórnych spoluhlások svojou distribúciou "najsonórnejšia".

Na základe týchto pravidiel boli definované nasledujúce transkripčné pravidlá:

$$| (SP+)r([SSPm\bar{v}n]) \rightarrow \backslash 1 \mathbf{r} = \backslash 2(tv - r - dil) \quad (3.244)$$

$$| (SP+)r | ([ḷ]SM) \rightarrow \backslash 1 \mathbf{r} = \backslash 2(mud - r - lant) \quad (3.245)$$

$$| (SP+)l([SSPm\bar{v}n]) \rightarrow \backslash 1 \mathbf{l} = \backslash 2(op - l - zlost) \quad (3.246)$$

$$\acute{r} \rightarrow \mathbf{r} =: (h - \acute{r} - stka) \quad (3.247)$$

$$\acute{l} \rightarrow \mathbf{l} =: (poz\acute{d} - \acute{l} - \acute{z}) \quad (3.248)$$

SSP označuje šumovú spoluhlásku, t. j.: b , c , \check{c} , d , \acute{d} , dz , $d\check{z}$, f , g , ch , j , k , p , s , \check{s} , t , \acute{t} , z , \check{z} ¹¹. Pravidlá (3.247) a (3.248) sú jednoznačné, keďže dlhé slabičné r a l sa v slovenčine označuje samostatnou grafémou.

V programovej realizácii bolo nutné definovať aj vyššie spomínané jednoduché pravidlá, aby boli prepísané všetky spoluhlásky. Ide o jednoduché pravidlá priameho prepisu tých grafém, na ktoré nebolo aplikované žiadne pravidlo, na fóny. Ako bolo spomenuté vyššie, po aplikácii všetkých pravidiel z predchádzajúcich častí zostali neprepísané už len grafémy, kde je proces prepisu jednoznačný. Pre ne boli definované nasledujúce transkripčné pravidlá:

$$\mathbf{b} \rightarrow \mathbf{b} (se - b - a) \quad (3.249)$$

$$\mathbf{c} \rightarrow \mathbf{ts} (sí - c - e) \quad (3.250)$$

$$\check{\mathbf{c}} \rightarrow \mathbf{tS} (sta - \check{c} - ilo) \quad (3.251)$$

$$\mathbf{dz} \rightarrow \mathbf{dz} (nú - dz - a) \quad (3.252)$$

$$\mathbf{d\check{z}} \rightarrow \mathbf{dZ} (ga - d\check{z} - o) \quad (3.253)$$

$$\mathbf{f} \rightarrow \mathbf{f} (š\acute{e} - f - a) \quad (3.254)$$

$$\mathbf{g} \rightarrow \mathbf{g} (a - g - \acute{a}t) \quad (3.255)$$

$$\mathbf{h} \rightarrow \mathbf{h} \backslash (je - h - o) \quad (3.256)$$

$$\mathbf{ch} \rightarrow \mathbf{x} (ti - ch - o) \quad (3.257)$$

$$\mathbf{k} \rightarrow \mathbf{k} (ru - k - a) \quad (3.258)$$

$$\mathbf{l} \rightarrow \mathbf{l} (ko - l - ok) \quad (3.259)$$

$$\mathbf{m} \rightarrow \mathbf{m} (hu - m - or) \quad (3.260)$$

¹¹Keďže tu nevznikajú nejasnosti, je použitá ortografická forma zápisu.

$$\mathbf{p} \rightarrow \mathbf{p} \quad (\text{zá} - p - \text{as}) \quad (3.261)$$

$$\mathbf{r} \rightarrow \mathbf{r} \quad (\text{op} - r - \text{ava}) \quad (3.262)$$

$$\mathbf{s} \rightarrow \mathbf{s} \quad (\text{mu} - s - \text{í}) \quad (3.263)$$

$$\mathbf{\check{s}} \rightarrow \mathbf{S} \quad (\text{du} - \check{s} - \text{a}) \quad (3.264)$$

$$\mathbf{v} \rightarrow \mathbf{v} \quad (\text{bojo} - v - \text{ať}) \quad (3.265)$$

$$\mathbf{w} \rightarrow \mathbf{v} \quad (\text{mega} - w - \text{att}) \quad (3.266)$$

$$\mathbf{x} \rightarrow \mathbf{ks} \quad (\text{se} - x - \text{u}) \quad (3.267)$$

$$\mathbf{z} \rightarrow \mathbf{z} \quad (\text{ná} - z - \text{or}) \quad (3.268)$$

$$\mathbf{\check{z}} \rightarrow \mathbf{Z} \quad (\text{mu} - \check{z} - \text{ovi}) \quad (3.269)$$

3.4.14 Výnimky

Vzhľadom na to, že v predchádzajúcich častiach pri definovaní pravidiel existovalo mnoho výnimiek, výnimkám — teda hlavne ich spracovaniu — bolo potrebné venovať špeciálnu pozornosť. Ciele takého postupu možno zhrnúť do dvoch bodov:

- Zabrániť zbytočnému nárastu slovníka výnimiek a súčasne pokryť čo najširší rozsah.
- Zabrániť aplikovaniu výnimky v prípadoch, keď to nie je potrebné.

Slová v slovenčine môžu mať veľké množstvo tvarov, rozsah tejto problematiky je v stručnosti načrtnutý v časti 3.4.15. V prípade, že je potrebné do zoznamu výnimiek zaradiť slovo, kde problematická časť, pre ktorú je slovo zaradené medzi výnimky, je obsiahnutá napríklad v začiatkovej časti slova, musia sa okrem daného slova pridať medzi zoznam výnimiek aj všetky jeho tvary a alternatívne formy. Napríklad, ak má byť zaradené medzi výnimky slovo *email*, ktoré je v dnešnej dobe dosť bežné, musia sa uvažovať aj slová odvodené od slova *email* a to *emaily*, *emailom*, *emailami*, *emailoch*, *emailový*, ... Obdobne by bolo potrebné postupovať v prípade, že výslovnostná výnimka je na konci slova, ako napríklad v prípade slova *dnes* a taktiež v strede slova. Posledný prípad umožňuje meniť nielen predpony, alebo len prípony, ale aj predpony, aj prípony. Z toho vyplýva, že zo súčasného rozsahu zoznamu výnimiek, ktorý obsahuje okolo 900 položiek, by sa hravo vytvoril súbor s viac než 10000 položkami, čo je vyše 10 násobný nárast, nehovoriac o tom, že už samotné definovanie všetkých možných tvarov nie je jednoduchá činnosť.

Z toho dôvodu bol definovaný zoznam výnimiek tým spôsobom, že obsahuje len minimálnu nutnú časť slova s informáciou o tom, ktorá časť nesmie byť ďalej menená. Spôsob uloženia výnimiek v slovníku a postup pri ich spracovaní demonštrujem na štyroch typických príkladoch zo zoznamu výnimiek:

1. **bosk** | **bOzk** – slovo *bosk* je v slovníku výnimiek zaradené bez akýchkoľvek obmedzení, to znamená, že túto výnimku je možné aplikovať na ľubovoľné slovo. Čiže v prípade ak budú na vstupe slová *bosk*, *boskávať*, *poboskať*,

vyboskávať, poprípade iná odvodenina, v takomto slove bude v prepise časť *bosk* nahradená postupnosťou *bOzk*.

2. **email** | **ImEI_1** ; **EmaI_1** – v prípade slova *email* je možné modifikovať len slová začínajúce postupnosťou *email*, čiže slová ako *emailový*, *emailovať* a podobne. V tomto prípade je možná dvojaká výslovnosť.
3. **dnes\$** | **dJE**s – v tomto prípade je to presne opačne. Toto pravidlo bude aplikované len na slová končiace na postupnosť *dnes*. Čiže napríklad *dodnes* avšak nie *nadnesené*, *odnesie* a iné.
4. **oni\$** | **OJI** – posledným typom sú slová, ktoré sa musia vyskytnúť presne v tom tvare, v akom sa nachádzajú v zozname výnimiek. Sem patria hlavne niektoré osobné zámená. Do tejto kategórie spadá aj väčšina dvojslovných výnimiek, ktoré tvorí väčšinou predložka a osobné zámeno, ako napríklad *s nami*, *k nej* a podobne.

Vo vyššie uvedenom zozname typov výnimiek si možno všimnúť, že obmedzenia sú zapísané vo forme regulárnych výrazov. Presne týmto spôsobom je zostavený zoznam výnimiek a celý systém je napísaný v jazyku Perl. S použitím regulárnych výrazov je – v prípade takejto štruktúry – samotný proces ošetrovania výnimiek veľmi jednoduchý.

Všetky výnimky zo zoznamu výnimiek sú aplikované na začiatku a ďalej sa pokračuje štandardným spracovaním. Takže ak v slove *vyboskávať* bola nahradená len časť *bosk*, predpona *vy* a prípona *ávať* sú ošetrené všeobecnými pravidlami.

Na začiatku tejto časti boli spomenuté dva problémy pri práci s výnimkami. Ten druhý – načrtnutý vyššie pri opise štruktúry uloženia dát – spočíva v citlivom "osekaní" slova pred tým, než je vložené do slovníka výnimiek. To znamená nielen skrátiť ho tak, aby pokrývalo všetky možné tvary, ale aj neskrátiť ho až príliš, aby nebol zasiahnutý nejaký tvar úplne iného slova. Z toho dôvodu bol slovník výnimiek po vytvorení ešte raz pozorne prekontrolovaný, aby sa do maximálnej možnej miery zabránilo nežiadúcim substitúciám.

3.4.15 Automatické určovanie morfematických hraníc

Na rozdiel od slabičných hraníc sa v prípade morfematických hraníc v slovenčine máme o čo oprieť, aj čo sa týka počítačového spracovania. Za základne dielo v tejto oblasti pre slovenčinu je možné považovať [49], z ktorého záverov bolo vychádzané.

Morfematická hranica sa v slovenčine vyskytuje na týchto miestach:

- na hranici slov,
- na hranici slovných základov v zložených slovách,
- na hranici predpony a koreňa alebo základu slova,

- na hranici koreňa alebo základu slova a odvodzovacej (výnimočne aj ohýbacej) prípony.

Prvý prípad možno detekovať bez problémov. V ostatných sa už bez pomocných slovníkov nedá zaobísť.

V [49] sa o morfeematickej analýze slovenčiny píše: *Pretože v slovníku nie sú uvedené celé slová, ale len korene a slovesné kmene, systém, aby mohol v slovníku identifikovať koreň, musí ho najprv izolovať od tvarových predpôn a prípon.*

V slovenskom slove možno oddeliť 0–5 znakov ako predponu, 0–4 znaky ako skloňovaciú (resp. časovaciú) príponu a 0–3 znaky ako relačnú príponu (pri slovesách). To dáva spolu $5 \times 4 \times 3 = 60$ možností rozštvrtienia tvaru na koreň a afixy. Keďže o koreni nevieme, či nie je pozmenený niektorou z 30-tich druhov alternácii, musíme sa pokúsiť odalternovať ho. Tým dostávame $60 \times 30 = 1800$ hypotetických koreňov pri jednom, jedinom tvare.

Na tomto mieste treba podotknúť, že SAPFO — systém autora vyššie citovaných slov — je napísaný v jazyku Prolog.

Slovenské tvarotvorné afixy sa nekombinujú ľubovoľne, ale ich spájateľnosť je veľmi špecifická. Navyše z veľkého počtu hláskových kombinácií nie všetky, ale iba niektoré tvoria slovenský afix.

SAPFO využíva skutočnosť, že slovenské tvary nemôžu byť zakončené ľubovoľným koncovým podreťazcom, ale iba podreťazcami z istej konečnej množiny. Ak slovo má byť napr. podstatným menom, musí byť zakončené niektorou z 36-tich substantívnych prípon. Ak to má byť navyše podstatné meno v určitom rode, čísle či páde, musí byť zakončené príponou, ktorá je pre tieto gramatické kategórie charakteristická.

*Ak by analyzovaný tvar mal byť napr. l-ovým slovesným tvarom, musí byť zakončený na **l**, **la**, **lo** alebo **li**. Inak ním nie je. Nie naopak; tvar sa môže končiť na **l**, **la**, **lo**, **li** a l-ovým slovesným tvarom byť nemusí. Preto je hypotézu potrebné overiť.*

Teda morfológický modul SAPFO sa správa ako konečný nedeterministický automat, ktorého vstupným reťazcom sú fonémy analyzovaného tvaru (sprava doľava smerom od konca slova). Podľa vstupujúcich foném sa automat dostáva do stavov, ktoré mu ukladajú overiť len malý počet hypotéz-možností (teda už nie všetky). Obdobným spôsobom sa postupuje zľava doprava od začiatku slova pri oddeľovaní prefixov.

Ako vidieť, vyššie uvedený problém nie je triviálny a okrem analýzy by k nemu musel byť vytvorený aj modul syntézy na overovanie hypotéz. Pri realizácii vyššie uvedeného postupu by sa vyskytlo viac — nie neriešiteľných — problémov, konkrétne:

- Slovník koreňov a slovesných kmeňov
- Návrh a implementácia konečného automatu na analýzu prefixov a afixov
- Systém spätnej syntézy na overovanie hypotéz

- Ošetrovanie alternovaných koreňov

V prípade alternovaných koreňov sa uvažuje skutočnosť, že niektoré majú v niektorých tvaroch modifikovaný koreň. Napríklad slovo *turkami*. Koreň slova *turkami* je *turek* a nie *turk*. V takýchto prípadoch je nutné po oddelení prípon modifikovať slovo za účelom získania správneho koreňa. Samozrejme tento proces podlieha tiež určitým pravidlám a v tomto prípade, keďže zakončenie obsahuje dve spoluhlásky, na základe zákonitostí v slovenčine sa medzi ne môže vložiť buď *e* alebo *o*, čím sa hneď na prvý pokus získa správny koreň.

Realizácia celého takehoto systému je námetom na samostatnú prácu, a preto sa realizovalo náhradné riešenie, ktoré síce nebude vykonávať morfológickú segmentáciu pre ľubovoľné slová, no umožňuje otestovanie celého systému fonetickej transkripcie aspoň v rámci obmedzení spôsobených práve morfológickým modulom.

Za týmto účelom bol navrhnutý jednoduchý morfológický slovník, ktorý obsahuje slova rozdelené na morfémy. Z dôvodu pokrytia čo najväčšieho počtu slov čo najmenším počtom záznamov v slovníku, neobsahuje celé slová, ale len časti slov, tak ako je to v prípade výslovnostných výnimiek opísaných v časti 3.4.14. V každom slove v slovníku nie je označená len morfológická hranica ako taká, ale je označené, či ide o zložené slovo, príponu alebo predponu, keďže táto informácia je dôležitá pre niektoré transkripčné pravidlá definované pre transkripciu spoluhlások.

3.4.16 Zhrnutie

Proces analýzy možnosti definovať transkripčné pravidlá na generovanie fonetického prepisu v slovenčine poukázal na niekoľko problémových oblastí:

- V prvom rade je to prepis cudzích a prebratých slov v slovenčine. Na tento očakávaný problém bolo poukázané už na začiatku pri definovaní cieľov týkajúcich sa fonetického prepisu, s tým, že v prípade cudzích a prevzatých slov budú pravidlá definované len vtedy, ak to bude možné, resp. nebude to vyžadovať špeciálne úsilie mimo rámca tejto práce. Cudzie, resp. prevzaté slová tvoria značnú časť slovníka výnimiek.
- Druhý vážny problém je potreba znalosti morfológických hraníc v prípade aplikovania niektorých pravidiel, hlavne zo skupiny znelostných asimilácií. Tejto problematike je venovaná časť 3.4.15. Zo záverov časti o morfológickom delení vyplýva, že na to, aby bolo možné vykonať spoľahlivú morfológickú analýzu, je potrebný slovník s koreňmi slov alebo morfológický slovník slovenčiny. Toto možno považovať za skutočnosť komplikujúcu vytvorenie autonómneho systému schopného generovať výslovnosť s minimálnym počtom pomocných slovníkov.
- Problém slabičných hraníc, ktoré sú potrebné napríklad v prípade pravidiel pre výslovnosť hlásky [j] a niektorých ďalších. Tomuto problému je venovaná

časť 3.5. Navrhnuté postupy opísané v časti 3.5 sú pre účely transkripcie vhodné a tvoria relatívne kompaktnú a samostatnú časť bez potreby externých slovníkov.

- Problematika "nespisovnej" výslovnosti je špecifický problém, ktorý bol riešený len na základe informácií z pravidiel výslovnosti o nespisovnej výslovnosti a na základe vlastných skúsenosti. Z toho vyplýva, že mnohé lokálne sa vyskytujúce výslovnostné špecifiká neboli zahrnuté medzi pravidlá.

Záverom možno skonštatovať, že v prípade vyriešenia vyššie spomínaných problémov je možné pre slovenčinu vytvoriť systém na spoľahlivé generovanie fonetickej transkripcie pomocou produkčných pravidiel.

3.5 Automatické určovanie hraníc slabík

Určovanie hraníc slabík komplikuje skutočnosť, že už samotná definícia slabiky nie je jednoznačná [53]. Na definíciu slabiky existuje niekoľko rôznych pohľadov. Ako východisko pre ďalšiu prácu bola zvolená charakteristika E. Paulinyho, vychádzajúca z jednoznačného fonologického hľadiska [44]. Presná definícia slabiky pre účely určovania hraníc nie je nevyhnutne dôležitá, keďže navrhnuté postupy pre určenie slabičných hraníc takúto definíciu nevyžadujú.

Tak ako neexistuje jednoznačná definícia slabiky, neexistujú ani presné pravidlá na určovanie hraníc slabík. Je možné zadefinovať sadu jednoduchých pravidiel na delenie slova na slabiky, avšak takéto pravidlá dosiahnu maximálne 60 %-nú úspešnosť. Pri použití zložitejších pravidiel je možné dosiahnuť 85 %-nú úspešnosť. Ďalšie zvyšovanie presnosti vedie hlavne k nárastu množstva výnimiek.

Druhým problémom je skutočnosť, že pre niektoré slová existuje viacero možných delení na slabiky. To vyplýva aj z vyššie spomínanej absencie jednoznačnej definície slabiky v slovenčine. Napríklad slovo *bystrý* je možné rozdeliť ako *by-strý*, *bys-trý* alebo aj *byst-rý*. Vo všetkých týchto troch prípadoch zostáva počet slabík nemenný a vo všetkých troch prípadoch je delenie na slabiky správne.

Pri návrhu spôsobu určovania hraníc slabík bolo nutné vychádzať z nasledujúcich skutočností:

- V slovenčine neexistujú presné pravidlá na určenie hraníc slabík.
- V niektorých prípadoch je možné delenie viacerými spôsobmi, pričom všetky sú správne.

V navrhnutom riešení je kombinovaná jednoduchá segmentácia založená na pravidlách a novonavrhnutý prístup aplikujúci teóriu jazykových modelov na slabičnú segmentáciu.

3.5.1 Určovanie hraníc slabík pomocou jednoduchých pravidiel

Pri hľadaní jednoduchých pravidiel na určovanie hraníc slabík sa vychádzalo hlavne z [44], [53], [61]. Medzi jednoduché pravidlá patria napríklad pravidlá

uvedené pri delení slov na konci riadku v pravidlách slovenského pravopisu [61], vzťahujúce sa na slabičnú hranicu.¹² Tieto je možné modifikovať pre potreby slabičnej segmentácie do nasledujúcej formy:

- Ak medzi dvoma samohláskami (dvojháskami, samohláskou a dvojháskou, slabičnými $r, l, \acute{r}, \acute{l}$ a samohláskou alebo dvojháskou) je jedna spoluhláska, slabičná hranica je pred spoluhláskou, napr. *že-na, pra-co-vať, bie-ly, vl-na, vr-tieť, vr-ba, Sl-ňa-va, ka-me-nár, pa-ra-bo-la, čia-ra, bie-lia-reň, zna-me-niu*.
- Ak medzi dvoma samohláskami (dvojháskami, samohláskou a dvojháskou, slabičnými $r, l, \acute{r}, \acute{l}$ a samohláskou alebo dvojháskou) je skupina dvoch spoluhlások, slabičná hranica je na rozhraní medzi dvoma spoluhláskami, napr. *všet-ci, všet-ky, žat-va, mas-lo, lás-ka, prch-ký, mĺk-vý, maš-lič-ka, ot-cami, chlap-cami, kviet-kami, chrb-tami, pas-ca, Pop-rad*.

Formálne tieto pravidlá možno zapísať nasledujúcim spôsobom:

$$h_{sp} \in \{ \text{Množina všetkých spoluhlások} \} \quad (3.270)$$

$$h_{sam} \in \{ \text{Množina všetkých samohlások} \} \quad (3.271)$$

$$h_{dif} \in \{ \text{Množina všetkých dvojhások} \} \quad (3.272)$$

$$h_{slsp} \in \{ \acute{r}, \acute{l} \} \quad (3.273)$$

$$h_{slspf} \in \{ r, l, \acute{r}, \acute{l} \} \quad (3.274)$$

$$h_{sam}h_{sp}h_{sam} \rightarrow h_{sam} - h_{sp}h_{sam} \quad (3.275)$$

$$h_{dif}h_{sp}h_{dif} \rightarrow h_{dif} - h_{sp}h_{dif} \quad (3.276)$$

$$h_{sam}h_{sp}h_{dif} \rightarrow h_{sam} - h_{sp}h_{dif} \quad (3.277)$$

$$h_{slsp}h_{sp}h_{sam} \rightarrow h_{slsp} - h_{sp}h_{sam} \quad (3.278)$$

$$h_{slsp}h_{sp}h_{dif} \rightarrow h_{slsp} - h_{sp}h_{dif} \quad (3.279)$$

$$h_{sam}h_{sp}h_{sp}h_{sam} \rightarrow h_{sam}h_{sp} - h_{sp}h_{sam} \quad (3.280)$$

$$h_{dif}h_{sp}h_{sp}h_{dif} \rightarrow h_{dif}h_{sp} - h_{sp}h_{dif} \quad (3.281)$$

$$h_{sam}h_{sp}h_{sp}h_{dif} \rightarrow h_{sam}h_{sp} - h_{sp}h_{dif} \quad (3.282)$$

$$h_{slspf}h_{sp}h_{sp}h_{sam} \rightarrow h_{slspf}h_{sp} - h_{sp}h_{sam} \quad (3.283)$$

$$h_{slspf}h_{sp}h_{sp}h_{dif} \rightarrow h_{slspf}h_{sp} - h_{sp}h_{dif} \quad (3.284)$$

kde $-$ označuje slabičnú hranicu. V (3.278) a (3.279) sa neuvažuje slabičné r, l z dôvodu vyhnutia sa zložitým pravidlám. V prípade r, l by bolo nutné určiť,

¹²Toto este bude podrobnejšie rozpracované.

či ide o slabičné alebo neslabičné. V (3.283) a (3.284) tento problém nevzniká, keďže v slovenčine sa nevyskytuje spoluhlásková postupnosť pozostávajúca z troch spoluhlások začínajúca sa neslabičným r . V (3.272) je potrebné rozlíšiť medzi dvojhláskou a samohláskovou skupinou, no keďže tomuto problému sa venuje časť 3.4.2, implementácia (3.272) bola jednoduchá.

Ako bolo uvedené vyššie, tieto pravidlá boli odvodené z pravidiel pre delenie slov na konci riadku. Slabičná segmentácia pomocou nich je síce správna, avšak nie jediná správna. Ak sa napríklad (3.280) aplikuje na slovo *hospodár*, výsledok bude *hos-podár*. Je to síce správny výsledok, avšak správne delenie je aj *ho-spodár*. Pre slovo *verbovať* je však segmentácia *ver-bovať* jediná správna.

Pravidlá (3.275) až (3.284) je možné rozpísať pre dvoj až päťslabičné slová a získať tak možné kombinácie pre jednoduché postupnosti [53], avšak na naše účely je táto jednoduchá slabičná segmentácia dostačujúca.

3.5.2 Štatistický prístup pri určovaní hraníc slabík

Keďže z vyššie uvedeného vyplýva, že delenie slov na slabiky pomocou pravidiel je pre slovenčinu problematické, použil sa štatistický prístup, kde všeobecne známe postupy, ktoré sa používajú pri jazykových modeloch, boli modifikované pre použitie na slabičnú segmentáciu.

Na rozdiel od jazykových modelov, kde základnou jednotkou je slovo, základnou jednotkou v tomto prípade je slabika. Každé slovo sa najprv rozdelí na postupnosti všetkých možných slabík. Pre každú takúto postupnosť slabík \mathbf{S} , kde

$$\mathbf{S} = s_1, s_2, \dots, s_n \quad s_i \in \mathcal{S} \quad (3.285)$$

a \mathcal{S} je množina všetkých možných slabík, je možné na základe Baysovoho kritéria výpočtu podmienených pravdepodobností definovať podmienenú pravdepodobnosť slabiky s_i pre každú postupnosť slabík \mathbf{S} ako

$$P(\mathbf{S}) = \prod_{i=1}^n P(s_i | s_1, \dots, s_{i-1}) \quad (3.286)$$

kde $P(s_i | s_1, \dots, s_{i-1})$ je podmienená pravdepodobnosť, že slabika s_i bude nasledovať po slabikách s_1, \dots, s_{i-1} . Postupnosť s_1, \dots, s_{i-1} je možné označiť ako históriu h_i a (3.286) zapísať do tvaru

$$P(\mathbf{S}) = \prod_{i=1}^n P(s_i | h_i) \quad (3.287)$$

Rovnica (3.286) hovorí, že pravdepodobnosť danej postupnosti slabík \mathbf{S} je rovná pravdepodobnosti prvej slabiky krát pravdepodobnosť druhej slabiky, za podmienky, že sa pred ňou nachádza prvá slabika, atď. krát pravdepodobnosť poslednej slabiky v prípade, že ju predchádzajú predchádzajúce slabiky.

V prípade, že sa uvažujú len dve predchádzajúce slabiky, možno (3.286) zapísať v tvare

$$P(\mathbf{S}) = \prod_{i=1}^n P(s_i | s_{i-2}, s_{i-1}) \quad (3.288)$$

Postupnosť $s_i | s_1, \dots, s_{i-1}$ je možné považovať za postupnosť generovanú konečným automatom, ktorý je v čase $i - 1$ v stave Φ_{i-1} a ďalšia slabika ho prevedie do stavu Φ_i . Rovnicu (3.286) potom možno zapísať ako

$$P(\mathbf{S}) = \prod_{i=1}^n P(s_i | \Phi_{i-1}) \quad (3.289)$$

Na odhad pravdepodobnosti $P(s_i | \Phi_{i-1})$ bol použitý "slabičný korpus", vytvorený z 10000 náhodne vybraných slovenských slov, ktoré boli rozdelené na slabiky. Každé slovo rozdelené na slabiky bolo spracované konečným automatom, ktorý kumuloval početnosť $C(s, \Phi)$, kde $C(s, \Phi)$ označuje početnosť výskytu slabiky s v stave Φ . Ak $C(\Phi)$ označuje koľkokrát bol automat v stave Φ ,

$$C(\Phi) = \sum_s C(s, \Phi) \quad (3.290)$$

potom odhad požadovanej pravdepodobnosti bude

$$P(s_i | \Phi_i = \Phi) = \frac{C(s_i, \Phi)}{C(\Phi)} \quad (3.291)$$

V prípade výpočtu pravdepodobností podľa (3.291) je dôležitá správne definovaná trénovacia množina, ktorá bude viesť k spoľahlivému odhadu požadovaných pravdepodobností.

Ako bolo uvedené vyššie, navrhnuté riešenie uvažuje len dve predchádzajúce slabiky. Potom z (3.288) a (3.290) (3.291) vyplýva

$$P(s_3 | s_1, s_2) \doteq P^*(s_3 | s_1, s_2) = \frac{C(s_1, s_2, s_3)}{C(s_1, s_2)} \quad (3.292)$$

kde $P(\cdot | \cdot)$ je podmienená pravdepodobnosť výskytu.

Bohužiaľ rovnica (3.292) nie je vhodná na výpočet pravdepodobnosti pre danú postupnosť slabík, a to z dvoch dôvodov:

- V slovenčine je bežný výskyt aj jednoslabičných a dvojslabičných slov.
- Nie všetky možné postupnosti slabík s_1, s_2, s_3 sa musia vyskytnúť v trénovej množine.

Na základe vyššie uvedeného je potrebné podmienenú pravdepodobnosť $P(s_3 | s_1, s_2)$ uvažovať ako interpoláciu podmienených pravdepodobností výskytov pre postupnosť trojíc, dvojíc a jednej slabiky:

$$P(s_3 | s_1, s_2) = \lambda_3 P(s_3 | s_1, s_2) + \lambda_2 P(s_3 | s_2) + \lambda_1 P(s_3) \quad (3.293)$$

Nezáporné váhy musia spĺňať podmienku $\lambda_1 + \lambda_2 + \lambda_2 = 1$.¹³ Ostáva určiť optimálne hodnoty pre λ_i .

Slabičný model správajúci sa podľa (3.293) môže byť považovaný za skrytý Markovov model (HMM). Zo začiatočného stavu $\tau_0(s_1, s_2)$ môže prejsť do jedného z troch nasledujúcich stavov $\tau_1(s_1, s_2), \tau_2(s_1, s_2), \tau_3(s_1, s_2)$ s prechodovými pravdepodobnosťami λ_1, λ_2 resp. λ_3 . Každý stav môže produkovať výstup $v \in \mathcal{G}$. Výstupné symboly zo stavov $\tau_1(s_1, s_2), \tau_2(s_1, s_2), \tau_3(s_1, s_2)$ majú pravdepodobnosti $P(v), P(v | s_2), P(v | s_1, s_2)$, (emisné pravdepodobnosti).

Výstupné (emisné) pravdepodobnosti sú v tomto prípade známe. Prechodové pravdepodobnosti je však potrebné určiť. Celkový HMM pre túto situáciu je však pre slabičnú segmentáciu neakceptovateľne veľký, keďže obsahuje $4 \times \mathcal{G}^2$ stavov. Našťastie podľa (3.293) pre $i \in \{1, 2, 3\}$ sú všetky prechodové pravdepodobnosti λ_i vedúce z $\tau_0(s_1, s_2)$ do $\tau_i(s_2, v)$ rovnaké bez ohľadu na aktuálnu kombináciu s_1, s_2, v .

Keďže (3.293) je HMM, odhad optimálnych hodnôt pre λ_i možno urobiť pomocou Baumovho–Welchovho (forward–backward) algoritmu [5], [6].¹⁴ Predtým, než bude opísaný odhad požadovaných parametrov, je potrebné určiť, aké tréningové dáta sú vhodné na odhad váh.

Z (3.293) vyplýva, že na odhad váh nie je možné použiť tie isté dáta, ktoré sú použité na výpočet početnosti výskytu $P^*(|)$, pretože v tomto prípade by bolo $\lambda_3 = 1$ a $\lambda_1 = \lambda_2 = 0$. Z toho vyplýva, že tréningová množina musí byť rozdelená na dve časti. Prvá, väčšia časť dát, bude slúžiť na výpočet početnosti výskytov $P^*(|)$ a druhá, menšia časť, bude použitá na odhad váh λ_i . Samozrejme po tomto kroku môže byť slabičný model vylepšený použitím všetkých dát na opätovný výpočet $P^*(|)$.

Táto technika je niekedy nazývaná "deleted interpolation". Pre zjednodušenie bola λ_i považovaná za konštantu. Je však jasné, že z $P^*(s_3 | s_1, s_2)$ je možné lepšie odhadnúť $P(s_3 | s_1, s_2)$ použitím väčšieho tréningového súboru, λ_i potom bude závisieť od početností $C(s_1, s_2)$ a $C(s_2)$. Jednoduchý spôsob ako to dosiahnuť je zmena štruktúry HMM, a to tak, že pred stavy τ_1 a τ_2 – nie však pred τ_3 – je vložený pomocný stav τ_4 . Tým pádom prechody t_3 a t_4 vedú zo stavu $\tau_0(s_1, s_2)$ do stavov $\tau_3(s_1, s_2)$ a $\tau_4(s_1, s_2)$, a prechody t_1 a t_2 vychádzajú zo stavu $\tau_4(s_1, s_2)$ a vedú do stavov $\tau_1(s_1, s_2)$ a $\tau_2(s_1, s_2)$. Keďže bola zmenená štruktúra HMM pridaním nových prechodových pravdepodobnosti λ'_i , pre λ'_i platí

$$\lambda_1 = \lambda'_4 \times \lambda'_1 \lambda_2 = \lambda'_4 \times \lambda'_2 \lambda_3 = \lambda'_3 \quad (3.294)$$

Samozrejme platí, že $\lambda'_2 = 1 - \lambda'_1$ a $\lambda'_4 = 1 - \lambda'_3$. Výhoda tohto postupu je, že $P(s_3 | s_1, s_2)$ sa ustáli v dvoch krokoch. Najprv sa získa

$$P^*(s_3 | s_2) = \lambda'_1 P^*(s_3) + \lambda'_2 P^*(s_3 | s_2) \quad (3.295)$$

¹³Na určenie pravdepodobnosti nemusí byť použitá početnosť výskytu v tréningovej množine. Je to len jeden z možných prístupov. Ďalšie metódy, ktoré sa používajú pre jazykové modely, možno nájsť napríklad v [30], [51].

¹⁴Tento postup je však v tomto prípade nepraktický a jednoduchší je priamy odhad napríklad pomocou "deleted interpolation" [4].

a potom

$$P(s_3 | s_1, s_2) = \lambda'_4 P^*(s_3 | s_2) + \lambda'_3 P^*(s_3 | s_1, s_2) \quad (3.296)$$

Hodnoty λ'_i je možné odhadnúť pomocou Baumovho–Welchovho algoritmu, avšak ako už bolo uvedené vyššie, tento postup je v tomto prípade zbytočne komplikovaný.

Nech váhy λ sú funkciami početnosti výskytov $C(s_1, s_2)$ a $C(s_2)$. Na základe tohto (3.295) môže byť zmenená na tvar

$$P^*(s_3 | s_2) = \xi(C(s_2)) \times P^*(s_3) + (1 - \xi(C(s_2))) \times P^*(s_3 | s_2) \quad (3.297)$$

a (3.296) na

$$P(s_3 | s_1, s_2) = \theta(C(s_1, s_2)) \times P^*(s_3 | s_2) + (1 - \theta(C(s_1, s_2))) \times P^*(s_3 | s_1, s_2) \quad (3.298)$$

kde koeficient ξ z rovnice (3.297) môže byť odhadnutý nezávisle na začiatku pre všetky rozdielne hodnoty $C(s_2)$. Hodnoty ξ by mali závisieť len na rozsahu, v ktorom sa vyskytuje $C(s_2)$, pretože len málo slabík s_2 bude mať vysoké hodnoty $C(s_2)$. Za týmto účelom je možné definovať $\psi(s_2)$, ktoré bude označovať rozsah, do ktorého patrí početnosť $C(s_2)$. Rozsahy sú určené experimentálne, aby sa dosiahlo pokrytie dostatočného množstva dát. Samozrejme, pre malé početnosti môže daný rozsah obsahovať len jeden výskyt a opačne.

Koeficienty ξ pre (3.297) boli zrátané nasledujúcim spôsobom:

1. Trénovacia množina bola rozdelená na dve časti, tak ako je opísané vyššie.
2. Zrátali sa relatívne početnosti $P^*(s_3 | s_2)$ a $P^*(s_3)$ z prvej časti dát.
3. Zrátal sa výskyt $N(s_2, s_3)$ bigramov (dvojíc) s_2, s_3 v druhej časti trénovacej množiny.¹⁵
4. Zrátala sa hodnota ξ nájdením maxima pre

$$\sum_{N(v) \in \psi} \sum_{s_3} N(v, s_3) \log[\xi \times P^*(s_3) + (1 - \xi) \times P^*(s_3 | v)] \quad (3.299)$$

Hodnota $\xi = \xi(\psi)$ môže byť samozrejme určená pomocou reestimačného procesu z Baumovho–Welchovho algoritmu. Na určenie maximálnej hodnoty ξ však stačí zrítať prvú deriváciu z (3.299)

$$\sum_{N(v) \in \psi} \sum_{s_3} N(v, s_3) \left[\xi + \frac{P^*(s_3 | v)}{P^*(s_3) - P^*(s_3 | v)} \right]^{-1} = 0 \quad (3.300)$$

a zrítať hodnotu ξ . Rovnica (3.300) má jedno riešenie, pretože druhá derivácia z (3.299) je

$$- \sum_{N(v) \in \psi} \sum_{s_3} N(v, s_3) \left[\xi + \frac{P^*(s_3 | v)}{P^*(s_3) - P^*(s_3 | v)} \right]^{-2} \quad (3.301)$$

¹⁵Početnosť v prvej časti trénovacej množiny bola označená ako $C()$ a v druhej časti ako $N()$.

čo je záporné pre všetky hodnoty ξ . Riešenie rovnice 3.300 môže byť nájdené ľubovoľným vhodným prehľadávaním intervalu.

Ako bolo uvedené vyššie, aplikovaný postup je len jeden z mnohých možných postupov. Ďalšie postupy možno nájsť napríklad v [30], [31]. Ako však vyplýva zo získaných výsledkov, presnosť odhadu váh λ_i nie je taká kritická ako v prípade jazykových modelov. Vyplýva to hlavne zo skutočnosti, že tréningová množina (zastúpenie slabík) pokrýva väčšie množstvo možných vstupov¹⁶, a z toho dôvodu je možný lepší odhad pravdepodobností $P(s_3 | s_2, s_1)$, $P(s_3 | s_2)$ a $P(s_3)$.

3.6 Segmentácia rečového signálu

Predtým, než sa pristúpi k samotnému procesu segmentácie, je potrebné definovať

- Aké segmentačné jednotky majú byť použité pri segmentácii.
- Akým spôsobom má byť predspracovaný rečový signál pre potreby vlastnej segmentácie.

Odpoveď na prvú otázku je v časti 3.2. Na predspracovanie rečového signálu bola použitá štandardná schéma, kde zo vstupného signálu (16 kHz, 12 bit) je zrátaný vektor príznakov. Ako vektor príznakov bol použitý 12-rozmerný vektor, ktorého prvky tvorili kepstrálne koeficienty v melovej frekvenčnej stupnici a ich prvé diferencie, čím bol na výstupe získaný 24-prvkový vektor príznakov, ktorý bol rátaný nad 10 ms vstupným oknom.

Tento 24-prvkový vektor bol vstupom pre vektorový kvantizátor, ktorý bol natrénovaný pomocou zhlukovacieho algoritmu typu k-mean [51]. Výstup z kvantizátora bol použitý ako vstup pre samotnú segmentáciu.

3.6.1 Proces segmentácie pomocou HMM

Ako bolo uvedené v časti 1.4, k segmentácii rečového signálu je možné pristupovať dvoma základnými spôsobmi: buď pomocou referenčného systému alebo iteračnými metódami. Keďže žiaden referenčný systém, pomocou ktorého by bolo možné urobiť segmentáciu rečového signálu, nebol k dispozícii, bola zvolená metóda z druhej skupiny, t. j. iteračná.

Proces segmentácie pomocou segmentačného algoritmu typu k-mean je opísaný v časti 1.4.1. Druhý spomínaný postup je pomocou Baumovho–Welchovho (forward–backward) algoritmu. Tento postup bol aplikovaný pri segmentácii slovenských viet použitých pri tréningu systémov na rozpoznávanie reči.

Výber vhodných segmentačných jednotiek, transkripcia tréningových viet, rovnako ako aj predspracovanie rečového signálu už boli opísané v predchádzajúcich

¹⁶Počet možných vstupov je počet rôznych slabík, ktorý je oproti počtu rôznych slov oveľa nižší.

častiach. Tu je uvedený len vlastný proces segmentácie, predpokladajúc, že problémy spočívajúce v príprave dát potrebných pre segmentáciu sú vyriešené.

Vlastná segmentácia pozostáva z dvoch krokov, ktoré sa opakujú, kým možno akceptovať rýchlosť konvergenzie tréningu. Prvým krokom je tréning HMM pre každú segmentačnú jednotku – v tomto prípade fónu – druhým krokom je segmentácia, čiže vyhľadávanie najpravdepodobnejších segmentov zodpovedajúcich postupnosti segmentačných jednotiek pre danú vetu. Na tréning v prvom kroku je použitý Baumov–Welchov algoritmus, na segmentáciu v druhom kroku je použitý Viterbiho algoritmus.

Predtým, než sa pristúpi k samotnému procesu segmentácie (tréningu a následnému segmentovaniu), je nutné definovať jednotlivé komponenty skrytého Markovovho modelu:

1. Počet stavov N pre HMM. Napriek tomu, že stavy sú skryté, mali by zodpovedať nejakým fyzikálnym vlastnostiam modelu (každý HMM modeluje jednu fónu). Ako vhodný počet sa javí použitie troch stavov, ktoré možno vnímať ako začiatočnú časť, centrálnu časť a záverečnú časť každej fóny. Množina všetkých stavov Markovho modelu je trojprvková množina $\mathcal{N} = \{1, 2, 3\} = \{q_t = i; i = 1, 2, 3\}$. Stav v čase t bude označovaný ako q_t .
2. Počet výstupov M v stave. Celkový počet výstupov nášho systému je počet segmentačných jednotiek plus výstupy modelov z HMM modelujúcich ticho. V prípade, kde je modelovaná každá segmentačná jednotka jedným HMM, daný HMM môže na výstupe generovať len jeden symbol o_i z množiny výstupných symbolov \mathcal{V} , zodpovedajúci danej segmentačnej jednotke. Nech je $\mathcal{V} = \{o_i; i = 1, 2, \dots, M\}$.
3. Maticu prechodových pravdepodobností $\mathbf{A} = a_{ij}$ s definovanou topológiou, kde

$$a_{ij} = P[q_{t+1} = j | q_t = i], \quad 1 \leq i, j \leq N, \quad i, j \in \mathcal{N} \quad (3.302)$$

4. Maticu výstupných pravdepodobností $\mathbf{B} = b_j(k)$, kde

$$b_j(k) = P[o_t = v_k | q_t = j], \quad 1 \leq k \leq M, \quad k \in \mathcal{V} \quad (3.303)$$

definuje pravdepodobnosť generovania symbolov v stave j , kde $j \in \mathcal{N}$.

5. Začiatočný stav $\pi = \pi_i$, v ktorom

$$\pi_i = P[q_1 = i], \quad 1 \leq i \leq N, \quad i \in \mathcal{N} \quad (3.304)$$

Z vyššie uvedeného vyplýva, že na úplnú špecifikáciu HMM je potrebné určiť parametre \mathbf{A} , \mathbf{B} a π . Pre zjednodušenie budú tieto parametre označované ako

$$\Theta = (\mathbf{A}, \mathbf{B}, \pi). \quad (3.305)$$

Baumov–Welchov algoritmus

Zatiaľ nie je známa metóda na analytický výpočet parametrov modelu maximalizujúca vierohodnosť (likelihood) výstupnej sekvencie symbolov modelu. Je však možné určiť $\Theta = (\mathbf{A}, \mathbf{B}, \pi)$ také, aby boli maximalizované pravdepodobnosti $P(\mathcal{O} | \Theta)$, kde \mathcal{O} je tréningová postupnosť. Na tento účel bol zvolený Baumov–Welchov (forward–backward) algoritmus.

Na opis spôsobu odhadu parametrov HMM je využitá pravdepodobnosť $P_t(i, j)$ toho, že model bol v čase t v stave i a v čase $t + 1$ v stave j , pre tréningovú postupnosť \mathcal{O} a parametre modelu Θ .

$$P_t(i, j) = P(q_t = i, q_{t+1} = j | \mathcal{O}, \Theta) \quad (3.306)$$

Nech $\alpha_t(i)$ je pravdepodobnosť, že v čase t bola pozorovaná postupnosť o_1, o_2, \dots, o_t a systém sa nachádzal v stave i :

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, q_t = i | \Theta) \quad (3.307)$$

Túto pravdepodobnosť je možné zrátať nasledujúcim postupom:

1. Inicializácia

$$\alpha_1(i) = \pi_i b_i(o_1), \quad 1 \leq i \leq N \quad (3.308)$$

2. Indukcia

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}), \quad 1 \leq t \leq T - 1; 1 \leq j \leq N \quad (3.309)$$

3. Koncová podmienka

$$P(\mathcal{O} | \Theta) = \sum_{i=1}^N \alpha_T(i) \quad (3.310)$$

Obdobným spôsobom je určená pravdepodobnosť $\beta_t(i)$ definovaná ako

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T | q_t = i, \Theta) \quad (3.311)$$

ktorá označuje pravdepodobnosť generovania výstupnej postupnosti v čase $t+1$ až T za predpokladu, že v čase t sa model nachádza v stave i . Túto pravdepodobnosť je možné zrátať nasledujúcim spôsobom:

1. Inicializácia

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (3.312)$$

2. Indukcia

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad t = T - 1, T - 2, \dots, 1; 1 \leq i \leq N \quad (3.313)$$

Na základe toho je možné (3.306) zapísať v tvare [30]

$$\begin{aligned} P_t(i, j) &= \frac{P(q_t = i, q_{t+1} = j, \mathcal{O} \mid \Theta)}{P(\mathcal{O} \mid \Theta)} = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{P(\mathcal{O} \mid \Theta)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)} \end{aligned} \quad (3.314)$$

Nech $P_t(i)$ je pravdepodobnosť, že v čase t je model v stave i , potom je možné vyjadriť $P_t(i)$ pomocou $P_t(i, j)$ ako

$$P_t(i) = \sum_{j=1}^N P_t(i, j). \quad (3.315)$$

Ak je zrátaná $P_t(i)$ za čas celej tréningovej postupnosti t , získa sa hodnota, ktorú je možné interpretovať ako očakávaný počet výskytu stavu i (počet prechodov zo stavu i). Podobne suma $P_t(i, j)$ môže byť interpretovaná ako očakávaný počet prechodov zo stavu i do stavu j .

$$\sum_{t=1}^{T-1} P_t(i) = \text{očakávaný počet prechodov zo stavu } i \text{ pre } \mathcal{O} \quad (3.316)$$

$$\sum_{t=1}^{T-1} P_t(i, j) = \text{očakávaný počet prechodov zo stavu } i \text{ do stavu } j \text{ pre } \mathcal{O} \quad (3.317)$$

Pomocou vyššie uvedených rovníc je možné odhadnúť parametre Markovovho modelu. Odhady parametrov π , \mathbf{A} a \mathbf{B} sú formulované nasledujúcim spôsobom

$$\bar{\pi}_j = \text{počet výskytov v stave } i \text{ v čase } (t = 1) = P_1(i) \quad (3.318)$$

$$\begin{aligned} \bar{a}_{ij} &= \frac{\text{očakávaný počet prechodov zo stavu } i \text{ do stavu } j}{\text{očakávaný počet prechodov zo stavu } i} \\ &= \frac{\sum_{t=1}^{T-1} P_t(i, j)}{\sum_{t=1}^{T-1} P_t(i)} \end{aligned} \quad (3.319)$$

$$\begin{aligned} \bar{b}_j(k) &= \frac{\text{očakávaný počet výskytov v stave } j \text{ ak bol generovaný symbol } o_k}{\text{očakávaný počet výskytov v stave } j} \\ &= \frac{\sum_{t=1}^T P_t(j)}{\sum_{t=1}^T P_t(j)} \cdot \end{aligned} \quad (3.320)$$

Ak je aktuálny model definovaný ako $\Theta = (\mathbf{A}, \mathbf{B}, \pi)$ a je použitý na výpočet pravých strán rovníc (3.318), (3.319) a (3.318) a novodefinovaný model $\bar{\Theta}, \bar{\Theta} = (\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\pi})$ je odhadnutý z ľavých strán týchto rovníc, tak buď inializačný model Θ predstavuje kritický bod pravdepodobnostnej funkcie, t. j. $\bar{\Theta} = \Theta$; alebo model $\bar{\Theta}$ je pravdepodobnejší než model Θ v zmysle $P(\mathcal{O} \mid \bar{\Theta}) > P(\mathcal{O} \mid \Theta)$ [5], t. j. bol nájdený model, ktorý s väčšou pravdepodobnosťou generuje pozorovanú výstupnú postupnosť.

Ak na základe vyššie opísaného postupu sú iteratívne použité parametre $\bar{\Theta}$ namiesto Θ a odhadnú sa nové parametre, je možné zvyšovať pravdepodobnosť, že postupnosť $\mathbf{O} = \mathbf{o}$ bola generovaná modelom, až kým sa nedosiahne hraničná hodnota. Finálny výsledok tohto postupu je najpravdepodobnejší odhad HMM. Na tomto mieste treba zdôrazniť, že Baumov–Welchov algoritmus vedie k určeniu lokálneho maxima a konverguje k optimálnym hodnotám rovnomerne.

Viterbiho algoritmus

Cieľom Viterbiho algoritmu je nájsť najpravdepodobnejšiu postupnosť stavov $\mathbf{q} = (q_1, q_2, \dots, q_T)$ pre generovanú výstupnú postupnosť $\mathbf{o} = (o_1, o_2, \dots, o_T)$. Nech

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P[q_1 q_2 \dots q_{t-1}, q_t = i, o_1 o_2 \dots o_t \mid \Theta] \quad (3.321)$$

kde $\delta_t(i)$ je najlepšie skóre (najvyššia pravdepodobnosť) pre jeden prechod v čase t , ktorý zodpovedá prvým t výstupom a končí v stave i . Z toho vyplýva, že

$$\delta_{t+1}(j) = [\max_i \delta_t(i) a_{ij}] \cdot b_j(o_{t+1}) \quad (3.322)$$

kde prvky sú dané parametrami modelu, t. j. maticami $\mathbf{A} = \{a_{ij}\}$ matica prechodových pravdepodobností príslušného HMM a $\mathbf{B} = \{b_{ij}\}$ matica výstupných pravdepodobností.

Nato, aby bola získaná postupnosť stavov je potrebné sledovať hodnoty, ktoré maximalizujú rovnicu (3.322) pre každé t a i . Kompletný postup pre nájdenie optimálnej postupnosti stavov môže byť rozpísaný do niekoľkých krokov. V ďalšom sú uvedené dve alternatívne metódy.

Štandardná implementácia

1. Inicializácia

$$\begin{aligned} \delta_1(i) &= \pi_i b_i(o_1), 1 \leq i \leq N \\ \psi_1(i) &= 0 \end{aligned}$$

2. Rekurzia

$$\begin{aligned} \delta_t(j) &= \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(o_t), 2 \leq t \leq T; i \leq j \leq N \\ \psi_t(j) &= \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], 2 \leq t \leq T; i \leq j \leq N \end{aligned}$$

3. Ukončenie

$$\begin{aligned} P^* &= \max_{1 \leq i \leq N} [\delta_T(i)] \\ q_T^* &= \arg \max_{1 \leq i \leq N} [\delta_T(i)] \end{aligned}$$

4. Backtracking

$$q_t^* = \psi_{t+1}(q_{t+1}^*), t = T - 1, T - 2, \dots, 1$$

Alternatívna implementácia Ak sú parametre modelu logaritmizované, Viterbiho algoritmus z predchádzajúcej časti je možné implementovať bez násobenia¹⁷:

1. Predspracovanie

$$\begin{aligned}\tilde{\pi}_i &= \log(\pi_i), & 1 \leq i \leq N \\ \tilde{b}_i(o_t) &= \log[b_i(o_t)], & 1 \leq i \leq N; 1 \leq t \leq T \\ \tilde{a}_{ij} &= \log(a_{ij}), & 1 \leq i, j \leq N\end{aligned}$$

2. Inicializácia

$$\begin{aligned}\tilde{\delta}_1(i) &= \log(\delta_1(i)) = \tilde{\pi}_i + \tilde{b}_i(o_1), & 1 \leq i \leq N \\ \psi_1(i) &= 0\end{aligned}$$

3. Rekúzia

$$\begin{aligned}\tilde{\delta}_t(j) &= \log(\delta_t(j)) = \max_{1 \leq i \leq N} [\tilde{\delta}_{t-1}(i) + \tilde{a}_{ij}] + \tilde{b}_j(o_t) \\ \psi_t(j) &= \arg \max_{1 \leq i \leq N} [\tilde{\delta}_{t-1}(i) + \tilde{a}_{ij}], & 2 \leq t \leq T; i \leq j \leq N\end{aligned}$$

4. Ukončenie

$$\begin{aligned}\tilde{P}^* &= \max_{1 \leq i \leq N} [\tilde{\delta}_T(i)] \\ q_T^* &= \arg \max_{1 \leq i \leq N} [\tilde{\delta}_T(i)]\end{aligned}$$

5. Backtracking

$$q_t^* = \psi_{t+1}(q_{t+1}^*), t = T - 1, T - 2, \dots, 1$$

¹⁷Pri niektorých implementáciách – napr. embedded systémy – to môže mať významný vplyv na rýchlosť.

Kapitola 4

Overenie navrhnutých riešení

Navrhnuté riešenia boli použité a overené v konkrétnych systémoch na rozpoznávanie reči, alebo boli testované na množine trénovacích dát. Jednotlivé riešenia sú podrobnejšie opísané v ďalších častiach, keďže možnosti overenia dosiahnutých výsledkov sú pre jednotlivé úlohy značne rozdielne.

4.1 Výber segmentačných jednotiek a ich kódovanie

Výber segmentačných jednotiek závisí od účelu, na ktorý majú byť použité. Z toho dôvodu je overenie správnosti výberu vhodných segmentačných jednotiek značne komplikované, keďže súvisí s overením celého systému, v rámci ktorého boli jednotky definované.

Keďže táto problematika bola v prvom kroku riešená v rámci medzinárodného projektu Copernicus [1], kde bolo potrebné spolu s výberom jednotiek vyriešiť aj ich kódovanie kompatibilné s inými jazykmi, boli ako segmentačné jednotky použité fonémy, tak ako je o tom pojednané v časti 3.3.1.

K fónam ako segmentačným jednotkám bolo možné prejsť až po tom, čo bola vyriešená problematika ich kódovania (slovenská verzia abecedy SAMPA)¹, ktorej sa venuje časť 3.3.2. Vhodnosť výberu týchto jednotiek nebola overená v systéme na rozpoznávanie reči tak ako tomu bolo v prípade foném, ale len na segmentácii množiny trénovacích viet.

Navrhnuté kódovanie fón, ktoré sa opiera o doterajšie fonetické výskumy slovenských hlások, v prípade problematických samohlások o umiestnenie v systéme CV a analýzu ich lokalizácie z hľadiska korešpondencie so symbolmi IPA aj v porovnaní s inými jazykmi [32], [34], 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,

¹Pri zvolení ľubovoľného iného kódovania prechodu na fóny ako na segmentačné jednotky nič nebránilo.

SK	→	DE	SK	→	DE	SK	→	DE	SK	→	DE
a		a	L [^] U\		j U	N		N	s		s
a:		a:	U ₋ [^] O		v O	r		r	z		z
{		E	p		p	r=		r	S		S
E		E	b		b	r=:		r r	Z		z
E:		E:	t		t	l		l	x		x
I		I	d		d	l=		l	h\		h
I:		I:	c		t j	l=:		l l	G		x
O		O	J\		d j	J		n j	j		j
O:		O O	k		k	L		l j	L [^]		I
U		U	g		g	f		f	ts		ts
U:		U U	m		m	v		v	dz		d z
L [^] a		j A	F		n	f ₋ v		f	tS		tS
L [^] E		j E	n		n	U ₋ [^]		v	dZ		dZ

Tabuľka 4.1: Prekódovacia SAMPA tabuľka slovenského fonetického repertoáru do nemeckého.

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 so 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. Vzhľadom na to, že na rozpoznávanie sa dnes používajú výlučne štatistické metódy, je možné o tom uvažovať ako o zväčšení rozptylu pre oblasť patriacu konkrétnym vokálom.

Správnosť voľby kódovania do systému SAMPA bola overená experimentom, v ktorom bol na rozpoznávanie slovenčiny použitý iný jazyk [25]. V tomto experimente bol použitý systém na rozpoznávanie nemčiny. Keďže išlo o rozpoznávač so SAMPA kódovaním výslovnostného slovníka, slovník rozpoznávača bol rozšírený o slovenské slová, a to tým spôsobom, že znaky fón, ktoré sú pre slovenčinu a nemčinu zhodné, boli ponechané bezo zmeny. Takýchto fón je v prípade slovenčiny a nemčiny 29 z 52. Pre fóny, ktoré nemčina nemá, bol použitý blízky ekvivalent pozostávajúci z jednej alebo dvoch fón. Napríklad pre palatály *c*, *J*\, *J* a *L* bolo použité priblíženie pre prepis palatalizovaných konsonantov *t*, *d*, *n* a *l*. Slovenské *r* nezodpovedá nemeckému *r*, nemčina nemá slabičné *r* a *l*, nemá niektoré dlhé vokály, ktoré boli zdvojené atď. Tabuľka 4.1 obsahuje výslednú prekódovaciu tabuľku. Je však potrebné poznamenať, že v prípade použitého rozpoznávača nemčiny boli aj v ňom na konverziu z originálnej fonetickej sady do SAMPA abecedy urobené určité zmeny. Napríklad fóna *dZ* sa v originále vôbec nevyskytovala

a bola vytvorená spojením *d* a *S*, čo je tiež kompromisné priblíženie.

Po pridaní takto prekódovaných slov do slovníka k nemeckým slovám systém bol schopný rozpoznávať aj slovenské slová. Podrobnejšie vyhodnotenie nebolo urobené, keďže išlo len o overenie predpokladu schodnosti takéhoto riešenia na jednoduché aplikácie v prípade, že dva jazyky od seba nie sú foneticky veľmi vzdialené.

Tabuľka 3.6 obsahuje konečnú verziu SAMPA pre slovenčinu, ktorá bola publikovaná v [27]. Okrem SAMPA znakov a znakov slovenskej transkripcie sú v nej uvedené aj príslušné IPA znaky.

4.2 Automatická transkripcia

Pravidlá definované v časti 3.4 boli programovo realizované v jazyku Perl. Z 225 výslovnostných pravidiel získaných z [33] a [47] bolo navrhnutých 257 formálnych transkripčných pravidiel. Dosiahnuté výsledky je možné zhrnúť do nasledujúcich bodov:

- Pravidla týkajúce sa samohlások (samohlások, dvojhlások, samohláskových skupín) boli implementované v celom opísanom rozsahu a ich otestovanie v praktickej prevádzke ukázalo vysokú spoľahlivosť pri generovaní fonetického prepisu.
- Pravidlá týkajúce sa prepisu spoluhlások boli implementované tiež v celom rozsahu, avšak kvalita výstupu závisí do značnej miery od kvality morfe-
matickej a slabičnej analýzy. Ďalší problém vplyvajúci na kvalitu výstupu je množstvo cudzích slov a slov považovaných z hľadiska výslovnosti za výnimky na vstupe. V systéme sú obsiahnuté výslovnostné výnimky, spomínané v pravidlách slovenskej výslovnosti, avšak aké percento výnimiek tieto výnimky pokrývajú, nie je známe.

Počas vlastnej programovej realizácie bolo dôležité správne zoradiť volanie jednotlivých transkripčných pravidiel. Volanie transkripčných pravidiel v tom poradí, v akom sú uvedené v časti 3.4, by viedlo k nesprávnemu prepisu. V prílohe B je uvedená časť kódu, z ktorého vyplýva skutočná postupnosť volaní jednotlivých transkripčných pravidiel.

Pred vlastnou aplikáciou transkripčných pravidiel na vstupné slová sú jednotlivé slová otestované na príslušnosť k výnimkám. V prípade, že k výnimkám nepatria, je urobená slabičná a morfematická segmentácia. Až po tomto kroku nasleduje vlastná fonetická transkripcia. Transkripcia prebieha slovo po slove so zohľadňovaním kontextu v rámci vety. V procese transkripcie súvislého textu je text v prvom kroku rozdelený na jednoduché vety. Jednoduché vety sú spracovávané sekvenčne. Slová v rámci jednoduchej vety sú prepisované podľa vyššie opísaného postupu.

Na otestovanie správnosti transkripcie bolo použitých 100 náhodne vybraných slov. Slová boli vybraté z testovacej množiny pre slabičnú segmentáciu a boli

vylúčené tie slová, ktoré neboli správne rozdelené na slabiky. Dôvodom pre túto voľbu bola snaha eliminovať chyby, ktoré by mohli vzniknúť pri delení na slabiky, a získať tak obraz o kvalite vlastnej transkripcie. Pre všetky slová z testovacej množiny bolo pridané potrebné morfeatické delenie do morfeatického slovníka. Vlastný test však prebehol najprv bez pridania slov do morfeatického slovníka a následne po pridaní slov do morfeatického slovníka. Snaha bola takisto zistiť vplyv morfeatického delenia na kvalitu vlastného prepisu. Generovaná ortoepická reprezentácia bola porovnaná s manuálne vytvoreným prepisom.

Prepis slov namiesto celých viet bol zvolený z toho dôvodu, že morfeatické hranice, ktoré sa javia ako najslabšie miesto v celom systéme, nie je problém detekovať na hranici slov, keďže tam sú jednoznačné. Problém vzniká pri detekcii morfeatických hraníc vnútri slov. To isté platí aj o slabičnej segmentácii a o slovách, ktoré patria medzi výslovnostné výnimky. V neposlednom rade hral dôležitú úlohu aj fakt, že percentuálna úspešnosť prepísaných viet nehovorí veľa o kvalite prepisu, keďže počet slov vo vetách je rôzny.

V nasledujúcej tabuľke je uvedená percentuálna úspešnosť pri prepise slov s a bez morfeatického slovníka:

	Prepis s morfeatickým slovníkom	Prepis bez morfeatického slovníka
1 spôsob prepisu	81 %	75 %
2 spôsoby prepisu	9 %	8 %
Spolu	90 %	83 %

Zoznam slov použitých na testovanie spolu s vygenerovaným ortoepickým prepisom je v prílohe A. Z analýzy vygenerovaného prepisu vyplynulo nasledujúce:

- Rozdiel medzi výsledkami dosiahnutými bez pomoci morfeatickej analýzy a s pomocou morfeatickej analýzy tvoria prevažne slová, v ktorých chyba pri prepise nastala na mieste znelostnej asimilácie. Znelostná asimilácia tvorí tú časť transkripcie, pre ktorú je potrebné správne morfeatické delenie. V jednom prípade išlo o výslovnosť zdvojených spoluhlások. Vo vykonanom teste išlo o nasledujúcich 7 slov: nepredpokladal (J E p r E t p O k l a d a l / J E p r E d p O k l a d a l), trikmi (t r I g m I / t r I k m I), podpíšu (p O t p I: S U / p O d p I: S U), odpálená (O t p a: l E n a:; O t p a: L E n a: / O d p a: l E n a:; O d p a: L E n a:), krúžkoch (k r U: S k O x / k r U: Z k O x), vládcom (v l a: t s t s O m / v l a: d t s O m), jesenná (j E s E n a: / j E s E n n a:).
- Druhú skupinu chýb — 10 % — tvorili chyby vzniknuté pri prepise výslovnosti *t, d, n, l* – *ť, ď, ň, ľ*. Ide o chyby, ktoré je nutné ošetriť slovníkom, keďže slovenčina v prípade výslovnosti *t, d, n, l* – *ť, ď, ň, ľ* má mnoho výnimiek. Išlo o nasledujúce slová: minifutbal (m I J I f U d b a l), kandidátska (k a n J Ī d a: t s k a), benevolenciu (b E J E v O l E n t s L ^ U, b E J E v O L E n t s L ^ U), kabinetu (k a b I J E t U), minimálnym (m I J I m a: l n I m), veterinárny (v E c E r I n a: r n I), Tibete (c I b E c E),

teroristických (c E r O r I s c I t s k I: x), jednej (j E d J E L ^), pevného (p E U ^ J E: hÖ). Z uvedených nesprávne prepísaných slov vyplýva, že ide prevažne o cudzie alebo prevzaté slová. Ak by cieľom bol prepis len slovenských slov, úspešnosť prepisu by prekročila 95 % v prípade použitia morfeomatického slovníka.

Ako príklad práce celého systému je uvedená transkripcia vety *Egyptská správa nehnuteľností zamestnáva tanečnicu s pedálom*. V nasledujúcej tabuľke je uvedený pre každé slovo z danej vety jeho ortoepický prepis spolu s číslami transkripčných pravidiel, ktoré boli pri prepise aplikované. Čísla defaultných pravidiel, t. j. (3.249) až (3.269), použitých v prípade, že žiadne z pravidiel vyplývajúcich z pravidiel slovenskej výslovnosti [33] nebolo použité, v tabuľke uvedené nie sú. Pravidlá boli aplikované v uvedenom poradí.

Grafémy	Fóny	Pravidlá
Egyptská	E g I p s k a:	(3.3) (3.5) (3.8) (3.240)
správa	s p r a: v a	(3.2) (3.8) (3.78)
nehnuteľností	J E h \ n U c E L n O s c I:	(3.3) (3.6) (3.7) (3.10) (3.170)
zamestnáva	z a m E s t n a: v a	(3.2) (3.3) (3.8) (3.78) (3.213)
tanečnicu	t a n d z m a j s t r a	(3.2) (3.213) (3.100) (3.170)
s	s	(3.105)
pedálom	p E d a: l I k O m p E d a: L I k O m	(3.3) (3.4) (3.6) (3.8) (3.170) (3.176) (3.3) (3.4) (3.6) (3.8) (3.170) (3.177)

4.3 Automatické určovanie hraníc slabík

Pri aplikácii postupov navrhnutých v časti 3.5 bolo na natréovanie systému použitých 11000 náhodne zvolených slovenských slov. 10000 slov bolo použitých na výpočet pravdepodobností $P(s_3 | s_2, s_1)$, $P(s_3 | s_2)$ a $P(s_3)$. Zvyšných 1000 bolo použitých pre výpočet váh λ zo vzťahu (3.293).

Pred vlastným výpočtom štatistík bolo potrebné urobiť slabičnú segmentáciu trénovacej množiny. Táto segmentácia bola v prvom kole realizovaná pomocou pravidiel uvedených v časti 3.5.1, čím bolo získaných 78 % správne segmentovaných slabík. Ostatné problematické delenia boli urobené ručne súčasne s kontrolou správnosti automatického delenia, čo potvrdilo správnosť segmentačných pravidiel navrhnutých v 3.5.1.

Priemerná dĺžka slova v trénovacej množine bola 3.374 slabiky a rozloženie monogramov, bigramov a trigramov bolo nasledujúce:

Monogramy : 3009
 Bigramy : 11258
 Trigramy : 9126

Počet monogramov je vlastne počet rôznych slabík v tréningovej vzorke.

Samotný proces segmentácie ľubovoľného slova na slabiky prebieha v dvoch nasledujúcich krokoch:

- Dané slovo je rozdelené na všetky možné (nemožné) postupnosti slabík (monogramov).
- Pre každú postupnosť slabík je zrátaná pravdepodobnosť podľa vzťahu (3.286) upraveného pre trigramový model.

Ako demonštračný príklad bolo zvolené slovo *programové*. V prvom kroku je slovo rozdelené na všetky možné postupnosti slabík²: *prog-ra-mo-vé*, *prog-ra-mov-é*, *pro-gra-mo-vé*, *prog-ram-o-vé*, *pro-gra-mov-é*, *p-ro-gra-mo-vé*, *prog-ram-ov-é*, *p-ro-gra-mov-é*. Pre každé slovo je zrátaná pravdepodobnosť (3.286) a výsledky sú zoradené zostupne. Pre zvolený príklad je generovaný nasledujúci výstup:

<i>prog-ra-mo-vé</i>	9.71614053467775	e-14
<i>prog-ra-mov-é</i>	6.8819726337521	e-15
<i>pro-gra-mo-vé</i>	1.23501771678877	e-15
<i>prog-ram-o-vé</i>	2.39035687120406	e-17
<i>pro-gra-mov-é</i>	1.45480521187513	e-18
<i>p-ro-gra-mo-vé</i>	3.34387569322274	e-20
<i>prog-ram-ov-é</i>	2.98421581923104	e-21
<i>p-ro-gra-mov-é</i>	3.93896194381078	e-23

Pri pohľade na uvedený výstup je zrejmé, že systém generuje aj úplne nereálne postupnosti, ktoré je možné jednoducho vylúčiť pravidlami z časti 3.5.1. Pri aplikovaní pravidiel z 3.5.1 by v tomto prípade z ôsmich vygenerovaných hypotéz bolo 5 vylúčených a zostali by:

<i>prog-ra-mo-vé</i>	9.71614053467775	e-14
<i>pro-gra-mo-vé</i>	1.23501771678877	e-15
<i>p-ro-gra-mo-vé</i>	3.34387569322274	e-20

Kombinácia obidvoch postupov sa javí byť výhodná, čo vyplýva aj zo skutočnosti, že pravidlá pre segmentáciu síce nedokážu rozdeliť každé slovo na správnu postupnosť slabík, avšak tam, kde sa slabičná hranica pomocou týchto pravidiel určí, je hranica 100 % správna.

V prípade kombinovania postupov uvedených v 3.5.1 a 3.5.2 je možné postupovať dvoma spôsobmi:

- Aplikovať pravidlá z 3.5.1 v prvom kroku, t. j. počas generovania všetkých možných postupností slabík.

²Všetky možné postupnosti slabík obsiahnutých v tréningovej množine.

- Aplikovať pravidlá z 3.5.1 v druhom kroku, t. j. ako filter na generovaný výstup.

V princípe nezáleží na tom, v ktorom kroku budú pravidlá aplikované. V prípade, že štatistická segmentácia je dostatočne spoľahlivá a pravidlá nemajú za úlohu korigovať prípadný nesprávny výstup, nie je potrebné ich uvažovať vôbec.

Z hľadiska výpočtovej náročnosti je vhodné použiť ich v prvom kroku a zredukovať tak prehľadávaný priestor pri generovaní všetkých možných slabík, čo je výpočtovo náročný krok, keďže sa generujú všetky možné postupnosti slabík pre dané slovo.

Na otestovanie spoľahlivosti celého systému bolo zvolených 203 náhodne vybratých slov, ktoré neboli súčasťou tréningovej množiny. Nad touto množinou testovacích dát bola urobená segmentácia a následne filtrácia generovaných postupností slabík. Na výstupe bola sledovaná úspešnosť správne vykonanej segmentácie pre najpravdepodobnejšiu postupnosť bez filtrácie a s filtráciou. Takisto bola sledovaná početnosť výskytu správne generovanej postupnosti na druhom mieste v prípade, že prvá postupnosť nebola správna. Ďalej bola sledovaná početnosť správnej postupnosti slabík pre prípad, že prvé aj druhé delenie je správne. Nakoniec bol sledovaný počet slov, kde správne delenie nebolo na prvom alebo druhom mieste, a počet slov, kde správne delenie nemohlo byť generované v dôsledku absencie slabiky v tréningovej množine. Dosiahnuté výsledky sú uvedené v nasledujúcej tabuľke:

	Delenie bez filtrácie	Delenie s filtráciou
1. delenie správne	66.99 %	71.92 %
2. delenie správne	7.88 %	2.95 %
1. aj 2. delenie správne	12.80 %	12.80 %
Nesprávne delenie	1.47 %	1.47 %
Absentujúce slabiky	8.37 %	8.37 %

Ako príklad na nesprávne delenie na prvom mieste je možné uviesť slovo *paneurópske* a na slovo, z ktorého slabika sa nenachádzala v tréningovej množine, je možné uviesť napríklad slovo *lúčoch*. Pre dané slová boli generované nasledujúce postupnosti:

pa-neu-róp-ske	lúč-och
pa-ne-u-róp-ske	lú-čo-ch
pan-eu-róp-ske	lúč-o-ch
pan-e-u-róp-ske	
pa-ne-u-ró-pske	
pa-ne-ur-ó-pske	
pa-neu-ró-pske	
pan-e-u-ró-pske	
pan-e-ur-ó-pske	
pan-eu-ró-pske	

Z dosiahnutých výsledkov vyplýva, že počet skutočne nesprávnych delení je relatívne malý. V prípade, že za nesprávne delenia považujeme tie, pri ktorých sa správna postupnosť slabík nachádza na druhom alebo niektorom ďalšom mieste, získame hodnoty 9.35 % resp. 4.42 % slov, kde sa správne delenie nenachádzalo na prvom mieste. 8.37 % chýb tvoria v oboch prípadoch slová, z ktorých slabiky neboli súčasťou tréningovej množiny. V prípade, že takéto slová budú z testu vylúčené, po prerátaní výsledkov dospejeme k nasledujúcej tabuľke:

	Delenie bez filtrácie	Delenie s filtráciou
1. delenie správne	87.09 %	92.47 %

V oboch prípadoch je výsledok lepší ako výsledky dosiahnuté len použitím pravidiel. Ako vyplynulo z analýzy nesprávne generovaných výsledkov, veľký rozdiel medzi filtrovaným a nefiltrovaným delením je spôsobený taktiež nedostatočnou tréningovou množinou. V prípade nesprávnych delení prevládali slová začínajúce sa predponou *vy-*, na ktorú sa vzťahujú pravidlá pre filtráciu. Pri zväčšovaní tréningovej množiny by mal rozdiel medzi filtrovaným a nefiltrovaným delením postupne zaniknúť.

Z výsledkov taktiež vyplýva, že 3009 rôznych slabík obsiahnutých v tréningovej množine je stále nedostatočný počet, keďže slová, ktoré neboli rozdelené z dôvodu absencie slabík, tvorili 8.37 % nesprávne nasegmentovaných slov z testovacej množiny.

Priemerný počet tréningových slov na slabiku sú približne 3 slová. Vzhľadom k celkovému počtu slabík v systéme — vyše 3000 — je toto číslo veľmi malé, čo oprávňuje k predpokladu, že rozšírenie tréningovej množiny povedie k ďalšiemu zlepšeniu celého systému.

4.4 Segmentácia rečového signálu

Postupy opísané v predchádzajúcej časti boli overené na tréningovej množine 2500 viet. Po predspracovaní rečového signálu a vytvorení príslušnej transkripcie pre zvolené segmentačné jednotky bola urobená vlastná segmentácia.

Proces segmentácie prebiehal v nasledujúcich krokoch:

1. V procese inicializácie bol každý samostatný HMM natrénovaný postupnosťou náhodných dát z vektorového kvantizátora.
2. Pomocou Viterbiho algoritmu a natrénovaných HMM sa vstupná veta nasegmentovala ako najpravdepodobnejšia postupnosť stavov. To znamená, že každý segment³ bol označený znakom z množiny výstupných symbolov.

³Keďže počas procesu segmentácie neboli povolené nulové prechody pre HMM, každý vygenerovaný výstupný symbol zodpovedá jednému prechodu, z čoho vyplýva, že každý symbol zodpovedá jednému vektoru príznakov, a teda označuje 10 ms okno reči.

Slovo	1. it	2. it	3. it	4. it	5. it	6.it
---	1	1	1	1	1	1
nie	7	38	42	42	42	42
som	19	50	125	80	80	80
v	31	62	133	139	113	113
Ostrave	35	136	137	143	124	123
som	99	276	325	226	226	223
v	156	288	333	265	255	261
Otrokoviciach	166	296	337	269	295	285
---	674	419	417	411	421	425

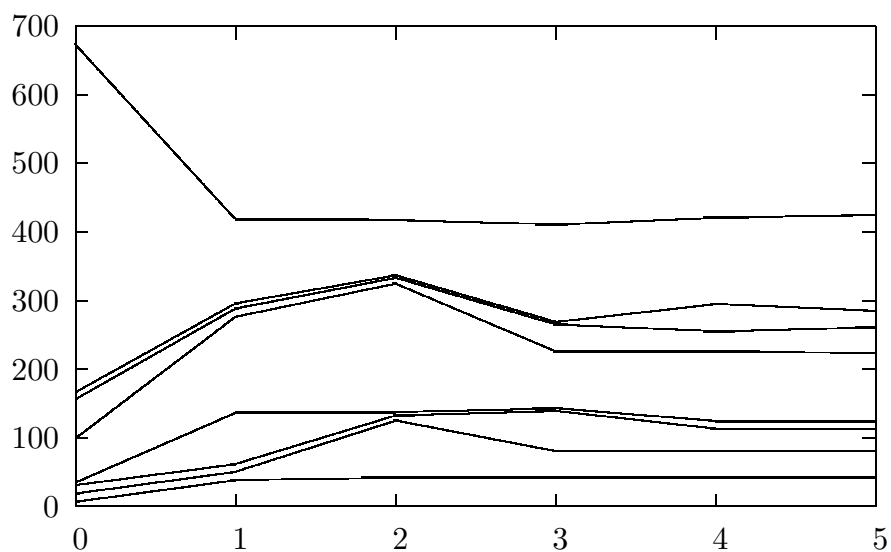
Tabuľka 4.2: Výsledky segmentácie vety *Nie som v Ostrave, som v Otrokoviciach* pre jednotlivé iterácie.

Ako bolo uvedené, HMM modelujúci jednu segmentačnú jednotku generuje len jeden druh výstupov, no keďže Viterbiho algoritmus pracuje nad zretazenými HMM, zodpovedajúcimi tomu, čo by malo byť v danom rečovom signále nahovorené, na výstupe bola pozorovaná postupnosť rôznych symbolov. Podrobnejšie bola metodika opísaná v časti 1.4.1.

- Po vykonaní segmentácie nad určitou skupinou viet boli segmenty zodpovedajúce jednému druhu segmentačnej jednotky použité na natrénovanie nového HMM pre danú jednotku. Toto bolo urobené pre všetky segmentačné jednotky.
- V prípade, že už bola použitá celá tréningová množina, proces segmentácie buď skončil, alebo prebehla ďalšia iterácia od kroku dva.

Výsledky dosiahnuté počas segmentácie sú demonštrované na vetách *Nie som v Ostrave, som v Otrokoviciach* a *Prosím Vás, kedy máme spoj z Levíc do Krakova*. V prvom kroku je potrebný prepis vety do segmentačných jednotiek, v tomto prípade fonetický prepis. Jeden z možných správnych prepisov bude vyzeráť takto: *JI_ ^E sOm v OstravE sOm v OtrOkovItsI_ ^ax a prOsI:m va:s kEdI ma:mE spOj z lEvI:ts dO krakOva*. Celá segmentácia bola použitá na tréningovej množine 2500 viet, kde prvá veta bola 3. v poradí a druhá veta bola 2002. v poradí. Tréningovanie HMM prebiehalo stále po spracovaní niekoľkých desiatok viet. Výsledky segmentácie po každej iterácii sú v tabuľke 4.2 a 4.3. Ich grafické znázornenie je na obrázkoch 4.1 a 4.2. Na osi y je poradové číslo vektora príznakov. Keďže každý vektor bol zrátaný nad 10ms oknom, vidieť, že prvá pokusná veta trvala približne 7 sekúnd a druhá 6. Čísla v tabuľke uvádzajú začiatky jednotlivých segmentov.

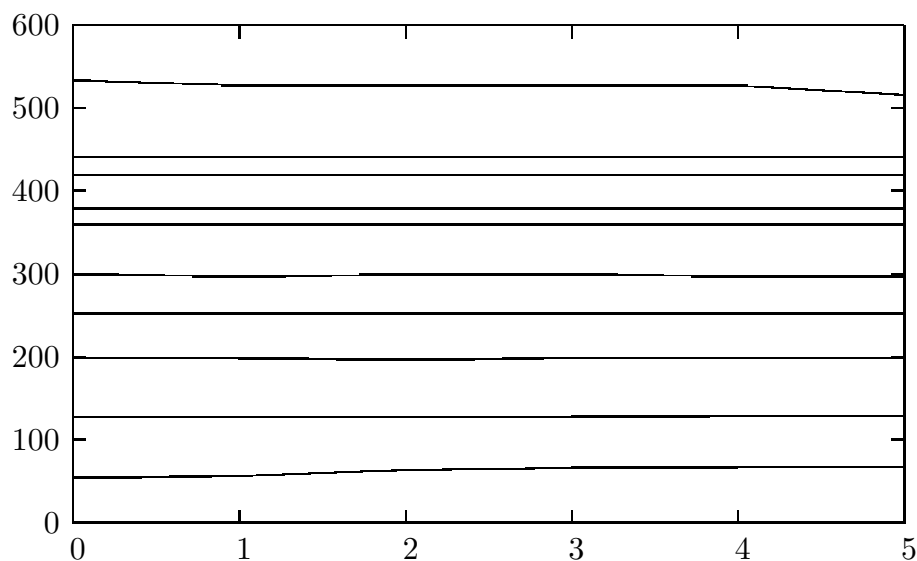
Pri porovnaní obidvoch viet vidieť veľký rozdiel pre prvú iteráciu v prvej a druhej vete. Keďže prvá veta bola tretia v poradí a jej prvá iterácia prebehla ešte pred prvým odhadom HMM, medzi prvou a druhou iteráciou prvej vety je veľký rozdiel. To však už neplatí pre druhú vetu. Napriek tomu, že išlo taktiež o



Obr. 4.1: Hranice slov po jednotlivých iteráciách pri segmentácii vety *Nie som v Ostrave, som v Otrokoviciach.*

Slovo	1. it	2. it	3. it	4. it	5. it	6.it
---	1	1	1	1	1	1
Prosím	55	56	64	66	67	67
Vás	127	127	127	127	129	129
kedy	198	198	197	198	198	198
máme	252	252	252	252	252	252
spoj	299	297	299	299	297	297
z	359	359	359	359	359	359
Levíc	379	379	379	379	379	379
do	419	419	419	419	419	419
Krakova	441	441	441	441	441	441
---	533	527	527	527	527	515

Tabuľka 4.3: Výsledky segmentácie vety *Prosím Vás, kedy máme spoj z Levíc do Krakova* pre jednotlivé iterácie.



Obr. 4.2: Hranice slov po jednotlivých iteráciách pri segmentácii vety *Prosím Vás, kedy máme spoj z Levíc do Krakova.*

prvú iteráciu, medzi prvou a druhou iteráciou je už rozdiel minimálny. Vyplýva to zo skutočnosti, že druhá veta bola 2002. v poradí a odhad parametrov HMM sa vykonával stále po niekoľkých desiatkach viet. V skutočnosti teda síce išlo o druhú iteráciu v rámci celej tréningovej množiny, avšak nie o druhú iteráciu pre odhad parametrov HMM.

Z tabuliek a grafov vyplýva, že už po štvrtej iterácii sa hranice segmentov pre prvú vetu ustálili. V prípade druhej vety možno považovať hranice za stabilné už po prvej iterácii. Z analýzy všetkých viet vyplynulo, že 4–5 iterácií je dostatočný počet na to, aby sa hranice segmentov ustálili. Zvyšovanie počtu iterácií nevedie k ďalšiemu výraznému zvýšeniu presnosti segmentácie.

V nasledujúcej tabuľke sú uvedené priemerné posuny začiatkov segmentov od predchádzajúcej iterácie. Hodnoty udávajú, o koľko vektorov príznakov – každý reprezentuje 10 ms okno – sa v priemere posunul začiatok každého segmentu od predchádzajúcej iterácie. Vo výpočte priemeru nebol uvažovaný prvý segment ticha (šumu), keďže ten je daný začiatkom testovacej vety a je vždy rovný jednej. Priemerné hodnoty boli rátané z absolutných hodnôt zmeny začiatku segmentu. Grafické znázornenie je na obrázku 4.3.

Iterácia	2.	3.	4.	5.	6.	7.	8.	9.
Priemerný posun	14.43	8.89	2.29	1.47	1.19	1.23	0.67	0.27

Postupnosť stavov pre jednotlivé slová po šiestej iterácii bola pre prvú vetu nasledujúca:

Kapitola 5

Výsledky dizertácie

V nasledujúcich častiach sú zhrnuté výsledky dosiahnuté v tejto práci a uvedené námety na ďalšie pokračovanie.

5.1 Výsledky dizertácie

Prínos tejto práce možno zhrnúť do nasledujúcich bodov:

- Návrh finálnej verzie SAMPA abecedy pre slovenčinu (časť 3.3). Slovenčina a čeština ako jedny z mála európskych jazykov doteraz nemali oficiálnu verziu SAMPA abecedy. Vyplývalo to aj z problémov spomínaných v tejto práci. Navrhnutá verzia už bola akceptovaná na domácom poli [27] a nič nebráni tomu, aby sa presadila aj na medzinárodnom.
- Analýza možnosti generovania fonetickej transkripcie pre slovenčinu s následným realizovaním navrhnutých pravidiel a ich overením pri reálnej segmentácii rečového signálu (časť 3.4). Táto úloha bola splnená len čiastočne. Bola urobená kompletná analýza problematiky, no z výsledkov analýzy problému vyplynula nemožnosť jeho riešenia bez externých zdrojov dát. Na základe toho bola samotná implementácia realizovaná len v obmedzenej miere.
- Aplikovanie teórie modelov jazyka na modely slabík za účelom automatického delenia slov na slabiky. Tento problém vyplynul z výsledkov analýzy pri riešení úloh opísaných v predchádzajúcom bode. Dá sa povedať, že táto úloha bola plne zvládnutá a bolo pre ňu navrhnuté inovatívne riešenie ¹.
- Segmentácia rečového signálu pomocou HMM. V procese segmentácie boli použité ako štartovacie údaje výsledky dosiahnuté pri segmentácii slovenských tréningových súborov na fonetickej úrovni (opísané v predchádzajúcich

¹Podľa poznatkov autora tento prístup k deleniu slov na slabiky ešte nebol nikde realizovaný.

bodoch). Význam tejto časti spočíva v aplikovaní modifikovaných stochastických postupov segmentácie na slovenčinu, ktorých jadrom boli súbory vytvorené na báze pôvodných výsledkov tejto práce.

5.2 Námety pre ďalšie pokračovanie práce

Námety pre ďalšie pokračovanie práce možno zhrnúť do troch bodov:

- Doriešenie problému morfematickej analýzy a vytvorenie kompaktného systému na generovanie fonetickej transkripcie.
- Po získaní dostatočného množstva slov s ich prepisom vytvoriť systém generujúci fonetický prepis nie na základe produkčných pravidiel, ale využívajúcich niektorý zo štatistických prístupov k problematike.
- Formalizácia štruktúr podporujúcich optimalizáciu úlohy delenia slov na slabiky. Súčasná implementácia nevyužíva žiadnu heuristiku pri generovaní všetkých možných postupností slabík, ale generuje ich "slepým" prehľadávaním. Tento časovo náročný proces by bolo možné výrazne urýchliť generovaním vhodnej stromovej štruktúry a následným prehľadávaním stromu pri generovaní možných výslovností. V ďalšom kroku by bolo možno týmto spôsobom generovať priamo správnu výslovnosť.

Ďakujem svojmu školiteľovi prof. Ing. Dušanovi Krokavcovi, CSc., za usmerenie, rady a pripomienky pri vypracovaní doktorandskej dizertačnej práce, pracovníkom Jazykovedného ústavu Ľudovíta Štúra SAV za poskytnutie potrebných lingvistických údajov. Za odbornú a priateľskú pomoc pri písaní práce ďakujem aj PhDr. Mire Nábělkovej.

Košice, december 2003

Ing. Jozef Ivanecký

Literatúra

- [1] 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: *Proceedings Int. Conf. on Spoken Language Processing*. Sydney, December 1998.
- [2] Bahl, L. R. – Das, S. – de Souza, P. V. – Epstein, M. – Mercer, R. L. – Merialdo, B. – Picheny, M. A. – Powel, J.: Automatic phonetic baseform determination. In: *Proceedings of the 1991 International Conference on Acoustics, Speech and Signal Processing*. Toronto 1991, s. 173–76.
- [3] Bahl, L. R.: Automatic phonetic baseform determination. In: *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*. 1991, s. 173–176.
- [4] Bahl, L. R. – Brown, P. F. – de Souza, P. V. – Mercer, R. L. – Nahamoo, D.: A fast algorithm for deleted interpolation. In: *Proceedings of Eurospeech 91*. Genova, September 1991, s. 1209–1212,
- [5] Baum, L.: An inequality and associated maximization technique in statistical estimation of probabilistic functions of a Markov process. *Inequalities*, vol. 3, 1972, s. 1–8.
- [6] Baum, L. E. – Petrie, T.: Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*, vol. 37, 1966, s. 1559–1563.
- [7] Černocký, J. – Pollák, P. – Hanžl, V. – Schwarz, P.: Pronunciation Transcription and Quality Control in Czech SpeechDat(E). URL: <http://www.fee.vutbr.cz/SPEECHDAT-E/public/conferences/euro-2001/cernocky.pdf>
- [8] Černý, M. – Matoušek, V. – Mautner, P.: An automatic creation of the pronunciation dictionary. In: *Proceedings of 3rd Slovenian-German and 2nd SDRV Workshop*. Ljubljana, 1996.
- [9] Damper, R. I. – Marchand, Y. – Adamson, M. J. – Gustafson, K.: Comparative evaluation of letter-to-sound conversion techniques for english text-to-

- speech synthesis. In: *3rd ESCA International Workshop on Speech Synthesis*. Jenelean Caves, 1998, s. 53–58.
- [10] Dankovičová, D.: Czech IPA. In: *Handbook of the International Phonetic Association*. Cambridge University Press, 1999.
- [11] Daržágín, S. – Franeková, Ľ. – Rusko, M.: *Konverzia a rečová syntéza slovenčiny*. Jazykovedný časopis, 45, 1994, 1, s. 31–43.
- [12] Dvončová, J. a kol.: *Fonetika a fonológia. Otázky transkripcie*. Bratislava, Univerzita Komenského v Bratislave 1988. 180 s.
- [13] Dvončová, J. – Jenča, G. – Kráľ, A.: *Atlas slovenských hlások*. Bratislava, SAV 1969. 105 s.
- [14] Fischer, V. – Günther, C. – Ivanecký, J. – Šedivý, J. – Ureš, L.: Towards Multi-Modal Interfaces for Embedded Devices. In: *Proceedings of the 13. Konferenz Elektronische Sprachsignalverarbeitung*. Dresden, 2002.
- [15] Fourcin, A. – Harland, G. – Barry, W. – Hazan, V.: *Speech input and output assessment-multilingual methods and standards*. Ellis Horwood Limited 1989.
- [16] Hála, B.: *Fonetické obrazy hlásek*. Praha, Státní pedagogické nakladatelství 1960, 127 s. + 43 s. obrazová příloha.
- [17] Hála, B.: *Fonetika v teorii a v praxi*. Praha, SPN 1975.
- [18] The International Phonetic Association: *Reproduction of The International Phonetic Alphabet*, <http://www2.arts.gla.ac.uk/IPA/fullchart.html>.
- [19] Isačenko, A. V. – Romportl, M.: *Návrh fonetické a fonologické transkripcie češtiny a slovenštiny*. Slovo a slovesnost, 27, 1966, s. 163–166.
- [20] Isačenko, A. V.: *Spektrografická analýza slovenských hlások*. Bratislava, SAV 1968.
- [21] Ivanecký, J.: *Metódy automatického rozpoznávania reči*. Písomná práca k dizertačnej skúške, Technická Univerzita Košice, 1996.
- [22] Ivanecký, J.: Problémovo orientovaná analýza prirodzeného jazyka. In: *Varia V*. Bratislava, SJS pri SAV 1996, s. 90-97.
- [23] Ivanecký, J. – Krokavec, D.: An Automatic Phonetic Transcription In Slovak Language. In: *Proceedings of The 2nd SQEL Workshop on Multi-Lingual Information Retrieval Dialogues*. Plzeň, University of West Bohemia 1997, s. 90-95.
- [24] Ivanecký, J.: Automatizácia fonematickej transkripcie slovenčiny. In: *Varia VII*. Ed. M. Nábělková. Bratislava, SJS pri SAV 1998, s. 48–52.

- [25] Ivanecký, J.: SAMPA v slovenčine a jej význam z pohľadu viacjazyčných systémov na rozpoznávanie reči. In: *Slovenčina a čeština v počítačovom spracovaní*. Bratislava, Veda 2001, s. 98–108.
- [26] Ivanecký, J.: Automatická transkripcia slovenčiny v počítačovom rozpoznávaní reči. In: *Slovenčina a čeština v počítačovom spracovaní*. Bratislava, Veda 2001, s. 109–116.
- [27] Ivanecký, J. – Nábělková, M.: *Fonetická transkripcia SAMPA a slovenčina*. Jazykovedný časopis, 53, 2002, s. 81–95.
- [28] Ivanecký, J. – Klehr, M. – Fischer, V. – Kunzmann, S.: Multi-Modal Voice Application Design in a Multi-Client Environment. In: *Proceeding of Text, Speech, Dialog 2003*. České Budejovice, 2003.
- [29] Ivanecký, J.: Štatistický prístup pri určovaní slabičných hraníc. In: *Slovko 2003*. Bratislava, Veda 2003.
- [30] Jelinek, F.: *Statistical Methods for Speech Recognition*. Cambridge – Massachusetts – London, The MIT Press 1998. 283 s.
- [31] Katz, S.: Estimation of probabilities from sparse data for the language model component of a speech recognizer. In: *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 3, 1987, s. 400–401.
- [32] Kráľ, Á.: Fonetická povaha a systém slovenských samohlások a dvojhĺások, In: *Studia Academica Slovaca*. 3. Ed. J. Mistrík. Bratislava, Alfa 1974, s. 51–66.
- [33] Kráľ, Á.: *Pravidlá slovenskej výslovnosti*. Bratislava, Slovenské pedagogické nakladateľstvo 1983. 632 s.
- [34] Kráľ, Á.: Slovenské samohlásky v konfrontácii so samohláskami iných jazykov. In: *Studia Academica Slovaca*. 3. Ed. J. Mistrík. Bratislava, Alfa 1974, s. 67–82.
- [35] Kráľ, A. – Sabol, J.: *Fonetika a fonológia*. Bratislava, Slovenské pedagogické nakladateľstvo 1987. 392 s.
- [36] Krokavec, D. – Filasová, A.: *Optimálne stochatické systémy*. Košice, Elfa 2002. 283 s.
- [37] 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.
- [38] Krokavec, D. – Ivanecký, J.: Semantic Processing of Sentences in Slovak Language. In: *Proceedings of The 2nd SQEL Workshop on Multi-Lingual Information Retrieval Dialogues*. Plzeň, University of West Bohemia 1997, Czech Republic, s.32-39.

- [39] Lucassen, J. M. – Mercer, R. L.: An information-theoretic approach to the automatic determination of phonetic baseforms. In: *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*. 1984, s. 42.5.1–42.5.4.
- [40] Mermelstein, P.: A Phonetic-Context Controlled Strategy for Segmentation and Phonetic Labeling of Speech. In: *IEEE Transactions on acoustic, speech, and signal processing*. V. ASSP-23, No. 1, 1975.
- [41] Mistrík, J.: *Frekvencia tvarov a konštrukcií v slovenčine*. Bratislava, VEDA 1985. 320 s.
- [42] Mistrík, J.: *Retrográdny slovník slovenčiny*. Bratislava, UK 1976. 735 s.
- [43] Ondruš, Š. – Sabol, J.: *Úvod do štúdia jazykov*. Bratislava, Slovenské pedagogické nakladateľstvo 1987. 344 s.
- [44] Pauliny, E.: *Fonológia spisovnej slovenčiny*. Bratislava, Slovenské pedagogické nakladateľstvo 1968. 128 s.
- [45] Pauliny, E.: *Slovenská fonológia*. Bratislava, Slovenské pedagogické nakladateľstvo 1979. 213 s.
- [46] Pauliny, E.: Typy slabík v slovenčine. *Slovenská reč*, 43, 1978, s 326–342.
- [47] Pauliny, E.: Slabičné [r][l] v slovenčine. *Slovo a slovesnosť*, 38, 1977, s 307–310.
- [48] Pauliny, E.: Vnútoraná väzba slabiky a jej hranice v slovenčine. *Slovenská reč*, 43, 1978, s. 193–209.
- [49] Páleš, E.: *Sapfo – Parafrázovač slovenčiny*, Bratislava, VEDA 1994. 305 s.
- [50] Psutka, J.: *Komunikace s počítačem mluvenou řečí*. Praha, Nakladatelství Akademie věd české republiky 1995.
- [51] Rabiner, L. – Juang, B. H.: *Fundamental of Speech Recognition*. New Jersey, Prentice Hall 1993. 507 s.
- [52] Sabol, J.: *Syntetická fonologická teória*. Bratislava, Jazykovedný Ústav Ľ. Štúra SAV 1988. 255 s.
- [53] Sabol, J.: Slovenská slabika (Náčrt problematiky). *Studia Academica Alovaca* 23. Bratislava, STIMUL 1994.
- [54] Sokolova, M. – Moško, G. – Šimon, F. – Benko, V.: *Morfematický slovník slovenčiny*. Prešov, Náuka 1999. 530 s.

- [55] Sorin, C.: Towards high-quality multilingual text-to-speech. In *Progress and Prospects of Speech Research and Technology*. Ed: Niemann H., de Mori R., and Hanrieder G. Sankt Augustin, Infix Publishing Co. 1994, s. 53–62.
- [56] Štefánik, J. – Rusko, M. – Považanec, D.: Frekvencia slov, grafém, hlások a ďalších elementov slovenského jazyka. *Jazykovedný časopis*, 50, 1999.
- [57] Viterbi, A. J.: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. In: *IEEE Transactions on Information Theory*. IT-13, 1967, s. 260–267.
- [58] Wells, J. C.: Computer-coded phonetic transcription. *Journal of the International Phonetic Association*, 17, No. 2, 1987, s. 94–114.
- [59] Wells, J. C.: Computer-coding the IPA: a proposed extension of SAMPA. URL: <http://www.phon.ucl.ac.uk/home/sampa/home.htm>, SAMPA – official WWW pages.
- [60] Wesenick, M. B.: Automatic Generation of German Pronunciation Variants. In: *Proceedings of the ICSLP 1996*. Philadelphia, 1996, s. 125-128.
- [61] *Pravidlá slovenského pravopisu*. Bratislava, VEDA 1998.

Použité skratky

HMM	skrytý Markovov model (Hidden Markov Model)
NN	neurónová sieť (Neural Network)
	slabičná hranica
∣	morfematická hranica
┌	hranica základu slova a prípony
└	hranica základu slova a predpony
~	hranica predpony alebo prípony
‡	hranica slov v zložených slovách
	hranica slov
[]	fonetický prepis
SM	samohláska
SP	spoluhláska
ZN	znelá spoluhláska
NZ	neznelá spoluhláska
SSP	šumová spoluhláska
ZNp	znelá párová spoluhláska
NZp	neznelá párová spoluhláska
ZNa	znelá spoluhláska vzniknuvšia asimiláciou neznej
NZa	neznelá spoluhláska vzniknuvšia asimiláciou znej
ZNH	znelá hláska
ZNs	znelá šumová spoluhláska
NZs	neznelá šumová spoluhláska
ZNsa	znelá šumová spoluhláska vzniknuvšia asimiláciou neznej
NZsa	neznelá šumová spoluhláska vzniknuvšia asimiláciou znej

Zoznam obrázkov

1.1	Schéma IPA z roku 1993.	12
1.2	Schéma IPA z roku 1993.	13
1.3	Schéma IPA z roku 1993.	14
1.4	Spôsob reprezentácie vety, slova a fonémy.	25
3.1	Rozloženie samohlások v niektorých slovanských jazykoch.	38
3.2	Vokalická schéma IPA z roku 1996.	40
3.3	Slovenské samohlásky a dvojhlasky v systéme kardinálnych vokálov. Čiarkovane je vyznačená poloha <i>e</i> , <i>i</i> po prepočte $(F2 + F3)/2$	41
4.1	Hranice slov po jednotlivých iteráciách pri segmentácii vety <i>Nie som v Ostrave, som v Otrokoviciach</i>	134
4.2	Hranice slov po jednotlivých iteráciách pri segmentácii vety <i>Prísím Vás, kedy máme spoj z Levíc do Krakova</i>	135
4.3	Priemerný posun začiatkov slov po jednotlivých iteráciách.	136

Zoznam tabuliek

1.1	ASCII/ISO 7-bitová znaková sada.	10
1.2	Prehľad počítačovo reprezentovateľných fonetických transkripcií.	15
1.3	SAM abeceda (SAMPA) pre fonetickú transkripciu I – 1987.	16
1.4	SAM abeceda (SAMPA) pre fonetickú transkripciu II – 1987.	17
1.5	Tabuľka znakov slovenskej fonetickej transkripcie	20
3.1	Fonematická verzia SAMPA abecedy pre slovenčinu.	35
3.2	Formantové frekvencie pre slovenské samohlásky.	38
3.3	Formantové frekvencie pre kardinálne vokály v systéme IPA.	39
3.4	Vzdialenosti slovenských samohlások a samohlások v schéme CV pre F_1 a F_2	40
3.5	Vzdialenosti slovenských samohlások a samohlások v schéme CV pre F_1 , F_2 a F_3	40
3.6	Porovnanie slovenskej transkripcie, IPA a SAMPA systému.	42
4.1	Prekódovacia SAMPA tabuľka slovenského fonetického repertoáru do nemeckého.	126
4.2	Výsledky segmentácie vety <i>Nie som v Ostrave, som v Otrokoviciach</i> pre jednotlivé iterácie.	133
4.3	Výsledky segmentácie vety <i>Prosím Vás, kedy máme spoj z Levíc</i> <i>do Krakova</i> pre jednotlivé iterácie.	134

Dodatok A

Testy

A.1 Zoznam testovacích slov pre automatickú transkripciu a získané výsledky

Slovo	Prepis s morf. delením	Prepis bez morf. delenia
hlavnými	h\ l a U_ ^ n I: m I	h\ l a U_ ^ n I: m I
uloví	U l O v I:	U l O v I:
urbanovou	U r b a n O v O U_ ^	U r b a n O v O U_ ^
rizika	r I z I k a	r I z I k a
zažili	z a Z I L I	z a Z I L I
	z a Z I I I	z a Z I I I
vybudovaním	v I b U d O v a J I: m	v I b U d O v a J I: m
technikov	t E x n I k O f_ v	t E x n I k O f_ v
mierový	m I_ ^ E r O v I:	m I_ ^ E r O v I:
ozývať	O z I: v a c	O z I: v a c
mladomanželom	m l a d O m a n Z E l O m	m l a d O m a n Z E l O m
ocitlo	O t s I t l O	O t s I t l O
vlada	v l a d a	v l a d a
hostovať	h\ O s c O v a c	h\ O s c O v a c
príbuzenskom	p r I: b U z E n s k O m	p r I: b U z E n s k O m
nepopierateľnou	J E p O p I_ ^ E r a c E L n O U_ ^	J E p O p I_ ^ E r a c E L n O U_ ^
náhrdelníky	n a: h\ r= J\ E l J I: k I	n a: h\ r= J\ E l J I: k I
spozorovala	s p O z O r O v a l a	s p O z O r O v a l a
cenzúra	t s E n z U: r a	t s E n z U: r a
nepredpokladal	J E p r E t p O k l a d a l	J E p r E d p O k l a d a l
ženám	Z E n a: m	Z E n a: m
minifutbal	m I J I f U d b a l	m I J I f U d b a l
povedala	p O v E d a l a	p O v E d a l a
pevného	p E U_ ^ J E: h\ O	p E U_ ^ J E: h\ O
šíriteľa	S I: r I c E L a	S I: r I c E L a
zachovali	z a x O v a l I	z a x O v a l I
	z a x O v a L I	z a x O v a L I
trikmi	t r I g m I	t r I k m I
donútila	d O n U: c I l a	d O n U: c I l a

Slovo	Prepis s morf. delenim	Prepis bez morf. delenia
polevu	p O l E v U p O L E v U	p O l E v U p O L E v U
nekompromisnú	J E k O m p r O m I s n U:	J E k O m p r O m I s n U:
podpíšu	p O t p I: S U	p O d p I: S U
kandidátska	k a n J\ I d a: t s k a	k a n J\ I d a: t s k a
lubrikovaného	l U b r I k O v a J E: h\ O	l U b r I k O v a J E: h\ O
splatenie	s p l a c E J L^E	s p l a c E J L^E
pôdou	p U_^O d O U_^	p U_^O d O U_^
sálu	s a: l U	s a: l U
dobu	d O b U	d O b U
kubiša	k U b I S a	k U b I S a
benevolenciu	b E J E v O l E n t s L^U b E J E v O L E n t s L^U	b E J E v O l E n t s L^U b E J E v O L E n t s L^U
jesenná	j E s E n n a:	j E s E n n a:
kabinetu	k a b I J E t U	k a b I J E t U
očakávaní	O t S a k a: v a J I:	O t S a k a: v a J I:
paláca	p a l a: t s a	p a l a: t s a
prerokovali	p r E r O k O v a l I p r E r O k O v a L I	p r E r O k O v a l I p r E r O k O v a L I
babej	b a b E L^	b a b E L^
šestákovou	S E s t a: k O v O U_^	S E s t a: k O v O U_^
mraziacich	m r a z I a t s I x	m r a z I^a t s I x
ružomberčan	r U Z O m b E r t S a n	r U Z O m b E r t S a n
svojvoľne	s v O j v O L J E s v O j U_^O L J E	s v O j v O L J E s v O j U_^O L J E
zdobia	z d O b L^a	z d O b L^a
oddieloch	O d J\ L^E l O x	O d J\ L^E l O x
sajulin	s a j U l I n s a j U L I n	s a j U l I n s a j U L I n
vyťaženie	v I c a Z E J L^E	v I c a Z E J L^E
odvtedy	O d f t E d I	O d f t E d I
závažnom	z a: v a Z n O m	z a: v a Z n O m
označila	O z n a t S I l a	O z n a t S I l a
úkonom	U: k O n O m	U: k O n O m
nepatrná	J E p a t r n a:	J E p a t r n a:
schybí	s x I b I:	s x I b I:
odpálená	O t p a: l E n a: O t p a: L E n a:	O d p a: l E n a: O d p a: L E n a:
palace	p a l a t s E	p a l a t s E
vydali	v I d a l I v I d a L I	v I d a l I v I d a L I
kopy	k O p I	k O p I
popálenú	p O p a: l E n U: p O p a: L E n U:	p O p a: l E n U: p O p a: L E n U:
oznámiť	O z n a: m I c	O z n a: m I c
stabilných	s t a b I l n I: x	s t a b I l n I: x
pripustenie	p r I p U s c E J L^E	p r I p U s c E J L^E
ľadového	L a d O v E: h\ O	L a d O v E: h\ O

Slovo	Prepis s morf. delenim	Prepis bez morf. delenia
mizernú	m I z E r n U:	m I z E r n U:
stane	s t a J E	s t a J E
minimálnym	m I J I m a: l n I m	m I J I m a: l n I m
drahej	d r a h \ E L ^	d !r! a h \ E L ^
veterinárny	v E c E r I n a: r n I	v E c E r I n a: r n I
konkurenčnou	k O N k U r E n t S n O U _ ^	k O N k U r E n t S n O U _ ^
útokom	U: t O k O m	U: t O k O m
kladiete	k l a J \ L ^ E c E	k l a J \ L ^ E c E
tibete	c I b E c E	c I b E c E
brada	b r a d a	b r a d a
krále	k r a: l E	k r a: l E
	k r a: L E	k r a: L E
krúžkoch	k r U: S k O x	k r U: Z k O x
jednej	j E d J E L ^	j E d J E L ^
ponoriť	p O n O r I c	p O n O r I c
teroristických	c E r O r I s c I t s k I: x	c E r O r I s c I t s k I: x
nezaradovali	J E z a r a J \ O v a l I	J E z a r a J \ O v a l I
	J E z a r a J \ O v a L I	J E z a r a J \ O v a L I
najlukratívnejší	n a j l U k r a c I: v J E L ^ S I:	n a L ^ l U k r a c I: v J E L ^ S I:
	n a j l U k r a c I: U _ ^ J E L ^ S I:	n a L ^ l U k r a c I: U _ ^ J E L ^ S I:
rozlúčkový	r O z l U: t S k O v I:	r O z l U: t S k O v I:
izolovalo	I z O l O v a l O	I z O l O v a l O
viceprimátora	v I t s E p r I m a: t O r a	v I t s E p r I m a: t O r a
dramaturgov	d r a m a t U r g O f _ v	d r a m a t U r g O f _ v
spomaľovaní	s p O m a L O v a J I:	s p O m a L O v a J I:
vládať	v l a: d a c	v l a: d a c
zvážila	z v a: Z I l a	z v a: Z I l a
rozopínajú	r O z O p I: n a j U:	r O z O p I: n a j U:
obsahovo	O b s a h \ O v O	O b s a h \ O v O
ničotu	J I t S O t U	J I t S O t U
reformným	r E f O r m n I: m	r E f O r m n I: m
predstavujú	p r E t s t a v U j U:	p r E t s t a v U j U:
vládcom	v l a: t s t s O m	v l a: d t s O m
zapaľoval	z a p a L O v a l	z a p a L O v a l
poňatí	p O J a c I:	p O J a c I:
nepodporiť	J E p O d p O r I c	J E p O d p O r I c

A.2 Zoznam testovacích slov pre slabičnú segmentáciu

Na otestovanie slabičnej segmentácie boli použité nasledujúce slová: *hlavnými, uloví, urbanovou, rizika, zažili, zabudlo, vybudovaním, technikov, mierový, parlamentnými, ozývať, mladomanželom, istotu, ocitlo, priaznivý, metropolitnom, vlada, hostovať, vyšetrujú, vyrastal, príbuzenskom, nepopierateľnou, náhrdelníky, spozorovala, cenzúra, nepredpokladal, ženám, minifutbal, mág, povedala, cvičenia, lúčoch, pevného, širitela, zachovali, trikmi, donútila, bčkami, polevu, poly-*

grafii, ukradomky, nekompromisnú, podpíšu, vtipný, kandidátska, presťahovaní, lubrikovaného, splatenie, vajíčko, pôdou, profesionálnejšie, senzácie, micháča, zmiatol, sálu, dobu, kubiša, benevolenciu, jesenná, kabinetu, očakávaní, paláca, prerokovali, babej, šestákovou, vykladateľ, mraziacich, ružomberčan, svojvoľne, zdočia, oddieloch, sadziab, sajulin, žiadnu, vyťaženie, odvtedy, závažnom, záznamom, cestárske, kábloviek, šmahelová, vyhradená, reprezentantky, vecné, zakladateľovi, označila, prípojok, úkonom, nepatrná, schybí, zasiahnutý, odpálená, palace, vydali, kopy, petnášťročného, popálenú, oznámiť, stabilných, pripustenie, pristane, ľadového, pripravovaných, sobote, neštandardnej, mizernú, stane, odpovedajú, minimálnym, hornorakúskej, drahej, veterinárny, konkurenčnou, útokom, cudzincom, kladiete, tibete, brada, kráľe, krúžkoch, jednej, ponoriť, teroristických, zaujímajú, nezaraďovali, najlukratívnejší, rozlúčkový, izolovalo, presadzoval, viceprimátora, unikátnych, dramaturgov, pridržať, spomaľovaní, vládať, dlžobám, zväžila, rozopínajú, obsahovo, ničotu, reformným, neutíchajúcich, povodňami, predstavujú, predsviatočných, vládcov, zapáľoval, poňatí, ústrednom, nepodporiť, pramenil, zneužitá, členenie, hokejovou, múdrych, tovarišstva, zdalo, rešpektovaním, nenastalo, krížoviek, nezlučiteľné, okresov, nerátajú, kútiky, účtovať, nedostával, prstencom, bytových, podpore, spadne, cerkve, ovocných, pozostávať, reťaz, súhlasím, rodiča, rozprávka, samostatnosti, pracovali, rozpadá, tiky, vyznel, vyvolený, elektrárne, voličov, maďarky, splnilo, presvedčivosť, rodinnými, posilou, domaľúvali, paneurópske, výtvarníctva, vytvarovala, reprezentačnom, nerozhoduje, zasnežovania, blížiacie, koliba, neublížili, podielali, povodne, určovania.

Zoznam slov ktoré neboli rozdelené v dôsledky chýbajúcej slabiky v trénovacej množine: *mág, cvičenia, lúčoch, zmiatol, sadziab, kábloviek, šmahelová, prípojok, cudzincom, , unikátnych, pridržať, múdrych, krížoviek, účtovať, spadne, vyznel, elektrárne.*

Zoznam slov v ktorých prvé ani druhé delenie nebolo správne: *profesionálnejšie, okresov, paneurópske.*

Dodatok B

Postupnosť transkripčných pravidiel

V nasledujúcom sú uvedené hlavné funkcie na prepis samohlások a spoluhlások. Keďže prepisy zložitejších fonetických javov sú v samostatných funkciách, ako príklad sú uvedené aj časti kódu na prepis *t*, *d*, *n*, *l* a znelostnú asimiláciu znelých spoluhlások na neznelé.

```
#!/usr/bin/perl

sub prepis_sam
{
    mess("Prepis samohlások a dvojhlások ...",3);
    my @spol=('!b!', '!c!', '!C!', '!d!', '!D!', '!f!', '!g!', '!h!', '!j!',
             '!k!', '!l!', '!ld!', '!L!', '!m!', '!n!', '!N!', '!p!', '!q!',
             '!r!', '!rd!', '!s!', '!S!', '!t!', '!T!', '!v!', '!w!', '!x!',
             '!z!', '!Z!', '!ch!', '!dz!', '!dZ!');
    my %sam=('!a!'=>'a', '!e!'=>'E', '!i!'=>'I', '!o!'=>'O', '!u!'=>'U',
            '!y!'=>'I', '!A!'=>'a:', '!E!'=>'E:', '!I!'=>'I:', '!O!'=>'O:',
            '!U!'=>'U:', '!Y!'=>'I:');
    my %lc_sam=('!a!'=>'a', '!e!'=>'E', '!i!'=>'I', '!o!'=>'O', '!u!'=>'U',
              '!y!'=>'I');
    my %dif=('!i! !a!'=>'I_~a', '!i! !e!'=>'I_~E', '!i! !u!'=>'I_~U',
            '!uo!'=>'U_~O');
    my %samsk=('!a! !i!'=>'a I', '!a! !u!'=>'a U', '!o! !i!'=>'O I',
              '!o! !o!'=>'O O', '!o! !u!'=>'O U', '!i! !i!'=>'I I',
              '!y! !u!'=>'I U', '!a! !o!'=>'a O', '!e! !a!'=>'E a',
              '!e! !A!'=>'E a:', '!E! !o!'=>'E: O', '!i! !E!'=>'I E:',
              '!i! !o!'=>'I O', '!i! !O!'=>'I O:', '!o! !e!'=>'O E',
              '!o! !i!'=>'O I', '!o! !o!'=>'O O', '!u! !u!'=>'U U',
              '!e! !a! !e!'=>'E a E', '!e! !e! !u!'=>'E E U',
```

```

'!e! !o! !a!'=>>'E O a', '!e! !o! !i!'=>>'E O I',
'o! !i! !o!'=>>'O I O');

my $spol=join(' | ', @spol);
#hranica slabiky uvazovana ako spoluhlaska koli impl.
$spol .= '|-';
my $es=" | - ";

my @samstack;
my %normal=(? !t!'=>>'t', '!d!'=>>'d', '!n!'=>>'n', '!l!'=>>'l');
foreach $wrđ (@stack)
{
  $_=$wrđ;
  mess("Aktualny stav: \"$_\",6);
  mess("TDNL tdnl pred yY ...",4);
  s/(? !t! !d! !n! !l!)( $es)!y!/$normal{$1} I/g;
  s/(? !t! !d! !n! !l!)( $es)!Y!/$normal{$1} I:/g;
  mess("Aktualny stav: \"$_\",6);
  mess("Jednoduche samohlasky...",4);
  foreach $sc (keys %sam)
  {
    s/^( $sc($es)($spol)/$sam{$sc} \2/; # a na zaciatku slova
    s/($spol)($es)$sc($es)($spol)/\1 $sam{$sc} \4/g; # a v strede
    s/($spol)($es)$sc($es)($spol)/\1 $sam{$sc} \4/g; # a v strede
    s/($spol)($es)$sc$/\1 $sam{$sc}/; # a na konci
  }
  s/!uo!/$dif{"!uo!"/g; #prepis ô DFR
  s/!u! !o!/$sam{"u"} $sam{"o"/g; #prepis uo
  foreach $sc (keys %sam)
  {
    #prepis vy?- ako vyexpedovat ...
    s/^(v! !y!($es)$sc/v $sam{"i"} $sam{$sc}/;
  }
  mess("Aktualny stav: \"$_\",6);
  #Tuna je osetrenie predpony pri
  mess("Osetrenie predpon pri ...",4);
  mess("Aktualny stav: \"$_\",5);
  #pria
  if(/^(p! !r! !i! !a!/)
  {
    my $ori=$_;
    my $rc=check_dic($dic_pri,"^pria",$wpos);
    $_=$ori;
    s/^(p! !r! !i! !a!/p r I_~a/ if($rc==1);
  }

```

```

    s/^!p! !r! !i! !a!/p r I a/ if($rc==0);
  }
#prie
if(/^!p! !r! !i! !e!/)
{
  my $ori=$_;
  my $rc=check_dic($dic_pri,"^prie",$wpos);
  $_=$ori;
  s/^!p! !r! !i! !e!/p r I E/ if($rc==1);
  s/^!p! !r! !i! !e!/p r I_^E/ if($rc==0);
}
#priu
if(/^!p! !r! !i! !u!/)
{
  s/^!p! !r! !i! !u!/p r I U/;
}
#Tuna bolo osetrenie predpony pri

#Dvojhlaskey
#Dvojhlaskey na konci
mess("Osetrenie dvojhlaskey na konci ...",4);
mess("Aktualny stav: \"$_\"",5);
foreach $sc (keys %lc_sam)
{
  mess("Test na \"!i! $sc\" ...",5);
  while(/!i! $sc$/)
  {
    mess("Aktualny stav: \"$_\"",6);
    my $ori=$_;
    my $rc=check_dic($dic_pri,"iaeu\\$", $wpos);
    $_=$ori;
    # diftong na konci
    s/($spol)($es)!i!($es)$sc$/\1 I $sam{$sc}/ if($rc==1);
    # diftong na konci
    s/($spol)($es)!i!($es)$sc$/\1 $dif{"!i! $sc"}/ if($rc==0);
  }
}
# V strede a na zaciatku. Tato postupnost nutna koli
# WB uvazovanom ako spoluhlaske
foreach $sc (keys %dif)
{
  # diftong v strede
  s/($spol)($es)$sc($es)($spol)/\1 $dif{$sc} \4/g;
  # diftong v strede

```

```

    s/($spol)($es)$sc($es)($spol)/\1 $dif{$sc} \4/g;
}

#Samohlaskove skupiny
foreach $sc (keys %samsk)
{
    s/($spol)($es)$sc($es)($spol)/\1 $samsk{$sc} \4/g; # a v strede
    s/($spol)($es)$sc($es)($spol)/\1 $samsk{$sc} \4/g; # a v strede
}

foreach $sc (keys %sam)
{
    # Vsetky ostatne samohlasky. Kedze len domace slova.
    s/$sc/$sam{$sc}/g;
}
push @samstack,$_;
}

@stack=@samstack;

if($alt==1) #Generuj kratke samohlasky
{
    $ori=$_;
    s/a:/a/g; s/E:/E/g; s/I:/I/g; s/O:/O/g; s/U:/U/g;
    push @stack,$_ if($_ ne $ori);
}
}

sub prepis_spol
{
    mess("Prepis spoluhlasok ...",3);
    my @spol=('!b!', '!c!', '!C!', '!d!', '!D!', '!f!', '!g!', '!h!', '!j!',
        '!k!', '!l!', '!ld!', '!L!', '!m!', '!n!', '!N!', '!p!', '!q!',
        '!r!', '!rd!', '!s!', '!S!', '!t!', '!T!', '!v!', '!w!', '!x!',
        '!z!', '!Z!', '!ch!', '!dz!', '!dZ!');
    my @sam=('a', 'E', 'I', 'O', 'U', 'a:', 'E:', 'I:', 'O:', 'U:');
    my @ksam=('a', 'E', 'I', 'O', 'U');
    my @samnoI=('a', 'E', 'O', 'U', 'a:', 'E:', 'O:', 'U:');
    my $spol=join('|', @spol);
    my $sam=join('|', @sam);
    my $ksam=join('|', @ksam);
    my $samnoI=join('|', @samnoI);
    my $nez="!p!|!f!|!t!|!s!|!c!|!S!|!C!|!T!|!k!|!ch!";
}

```

```

my $zne="!b!|!v!|!d!|!z!|!dz!|!Z!|!dZ!|!D!|!g!|!h!";
my $znep="b|f_v|d|z|dz|Z|dZ|!J\\|g|h|v|m|j";
my $es=" | - | < | > | ; "; # Ziadna (Empty) alebo slabicna hranica
my $ms=" < | > | ; "; # Ziadna (Empty) alebo slabicna hranica

mess("Osetrenie j a I_^",4);
rule(\@stack,"- !j! ($sam)", "- j \\1", "Vyslovnost j");
rule(\@stack,"($samnoI) !j! -", "\\1 I_^ -", "Vyslovnost I_^");
rule(\@stack,"(I|I:) !j! -", "\\1 j -::\\1 I_^ -",
      "Vyslovnost j aj I_^");
rule(\@stack,"- ($spol) !j! ($sam)", "- \\1 j \\2::- \\1 I_^ \\2",
      "Vyslovnost j aj I_^");
rule(\@stack,"!j!", "j", "Vyslovnost j default");
mess("Vyslovnost f, f_v",4);
rule(\@stack,"!f! ($es) ($sam)", "f \\2", "Vyslovnost f");
rule(\@stack,"- !v! ($es) (!n!|!N!)", "- v \\2::- f_v \\2",
      "Vyslovnost -vn");
rule(\@stack,"^- !v! ($es) !z! ($es) ($nez)", "- f z \\2\\3",
      "Vyslovnost f");
rule(\@stack,"- !v! ($es) ($nez)", "- f \\2", "Vyslovnost f");
rule(\@stack,"^- !v! ($es) !z!", "- f_v z", "Vyslovnost f_v");
rule_fv(\@stack);
rule(\@stack,"- !v! ($es) ($zne)", "- f_v \\2", "Vyslovnost f_v");
mess("Osetrenie v a U_^",4);
# rule(\@stack,"^- !v! -\$", "- f -", "vyslovnost predlozky 'v'");
rule(\@stack,"- !v! ($ksam|!r!|!rd!|!l!|!ld!|!L!|j)", "- v \\1",
      "Vyslovnost v");
rule(\@stack,"- ($spol) !v! ($ksam|!r!|!rd!|!l!|!ld!|!L!|j)",
      "- \\1 v \\2", "Vyslovnost v");
rule(\@stack,"U !v! -", "U v -::U U_^ -", "Vyslovnost v aj U_^");
rule(\@stack,"($ksam|!r!|!rd!|!l!|!ld!) !v! -", "\\1 U_^ -",
      "Vyslovnost U_^");
rule(\@stack,"O U -\$", "O U_^ -", "Vyslovnost U_^");
rule(\@stack,"!v!", "v::U_^", "Vyslovnost v aj U_^");
mess("Osetrenie m, F, n, N",4);
rule(\@stack,"!n! ($es) !(k|g)!", "N \\2", "Vyslovnost N");
rule(\@stack,"!n! ($es) !x!", "N k s", "Vyslovnost N");
rule(\@stack,"!n! ($es) !([szSZL])!", "n \\2", "Vyslovnost n skrt");
rule(\@stack,"!n! ($es) !ch!", "n x", "Vyslovnost n skrt");
rule(\@stack,"!m! ($es) !?(v|f) !?", "F \\2", "Vyslovnost F");
rule(\@stack,"!n! ($es) !b!", "m b", "Vyslovnost m");
rule(\@stack,"!m!", "m", "Vyslovnost m");
mess("Spoluhlaskove skupiny",4);
rule(\@stack,"!n! ($es) ![td]! ($es) !s! ($es) !k!", "n ts k::n s k",

```

```

    "Vyslovnost ntsk, ndsk");
rule(\@stack,"!s!($es)!t!($es)!s!($es)!t!($es)!v!","s s t v",
    "Vyslovnost ststv");
rule(\@stack,"!p!($es)!t!($es)!s!($es)!k!","p s k",
    "Vyslovnost ptsk");
rule(\@stack,"!s!($es)!t!($es)!s!($es)!k!","s s k",
    "Vyslovnost stsk");
rule(\@stack,"!s!($es)!T!($es)!s!($es)!t!","s s t",
    "Vyslovnost sTst");
rule(\@stack,"!(t|d|dz)!($es)!s!($es)!k!","ts k",
    "Vyslovnost tsk, dsk, dzsk");
rule(\@stack,"!d!($es)!s!($es)!t!","ts t","Vyslovnost dst");
rule(\@stack,"!(t|D)!($es)!s!($es)!t!($es)!v!|v)","ts t v",
    "Vyslovnost tstv, Dstv");
rule(\@stack,"!(t|D)!($es)!s!($es)!t!($es)U_\\^","ts t U_^",
    "Vyslovnost tstv, Dstv");
rule(\@stack,"!z!($es)!s!($es)!k!($es)(I|I:|O) -\\$","s k \\4 -",
    "Vyslovnost zsk");
rule(\@stack,"!z!($es)!S!($es)!t!($es)I($es)!n!($es)a -\\$","
    S c I n a -","Vyslovnost zST");
rule(\@stack,"![dt]!($es)!S!($es)!t!","tS c","Vyslovnost dST, tST");
rule(\@stack,"!p!($es)!t!($es)!C!($es)!n!","p tS J",
    "Vyslovnost ptCN");
rule(\@stack,"!S!($ms)!s!","S s","Vyslovnost Ss");
rule(\@stack,"!Z!($ms)!s!","Z s","Vyslovnost Zs");
rule(\@stack,"!s!($ms)!S!","s S","Vyslovnost sS");
rule(\@stack,"!z!($ms)!S!","z S","Vyslovnost zS");
rule(\@stack,"!z!($ms)!Z!","z Z","Vyslovnost zZ");
rule(\@stack,"!s!($ms)!C!","s tS","Vyslovnost sC");
rule(\@stack,"!C!($ms)!s!","tS s","Vyslovnost Cs");
rule(\@stack,"!z!($ms)!C!","z tS","Vyslovnost zC");
mess("Zdvojené spoluhlásky",4);
rule_dbl_SP(\@stack);
mess("Asimilacia znelej na neznelu",4);
rule_ZN_NZ(\@stack);
mess("Asimilacia neznelej na znelu",4);
rule_NZ_ZN(\@stack);
mess("Asimilacia h, ch",4);
rule_hch(\@stack);
mess("Specialitky",4);
rule(\@stack,"^- !d!($es)I($es)!s!($es)($zne|$znep|$sam)",
    "- d I z \\4","Vyslovnost d[iy]s-");
rule(\@stack,"^- E($es)!x!($es)($zne|$znep|$sam)","- E g z \\3",
    "Vyslovnost ex-");

```

```

    rule(\@stack,"!x!","k s","Vyslovnost x");
#Tuna prepis secky okrem d,t,n,l,D,T,N,L,r,l,
my @spolr=('!b!','!c!','!C!','!dz!','!dZ!','!f!','!g!','!h!','!ch!','
    '!k!','!ld!','!m!','!p!','!rd!','!s!','!S!','!v!','!w!','
    '!x!','!z!','!Z!');
my %neasimr=('!b!'=>!b','!c!'=>!ts','!C!'=>!tS','!dz!'=>!dz',
    '!dZ!'=>!dZ','!f!'=>!f','!g!'=>!g','!h!'=>!h\\',
    '!ch!'=>!x','!k!'=>!k','!ld!'=>!l=:','!m!'=>!m',
    '!p!'=>!p','!rd!'=>!r=:','!s!'=>!s','!S!'=>!S',
    '!v!'=>!v','!w!'=>!v','!x!'=>!k s','!z!'=>!z',
    '!Z!'=>!Z');
my $spolr=join('|',@spolr);
foreach $_ (@stack)
    {s/($spolr)/$neasimr{$1}/g;}

mess("Prepis dtnl",4);
rule_dtnl(\@stack);

foreach $_ (@stack)
    {s/[><]//g; s/[ ]+ /g;}
}

sub rule_dtnl
{
    my $stack=@_[0];
    my @lstack=@$stack;
    my @ostack=();
    my $ms=" < | > | "; # Morfematicka hranica
    my $es=" | - "; # Ziadna (Empty) alebo Slabicna hranica
    my $as="$es|$ms";
    my @spol=('b','ts','tS','dz','dZ','f','f_v','g','h\\','x','k','l=:','
        'm','F','N','p','r=:','s','S','v','z','Z');
    my $spol=join('|',@spol);
    my $tdnl="!t!|!d!|!n!|!l!";
    my $TDNL="!T!|!D!|!N!|!L!";
    my %normal=('!t!'=>!t','!d!'=>!d','!n!'=>!n','!l!'=>!l',
        '!T!'=>!c','!D!'=>!J\\','!N!'=>!J','!L!'=>!L');
    my %makko=('!t!'=>!c','!d!'=>!J\\','!n!'=>!J','!l!'=>!L');
    foreach $wrđ (@lstack)
    {
        my $owrd=$wrđ;
        mess("Vyslovnost TDNL",5);
        mess("Aktualny stav 1: \"$wrđ\"",6);
        $wrđ=~s/($TDNL)/$normal{$1}/g;
    }
}

```

```

mess("Aktualny stav 2: \"$wrd\"",6);
mess("Vyslovnost tdnl pre aou",5);
mess("Aktualny stav 1: \"$wrd\"",6);
$wrd=~s/($tdnl)($as)(a|O|U|a:|O:|U:)/$normal{$1} $3/g;
mess("Aktualny stav 2: \"$wrd\"",6);
mess("Dalsie pravidla ...",5);
mess("Aktualny stav 1: \"$wrd\"",6);
$wrd=~s/($tdnl)($as)(a|O|U|a:|O:|U:|E:|U_\\^ O|$spol)$\\
/$normal{$1} $3/g;
$wrd=~s/($tdnl) < (I|E|I:|E:)/$normal{$1} $3/;
$wrd=~s/($tdnl)($as)I($as)ts($as)k($as)I:/$normal{$1} I ts k I:/;
$wrd=~s/($tdnl)($as)I($as)z($as)O($as)v($as)a($as)c$\\
/$normal{$1} I z O v a c/;
$wrd=~s/($tdnl)($as)I($as)z($as)m($as)U($as)s$\\
/$normal{$1} I z m U s/;
$wrd=~s/($tdnl)($as)I($as)s($as)t($as)a/$normal{$1} I s t a/;
$wrd=~s/($tdnl)($as)E($as)n($as)t/$normal{$1} E n t/;
$wrd=~s/($tdnl)($as)E($as)(r|!r!)/$normal{$1} E r/;
$wrd=~s/($tdnl)($as)E:($as)(r|!r!)/$normal{$1} E: r/;
$wrd=~s/($tdnl)($as)E($as)s/$normal{$1} E s/;
my $ttt=$wrd;
$wrd=~s/($tdnl) > I($as)k/$makko{$1} I k/;
$wrd=~s/($tdnl) > I:($as)k/$makko{$1} I: k/;
$ttt=~s/!l! > I($as)k/l I k/;
$ttt=~s/!l! > I:($as)k/l I: k/;
mess("Aktualny stav 2: \"$wrd\"",6);
push @ostack,$ttt if($wrd ne $ttt);
$wrd=~s/(!t!!|!d!!|!n!)($as)(I|E|I:)/$makko{$1} $3/g;

# l vseobecne
$ttt=$wrd;
$wrd=~s/!l!($es)(I|E|I:|E:)/L $2/g;
$ttt=~s/!l!($es)(I|E|I:|E:)/l $2/g;
push @ostack,$ttt if($wrd ne $ttt);
mess("Aktualny stav 2: \"$wrd\"",6);
#Default standard
mess("Default tdnl ...",5);
mess("Aktualny stav 1: \"$wrd\"",6);
$wrd=~s/($tdnl|$TDNL)/$normal{$1}/g;
mess("Aktualny stav 2: \"$wrd\"",6);
mess("Aplikovane pravidlo: vyslovnost dtnl",4) if($owrd ne $wrd);
push @ostack,$wrd;
}
@$stack=@ostack;

```

```

}

sub rule_ZN_NZ
{
my $stack=@_[0];
my @lstack=@$stack;
my @ostack=();
my $zne="!b!|f_v|!d!|!z!|!dz!|!Z!|!dZ!|!D!|!g!|!h!";
my $nez="!p!|!f!|!t!|!s!|!c!|!S!|!C!|!T!|!k!|!ch!";
my %asim_zn=( 'b!'=>'p', 'f_v'=>'f', 'd!'=>'t', 'z!'=>'s', 'dz!'=>'ts',
              'Z!'=>'S', 'dZ!'=>'tS', 'D!'=>'c', 'g!'=>'k', 'h!'=>'x' );
my %neasim=( 'b!'=>'b', 'f_v'=>'f_v', 'd!'=>'d', 'z!'=>'z', 'dz!'=>'dz',
              'Z!'=>'Z', 'dZ!'=>'dZ', 'D!'=>'J\\', 'g!'=>'g',
              'h!'=>'h\\', 'p!'=>'p', 'f!'=>'f', 't!'=>'t', 's!'=>'s',
              'c!'=>'ts', 'S!'=>'S', 'C!'=>'tS', 'T!'=>'c', 'k!'=>'k',
              'ch!'=>'x', 'm'=>'m', 'v'=>'v' );
my $ses=" | - "; # Ziadna (Empty) alebo slabicna hranica
foreach $wrđ (@lstack)
{
mess("Asimilacia na hranici slova",5);
mess("Aktualny stav 1: \"$wrđ\" RW: \"$rword\"",6);
my $owrd=$wrđ;
#Na hranici dvoch slov
if($wrđ=~/$zne$/ and $rword=~/$nez/)
{
$wrđ=~s/$zne)/$asim_zn{$1}/;
$rword=~s/^(($nez)/$neasim{$1}/;
}
mess("Aktualny stav 2: \"$wrđ\" RW: \"$rword\"",6);
#Predpony
mess("Asimilacia na hranici predpony a korena",5);
mess("Aktualny stav 1: \"$wrđ\" RW: \"$rword\"",6);
$wrđ=~s/$zne) < ($nez)/$asim_zn{$1} $neasim{$2}/;
mess("Aktualny stav 2: \"$wrđ\" RW: \"$rword\"",6);
#Zlozene slova
mess("Asimilacia v zlozenych slovach",5);
mess("Aktualny stav 1: \"$wrđ\" RW: \"$rword\"",6);
$wrđ=~s/$zne) ; ($nez)/$asim_zn{$1} $neasim{$2}/;
mess("Aktualny stav 2: \"$wrđ\" RW: \"$rword\"",6);
#Na konci slova
mess("Asimilacia na konci slova",5);
mess("Aktualny stav 1: \"$wrđ\" RW: \"$rword\"",6);
$wrđ=~s/$zne)($ses)($nez)/$asim_zn{$1} $neasim{$3}/;
mess("Aktualny stav 2: \"$wrđ\" RW: \"$rword\"",6);
}
}

```

```

#Pripony
mess("Asimilacia na hranici korena a pripony",5);
mess("Aktualny stav 1: \"$wrd\" RW: \"$rword\"",6);
$wrd=~s/($zne) > ($nez)/$asim_zn{$1} $neasim{$2}/;
mess("Aktualny stav 2: \"$wrd\" RW: \"$rword\"",6);
mess("Aplikovane pravidlo: Prepis znej spoluhlasky na neznelu",4)\
  if($owrd ne $wrd);
  push @ostack,$wrd;
}
@$stack=@ostack;
}

```

```

# aääbcčdđdzdeéefghchiiijklíłmnňoóôpqrřssttúúvxyýžž
# AĀBCČDĎDzDžEĚFGHChIĪJKLĹLMNŇOŌPQRŘSŠTŤUÚVXYÝŽŽ

```