

Razpoznavnik tekočega govora UMB Broadcast News 2014: kakšno vlogo igra velikost učnih virov?

Andrej Žgank, Gregor Donaj, Mirjam Sepesy Maučec

Inštitut za elektroniko in telekomunikacije, Fakulteta za elektrotehniko, računalništvo in informatiko,
Univerza v Mariboru
Smetanova ul. 17, 2000 Maribor
andrej.zgank@um.si, gregor.donaj@um.si, mirjam.sepesy@um.si

Povzetek

V članku bomo predstavili rezultate analize, kako na uspešnost razpoznavanja tekočega govora vpliva velikost učnih virov, ki jih uporabimo pri izdelavi akustičnih in jezikovnih modelov. Analizo smo zasnovali na sistemu za razpoznavanje slovenskega tekočega govora UMB Broadcast News, katerega domena so dnevno-informativne televizijske oddaje. Za izhodišče smo uporabili verzijo sistema predstavljeno leta 2010, ki smo jo nadgradili z akustičnimi in jezikovnimi modeli naučenimi na dodatnih virih. Z nadgrajenim sistemom smo dosegli najboljšo pravilnost razpoznavanja besed 73,30%, kar sicer predstavlja majhno izboljšanje, vendar je bilo za to potrebno uporabiti bistveno obsežnejše vire. Z večanjem obsega virov se izboljšanje sistema zmanjšuje. Zato je razen obsega uporabljenih virov smiselno razmišljati tudi o učinkovitejših, jeziku prilagojenih postopkih razpoznavanja govora.

UMB Broadcast News 2014 continuous speech recognition system: what is the influence of language resources' size?

This paper presents the results of an analysis, how the size of language resources for training acoustic and language models influences the speech recognition accuracy. The Slovenian continuous speech recognition system UMB Broadcast News was used for the experiments. Its speech recognition domain are TV news shows. As baseline, the system presented in 2010 was used. The acoustic and language models training procedure for the experiments applied additional language resources. The improved speech recognition system achieved 73.30% word accuracy. The best speech recognition result presents a small accuracy improvement but significantly larger language resources were needed to achieve this. Enlarging language resources after a certain size brings only small improvements. This indicates that further research into language adapted methods for speech recognition is needed.

1. Uvod

Za razliko od nekaterih drugih sodobnih informacijsko komunikacijskih tehnologij, ima pri govornih tehnologijah še vedno zelo pomembno vlogo jezik. Slovenščina sodi v kategorijo visoko pregibnih jezikov, hkrati pa je zaradi relativno majhnega števila govorcev tudi tržno nezanimiva. Tako jo lahko z vidika trenutnega razvoja področja razpoznavanja govora ponovno prištevamo med jezike s premajhnim obsegom govornih virov, ki so na voljo za uporabo, hkrati pa doseženi rezultati razpoznavanja tekočega govora zaostajajo za rezultati doseženimi za nekatere jezike z velikim številom govorcev.

Za doseganje boljših rezultatov razpoznavanja govora se običajno uporabljata dve metodi. Prva je povečevanje obsega virov uporabljenih za učenje akustičnih in jezikovnih modelov. Druga metoda pa je uporaba izboljšanih algoritmov, ki lahko dodatno upoštevajo katero izmed specifik jezika (Rotovnik et al., 2007; Dobrišek & Mihelič, 2010), kot sta na primer v slovenskem jeziku visoka pregibnost besed in relativno prosti vrstni red besed v stavku.

Raziskovalno delo na področju razpoznavanja tekočega slovenskega govora je poskušalo slediti obema možnima metodama za izboljšanje rezultatov. Konec aprila 2014 je v Ljubljani potekala delavnica, kjer so predstavili rezultate evropskega projekta za avtomatsko podnaslavljanje in prevajanje predavanj, kjer je bila kot eden izmed jezikov vključena tudi slovenščina. V razpravi po predstavitvi je bila ena izmed tem tudi, ali bi za doseganje boljših rezultatov za slovenski jezik zadostovalo samo povečevanje učnih virov ali pa so

potrebni tudi kakšni posebni algoritmi, ki naslavlajo značilnosti visoko pregibnih jezikov.

V članku¹ bomo poskusili analizirati, koliko lahko k izboljšanju rezultatov razpoznavanja govora doprinese povečevanje učnih virov, v našem primeru govorne baze za učenje akustičnih modelov in besedilnega korpusa za učenje statističnega jezikovnega modela. Eksperimente smo zasnovali na sistemu avtomatskega razpoznavanja tekočega slovenskega govora UMB Broadcast News (Žgank & Sepesy Maučec, 2010; Žgank et al., 2008), ki razpozna govor v dnevno-informativnih televizijskih oddajah. Povečanje učnih virov bomo izvedli izključno z uporabo dodatnih slovenskih virov. V preteklosti smo že izvajali križno-jezične eksperimente (Žgank et al., 2004/2) - sicer samo z razpoznavanjem izoliranih in vezanih besed - kjer smo za razpoznavanje slovenskega govora uporabili akustične modele naučene na govorni bazi slovaškega jezika, ki predstavlja enega izmed slovenščini sorodnejših jezikov. Rezultati takšnega križno-jezičnega eksperimenta so pokazali, da je za preproste scenarije sicer možno doseči zelo dobre rezultate, z večanjem kompleksnosti testnih scenarijev pa se rezultati takšnega razpoznavanja govora bistveno poslabšajo.

V nadaljevanju članka bomo najprej predstavili jezikovne vire, ki smo jih uporabili pri izgradnji sistema avtomatskega razpoznavanja govora. V tretjem poglavju bo sledila kratka predstavitev izdelave akustičnih in jezikovnih modelov eksperimentalnega sistema. Rezultate in analizo vrednotenja razpoznavanja govora bomo predstavili v četrtem poglavju. Zaključek in smernice za nadaljnje delo bomo podali v petem poglavju.

¹ Raziskovalno delo je bilo delno sofinancirano s strani ARRS po pogodbi št. P2-0069.

2. Jezikovni viri

Jezikovni viri predstavljajo eno izmed ključnih točk pri izdelavi razpoznavnika tekočega govora in igrajo še posebej pomembno vlogo pri jezikih z manjšim številom govorcev, kamor lahko prištevamo tudi slovenski jezik.

2.1. Govorna baza BNSI Broadcast News

Kot osnovo za izvedbo eksperimentov smo uporabili slovensko govorno bazo BNSI Broadcast News (Žgank et al., 2004/1), v enakem obsegu kot za zadnje razviti sistem razpoznavanja tekočega govora (Žgank & Sepesy Maučec, 2010). Baza je na voljo pri organizaciji ELRA/ELDA (ELRA, 2014). Akustični modeli naučeni na takšni bazi so služili kot osnova za nadaljnje eksperimente.

Za izvedbo analize vpliva velikosti učnega korpusa smo uporabili dva dodatna vira transkribiranih posnetkov. Prvi vir so bili preostali posnetki iz govorne baze BNSI Broadcast News, katerih obseg je bil 8 ur. Glede na tip oddaj so se ti posnetki ujemali z do sedaj uporabljenimi posnetki iz govorne baze BNSI Broadcast News. Kot drugi dodatni vir smo uporabili transkribirane posnetke iz interne baze televizijskih oddaj, ki jih bomo v nadaljevanju označevali z IETK-TV. Časovno obdobje teh dodatnih posnetkov ustreza obdobju posnetkov, ki so vključeni v govorno bazo BNSI Broadcast News. Za segmentacijo, označevanje in transkribiranje smo uporabili enake postopke kot za izdelavo baze BNSI Broadcast News (Žgank et al., 2004/1). Oddaje v tem dodatnem delu učne baze po tipu delno odstopajo od oddaj iz govorne baze BNSI Broadcast News, saj posnetki vsebujejo tudi različne intervjuje in omizja, ter s tem pokrivajo bistveno širši spekter televizijskih vsebin in načinov govora. Iz govorne baze IETK-TV smo kot drugi vir učnega materiala uporabili 29 ur dodatnih posnetkov. Baza BNSI Broadcast News v nasprotju z bazo IETK-TV vsebuje samo večerne in nočne dnevno-informativne oddaje, kjer je v večji meri prisoten bran kot spontan govor (Schwartz et al., 1997). Primerjavo lastnosti med govorno bazo BNSI Broadcast News in govorno bazo IETK-TV podajamo v tabeli 1.

<i>Lastnost</i>	<i>BNSI Broadcast News</i>	<i>IETK-TV</i>
<i>dolžina</i>	29,86 ure	28,97 ure
<i>število oddaj</i>	34	30
<i>število tipov oddaj</i>	2	7
<i>število govorcev</i>	2073	784
<i>delež spontanega govora</i>	30,74%	68,37%

Tabela 1: Primerjava lastnosti učnega nabora govorne baze BNSI Broadcast News in IETK-TV.

Opravljen primerjava kaže, da vsebuje govorna baza IETK-TV bistveno večji delež spontanega govora pri hkrati manjšem številu različnih govorcev, kar je posledica vključitve različnih intervjujev in omizij v nabor oddaj.

Da bi lahko pravilno ovrednotili vpliv količine posnetkov v učni govorni bazi, smo kot izhodišče vzeli učni govorni korpus iz predhodnih eksperimentov, nato pa

smo iz dodatnega materiala korakoma dodajali nove posnetke za učenje akustičnih modelov, in sicer v obsegu: 25%, 50%, 75% in 100%. Na takšen način bomo v nadaljevanju tudi označevali pripadajoče akustične modele.

2.2. Tekstovne baze

Za učenje jezikovnih modelov smo obstoječim besedilnim korpusom v dveh korakih dodali še korpus FidaPLUS.

Korpus FidaPLUS (Arhar & Gorjanc, 2007) je referenčni korpus slovenskega pisanega jezika, ki obsega 621 mio. besed in vsebuje besedila iz tekstovnih virov iz obdobja 1990-2006. Večji del korpusa predstavljajo časopisni in revijalni članki ter knjige. Nekaj besedil izvira tudi iz Spleta. Korpus je lematiziran in označen z morfosintaktičnimi značkami, ki pa v pričujočem članku niso bile uporabljene.

Obstoječa korpusa BNSI-Speech in BNSI-Text sta korpusa govornega jezika, obstoječi korpus Večer in novo dodani korpus FidaPLUS pa sta korpusa pisanega jezika. Med govornim (predvsem spontano govornim) in pisanim jezikom je velika razlika (Stouten et al., 2006). Številnih pojavov, ki so značilni za govor (npr. krajše izjave, uporaba mašil, ponavljanje, napačni starti ipd.), v pisanem jeziku ne zasledimo (Žgank et al., 2008). Ker je FidaPLUS po obsegu neprimerno večja od korpusov govornega jezika, bi se z enostavnim razširjanjem učnega korpusa lastnosti govornega jezika izgubile. Uravnotežen vpliv različnih jezikovnih virov smo dosegli z linearno interpolacijo na korpusu BNSI-Devel. BNSI-Devel korpus je po strukturi enak korpusu BNSI-Speech in obsega 4 oddaje. Poskrbeli smo, da med korpusi BNSI-Speech, BNSI-Devel in BNSI-Eval (ki je namenjen vrednotenju) ni vsebinskega prekrivanja.

3. Razpoznavnik govora UMB BN 2014

Sistemi za razpoznavanje tekočega govora so še vedno eni izmed najkompleksnejših na področju govornih tehnologij. V eksperimentih smo kot izhodišče uporabili konfiguracijo iz predhodnih eksperimentov (Žgank & Sepesy Maučec, 2010), ki smo ji delno spremenili modul za izločanje značilnik, kjer smo postopku mel-kepralnih koeficientov dodali normalizacijo srednjih vrednosti keprtra ter normalizacijo energije signala. Za vpeljavo normalizacije smo se odločili zato, ker smo razširili nabor oddaj v učni bazi, ki so imele med seboj delno različne akustične značilnosti.

3.1. Izdelava akustičnih modelov

Osnovo razpoznavnika tekočega govora predstavljajo zvezni prikriti modeli Markova (HMM) s tristanjsko levo-desno topologijo ter Gaussovimi porazdelitvami funkcije verjetnosti. Kot osnovno enoto akustičnih modelov smo uporabili grafem, ki je že v predhodnih eksperimentih (Žgank et al., 2008; Žgank & Kačič, 2005/1) pokazal učinkovitost delovanja.

Za učenje akustičnih modelov smo uporabili postopek, predstavljen v (Žgank & Sepesy Maučec, 2010). Osnovne značilnosti tako razvitega sistema razpoznavanja govora so podane v tabeli 2.

UMB BN ASR	
<i>Izloč. značilnik</i>	MFCC z normalizacijo
<i>Karakteristike značilnik</i>	Okno 25 ms, korak 10 ms, MFCC 12 koef., energija, 1. in 2. odvod, 26 filtrov, normalizacija kepstra in energije
<i>Akustični model</i>	Medbesedni (Odell, 1995) trigrafemi
<i>Kompleksnost AM</i>	utežena vsota 16 Gaussovih porazdelitev verjetnosti na stanje
<i>Združevanje AM</i>	odločitveno drevo na osnovi grafemskih razredov (Žgank et al., 2005/2)
<i>Jezik. modeli</i>	Interpolirani trigrami
<i>Vel. slovarja</i>	64.000 besed

Tabela 2: Značilnosti razpoznavalnika tekočega govora.

Sistem, predstavljen v tabeli 2, je služil kot izhodiščni sistem za vrednotenje razpoznavanja govora. Nato smo štirikrat v celoti ponovili postopek učenja akustičnih modelov, vsakič z večjim naborom učnih posnetkov, predstavljenim v poglavju 2.1. Kompleksnost akustičnih modelov smo kontrolirali s konstantno nastavitvijo praga povečanja logaritemske verjetnosti v postopku združevanja z odločitvenim drevesom. Za vrednotenje razpoznavalnika govora smo tako imeli na voljo 5 različnih naborov akustičnih modelov. Razpoznavanje govora smo vedno izvedli z identičnim naborom parametrov dekodirnika.

3.2. Izdelava jezikovnih modelov

Za gradnjo jezikovnih modelov smo uporabili orodje SRI Language Modeling Toolkit (Stolcke, 2002). Na osnovi besedilnih korpusov smo zgradili dva trigramska jezikovna modela. Oba sta bila sestavljena iz štirih komponent: prvo komponento smo zgradili na korpusu BNSI-Speech, drugo na korpusu BNSI-Text, tretjo na korpusu Večer in četrto na korpusu FidaPLUS. V prvem modelu (JMx1) smo uporabili polovico FidePLUS, v drugem pa celo (JMx2). Dokumente za prvi model smo izbirali naključno. V obeh modelih smo v prvih treh komponentah ohranili vse bigrame in trigrame iz učnih korpusov, v četrti komponenti pa smo izločili vse n-grame s frekvenco 1.

Slovar je obsegal 64.000 besed. Vseboval je vse besede korpusov BNSI-Speech in BNSI-Text. Do velikosti 64.000 smo ga dopolnili z najpogostejšimi besedami iz korpusa Večer.

Uporabili smo Good-Turingovo glajenje in sestopanje po Katz-u. Interpolacijske koeficiente komponent smo določili tako, da smo minimizirali perpleksnost jezikovnega modela na korpusu BNSI-Devel. Interpolacijski koeficienti po komponentah za oba modela so predstavljeni v tabeli 3. Utež četrte komponente se je po podvojitvi velikosti zmanjšala.

Komponenta	JMx1	JMx2
BNSI-Speech	0,18	0,18
BNSI-Text	0,23	0,24
Večer	0,11	0,12
FidaPLUS	0,48	0,46

Tabela 3: Koeficienti λ komponent v interpoliranih trigramskih modelih.

Perpleksnost prvega modela na BNSI-Eval je znašala 258, drugega modela pa 246. Delež besed izven slovarja (OOV) je bil 4,22%. Besed, ki so bile izven slovarja, nismo posebej modelirali na nivoju akustičnega modela. Podvojitve velikosti četrte komponente je perpleksnost izboljšala le za 4,56%. Pri tem se je število bigramov v modelu povečalo za 24%, trigramov pa kar za 45%.

4. Rezultati eksperimentov

Vrednotenje sistema za razpoznavanje govora smo izvedli z namenskim evalvacijskim naborom baze BNSI Broadcast News, ki vsebuje 4 različne dnevno-informativne oddaje s 1898 stavki. Rezultate razpoznavanja govora smo podali v odstotku pravilno razpoznanih besed, kjer smo upoštevali tudi vrinjene oz. izbrisane besede, ki posledično dodatno poslabšajo rezultat.

Akustični modeli	Pravilno razpoznane besede (%)
+00%	72,44
+25%	73,10
+50%	72,87
+75%	73,00
+100%	73,19

Tabela 4: Rezultati razpoznavanja govora za večanje učne baze akustičnih modelov z jezikovnim modelom JMx1.

Izhodiščni sistem razpoznavanja govora, ki je uporabljal osnovne akustične modele (+00%), ter osnovni jezikovni model (JMx1) je dosegel 72,44% pravilnost razpoznavanja besed na evalvacijskem naboru posnetkov baze BNSI Broadcast News. Doseženi rezultat je primerljiv z razpoznavalniki govora za slovenski jezik, ki uporabljajo podobno kompleksno zasnovo (Žgank in Sepesy Maučec, 2010).

V naslednjem koraku smo z uporabo osnovnega jezikovnega modela JMx1 ovrednotili različne akustične modele (+25% ... +100%), naučene na večjem naboru učnih posnetkov. Rezultati razpoznavanja tekočega govora so se izboljšali na 72,87% (+50%) do največ 73,19% z akustičnimi modeli +100%. Tako smo s povečanjem učne baze posnetkov za približno dvakrat dosegli izboljšanje rezultatov razpoznavanja govora za največ 0,75% absolutno. Glede na količino dodatnega učnega materiala, ter potrebni finančni in časovni vložek za njegovo pripravo, lahko trdimo, da je izboljšanje rezultatov razpoznavanja govora relativno majhno.

Akustični modeli	JMx1, Pravilnost razpoznavanja besed (%)	JMx2, Pravilnost razpoznavanja besed (%)
+00%	72,44	72,53
+25%	73,10	72,97
+50%	72,87	73,02
+75%	73,00	73,15
+100%	73,19	73,30

Tabela 5: Rezultati razpoznavanja govora z jezikovnim modelom naučenim na večjem korpusu.

V zadnjem koraku vrednotenja smo analizirali, kako na rezultate razpoznavanja govora vpliva povečanje

učnega korpusa jezikovnega modela za dvakrat (JMx2). Rezultati so predstavljeni v tabeli 5.

V izhodišču smo s povečanjem učnega korpusa za pripravo jezikovnega modela za dvakratnik (JMx2) uspeli doseči rezultat razpoznavanja govora 72,53%. Izboljšanje rezultata je znašalo 0,09% absolutno, kar predstavlja minimalno razliko, še posebej če upoštevamo, kako veliko povečanje besedilnega korpusa in posledično jezikovnega modela je bilo potrebno za doseg tega rezultata.

Podobna minimalna izboljšanja rezultatov z novim jezikovnim modelom JMx2 smo dosegli tudi v kombinaciji z različnimi akustičnimi modeli naučenimi na povečani govorni bazi. Edina izjema so bili akustični modeli +25%, kjer je z večjim jezikovnim modelom JMx2 prišlo celo do rahlega poslabšanja rezultatov, in sicer za 0,13% absolutno.

V kombinaciji akustičnih modelov +100% in jezikovnega modela JMx2, kjer smo obakrat uporabili največje razpoložljive govorne in jezikovne vire, smo dosegli skupno najboljši rezultat razpoznavanja govora 73,30%, kar sicer predstavlja 0,86% absolutno izboljšanje, vendar je pri tem potrebno upoštevati, koliko večji viri so bili potrebni za doseg takšnega rezultata.

Na dobljenih rezultatih smo izvedli test statistične značilnosti, pri čemer smo za mejo statistične značilnosti izbrali vrednost $\alpha=0,05$. Izkazalo se je, da je statistično značilna (0,012) le razlika pri povečanju govorne baze (+100%), kadar uporabljamo jezikovni model JMx2. Vse ostale primerjave so pokazale, da izboljšanja niso statistično značilna za izbrano mejo. Ti rezultati dodatno potrjujejo našo domnevo, da samo s povečevanjem učnega materiala ne moremo doseči bistvenih izboljšav razpoznavanja slovenskega tekočega govora.

Če hipotetično predpostavimo, da bi se ohranil takšen trend povečevanja pravilnosti razpoznavanja besed z večanjem obsega učnih virov (v kar avtorji sicer dvomimo), bi za doseganje pravilnosti razpoznavanja besed vsaj 90%, po oceni potrebovali tematsko ustrezne vire v obsegu več kot 1100 ur transkribiranega govora in jezikovne vire z več kot 12,1 giga besed. Takšen predpostavljen obseg virov za večkratni faktor presega vse do sedaj ustvarjene vire za slovenski jezik.

5. Zaključek

V članku smo poskusili odgovoriti na vprašanje, kako pomembna je velikost govornih in jezikovnih virov za izboljšanje rezultatov razpoznavalnika govora za slovenski jezik. Analiza eksperimentov je pokazala, da je s povečanjem virov sicer možno doseči minimalno izboljšanje rezultatov, vendar so za doseg tega cilja potrebna velika vlaganja v izdelavo virov. Pokazalo se je, da je pri tem izrednega pomena tudi ujemanje virov v žanru oz. domeni.

Na osnovi doseženih rezultatov lahko z dokaj veliko verjetnostjo predpostavimo, da za visoko pregibni slovenski jezik ne zadostuje samo večanje obsega govornih in jezikovnih virov, temveč da je hkrati potrebno tudi nadaljevati z raziskovalnim delom na področju algoritmov, ki bi ustrezno naslavljali specifične lastnosti visoko pregibnega slovenskega jezika.

Zahvala

Zahvaljujemo se avtorjem besedilnega korpusa FidaPLUS, ki so nam omogočili njegovo uporabo za

jezikovno modeliranje avtomatskega razpoznavalnika govora.

6. Literatura

- Arhar, Š., Gorjanc, V., (2007). Korpus FidaPLUS: nova generacija slovenskega referenčnega korpusa. *Jezik in slovnstvo* 52/2., 95--110.
- Dobrišek, S., Mihelič, F., (2010). Zmanjševanje odvečnosti končnih pretvornikov za učinkovito gradnjo razpoznavalnikov slovenskega govora z velikim besednjakom. *Jezikovne tehnologije*, Ljubljana, Slovenija.
- ELRA (2014). *BNSI Catalog Reference : S0275*: www.elra.info.
- Odell, J.J., (1995). *The Use of Context in Large Vocabulary Speech Recognition*. Doktorska disertacija, Univerza v Cambridgeu, Velika Britanija.
- Rotovnik, T., Sepesy Maučec, M., Kačič, Z. (2007). Large vocabulary continuous speech recognition of an inflected language using stems and endings. *Speech communication*, 2007, vol. 49, iss. 6, 437--452.
- Schwartz, R., Jin, H., Kubala, F., Matsoukas, S., (1997). Modeling those F-Conditions - or not. *Proc. DARPA Speech Recognition Workshop*, Chantilly, ZDA.
- Stolcke, A. (2002). SRILM - an extensible language modeling toolkit. *International Conference on Speech and Language Processing*, II: 901--904.
- Stouten, F., Duchateau, J., Martens, J.-P., Wambacq, P., (2006). Coping With Disfluencies In Spontaneous Speech Recognition: Acoustic Detection And Linguistic Context Manipulation, *Speech Communication* vol. 48, issue 11, 1590--1606.
- Žgank, A., Rotovnik, T., Sepesy Maučec, M., Verdonik, D., Kitak, J., Vlaj, D., Hozjan, V., Kačič, Z., Horvat, B., (2004/1). Acquisition and annotation of Slovenian broadcast news database. *Fourth international conference on language resources and evaluation, LREC 2004*, Lizbona, Portugalska.
- Žgank, A., Kačič, Z., Vicsi, K., Szaszak, G., Diehl, F., Juhar, J., Lihan, S., (2004/2). Crosslingual transfer of source acoustic models to two different target languages. *Robust2004 : COST278 and ISCA Tutorial and Research Workshop (ITRW) on Robustness Issues in Conversational Interaction*, Norwich, Velika Britanija.
- Žgank, A., Kačič, Z., (2005/1). Primerjava treh tipov akustičnih osnovnih enot razpoznavalnika slovenskega govora. *Elektrotehniški vestnik*, 2005, Ljubljana, Slovenija.
- Žgank, A., Horvat, B., Kačič, Z., (2005/2). Data-driven generation of phonetic broad classes, based on phoneme confusion matrix similarity. *Speech Communication*, vol. 47, issue 3, 379--393, november 2005.
- Žgank, A., Kos, M., Kotnik, B., Sepesy Maučec, M., Rotovnik, T., Kačič, Z. (2008). Nadgradnja sistema za razpoznavanje slovenskega tekočega govora UMB Broadcast news. *Jezikovne tehnologije 2008*, Ljubljana, Slovenija.
- Žgank, A., Sepesy Maučec, M., (2010). Nadgradnja sistema Razpoznavalnik tekočega govora UMB Broadcast News 2010: nadgradnja akustičnih in jezikovnih modelov. *Jezikovne tehnologije 2010*, Ljubljana, Slovenija.