Razpoznavanje tekočega slovenskega govora z velikim slovarjem

Janez Kaiser, Mirjam Sepesy Maučec, Zdravko Kačič in Bogomir Horvat

Laboratorij za digitalno procesiranje signalov,
Fakulteta za elektrotehniko, računalništvo in informatiko,
Univerza v Mariboru
Smetanova 17, 2000 Maribor, Slovenija
e-mail: *janez.kaiser@uni-mb.si*

Abstract

V pričujočem članku bomo predstavili zasnovo sistema za razpoznavanje tekočega govora z velikim slovarjem v slovenščini. Predstavili bomo potrebno infrastrukturo za izdelavo avtomatskega razpoznavalnika tekočega jezika in podrobneje predstavili težave, na katere naletimo pri izgradnji takega sistema zaradi specifičnih lastnosti slovenskega jezika. Predstavili bomo zgradbo in prve dosežene rezultate eksperimentalnega razpoznavalnika, ki ga gradimo v našem laboratoriju.

Abstract

In this paper, we present the structure of a Slovenian large-vocabulary, continuous-speech recogniser. We present the needed infrastructure for the development of such a recogniser and difficulties, which are caused by specific properties of Slovenian language. We also present the first results with the experimental recogniser for Slovenian language, which is being developed at our laboratory.

1. Uvod

Govor je človeku najbolj naraven način komuniciranja, zato se je pojavila težnja, da bi ga lahko uporabljali tudi pri sporazumevanju človeka s strojem. Ta ideja je zadnjem času v svetu povzročila silovit razvoj jezikovnih in govornih tehnologij, ki vključujejo sisteme avtomatskega razpoznavanja, razumevanja in sinteze govora.

V svetu so se prvi sistemi, ki so omogočali razpoznavanje tekočega govora z velikim slovarjem, pojavili ob koncu osemdesetih let, najprej za angleški jezik, nato pa tudi za druge, kot so nemščina, francoščina ali kitajščina. V zadnjih nekaj letih so najboljši razpoznavalniki dosegli raven uspešnosti razpoznavanja, ki omogoča njihovo uporabo v komercialnih izdelkih, kot so narekovalniki in telefonski sistemi za poizvedovanje.

Razvoj razpoznavalnikov tekočega govora v slovenščini na žalost močno zaostaja za dosežki za „velike" jezike. Edini nam znan in v literaturi dokumentirani razpoznavalnik slovenskega tekočega govora so izgradili na Fakulteti za elektrotehniko v Ljubljani (Ipšić et al., 1999) kot del sistema za poizvedovanje o letalskih letih. Razpoznavalnik ima slovar približno 800 besed in je omejen na omenjeno domeno.

Z namenom, da bi izpolnili praznino, ki zeva na področju avtomatskega razpoznavanja tekočega slovenskega govora, smo se v našem laboratoriju odločili izdelati ustrezen razpoznavalnik z slovarjem nekaj deset tisoč besed, ki bi bil zmožen razpoznavati vsebinsko neomejen tekoč slovenski govor. Poudariti velja, da predstavlja vsaka besedna oblika samostojno besedo v slovarju, kar pri pregibnem slovenskem jeziku še vedno predstavlja pereč problem.

V prvem delu članka bomo predstavili zasnovo avtomatskega razpoznavalnika govora, ki smo jo izbrali za naš sistem. Zasnova temelji na statističnih modelih, kar je danes praktično edini uporabljan način izvedbe razpoznavanja. V drugem delu bomo najprej opisali težave, na katere naletimo pri izdelavi statističnih jezikovnih modelov za slovenščino. Nato bomo predstavili praktično implementacijo eksperimentalnega razpoznavalnika, ki ga razvijamo v našem laboratoriju in prve meritve uspešnosti razpoznavanja, ki smo jih dosegli.

2. Statistično razpoznavanje tekočega govora

Pri razpoznavanju govora želimo tekoč govor samodejno pretvoriti v tekstovno obliko. To nalogo opravlja razpoznavalnik, ki govorni signal pretvori v zaporedje besed. Zaporedje besed označimo z vektorjem $W = [w_1, w_2, ..., w_n]$, kjer n označuje število besed v zaporedju. Razpoznavalnik najprej z akustično analizo govorni signal pretvori v vektor značilk $Y = [y_1, y_2, ..., y_m]$. Če s $P(W|Y)$ označimo verjetnost, da je bil izgovorjen niz besed W ob znanih značilkah Y, lahko odločitev razpoznavalnika predstavimo s formulo (Duda and Hart, 1973):

$$\hat{W} = \arg\max_{W} P(W|Y). \qquad (1)$$

Razpoznavalnik izbere niz besed $\hat{W}$, ki je najbolj verjeten pri danem zaporedju značilk Y, ki smo ga izluščili iz govornega signala. Predvsem pri razpoznavanju tekočega govora akustične značilke izgovorjenih besed ne določajo dovolj zanesljivo, zato je pogojno verjetnost $P(W|A)$ potrebno razčleniti. Če v enačbi (1) uporabimo Bayesovo pravilo, lahko zapišemo

$$\hat{W} = \arg\max_{W} \frac{P(Y|W)\,P(W)}{P(Y)}. \qquad (2)$$

Verjetnost $P(Y)$ na odločitev razpoznavalnika ne vpliva, zato jo lahko izločimo. $P(Y|W)$ je verjetnost,

da smo zaznali zaporedje značilk Y, če je bil izgovorjen niz besed W. Te verjetnosti so določene z akustičnim modelom razpoznavalnika. Če bi značilke dovolj dobro določale izgovorjene besede, bi lahko iz formule (2) izločili tudi verjetnost $P(W)$. Ko si značilke različnih besed postanejo zelo podobne in začne razpoznavalnik z odločitvenim pravilom $P(Y|W)$ izbirati napačne besede, lahko z vpeljavo verjetnosti $P(W)$ odpravimo te pomanjkljivosti. $P(W)$ določa verjetnost, da je bil izgovorjen niz besed W. Te verjetnosti so določene z jezikovnim modelom razpoznavalnika. Opisano zgradbo razpoznavalnika prikazuje slika 1.

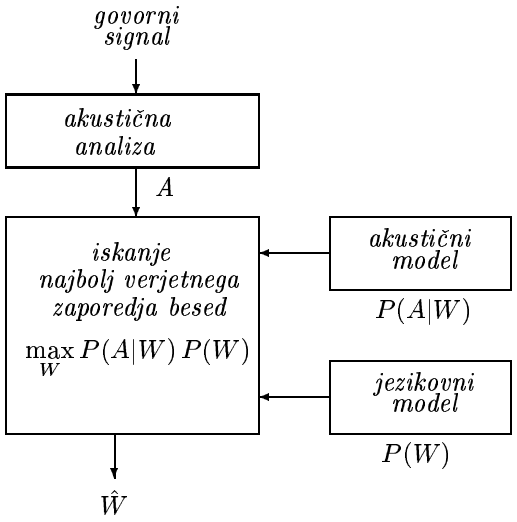

Figure 1: Model razpoznavalnika govora

2.1. Akustična analiza

Osnovna zahteva za akustično analizo za razpoznavanje govora je izločiti informacije o fonetični identiteti danega govora, hkrati pa mora biti analiza neobčutljiva na faktorje, kot so razlike med govorci in efekti zaradi različnih komunikacijskih kanalov.

Osnovna predpostavka, na kateri temelji običajna akustična analiza je, da lahko govor obravnavamo kot lokalno stacionaren. Zato lahko govor najprej razrežemo na kratka okna (običajno približno 20 ms), nad katerimi nato izvajamo nadaljnjo analizo.

V današnjih razpoznavalnikih je akustična analiza usmerjena na ugotavljanje lastnosti govora, ki so pogojene z obliko vokalnega trakta. Najpogostejši danes uporabljani zvrsti akustične analize sta LPC (*Linear Prediction Coefficients*) analiza in kepstralna analiza (Picone, 1993).

LPC analiza modelira vokalni trakt kot akustično cev brez izgub in vejitev. S tako predpostavko je mogoče vokalni trakt modelirati kot filter s samimi poli. Metode, kot je avtoregresijsko modeliranje, prilagodijo parametre filtra značilnostim govornega signala, te parametre pa nato uporabimo kot značilke za razpoznavanje.

Pri kepstralni analizi obravnavamo govor kot vzbujanje, ki je modulirano pri prehodu skozi vokalni trakt. Kepstralna analiza nam daje možnost, da ločimo vzbujanje od karakteristike vokalnega trakta. To dosežemo z logaritmiranjem preslikave govora v frekvenčnem prostoru. Posebna prednost kepstralne analize je v nekoreliranosti kepstralnih koeficientov, kar poenostavi akustično modeliranje.

2.2. Akustični model

Akustični model opisuje verjetnost poljubnega zaporedja akustičnih značilk Y pri danem nizu besed W. Danes so najpogosteje uporabljani akustični modeli, zasnovani na prikritih modelih Markova (HMM – *Hidden Markov Models*). HMM so bili uspešno uporabljeni v najrazličnejših tipih razpoznavalnikov, od preprostih za razpoznavanje malega števila osamljenih besed, do najkompleksnejših za razpoznavanje pogovornega jezika.

HMM so v osnovi končni avtomati, ki lahko generirajo zaporedja značilk. Ob vsaki časovni enoti t HMM preide iz stanja i v stanje j z verjetnostjo a_{ij} in pri tem generira značilko y_t z verjetnostjo $b_j(y_t)$. Primer strukture HMM, ki jo uporabljamo za modeliranje fonemov, prikazuje slika 2.

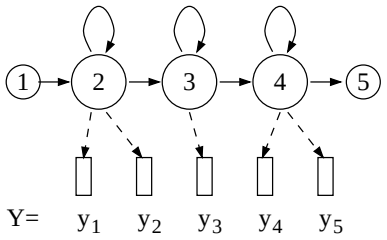

Figure 2: Primer HMM za fonem s tremi stanji, ki lahko generirajo značilke.

Skupna verjetnost zaporedja značilk Y in danega zaporedja stanj X danega modela M je tako:

$$P(Y, X|M) = a_{x(0)x(1)} \prod_{t=1}^{T} b_{x(t)}(y_t) a_{x(t)x(t+1)} \qquad (3)$$

V praksi je znano le zaporedje značilk Y, ne pa tudi zaporedje stanj X (od tod ime prikriti modeli Markova). Vseeno pa lahko pri modelu M danega niza besed W zahtevano verjetnost $P(Y|W) = P(Y|M)$ izračunamo kot vsoto enačbe (3) po vseh možnih poteh:

$$P(Y|W) = P(Y|M) = \sum_{\mathcal{X}} P(Y, X|M) \qquad (4)$$

Za izračun $P(Y|M)$ se v praksi uporabljata dve metodi: metoda naprej-nazaj in Viterbijeva metoda. Metoda naprej-nazaj daje točno vrednost $P(Y|M)$, hkrati pa je tudi osnova za Baum-Welchev algoritem, ki se uporablja za določanje parametrov HMM. Viterbijeva metoda daje približek prave vrednosti, tako da najde le najbolj verjetno zaporedje stanj v enačbi (4). Viterbijevo metodo uporablja večina razpoznavalnikov, saj je hitrejša in lažje praktično izvedljiva od metode naprej-nazaj.

Za zasnovo razpoznavalnika je ključnega pomena izbira oblike verjetnostne porazdelitve $b_j(y_t)$. Najpogosteje se uporabljajo mešanice Gaussovih porazdelitev:

$$b_j(y_t) = \sum_{m=1}^{M} \mathcal{N}(y_t; \mu_{jm}, \Sigma_{jm}) \quad (5)$$

Teoretično bi lahko izdelali HMM za poljubne enote govora – cele stavke, besede ali podbesedne enote kot so fonemi. Daljše kot so enote, bolj precizno je modeliranje, hkrati pa je težje dobro določiti parametre HMM. Tako se pri razpoznavanju tekočega govora z velikim slovarjem skoraj izključno uporabljajo samo podbesedne enote, najpogosteje fonemi. Za doseganje visoke uspešnosti razpoznavanja pa ne modeliramo samo nekaj nad 30 osnovnih fonemov, ampak upoštevamo tudi močno kontekstno odvisnost fonemov – fonemi so različno izgovorjeni glede na okoliške foneme. Najprimernejši način za upoštevanje kontekstnih pojavov je uporaba trifonov, kjer upoštevamo vpliv predhodnega in sledečega fonema. Tako bi stavek ,,Krava Liska", ki ga fonetično zapišemo

sil k r a: v a l i: s k a sil

zapisali s trifoni kot

sil sil-k+r k-r+a: r-a:+v a:-v+a v-a+l a-l+i: l-i:+s i:-s+k s-k+a k-a+sil sil

Tukaj je sil oznaka za tišino.

V praksi uporaba trifonov pomeni, da bi v najslabšem primeru morali modelirati 30^3 trifonov, kar ni ekonomično izvedljivo. Ta problem rešujemo z deljenjem skupnih parametrov med posameznimi trifoni.

2.3. Jezikovni model

Z jezikovnim modelom želimo opisati značilnosti jezika. Razlikujemo dve skupini: slovnične modele in verjetnostne modele.

Slovnične modele oblikujejo predvsem jezikoslovci. Ti modeli so zasnovani na slovnici jezika. Slovnica je predstavljena z množico pravil, ki opisujejo strukturo jezika. Definiranje pravil zahteva od snovalca poglobljeno znanje o jeziku. Slovnični modeli nam omogočajo vpogled v strukturo stavkov in hkrati nudijo podporo pri določanju pomena oz. vsebine stavkov. Kljub dolgoletnim naporom še nikomur ni uspelo sestaviti pravil, ki bi dovolj robustno opisovala dinamiko jezika kot celote. Pri oblikovanju slovnic se običajno omejimo na izbrano strukturo stavkov in njihovo pragmatično vrednost.

Predvsem pri aplikacijah, ki ne zahtevajo bogatih modelov jezika, so se uveljavili verjetnostni modeli (Jelinek, 1997). Verjetnostni modeli določajo verjetnosti segmentov besedila, za razliko od prej omenjenih modelov, ki dajejo le dva odgovora: izbrani segment je ali ni element jezika.

Pri razpoznavanju govora so se uveljavili predvsem verjetnostni modeli, še posebej besedni trigramski modeli.

2.3.1. Verjetnostni model

Verjetnost $P(W)$, da bo izgovorjeno zaporedje W lahko zapišemo kot

$$P(W) = P(w_1, ..., w_n) = \prod_{i=1}^{n} P(w_i|w_1, ..., w_{i-1}), \quad (6)$$

kjer je $P(w_i|w_1, ..., w_{i-1})$ verjetnost, da bo izgovorjena beseda w_i, če so bile predhodno izgovorjene besede $w_1, ..., w_{i-2}$ in w_{i-1}.

Besede w_i so elementi slovarja ϑ, ki določa besedišče jezikovnega modela. Slovar jezikovnega modela in slovar akustičnega modela morata biti enaka.

Zaporedje vseh predhodno izgovorjenih besed običajno omejimo na zadnji dve besedi. Tako pridemo do približka

$$P(W) = \prod_{i=1}^{n} P(w_i|w_{i-2}, w_{i-1}), \quad (7)$$

ki mu pravimo trigramski jezikovni model. Pri bigramskih jezikovnih modelih pa upoštevamo le zadnjo izgovorjeno besedo.

Verjetnosti bi lahko ocenili z relativnimi frekvencami besed in besednih nizov v učnem besedilu. Te ocene se izkažejo kot pomanjkljive (to bomo pokazali v poglavju z eksperimenti), saj jezikovni model, katerega ocene izhajajo iz formule (6), nizom besed W, ki kot podnize vsebujejo besedne trojice, ki se v učnem besedilu niso pojavile, priredi oceno verjetnosti $P(W) = 0$. Razpoznavalnik z odločitvenim kriterijem (2) bo v teh primerih zaveden v napačno odločitev. Izračunane frekvence je nujno na nek način "zgladiti", za kar smo uporabili Good Turingovo oceno verjetnosti:

$$P_{GT} = \frac{r^*}{N}, \quad (8)$$

$$r^* = (r+1)\frac{N_{r+1}}{N_r}, \quad (9)$$

N označuje velikost učnega korpusa, r^* pa glajeno frekvenco. Z N_r smo označili število n-gramov, ki so se v učnem korpusu pojavili r-krat. V naših eksperimentih smo uporabili Good Turingovo glajenje v kombinaciji s sestopanjem (ang. "*Katz's backing-off*") (Katz, 1987):

$$\begin{aligned} P(w_3|w_1, w_2) \quad = \quad & \tilde{P}(w_3|w_1, w_2) + \theta(\tilde{P}(w_3|w_1, w_2)) \\ & \cdot \alpha(w_1, w_2)\, P(w_3|w_2), \end{aligned} \quad (10)$$

kjer je

$$\theta(x) = \begin{cases} 1, & \text{če } x = 0 \\ 0, & \text{sicer.} \end{cases} \quad (11)$$

α označuje normalizacijsko konstanto.

Jezikovne modele neodvisno od ostalih komponent razpoznavalnika ocenimo s perpleksnostjo PP, ki je definirana kot

$$PP = P(w_1, w_2, ..., w_m)^{-\frac{1}{m}}. \quad (12)$$

Jezikovni model je tem boljši, čim manjša je perpleksnost pri enaki velikosti slovarja.

3. Eksperimenti

V tem poglavju bomo predstavili eksperimente, ki smo jih izvedli v okviru prvih poizkusov izdelave razpoznavalnika tekočega slovenskega govora z velikim slovarjem. Posebno pozornost smo posvetili jezikovnemu modelu, ki se je izkazal kot najobčutljivejši del razpoznavalnika.

3.1. Jezikovni model

Eksperimentalne besedne trigramske jezikovne modele za slovenščino smo gradili s korpusom besedil člankov časopisa Večer. Korpus je obsegal 2 milijona besed. Razdelili smo ga na dva dela: učni in testni del. Testni del je predstavljal 10 % celotnega korpusa, ostalo pa učni del. Slovar je vseboval V najbolj pogostih besed v učnem korpusu. Vse besede, ki niso bile v slovarju, smo preslikali v enotni simbol OOV (ang. "*Out Of Vocabulary*"). V učnem delu smo izračunali relativne frekvence besed in besednih nizov dolžine 2 (bigramov) in dolžine 3 (trigramov). Delež testnih trigramov in bigramov, ki so se pojavili že v učnem korpusu prikazuje slika 3.

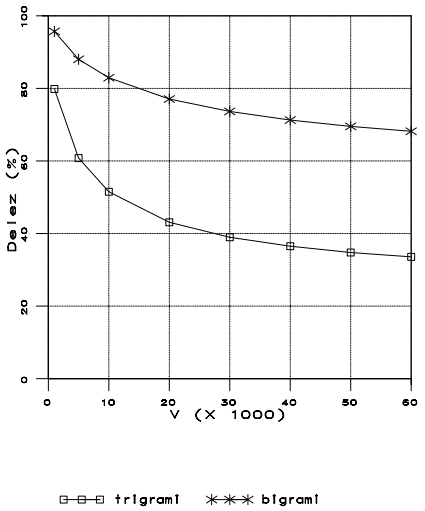

Figure 3: Delež bigramov in trigramov testnega vzorca, ki so se pojavili že v učnem vzorcu.

Pri velikem slovarju (V = 60.000) se več kot 60 % testnih trigramov v učnem korpusu ni pojavila niti enkrat. S še tako majhnim slovarjem (V = 1.000) nismo dosegli zadovoljive pokritosti testnega vzorca, saj je delež testnih trigramov, ki smo jih zasledili že v učnem korpusu je 80 %, delež testnih bigramov pa 96 %.

Oglejmo si primer stavka iz testnega korpusa:

Ko bi me ne bilo nazaj.

Trigrami [ko bi me], [bi me ne] in [me ne bilo] so se pojavili tudi v učnem korpusu, trigram [ne bilo nazaj] pa ne.

Uporabili smo Good Turingovo glajenje v kombinaciji s sestopanjem. Zgradili smo nekaj vzorčnih jezikovnih modelov s slovarji različnih velikosti. Ocenili smo jih s perpleksnostjo. Rezultati so zbrani v tabeli 1. Perpleksnosti modelov z različno velikimi slovarji ne moremo direktno primerjati. Na primer jezikovni model s slovarjem velikosti $V = 1000$ in perpleksnostjo $PP = 1000$ ne vsebuje nobene informacije, jezikovni model s slovarjem velikosti $V = 60.000$ in isto perpleksnostjo $PP = 1000$ pa vsebuje informacijo o lastostih jezika. Za lažjo primerjavo jezikovnih modelov smo v zadnjem stolpcu dodali relativno perpleksnost, ki označuje razmerje med perpleksnostjo in velikostjo slovarja.

V	PP	$\frac{PP}{V} \cdot 100$
1.000	145	14,50
5.000	292	5,84
10.000	417	4,17
20.000	599	3,00
30.000	732	2,44
40.000	832	2,08
50.000	918	1,84
60.000	988	1,65

Table 1: Perpleksnosti jezikovnih modelov z različno velikimi slovarji

Relativna perpleksnost kaže, da so jezikovni modeli z večjim slovarjem boljši. Razlog je predvsem velik delež neznanih besed v testnem vzorcu. Primerjava deleža neznanih besed v testnem vzorcu pri različno velikih slovarjih prikazuje slika 4.

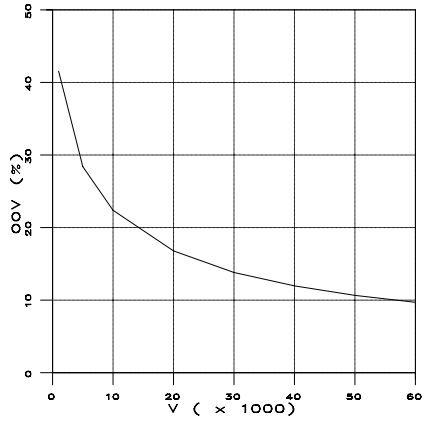

Figure 4: Delež neznanih besed v testnem vzorcu.

Pri slovarju 1000 besed je skoraj polovica besed v testnem vzorcu neznanih. Pri slovarju 60.000 besed se je delež neznanih besed zmanjšal na 10%, kar je še vedno nesprejemljivo, saj to pomeni, da je v testnem vzorcu neznana vsaka deseta beseda.

Razlog omenjenih težav je bogata fleksijska morfologija slovenskega jezika. Neznana beseda je v večini primerov le druga besedna oblika besede v slovarju.

Na podobne težave naletimo tudi pri drugih slovanskih jezikih. Tako sta Whittaker in Woodland

(1998) primerjala lastnosti besednih trigramskih modelov za ruščino in angleščino. Ugotovila sta, da bi moral biti ruski slovar za en velikostni razred večji od angleškega slovarja, da bi dobili enako pogostost neznanih besed. Pri enako velikem slovarju je bila perpleksnost za ruščino opazno večja od perpleksnosti za angleščino.

3.2. Testi razpoznavanja

Teste razpoznavanja smo izvedli na studijskem delu baze izgovarjav SNABI (Kačič et al., 1994). Učni del baze je vseboval 16401 izgovarjav (stavkov in osamljenih besed) 52 različnih govorcev. Testni del baze je vseboval 759 izgovarjav štirih govorcev iz dela base SNABI, imenovanega LINGUA. Slovar tega dela baze je bil zbran iz slovenskih literarnih del in časopisov in je tako primeren za testiranje razpoznavalnika, ki ni omejen na določeno domeno. Testni del baze ni vseboval izgovarjav govorcev iz učnega dela.

Posnetke govora smo najprej predpovdarili s filtrom s prenosno funkcijo $s'_n = s_n - 0.95 s_{n-1}$. Signal smo nato oknili s Hammingovim oknom dolžine 20ms. Za vsako okno smo izračunali 12 mel kepstralnih koeficientov in energijo. Ti so skupaj s prvimi in drugimi odvodi tvorili vektor značilk dolžine 39.

Razpoznavalnik je bil realiziran z uporabo paketa HTK 2.2 (Young et al., 1999). Zasnovan je bil na trifonskih HMM. Uporabili smo zvezne levo-desne HMM s tremi stanji. Za verjetnostne porazdelitve $b_j(y_t)$ smo uporabili mešanico 16 Gaussovih porazdelitev. Deljenje parametrov med stanji smo izvedli z uporabo združevanja s fonetičnimi drevesi (Young et al., 1999). Kot rezultat smo dobili 5654 različnih modelov trifonov s skupaj 1642 različnimi stanji.

Slovar za razpoznavanje je bil zgrajen avtomatsko s pravili in je vseboval 29 različnih fonemov. To število je manjše od običajnega, saj nismo ločili dolgih in kratkih samoglasnikov, izločili pa smo tudi nekatere redke foneme.

V okviru preizkusov uspešnosti razpoznavanja smo izdelali tri jezikovne modele, imenovane JM1, JM2 in JM3. Ker nam ni bila dosegljiva infrastruktura, ki bi nam omogočala gradnjo jezikovnih modelov, prilagojenih slovničnim lastnostim slovenskega jezika, smo jezikovne modele zgradili po jezikovno neodvisnih metodah, opisanih v razdelku 2.3.1.. Zaradi posebnosti HTK razpoznavalnika smo morali uporabiti bigramske modele. Prav tako smo velikost slovarja morali omejiti na največ kakih 20000 besed zaradi omejitev pri uporabi računalniškega pomnilnika med razpoznavanjem.

Jezikovni model JM1 je bil izdelan na tekstovnih transkripcijah dela baze LINGUA in je vseboval okoli 4100 besed. Izdelan je bil predvsem za uporabo pri ocenjevanju kakovosti akustičnega modela. Tak model ni reprezentativen za ocenjevanje jezikovnega modela, ker je zgrajen na testni bazi. Jezikovni model JM2 je bil izdelan na korpusu besedil člankov časopisa Večer in je vseboval 20000 najpogostejših besed v korpusu. Jezikovni model JM3 je bil podoben JM2, le da smo mu dodali stavke iz korpusa LINGUA in tako dosegli, da pri razpoznavanju nismo imeli neznanih besed.

Perpleksnost in pogostost neznanih besed (OOV), merjenih na korpusu LINGUA, ter velikost slovarja za posamezne jezikovne modele prikazuje tabela 2. Kot

Tip JM	V	PP	$\frac{PP}{V} \cdot 100$	OOV [%]
JM1	4142	15,97	0,39	0,00
JM2	20000	345,79	1,75	14,92
JM3	21604	50,52	0,23	0,00

Table 2: Lastnosti uporabljenih jezikovnih modelov.

je bilo lahko pričakovati, smo dobili najmanjšo perpleksnost z jezikovnim modelom JM1, saj je bil učen samo na korpusu LINGUA. Nekaj večjo perpleksnost dobimo z JM3, ki je bil učen na korpusih Večer in LINGUA. Z JM2, ki je bil učen samo na korpusu Večer, dobimo mnogo večjo perpleksnost kot z drugima dvema jezikovnima modeloma. Dodatno dobimo dokaj večji delež neznanih besed – 14,92 %.

Rezultate razpoznavanja besed pri uporabi posameznih jezikovnih modelov prikazuje tabela 3. V vseh treh primerih smo uporabljali isti akustični model.

Tip JM	Pravilno [%]	Točno [%]
JM1	94,88	93,07
JM2	44,38	33,56
JM3	84,82	82,86

Table 3: Rezultati razpoznavanja s posameznimi jezikovnimi modeli.

Rezultati so izračunani po običajnih formulah:

$$\text{Pravilno } [\%] = \frac{N - D - S}{N} \times 100\% \qquad (13)$$

$$\text{Točno } [\%] = \frac{N - D - S - I}{N} \times 100\% \qquad (14)$$

kjer je N število besed v referenčnem zapisu, D število izpuščenih besed, S število zamenjanih besed in I število vrinjenih besed pri razpoznavanju.

Uspešnost razpoznavanja pri uporabi jezikovnega modela JM1 nam daje dobro sliko o kakovosti akustičnega modela, saj ima ta v tem primeru večji vpliv na razpoznavanje kot jezikovni model. Uspešnost razpoznavanja je tukaj primerljiva z rezultati, ki so bili doseženi za druge jezike pri podobno zastavljenih eksperimentih. Lahko zaključimo, da smo uspeli izdelati dober akustični model za slovenščino in tako ta del razpoznavalnika ne predstavlja posebnega problema.

Pri uporabi jezikovnega modela JM2 uspešnost razpoznavanja zelo pade. Vzrok temu je verjetno v dejstvu, da JM2 v primerjavi z JM1 dosti slabše opisuje lastnosti stavkov v testni bazi (za 20× večja perpleksnost!). Dodatno pa imamo tudi veliko pogostost neznanih besed, kar tudi močno vpliva na poslabšanje uspešnosti razpoznavanja.

Ko smo pri izgradnji JM3 dodali stavke iz korpusa LINGUA, smo se izognili pojavljanju neznanih

besed, hkrati po smo tudi zmanjšali perpleksnost. To se je odrazilo tudi na uspešnosti razpoznavanja, ki je le kakih 10 % slabše kot pri uporabi JM1. To dokazuje, da sama velikost slovarja nima posebej velikega vpliva na uspešnost razpoznavanja; veliko bolj pomembni sta perpleksnost in pogostost neznanih besed.

4. Zaključek

V pričujočem članku smo predstavili zasnovo razpoznavalnika tekočega slovenskega govora z velikim slovarjem in nekatere težave, na katere smo naleteli pri njegovi izdelavi. Kot najbolj pereč problem se je izpostavilo modeliranje slovenskega jezika, pri katerem smo morali uporabiti jezikovno neodvisne metode, saj nam ni bila dosegljiva infrastruktura, ki bi omogočala izgradnjo boljših modelov. Uspešnost teh metod je žal v veliki večini slabša od jezikovno odvisnih postopkov modeliranja jezika. Tudi v našem primeru se je izkazalo, da z jezikovno neodvisnim načinom gradnje jezikovnega modela za slovenski jezik ne dobimo zadovoljivih rezultatov. V prihodnosti se bomo zato posvetili problemu definiranja slovarja pri fleksijskih jezikih. Vpeljali bomo skladenjsko odvisno dvodelno predstavitev besede s krnom (Popovič, 1992) (tj. besedo, ki smo ji odvzeli slovnične lastnosti) in obrazilom.

5. References

R. O. Duda in P. E. Hart. 1973. *Pattern Classification and Scene Analysis.* Wiley.

I. Ipšić, F. Mihelič, S. Dobrišek, J. Gros, in N. Pavešić. 1999. A slovenian spoken dialog system for air flight inquiries. V: *Proc. Eurospeech '99*, str. 2659–2662.

F. Jelinek. 1997. *Statistical Methods for Speech Recognition.* MIT Press.

S. Katz. 1987. Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recogizer. *IEEE Transactions on Acoustics, Speech, and Signal Processing.*

Z. Kačič, B. Horvat, in R. Derlič. 1994. Zasnova baze izgovarjav slovenskega jezika snabi. V: *Zbornik 3. Elektrotehniške in računalniške konference ERK'94*, str. B:327–330.

J.W. Picone. 1993. Signal modeling techniques in speech recognition. *Proceedings of the IEEE*, 81(9).

M. Popovič. 1992. The Effectiveness of Stemming for Natural-Language Access to Slovene Textual Data. *Journal of the American Society for Information Science.*

E.W.D Whittaker in P.C. Woodland. 1998. Comparison of language modelling techniques for russian and english. V: *Proc. ICSLP '98.*

S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev, in P. Woodland. 1999. *The HTK Book.* Entropic.