

Transkripce psaného českého textu do fonetické podoby

Dana Nejedlová

Laboratoř počítačového zpracování řeči
Technická univerzita v Liberci

1. ZAŘAZENÍ TÉMATU DO PROBLEMATIKY KOMUNIKACE S POČÍTAČEM MLUVENOU ŘEČÍ

Psaná řeč je posloupností znaků a mluvená řeč je posloupností zvuků. Mezi psanou řečí a její fonetickou (zvukovou) podobou je potřeba najít korespondenci, která se využívá při automatickém rozpoznání řeči a při počítačové syntéze řeči.

Automatické rozpoznání řeči se provádí pomocí uzavřeného (s určitou konečnou množinou slov) slovníku, ve kterém má každé slovo jednu nebo více alternativních fonetických transkripcí, což je vyjádření zvukové podoby slova pomocí znaků zvaných fonémy. Fonetické transkripce slov se porovnávají s výsledkem analýzy akustického signálu.

Počítačová syntéza řeči má jako vstup psaný text, který je třeba převést na posloupnost fonémů, které budou dále syntetizovány do zvuků. Na rozdíl od automatického rozpoznání řeči, kde slova ve slovníku mají statické předem připravené fonetické transkripce, počítačová syntéza řeči by měla pořizovat fonetické transkripce pro libovolný předem neznámý text za běhu.

2. ROZDÍLY MEZI MLUVENOU A PSANOU FORMOU JAZYKA

Při automatickém převodu mluvené řeči na text nebo při opačném procesu syntézy řeči podle textu je třeba přihlédnout k jistým odlišnostem mluveného a psaného jazyka. Toto téma je z převážně akustického hlediska zpracováno v literatuře [1].

Při provádění počítačové syntézy řeči, která reprodukuje například text novinového článku, nebo při pořizování testovacích dat pro rozpoznávač je třeba přepsat velké množství textu (takzvaný textový korpus) do fonetické podoby. Zdrojem textového korpusu je většinou kolekce převážně zpravodajských textů dostupných v elektronické podobě na Internetu. Při tomto úkolu je třeba řešit tři hlavní problémy a to zkratky, číslovky, homofony a homografy. Popis této problematiky je čerpán z práce [2].

2.1 Zkratky

Zkratky je třeba přepsat na nezkrácená slova, která se obvykle vyslovují místo zkratk v mluvené formě jazyka. V češtině je to komplikováno její vysokou flektivností. Rozlišují se tedy zkratky, které lze přepsat jediným tvarem (*např., apod., t.j.*) a zkratky jejichž rozvinutý tvar záleží na kontextu, protože podléhá skloňování nebo časování. V takových případech lze rozvinutý tvar někdy určit podle předložky, což umožňuje stanovit jednoduchá pravidla pro přepis. V mnoha případech však správné stanovení rozvinutého tvaru zkratky vyžaduje sémantickou analýzu textu a ta vyžaduje velké množství složitých pravidel.

2.2 Číslovky

Číslovky mají podobnou problematiku jako zkratky. Některé číslovky mají alternativní způsob výslovnosti. Například 25 lze vyslovit jako *dvacet pět* nebo jako *pěťadvacet*. V takových případech se zvolí častěji užívaný styl (*pěťadvacet*). Třída slov jako je *x-letý, x-dvěřový, x-místný, x-patrový* a podobně je rozepisována se stejnou koncovkou a místo „x-“ je doplněna číslovka ve 2. pádu.

2.3 Homofony

Homofony jsou slova, která se jinak píší a stejně vyslovují. Příkladem jsou slovesa *bít* a *být* nebo slovo *NATO* a sousloví *na to*. Velmi četnou třídou jsou slovesa lišící se v koncovce *i/y* podle jejich rodu, například *žili* a *žily*. Další významnou kategorií jsou slova lišící se v kapitálkách, například *donutil* a *Donutil*, tedy slovo je někdy slovem majícím svůj běžný význam a někdy osobním jménem nebo místním názvem. Slova ze všech těchto kategorií jsou problémem při automatickém rozpoznání řeči, kdy chceme, aby rozpoznávač napsal rozpoznané slovo správně. Jeden z možných způsobů jeho alespoň částečného řešení je použití *n*-gramového jazykového modelu.

2.4 Homografy

Homografy patříci spolu s homofony mezi homonyma [3, str. 592] jsou slova, která se stejně píší a jinak vyslovují. Příkladem je slovo *byty*, které může znamenat obydlí a potom se vyslovuje [*bity*], a nebo jednotky informace a potom se vyslovuje [*bajty*]. Určení správné fonetické transkripce záleží na sémantickém kontextu. Zatímco pro účely automatického rozpoznání řeči stačí do slovníku přidat k homografu alternativní fonetické transkripce, při počítačové syntéze řeči je nutné provádět sémantickou analýzu za běhu, při které lze využít metody automatického určování tématu (topic detection).

3. SEGMENTACE

Za účelem sestavení slovníku a jazykového modelu pro rozpoznávač je nutné přepsat psanou podobu řeči do fonetické podoby a problémy s tím spojené byly probrány v předchozí kapitole. Za účelem trénování akustických modelů pro rozpoznávač a syntezátor je nutné mapovat segmenty akustického záznamu lidské řeči na fonetické jednotky.

3.1 Jednotky zvuků řeči

Nejčastěji volenou jednotkou pro účely počítačového zpracování řeči je foném. **Foném** je nejmenší jednotka zvukové stránky řeči, která může rozlišovat jednotlivá slova. Například, když ve slově *les* s fonetickým přepisem [*les*] zaměníme foném [*l*] za foném [*p*], tak se toto slovo změní na *pes*. Zvolíme-li jako segmentační jednotku foném, dostaneme nejmenší možný inventář segmentů řeči. To je nesporná výhoda segmentování podle fonémů. Nevýhodou fonémů je však to, že neobsahují informaci o koartikulaci s jejich fonetickým okolím. **Koartikulace** je jev, při kterém určitý foném nabývá různých variant v závislosti na předcházejícím a následujícím zvuku a na tempu a intonaci řeči. Pro odlišení různých variant fonémů byl zaveden termín **fón**, což je jakákoli varianta jakéhokoli fonému, a **alofón**, což je jakákoli varianta určitého fonému. Abychom se při segmentaci vyhnuli problémům s koartikulací, můžeme jako segmentační jednotky využít některé jiné možnosti, mezi něž dle [4] patří:

- **alofóny**, jejichž nevýhodou je velikost jejich inventáře. Ten může nabývat takového počtu, kolik je různých variant, jak lze každý foném obklopit jinými fonémy,
- **difóny**, což je posloupnost samohláska-souhláska, jejichž nevýhodou je opět velikost inventáře,
- **slabiky**, což jsou fonetické útvary, které obsahují samohláskové jádro plus volitelné počáteční a koncové souhlásky nebo skupiny souhlásek. Jejich nevýhodou je vedle velkého inventáře (v mluvené češtině to je více než 10 000) i fakt, že neexistuje obecný souhlas z fonetického i lingvistického hlediska o tom, kde mají být umístěny hranice slabiky. V češtině existují také slabiky tvořené výhradně souhláskami, například ve slovech *prst*, *smrk* nebo *blb*.

- **poloslabiky**, což jsou kompozice počáteční poloslabiky, obsahující počáteční souhláskový shluk a příslušné části samohláskového jádra, plus koncové poloslabiky, obsahující zbývající část samohláskového jádra a koncový souhláskový shluk. Jejich výhodou je, že až pětkrát redukuje rozměr inventáře vyžadovaného pro reprezentaci promluv ve srovnání s celými slabikami. Nevýhodou je složitost s rozdělováním samohláskového jádra.
- **demislabiky**, které podstatně eliminují obtíže vyskytující se při dělení na poloslabiky. Od poloslabik se liší umístěním hranice mezi počáteční a koncovou demislabikou. Počáteční demislabika je vymezena zcela krátce tak, že rozdělovací hranice je umístěna těsně za přechod souhláska-samohláska (tj. počáteční demislabika obsahuje koartikulační jev, ale neobsahuje podstatnou část samohlásky). Umístění hranice tímto způsobem je efektivní tím, že uvolňuje omezení na obou stranách hranice a má i potenciál v redukci inventáře, podobně jako u poloslabik.
- **slova**, jejichž výhodou je eliminace nutnosti zacházet s komplikovanými algoritmy pro segmentaci a identifikaci jednotek nižších. Při rozpoznávání plynulé promluvy však dochází ke koartikulačním jevům i mezi krajními fonémy sousedních slov, což velmi komplikuje problém vymezení hranic slov.

3.2 Příprava dat pro trénování akustických modelů fonémů

Na začátku je třeba mít akustický záznam řeči a její přesný přepis do fonetické podoby tvořený posloupností fonémů nebo jiných jednotek zmíněných v předchozí kapitole. První data je třeba připravit neautomatizovaně. Dále je třeba ručně projít tento záznam a přiřadit jeho úsekům fonémy. Tento proces se nazývá **segmentace**. Poté, co je všem druhům fonémů přiřazen dostatečný počet úseků akustického záznamu řeči, lze natrénovat akustické modely jednotlivých fonémů a v segmentaci pokračovat buď plně automatizovaně nebo s následnou opravou hranic segmentů člověkem. Problém koartikulace je v případě použití fonémů jako jednotky řeči řešen maximalizací množství segmentovaných záznamů, aby se akustické modely fonémů staly robustními.

Výsledkem je, že jsou-li ve slovníku rozpoznávače slova přepsaná na posloupnost fonémů, rozpoznávač má pro každý foném jeho akustický model, který může porovnávat se skutečným akustickým signálem rozpoznávané promluvy.

4. FONETICKÁ TRANSKRIPCE ČEŠTINY

Fonetická transkripce zmiňovaná již v předchozích kapitolách tohoto příspěvku je definována v [4] jako přesný a nedvojznačný přepis textu či zvuků na posloupnost fonémů. Existují dva hlavní druhy přístupů k její automatizované realizaci a to metody založené na pravidlech a metody strojového učení. V obou metodách je třeba zvolit nějakou fonetickou abecedu, která přiřazuje každému fonému nějaký grafický symbol.

4.1 Fonetická abeceda češtiny

Pro zápis fonémů je třeba využít nějakou fonetickou abecedu. Mezinárodní fonetická abeceda (*International Phonetic Alphabet – IPA*), zveřejněná v článku [5], se pro národní účely může nahradit abecedami lépe vystihujícími místní fonetická pravidla. V článku [6] byla navržena fonetická abeceda pro češtinu (*Phonetic Alphabet for Czech – PAC*), kterou jsou zapisovány příklady fonetické transkripce v celém tomto příspěvku. Inventář fonémů podle této abecedy je zachycen v tabulce č. 1. Později byly k abecedě fonémů přidány symboly pro pauzu, různé druhy ruchů, šva (shwa - vyslovované typicky při hláskování souhlásek tak, že je nevyslovujeme jako slabiku, ale například řekneme *b* místo *bé*) a ráz

(glottal stop - vyslovované někdy mezi dvěma samohláskami, které spolu netvoří dvojhlásku nebo před slovy začínajícími na samohlásku).

Důvody pro zavedení fonetické abecedy pro češtinu byly následující:

- Některé české fonémy v mezinárodních abecedách chybí.
- České fonémy se mohou snadno vyjádřit českými národními znaky, což na rozdíl od používání mezinárodních fonetických abeced, ulehčuje českým badatelům práci v tom, že se snadno naučí představit si zvuk fonému podle jeho zápisu.
- S českými národními znaky, které vedle znaků anglické abecedy byly pro zápis fonémů výhradně použity, se lépe operuje v počítačových operačních systémech s českým národním prostředím než s těmi mezinárodními.

Tab. 1. *Symbols pro české fonémy z PAC dle [6]*

Číslo	Foném vyjádřený českými hláskami	Foném dle PAC	Příklad	Číslo	Foném vyjádřený českými hláskami	Foném dle PAC	Příklad
1	„a“	a	táta	21	„m“	m	máma
2	„á“	á	táta	22	„m“	M	tramvaj
3	„b“	b	bába	23	„n“	n	víno
4	„c“	c	ocel	24	„n“	N	banka
5	„dz“	C	leckde	25	„ň“	ň	koně
6	„č“	č	čichá	26	„o“	o	kolo
7	„dž“	Č	rádža	27	„ó“	ó	óda
8	„d“	d	jeden	28	„p“	p	pupen
9	„d“	d'	dělat	29	„r“	r	bere
10	„e“	e	lev	30	„ř“	ř	moře
11	„é“	é	méně	31	„ř“	Ř	keř
12	„f“	f	fauna	32	„s“	s	sud
13	„g“	g	guma	33	„š“	š	duše
14	„h“	h	aha	34	„t“	t	dutý
15	„ch“	X	chudý	35	„t“	t'	kutil
16	„i“ nebo „y“	i	bil, byl	36	„u“	u	duše
17	„í“ nebo „ý“	í	vítr, lýko	37	„ú“ nebo „ů“	ú	růže
18	„j“	j	dojat	38	„v“	v	láva
19	„k“	k	kupec	39	„z“	z	koza
20	„l“	l	dělá	40	„ž“	ž	růže

4.2 Fonetická transkripce založená na pravidlech

Protože si nelze zapamatovat všechny tvary výslovnosti pro každou promluvu, je nutné hledat obecná pravidla, podle nichž by bylo možné fonetický přepis automaticky vytvářet. Tato obecná pravidla mohou být formulována jako produkční pravidla a nazývají se **fonologická pravidla**. [4] Protože většina fonémových změn může být vysvětlena levým a pravým kontextem daného fonému, lze definovat obecné produkční pravidlo ve tvaru

JESTLIŽE řetězci znaků *A* bezprostředně předchází řetězec znaků *C* a je bezprostředně následován řetězcem znaků *D*,

PAK se *A* přepíše na řetězec znaků *B*.

Pro jednoduchost budeme v dalším výkladu toto pravidlo zapisovat ve tvaru

$A \rightarrow B / C _ D$.

Základní fonologická pravidla pro český jazyk jsou uvedena v literatuře [4]. V kapitolách 4.2.1 až 4.2.4 uvádíme jejich přehled. Tato pravidla využívají označení pro fonémy z tabulky č. 1 a zkratky vysvětlené v tabulce č. 2.

Tab. 2. Dělení českých hlásek dle [4]

Samohlásky (SA)	a, á, e, é, i, í, o, ó, u, ú									
Znělé párové souhlásky (ZPS)	b	d	d'	g	z	ž	v	h	dz (C)	dž (Č)
Neznělé párové souhlásky (NPS)	p	t	t'	k	s	š	f	ch (X)	c	č
Jedinečné souhlásky (znělé) (JS)	m, n, ň, l, j, r, ř									

4.2.1 Základní přepisovací pravidla pro česká slova

Pro české *ch* (pozůstatek spřažkového pravopisu) se používá označení [X]

$$ch \rightarrow X / _ \quad (1)$$

České *ů* přepisujeme jako [ú]

$$ů \rightarrow ú / _ \quad (2)$$

Písmeno *w* přepisujeme na [v]

$$w \rightarrow v / _ \quad (3)$$

Písmeno *q* se přepisuje na [kv]

$$q \rightarrow kv / _ \quad (4)$$

Samohlásky *y/ý* přepisujeme na [i/í]

$$y \rightarrow i / _ \quad (5)$$

$$ý \rightarrow í / _ \quad (6)$$

Následuje-li *ě* po *b, p, f, v*, přepisuje se na [je]

$$ě \rightarrow je / <b, p, f, v> _ \quad (7)$$

Spojení *dě, tě, ně, mě* přepisujeme na [d'e], [t'e], [ňe], [m'ne]

$$dě \rightarrow d'e / _ \quad (8)$$

$$tě \rightarrow t'e / _ \quad (9)$$

$$ně \rightarrow ňe / _ \quad (10)$$

$$ě \rightarrow ňe / m_ \quad (11)$$

Spojení *di, ti, ni* přepisujeme na [d'i], [t'i], [ňi]

$$d \rightarrow d' / _ <i, í> \quad (12)$$

$$t \rightarrow t' / _ <i, í> \quad (13)$$

$$n \rightarrow ň / _ <i, í> \quad (14)$$

Jestliže x stojí před znělou souhláskou, přepisuje se na $[gz]$, jestliže stojí před neznělou souhláskou či na konci slova, přepisuje se na $[ks]$

$$x \rightarrow gz / _ < \text{ZPS, JS} > \quad (15)$$

$$x \rightarrow ks / _ < \text{NPS, -} > \quad (16)$$

Písmeno x přepisujeme na $[ks]$ na počátku slova před samohláskou a mezi samohláskami

$$x \rightarrow ks / - _ \text{SA} \quad (17)$$

$$x \rightarrow ks / \text{SA}_1 _ \text{SA}_2 \quad (18)$$

Z pravidla (18) existuje výjimka. Jestliže na počátku slova je dvojice ex a následuje-li po ní samohláska, přepisuje se na $[egz]$

$$ex \rightarrow egz / - _ \text{SA} \quad (19)$$

4.2.2 Spodoba znělosti

Spodoby znělosti se zúčastňují jen souhlásky ze skupiny souhlásek párových. Spojení takovýchto dvou souhlásek je buď celé znělé, nebo neznělé podle toho, je-li poslední souhláska znělá či neznělá. Abychom mohli formalizovat zápis přepisovacího pravidla označíme \neg ZPS jako neznělý protějšek ke znělé souhlásce ZPS, tj. $\neg b = p$, $\neg d = t$, $\neg d' = t'$, $\neg g = k$, $\neg v = f$, $\neg z = s$, $\neg \check{z} = \check{s}$, $\neg h = ch$, $\neg C = c$, $\neg \check{C} = \check{c}$. Obdobně označíme \neg NPS znělým protějškem k neznělé souhlásce NPS. Jestliže půjde při přepisu o stejnou souhlásku, přidělíme jí stejný index, různý index však nemusí znamenat, že jde o jinou souhlásku.

Základní pravidla spodoby znělosti pak mají tvar

$$\text{ZPS}_1 \rightarrow \neg \text{ZPS}_1 / _ < -, \text{NPS, ZPS}_2 - > \quad (20)$$

$$\text{NPS}_1 \rightarrow \neg \text{NPS}_1 / _ \text{ZPS} \quad (21)$$

Spodoba znělosti se může určitým způsobem projevit i přes hranice slov. Její účinek může nastat jen při plynulém vyslovení příslušného slovního spojení (bez zřetelné pauzy mezi slovy). Platí zde základní pravidlo, že znělá souhláska ztrácí na konci slova znělost a může ji nabýt jen tehdy, když po ní následuje v počátku následujícího slova znělá souhláska párová

$$\text{ZPS}_1 \rightarrow \text{ZPS}_1 / _ - \text{ZPS}_2 \quad (22)$$

$$\text{ZPS}_1 \rightarrow \neg \text{ZPS}_1 / _ < - \text{NPS, - SA, - JS} > \quad (23)$$

Obdobné pravidlo platí pro neznělou párovou souhlásku na konci slova, začíná-li následující slovo znělou párovou souhláskou a nebo souhláskou \check{r}

$$\text{NPS}_1 \rightarrow \neg \text{NPS}_1 / _ < - \text{ZPS, - } \check{r} > \quad (24)$$

Z pravidel o spodobě znělosti (20) až (24) existují tyto výjimky:

Vícehláskové předložky zakončené znělou párovou souhláskou (*před, pod, nad, bez, ob, od*) si zachovávají svou znělost, začíná-li následující slovo znělou párovou nebo jedinečnou souhláskou

$$\text{ZPS}_1 \rightarrow \text{ZPS}_1 / _ < - \text{ZPS}_2, - \text{JS} > \quad (25)$$

Předložky z , v před znělou párovou či jedinečnou souhláskou zůstávají znělé

$$z \rightarrow z / _ < - \text{ZPS, - JS} > \quad (26)$$

$$v \rightarrow v / _ < - \text{ZPS, - JS} > \quad (27)$$

Předložka k se před nepárovými souhláskami a souhláskou v nemění v souhlásku znělou

$$k \rightarrow k / - _ < - \text{JS}, - v > \quad (28)$$

Znělá souhláska *v* se spodobuje, ale sama spodobu nezpůsobuje

$$v \rightarrow f / _ \text{NPS} \quad (29)$$

$$\text{NPS} \rightarrow \text{NPS} / _ v \quad (30)$$

Spojení souhlásek *s* a *h* se spodobuje podle uvedeného základního pravidla (21) jestliže mezi *s-h* je zřetelný prefixový šev ↓

$$s \rightarrow z / _ \downarrow h \quad (31)$$

V ostatních případech se ve spojení *sh* může postupovat podle základního pravidla (zejména ve výslovnosti na Moravě)

$$s \rightarrow z / _ h \quad (32)$$

a nebo dochází k tzv. postupné asimilaci (zejména v Čechách)

$$h \rightarrow X / s _ \quad (33)$$

Souhláska *ř* se v postavení před souhláskou párovou řídí základním pravidlem o spodobě znělosti

$$\check{r} \rightarrow \check{r} / _ \text{ZPS} \quad (34)$$

$$\check{r} \rightarrow \check{R} / _ < -, \text{NPS}, \text{ZPS} - > \quad (35)$$

V postavení po párové souhlásce podléhá *ř* postupné asimilaci

$$\check{r} \rightarrow \check{r} / \text{ZPS} _ \quad (36)$$

$$\check{r} \rightarrow \check{R} / \text{NPS} _ \quad (37)$$

4.2.3 Spodoba artikulační

Při spojení dvou souhlásek s rozdílnou artikulací se vyrovnává rozdíl mezi jejich výslovností artikulační spodobou. Rozeznáváme přitom asimilaci postupnou (předcházející souhláska ovlivňuje následující) a asimilaci zpětnou (následující souhláska ovlivňuje souhlásku předcházející).

Artikulační spodobu lze pozorovat u mnoha spojení souhlásek. Podle toho, jak těsnou navrhujeme fonetickou transkripci, tj. jak přesně chceme přepsat či zaznamenat vyslovovanou řeč, tak bohatá budou přepisovací pravidla. Pokusíme se nyní definovat nejdůležitější z těchto pravidel.

Jestliže souhláska *n* stojí před okluzívy *k* nebo *g*, spodobuje se v [N]

$$n \rightarrow N / _ < k, g > \quad (38)$$

Jestliže nazála *m* stojí před frikativami *v* nebo *f*, spodobuje se v [M]

$$m \rightarrow M / _ < v, f > \quad (39)$$

Jestliže nazála *n* stojí před okluzívy *t'*, *d'*, spodobuje se často v [ň]

$$n \rightarrow \check{n} / _ < t', d' > \quad (40)$$

Jestliže nazála *ň* stojí za souhláskami *d*, *t*, dochází k jejich spodobě na [d'], [t']

$$d \rightarrow d' / _ \check{n} \quad (41)$$

$$t \rightarrow t' / _ \check{n} \quad (42)$$

Připouští se (tj. může i nemusí být) zjednodušená výslovnost závěrových souhlásek *t, d* ve spojení s úžinovými *s, z, š, ž*. Tato výslovnost může být realizována pomocí polouzávěrových protějšků [c], [č], popřípadě [C], [Č]

$$ts \rightarrow c / _ \quad (43)$$

$$tš \rightarrow č / _ \quad (44)$$

$$ds \rightarrow c / _ \quad (45)$$

$$dš \rightarrow č / _ \quad (46)$$

$$dz \rightarrow C / _ \quad (47)$$

$$dž \rightarrow Č / _ \quad (48)$$

V případě prefixového či mezislovního švu ve spojení *ts, tš, ds, dš, dz, dž* se dává přednost zachování výslovnosti obou souhlásek. Výslovnost obou souhlásek se zachová i na hranici předložek a jmen. Přitom ve spojení *ds, dš* se obvykle uplatní pravidlo o spodobě znělosti.

Dvě stejné souhlásky *cc, čč, dd, jj, kk, ll, nn, mm, ss, šš, tt, zz*, které se nacházejí na prefixovém či sufixovém švu (což je naprostá většina případů), se při vyslovení převážně redukuje na souhlásku jedinou. Pouze chce-li řečník zdůraznit šev, nebo se zdvojená souhláska nachází na mezislovním švu, vysloví souhlásku zdvojenou. Podle tohoto pravidla lze postupovat i v případě, kdy vedle sebe stojí dvojice hlásek *NPS → NPS* nebo *ZPS → ZPS*. Většinou zde dochází k spodobě znělosti a vedle sebe jsou dvě stejné znělé či neznělé párové souhlásky, u kterých může, ale nemusí dojít ke splynutí v hlásku jedinou.

4.2.4 Slovní přízvuk a ráz

Důležitou složkou fonetické transkripce je i umístění slovního přízvuku a rázu. Protože korektní popis všech pravidel a všech výjimek k těmto pravidlům je rozsáhlý, jsou zde uvedeny jen podstatné znalosti nezformulované do pravidel. Ráz se v češtině vytváří automaticky a pravidelně po každé delší pauze, pokud další promluva začíná samohláskou. Spisovná výslovnost vyžaduje užití rázu i po neslabičných předložkách *v, z, s* a *k*, začíná-li následující slovo samohláskou. Slovní přízvuk je v mluvené češtině v zásadě vázán na první slabiku přízvukového taktu. Přízvukový takt je úsek promluvy s jedním přízvukovým vrcholem. Přízvukový takt má v češtině tyto vlastnosti:

- Přízvučná slabika je zpravidla první slabikou přízvukového taktu.
- V neutrální promluvě leží hranice taktů v místě hranic slov.
- Přízvukový takt může obsahovat několik slov.
- Hranice mezi jednotlivými přízvukovými takty je obvykle charakterizována kontrastem stupně přízvučnosti (nepřízvučná – přízvučná).

Některá dílčí upřesnění a časté výjimky:

- Původní předložky jednoslabičné, jako je *bez, na, do, ke, o, od, pod, po, přes, u* apod., přejímají obvykle přízvuk následujícího slova a tvoří s ním jeden takt.
- Předložky nepůvodní, např. *blíž, dle, kol, krom, skrz* apod., přízvuk následujícího slova nepřijímají.
- Některá slova, obvykle jednoslabičná, přízvuk nemají a vytvářejí se slovem předcházejícím jednoslabičný takt.

- Některá slova nemají přízvuk zcela pravidelně a jejich základní podoba je nepřívětivá. Tato slova se nazývají příklonkami a jsou to například zájmena *se, si, mně, mi, ho, mu*, částice *-li* apod.

U výslovnosti několika frekventovaných slov se připouští i její nespisovná varianta:

jsem [jsem] i [sem], jsi [jsi] i [si], jste [jste] i [ste], jsme [jsme] i [sme], sedm [sedm] i [sedum], osm [osm] i [osum] apod.

Nepřípustná nespisovná výslovnost se může vyskytnout například u slov:

kdyby [dibi], když [diš], vždycky [dicki], švestka [šveska], jedenáct [jedenác] apod.

4.2.5 Realizace automatické fonetické transkripce založené na pravidlech

Algoritmus využitelný pro naprogramování fonetické transkripce se dá zjednodušeně popsat dle [4] takto:

- daný text se zpracovává znak po znaku zleva doprava;
- u každého znaku se nejprve zjišťuje, zda u něj nelze uplatnit nějakou výjimku; pokud ano, výjimka se prioritně uplatní;
- pokud nelze na daný text uplatnit některou z výjimek, aplikuje se vhodné základní pravidlo;
- jestliže nelze na daný znak (písmeno) aplikovat žádnou z výjimek ani žádné z pravidel, znak se jednoduše opíše do vytvářeného fonetického řetězce.

První program pro fonetickou transkripci používaný v Laboratoři počítačového zpracování řeči TUL vyvinul ing. Volejník a popisuje jej ve své diplomové práci [7] takto: Program v přepisovaném textu nejprve nalezne větu, tj. text oddělený na začátku a na konci některým z interpunkčních znamének. Ve větě algoritmus programu vyhledává slova vyskytující se v souboru výjimek, do kterého jsou zařazena cizí slova mající odlišnou výslovnost, než by odpovídalo již zmíněným fonologickým pravidlům, a také jména, kterých je příliš mnoho a tudíž v souboru nejsou všechna. V případě nalezení slova - výjimky si zapamatuje jeho pozici ve větě a pořadové číslo slova v souboru výjimek a poté zavolá funkci, která provádí vlastní fonetickou transkripci. Provede se sekvenčním způsobem písmeno po písmenu fonetická transkripce a poté jsou vyměněna slova nalezená jako výjimky na začátku algoritmu. Tímto způsobem se postupně převede celý text.

Míra úspěšnosti transkripce tímto programem se blíží 100 %. Vyskytne-li se v transkripci chyba, lze ji většinou odstranit přidáním nějaké výjimky. Soubor výjimek je tabulka, jejíž 1. sloupec je úsek slova, 2. sloupec je jeho transkripce podle pravidel a 3. sloupec je transkripce platící pro výjimku. Chceme-li dosáhnout 100% úspěšnosti, musíme do tohoto souboru zadávat úseky slov co nejspeciálnější, což znamená, že záznamů zde bude více než v případě, kdybychom volili kratší univerzálnější úseky slov. Například pokud jsou v souboru výjimek tyto záznamy:

*fini fiňi fini,
manit maňit manit,
moni moňi moni,*

program pro fonetickou transkripci nám dá tyto výsledky:

finiš je správně přepsáno na [finiš], ale chybně je přepsáno *Fini* na [fini].

humanitární je správně přepsáno na [humanitární], ale chybně je přepsáno *vymanit* na [vimanit].

Monika je správně přepsáno na [monika], ale chybně je přepsáno *Mormoni* na [mormoni].

V češtině je mnohdy nutné výjimky realizovat jako výčet všech gramatických tvarů slova s výjimečnou fonetickou transkripcí. Někdy však ani to nepomůže a potom je nutné zvolit pravidlo, které je úspěšné ve většině případů. Například slovo *unikáte* je jednak 6. pád slova *unikát* a zároveň tvar slovesa *unikat*. V tomto případě je jednoznačně používanější 2. varianta a proto je nutné pravidla nastavit tak, aby slovo *unikáte* bylo přepisováno jako [uňikáte] [2], viz též kapitola 2.4 o homografech.

4.3 Metody strojového učení využívané pro fonetickou transkripci

Společným prvkem metod strojového učení taktéž nazývaných **statistické metody** je nutnost sběru textového korpusu opatřeného správnou fonetickou transkripcí, což je spojeno se značným objemem manuální práce. Rozhodnutí se zda zvolit metodu založenou na pravidlech nebo nějakou metodu strojového učení závisí na počtu pravidel pro fonetickou transkripci platících pro daný jazyk. Například fonetická pravidla pro češtinu jsou relativně jednoduchá, protože ve většině případů jednotlivé znaky textu, takzvané grafémy, korespondují s jednotlivými fonémy. U jazyků se složitějšími pravidly, jako je například angličtina, stojí za to uvažovat možnosti použití metod strojového učení.

4.3.1 Neuronová síť

V 80. letech došlo na vědeckých pracovištích zabývajících se umělou inteligencí k obnovení zájmu o v průběhu 70. let podceňovaný obor zvaný neuronové sítě, který se rozvíjel od začátku 50. let. Jedním z produktů této doby byl experiment s neuronovou sítí provedený Američany Terrencem Sejnowskim z Johns Hopkins University a Charlesem Rosenbergem z Princetonu publikovaný v článcích [8] a [9]. Jejich síť NETtalk četla tisíce příkladů anglických slov, pomocí řečového syntezátoru je vyslovovala a výsledek své fonetické transkripce korigovala podle správné fonetické transkripce, kterou měla spolu se slovy k dispozici. Úspěšnost fonetické transkripce sítě se postupně zlepšovala až na přibližně 90 % správně přepsaných dvojic písmeno-foném. Nakonec síť dokázala správně vyslovovat i některá slova, u kterých neměla k dispozici jejich správnou fonetickou transkripci, což znamená, že zobecnila dříve naučená pravidla z jedněch příkladů na jim podobné. Perličkou bylo, že syntetizovaná řeč sítě NETtalk se zpočátku podobala žvatlání malého dítěte.

Ing. Nejedlová provedla velmi podobný pokus pro případ fonetické transkripce češtiny, který zdokumentovala ve výzkumné zprávě [10]. Naprogramovala třívrstvou neuronovou síť učící se podle algoritmu **back-propagation**. Na rozdíl od sítě NETtalk byl výstup této sítě pouze textový. Neuronová síť pro fonetickou transkripci češtiny přepisovala izolovaná slova. Každé slovo četla pomocí pomyslného 5-ti znakového okénka, kterým postupně procházel řetězec písmen tohoto slova, a snažila se uhodnout foném patřící písmenu, které se právě nacházelo uprostřed. Po každém odhadu síť následovalo přepočítání vah podle správné hodnoty fonému, který tvořil druhou část vstupu, algoritmem back-propagation.

Po několika pokusech s touto sítí bylo zřejmé, že úspěšnost jejích předpovědí nikdy nepředčí výkon programu Ing. Volejníka. Síť byla schopná se naučit statisticky nejčastější fonologická pravidla, ale některá statisticky významná pravidla v jejím repertoáru vždy chyběla. Maximální počet pravidel, která se síť kdy naučila, bylo 86. Naprosto nemožné by bylo chtít, aby se neuronová síť naučila výjimky.

4.3.2 Rozhodovací stromy

Rozhodovací stromy dokážou z příkladů odvodit sadu pravidel, které mohou mít formu produkčních pravidel z kapitoly 4.2. Jejich použití v oblasti fonetické transkripce popisují například články [11] a [12].

Rozhodovací stromy pracují tak, že všechna slova testují sadou regulárních výrazů. Podle toho, zda slovo regulárnímu výrazu vyhovuje či nevyhovuje, je zařazeno do jedné či druhé větve rozhodovacího stromu. Vzniklý strom má větve odpovídající fonologickým pravidlům.

Nevýhodou rozhodovacích stromů je velká výpočetní náročnost a velké množství vzniklých pravidel, které se musí následně čistit od pravidel vzniklých chybami v trénovacích datech a výjimkami.

4.3.3 Konečné automaty

Konečné automaty také dokážou z příkladů odvodit sadu pravidel, které jsou ve formě grafu s přechody mezi uzly popisujícími, jak se má sekvence grafémů přepsat na sekvenci fonémů. Jejich použití v oblasti fonetické transkripce popisuje například článek [13].

Jejich vstupem je trénovací množina dvojic párů slov a jejich fonetických transkripcí. Algoritmus začne konstrukcí převodníku ve formě stromu, který pokrývá všechny trénovací vzorky. Kořen stromu je počáteční stav převodníku a každý list stromu koresponduje s koncem vzorku. Výstupní symboly – fonémy jsou umísťovány co nejbližší ke kořeni stromu tak, aby nevznikaly konflikty ve výstupu dané hrany grafu. [13]

Nevýhodou konečných automatů dle [13] je malá zobecňovací schopnost a vznik lingvisticky nepřirozených pravidel, kdy například jedna hrana stromu vymaže sekvenci fonémů a následující hrana tu samou sekvenci vloží, takže je nutné vyvíjet metody pro zdokonalování vzniklých pravidel.

4.3.4 Genetické algoritmy

Genetické algoritmy se ve fonetické transkripci používají jednak pro redukci pravidel vzniklých například při použití rozhodovacích stromů [11], jednak pro hledání nových pravidel, která se dříve řešila výjimkami [14], [15].

V článku [14] se hledala nová fonologická pravidla pro přepis slabik *di*, *ti*, *ni* v českých slovech, které se ve slovech s českým původem přepisují většinou jako *[dʲi]*, *[tʲi]*, *[nʲi]* a v přejatých slovech jako *[di]*, *[ti]*, *[ni]*.

Genetický algoritmus v článku [14] reprezentoval fonologická pravidla jako jedince žijící v prostředí slov a jejich správných fonetických transkripcí. Fonologická pravidla $A \rightarrow B / C _ D$ z kapitoly 4.2 mají v genetickém algoritmu stabilní části *A* a *B*, tedy slabiky *di*, *ti*, *ni*, a variabilní, mutující části *C* a *D*, tedy okolní grafémy slabik *di*, *ti*, *ni*. Každé nové pravidlo, které je výsledkem mutace, je individuálně testováno na slovech v trénovací množině. Pokud se testované pravidlo dopustí méně chyb na testovacích datech než původní sada pravidel, tak je vybráno do další generace jedinců genetického algoritmu nebo jako konečné nové pravidlo, když je po určitém počtu generací genetický algoritmus ukončen. V průběhu genetického algoritmu jsou do dalších generací vybírána také některá pravidla, která kazí transkripce některých slov.

5. ZÁVĚR

Fonetickou transkripci českého textu lze pravděpodobně nejlépe realizovat pomocí lidmi sepsaných pravidel a výjimek. Metody strojového učení jsou využitelné například pro objevování nových pravidel ze statisticky významných výjimek. Dále je třeba vyvíjet algoritmy, které zvýší konzistenci sady pravidel a výjimek, protože při jejich vysokém počtu nelze vyloučit, že některá pravidla a výjimky jsou přebytné a tudíž jen zbytečně zpomalují systém pro fonetickou transkripci. Díky flektivnímu charakteru českého jazyka se počítačové metody zpracování neobejdou bez velkého objemu manuální práce spočívající například v anotaci českých textů.

Literatura:

- [1] Palková, Z.: Fonetika a fonologie češtiny. Karolinum. Upravené vydání – dotisk. Praha 1997. ISBN 80-7066-843-1.
- [2] Nejedlová, D.: Creation of Lexicons and Language Models for Automatic Broadcast News Transcription. [Ph.D. Thesis]. Liberec 2006. Technická univerzita v Liberci. Fakulta mechatroniky a mezioborových inženýrských studií.
- [3] Jurafsky, D., Martin, J. H.: Speech and Language Processing. Prentice Hall, Inc., New Jersey, 2000, ISBN 0-13-095069-6.
- [4] Psutka, J.: Komunikace s počítačem mluvenou řečí. Academia. Praha 1995. ISBN 80-200-0203-0.
- [5] The International Phonetic Alphabet. Journal of the Phonetic Association, vol. 19, no. 12, December 1989.
- [6] Nouza, J., Psutka, J., Uhlíř, J.: Phonetic Alphabet for Speech Recognition of Czech. In: Radio Engineering, vol. 6, no. 4, December 1997, pp. 16-20.
- [7] Volejník, M.: Fonetická transkripce psané a mluvené češtiny pro účely automatického zpracování řeči. [Diplomová práce]. Liberec 1999. Technická univerzita v Liberci. Fakulta mechatroniky a mezioborových inženýrských studií.
- [8] Sejnowski, T. J., Rosenberg, C. R.: NETtalk: a Parallel Network That Learns to Read Aloud. In: Cognitive Science, 14, 1986, pp. 179-211.
- [9] Sejnowski, T. J., Rosenberg, C. R.: Parallel Networks That Learn to Pronounce English Text. In: Complex Systems, 1, 1987, pp. 145-168.
- [10] Nejedlová, D.: Fonetická transkripce češtiny pomocí třívrstvé neuronové sítě. [Výzkumná zpráva]. Liberec 2000. Technická univerzita v Liberci. Fakulta mechatroniky a mezioborových inženýrských studií.
- [11] Zelinka, J., Müller, L.: Automatic General Letter-to-Sound Rules Generation for German Text-to-Speech System. In: Text, Speech and Dialogue, Lecture Notes in Artificial Intelligence, 3206, Springer, 2004, pp. 537-543.
- [12] Wolters, M., van den Bosh, A.: Automatic Phonetic Transcription of Words Based on Sparse Data. In: Workshop Notes of the ECML/MLnet Workshop on Empirical Learning of Natural Language Processing Tasks, Prague, Czech Republic (1997), pp. 61-70.
- [13] Gildea, D., Jurafsky, D.: Automatic induction of finite state transducers for simple phonological rules. In: Proceedings of the Meeting of the Association for Computational Linguistics, pages 9-15, 1995.
- [14] Kolorenč, J.: Evolving Phonological Rules Using Grammatical Evolution. In: POSTER 2004 [CD-ROM]. ČVUT FEL, Prague 2004.
- [15] Kolorenč, J.: Tvorba a adaptace lingvistické vrstvy pro systém rozpoznávání mluvené češtiny. [Disertační práce]. Liberec 2008. Technická univerzita v Liberci. Fakulta mechatroniky a mezioborových inženýrských studií.