NEJČ ROBIDA
Sinteza govora in Govorec 3
Diplomsko delo

Mentor: doc. dr. Hotimir Tivadar

Ljubljana, september 2013
ZAHVALA

Za pomoč pri pripravi diplomske naloge se zahvaljujem mentorju doc. dr. Hotimirju Tivadarju, ki je z veseljem sprejel temo in mi jo pomagal razviti, sodelavcem na Amebisu, predvsem Simonu Rozmanu za prijetno delovno izkušnjo, svetovanje pri tehničnem delu naloge in zaupanje v moje jezikoslovne rešitve pri Govorcu 3, in svoji družini, ki me je podpirala in vodila do konca študija.
IZVLEČEK

V diplomski nalogi Sinteza govora in Govorec 3 je povzeta zgodovina razvoja sintetizatorja govora, predstavljene so različne vrste sintez, posebej pa je izpostavljen jezikoslovni vidik razvoja sintetizatorja. Podrobneje so opisane izkušnje z delom pri Amebisovem Govorcu 3 in težave, ki so se pojavljale pri transkribiranju in urejanju slovarja izgovarjav, podana je tudi analiza nekaterih posebnih akustičnofonetičnih pojavov v govorni zbirki.

Ključne besede: sinteza govora, sintetizator, govorna zbirka, transkripcija, Govorec 3

ABSTRACT

The diploma Speech synthesis and the speech synthesizer Govorec 3 summarizes the history of speech synthesizer development. Furthermore, different types of syntheses are presented, and particularly highlighted is the linguistic view of synthesizer development. Experiences when working with the speech synthesizer Govorec 3 from Amebis and the difficulties that have emerged while making phonetic transcriptions and editing the pronunciation lexicon are described in detail. Finally, several analyses of the distinct acoustic-phonetic phenomenon in a speech corpus are presented.

Key words: speech synthesis, speech synthesizer, speech corpus, phonetic transcription, Govorec 3
<table>
<thead>
<tr>
<th>Kazalo</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Uvod</td>
<td>1</td>
</tr>
<tr>
<td>2 Tvorjenje glasov in govorjeni jezik</td>
<td>2</td>
</tr>
<tr>
<td>3 Zgodovina sintetizatorjev govora</td>
<td>5</td>
</tr>
<tr>
<td>4 Sintetizator govora</td>
<td>7</td>
</tr>
<tr>
<td>5 Sinteza</td>
<td>9</td>
</tr>
<tr>
<td>5.2 Sintetizator govora Govorec</td>
<td>13</td>
</tr>
<tr>
<td>5.2.1 Sintetizatorja Govorca 1 in 2</td>
<td>13</td>
</tr>
<tr>
<td>5.2.3 Sintetizator Govorec 3</td>
<td>16</td>
</tr>
<tr>
<td>5.2.4 Govorne zbirke</td>
<td>17</td>
</tr>
<tr>
<td>5.2.5 Slovenske govorne zbirke</td>
<td>18</td>
</tr>
<tr>
<td>6 Vhodno besedilo</td>
<td>20</td>
</tr>
<tr>
<td>6.1 Slovar izgovarjav kot del Govorca</td>
<td>20</td>
</tr>
<tr>
<td>6.1.2 Slovar izgovarjav in Govorec 3</td>
<td>22</td>
</tr>
<tr>
<td>6. 2 Slovnična analiza pri Govorcih 2 in 3</td>
<td>23</td>
</tr>
<tr>
<td>6.2.1 Slovnična analiza in transkripcije pri Govorcu 3</td>
<td>24</td>
</tr>
<tr>
<td>6.3 Grafemska pretvorba</td>
<td>27</td>
</tr>
<tr>
<td>6.3.2 Grafemska pretvorba in Govorca 2 in 3</td>
<td>27</td>
</tr>
<tr>
<td>6.4 Transkripcija</td>
<td>28</td>
</tr>
<tr>
<td>6.4.1 Transkribiranje in Govorec 3</td>
<td>30</td>
</tr>
<tr>
<td>7 Anketa in rezultati</td>
<td>41</td>
</tr>
<tr>
<td>8 Zaključek</td>
<td>43</td>
</tr>
<tr>
<td>9 Povzetek</td>
<td>45</td>
</tr>
<tr>
<td>10 Viri in literatura</td>
<td>47</td>
</tr>
<tr>
<td>11 Priloge</td>
<td>49</td>
</tr>
<tr>
<td>11.1 Anketa o kakovosti in uporabnosti Govorca 2</td>
<td>49</td>
</tr>
</tbody>
</table>
11.1.1 Besedilo 1

11.1.2 Besedilo 2

11.2 Zgoščenka s posnetki sintetiziranega govora
Kazalo slik

Slika 1: Shematski prikaz postopka učenja in sinteze govora z uporabo prikritih modelov Markova (Vesniger idr. 2004) ........................................................................................................ 10
Slika 2: Dekompozicija glasu u na harmonski in šumn del (Rozman 2005: 10) ......... 11
Slika 3: Dekompozicija glasu z na harmonski in šumn del (Rozman 2005: 11) ........... 12
Slika 4: Shema sinteze besede lipa po postopku TD-PSOLA ........................................ 14
Slika 5: Prelivanje difonov ................................................................................................... 15
Slika 6: Naglas in analiza besede zavóhala ...................................................................... 15
Slika 7: Slovar izgovarjav (delovno okno) ........................................................................... 20
Slika 8: Slovar izgovarjav (primer dvojnic) ....................................................................... 22
Slika 9: Delovno okolje v Transcriberju .............................................................................. 31
Slika 10: Analiza besede cianid (bralka) ............................................................................ 33
Slika 11: Analiza besede socialnoekonomski (bralka) ..................................................... 34
Slika 12: Analiza besede cianid (bralec) ............................................................................ 34
Slika 13: Analiza besede socialnoekonomski (bralec) ..................................................... 35
Slika 14: Analiza zvenečega γ (bralka). ............................................................................ 37
Slika 15: Analiza zvenečega γ (bralka) ............................................................................ 38
Slika 16: Analiza zvenečega γ (bralec) ............................................................................ 38
Slika 17: Analiza zvenečega γ s primerom nezvenečega x (bralec) .................. 39
Slika 18: Analiza primera [sl"a:bix tr"u:Sp] ........................................................................ 40
Kazalo tabel

Tabela 1: Seznam glasov ................................................................. 5
Tabela 2: Primeri napačnih slovničnih analiz ................................ 25
Tabela 3: Primerjava izgovora členka že pri bralcih ........................ 32
Tabela 4: Primeri izgovora i pred samoglasnikom ....................... 35
Tabela 5: Primeri besed, v katerih si v verigi sledita dva soglasnika .......................... 36
Tabela 6: Časovna analiza zlitih glasov (bralka) ............................ 36
Tabela 7: Časovna analiza zlitih glasov (bralec) ........................... 37
Tabela 8: Številčni rezultati ankete ............................................. 41
Človeški govor je najpomembnejši člen medsebojnega sporazumevanja, zato je prizadevanje za strojno posnemanje glasu v času vsesplošne informatizacije pomembna znanstvenoraziskovalna dejavnost, ki je že in bo vedno bolj uporabljana tudi v tržne namene. Pri ustvarjanju sintetizatorja govora se srečamo z različnimi znanstvenimi področji v jezikoslovju, računalništvu in elektrotehnik, kar je pristen primer interdisciplinarnosti. „Sinteza govora je proces umetnega ustvarjanja govora, ki ga lahko uporabljamo za najrazličnejše namene“ (Žganec Gros 2001: 21), lahko je pripomoček slepim, samodejni odzivnik v različnih telekomunikacijskih sistemih, interaktivni sogovornik ali pa samo strojni bralec različnih besedil.


Pri delu z Govorcem smo se veliko ukvarjali tudi z notacijami, to so enotne in standardizirane računalniške abecede, ki s posebnimi simboli zapisujejo glasove. V diplomskem delu smo za fonetične zapise uporabljali notacijo MRPA, to smo uporabljali tudi pri transkribiranju govorne zbirke za Govorca 3. Te zapise nakazujejo oglati oklepaji ([]), ti pa se žal pojavljajo tudi v citatih, kjer smo jih zamenjali s poševnicami (//). Današnjo verzijo Govorca, ki je na spletu na Amebisovi strani, bomo imenovali Govorec 2 in novo Govorec 3, saj so nekateri postopki in izzivi pri ustvarjanju sinteze precej različni. Govorec 1 pa je bil last Instituta „Jožef Stefan“. Zaradi jasnosti besedo govorec uporabljamo samo za računalniški program, osebe, ki so bile posnete pri ustvarjanju govornih zbirk, pa se imenujejo bralci, saj so vsa besedila, ki so uporabljena za sintetizator govora, prebrana in niso nikoli prosto
govorjena. Diplomskega dela nismo izrecno delili na teoretični in praktični del, vse težave in ugotovitve pri delu z novim Govorcem smo navajali sproti pri vsakem poglavju.

2 Tvorjenje glasov in govorjeni jezik

Za uspešen in realističen sintetizator govora je pomembno dobro poznavanje akustične in izgovorne fonetike ter delovanja govoril – ti osnovni procesi so opisani že v Toporišičevi slovnici (Toporišič 2004) in jih tudi na kratko povzemamo. Organi, ki sodelujejo pri tvojenju glasov, so: trebušna prepona, rebra z mišicami prsnega koša, pljuča, sapnici s sapnikom, grlo s pokrovko, žrelna stena, ustna in nosna votlina, jezik, zobje in ustnice. Za tvojenje sta pomembna predvsem izdišni zrak in nihanje glasilik, ki z enakomernim tresenjem tvorijo zven oziroma osnovni ton \(F_0\), tega bomo iskali tudi pri zvenečem \(\gamma\). Glasiliki se treseta pri vsem zvenečih glasovih, pri nezvenečih \(p, f, t, s, c, \check{c}, \check{s}, k, h\) pa ne. Vsi zašepetani glasovi so nezveneči. Ob razprtih glasilkah in odprti nadgasilčni poti slišimo na primer nezveneči grlni \(x\), ko se pripreta in zanihata, pa nastane zveneči grlni \(\gamma\). Za raziskovanje glasov so pomembni predvsem sonogrami, iz katerih so vidni tudi formanti vsakega glasov. Formanti so frekvenčne komponente govora, s katerimi lahko izgoverimo in razločimo različne glasove. O njih so pisali že Lehiste, Petek, Toporišič, Ozbič in Tivadar.\(^1\) Najkrepkejši so pri samoglasnikih, srednja odprtostna stopnja je pri zvočnikih in najmanjša pri nezvočnikih. Nezvočnike delimo na zapornike \(p, t, k, b, d, g\), zlitnike \(c, \check{c}, d\check{z}\) in pripornike \(f, s, \check{s}, h, z, \check{z}\). Glasove delimo še na netrajne (zlitniki in zaporniki) in trajne (Toporišič 2004: 41–45). Fonetično-fonološke zakonitosti so pri sami sintezi zelo pomembna osnova. Ob tem imejmo „v mislih, da je število možnih glasov neskončno ter da je odločitev fonetika, kdaj bo dele glasovne verige prepoznaval kot različne glasove in kdaj kot ponovitve istega glas, arbitrama“ (Golden 1996: 131). Zato je pomembno, da se s transkripcijami govornih zbirk, namenjenimi sintetizatorju, ki morajo biti zaradi poznejše samodejne programske analize enotne, ukvarja čim manj raziskovalcev hkrati, najbolje eden.

\(^1\) O poglobljenih analizah formantov slovenskih samoglasnikov več tudi v članku Petra Jurgea Formant frequencies of standard slovene vowels (slov. Formantne frekvence samoglasnikov standardne slovenščine). Tam svoje raziskave primerja s prejšnjimi in poda v prilogah tudi številčne rezultate. Več o odčitavanju formantov piše Jurgec v članku Natančnost odčitavanja formantov pri digitalnem spektografiранju na podlagi LPC-analize. Hotimir Tivadar piše o formantih v prispevku z naslovom Normativni vidik slovenščine v 3. tisočletju – knjižna slovenščina med realnostjo in idealnostjo.
Cilj same sinteze je, da je umetno tvorjen govor čim bolj podoben naravnemu. „Za človeški govor je značilno, da je zelo spremenljiv. Načini izgovarjanja iste besede si niso nikoli povsem enaki“ (Žganec Gros 2000: 8). „Pri moških je osnovna frekvenca od 60 do 230 Hz, pri ženskah od 180 do 400 Hz /…/ Tako kot hipno ne moremo spremeniti frekvence glasilk, ne moremo hipno preoblikovati govorne cevi. Če želimo sintetizirati sintetizirani govor čim bolj naravno, moramo zagotoviti, da so zvezne spremembe tudi v sintetiziranem govoru“ (Rozman 2005: 2). Medtem ko je sinteza govora sama že zelo dolga, ostaja največja težava „počlovečenje“ in s tem „poslabšanje“ idealnega signala do te mere, da se sliši naravno fluenten in človeški. Pri naravnosti govora se ukvarjamo tudi z lastnostmi govorjenega jezika in njegovim razmerjem do pisnega jezika; pri delu z Govorcem (1, 2 in 3) se s temi vprašanj in redno srečujemo. Govornica, s katero smo se ukvarjali, temeljila je na istih besedilih, ki sta jih bralca brala z različnimi izgovori in naglaševanji besed. Vsaka izgovorjena beseda pa je zaradi samega sistema delovanja Govorca oblikoslovno in izgovorno določena tudi v slovarju izgovarjav, zato smo morali poleg pravilnega fonetičnega zapisa posnetkov skrbeti tudi za pravilni zapis izgovora v slovarju. Tu smo se večinoma zanašali na pravopis, saj je „v slovenskem prostoru pogosto sinonim za reševanje kodifikacijskih reševanj v celoti. Pravzaprav je kot kodifikacijski priročnik obravnaval še pravilen izgovor in tudi vsa slovnično težja mesta slovenskega jezika“ (Tivadar 2012: 203). A smo vseeno od te norme velikokrat zaradi praktičnih razlogov odstopili. Rešitve v pravopisu so rojenemu govorcu velikokrat tuje in neznane. Pri samih bralcih se je v končni analizi potrdilo splošno dejstvo, da za raziskovanje govorjenega jezika ni pomembna samo kvantiteta posnetkov, ampak tudi številčnost govorcev in njihova socialno-geografska raznolikost. Bralca sta imela kljub isti pravorečni šoli na nacionalnem radiu različne glasovne značilnosti. „Tako je npr. reduciranje nenaglašenih in tudi naglašenih samoglasnikov /…/ v knjižnem govoru nezaželeno“ (Tivadar 2012: 209), a se v posnetkih pri besedah s kratkim naglašenim a redukcija vseeno pojavlja. Notacijo, ki smo jo od začetka uporabljali pri transkripcijah za Govorca, smo morali zaradi razkoraka med teorijo in samimi posnetki naknadno spremeniti. Vse simbole glasov, ki smo...

---

2 V slovarju izgovarjav smo najprej popravili večino glagolov. Napačno so bili naglašeni na primer predvsem deležniki moškega spola na –l, ki imajo predpisan naglas na korenskem samoglasniku, velikokrat so manjkale tudi dvojnice. Večkrat smo spreminjali tudi izgovore tujk in izgovor črke l kot y.

3 Take spremembe so: uvedba ozkega izgovora o v prihodniku bo, čeprav je nenaglašen in bi moral biti izgovorjen široki ɔ, uvedena dvojnica pri glavnih števnikih na -ajst in na primer popravek izgovora džójnt v džôjnt.

4 O razmerju med govorjenim in pisnim jezikom več v člankih Hotimirja Tivadarja Vzpostavitev razmerij med govorom in branjem, recitacijo in igranjem, Gradiva utemeljenost opisa slovenskega govorjenega (knjižnega) jezika, Nevarna razmerja med pisnim in govorjenim jezikom in Nove usmeritve pri raziskavah govora s pogledom v preteklost. Pri delu z Govorcem je izredno pomembno, da se zavedamo nedoločenosti tega razmerja, saj se s takšnimi projektji vendarle (ne)hote vrednoti in usmerja govorjeni knjižni jezik.
jih uporabljali, z opisi in primeri položajev v besedah navajamo v Tabeli 1, saj so tudi fonetični zapisi v diplomskem delu v notaciji MRPA. Poleg pisave MRPA, ki smo jo uporabljali pri Govorcu, navajamo tudi mednarodno fonetično abecedo IPA.

<table>
<thead>
<tr>
<th>Glas⁵</th>
<th>MRPA</th>
<th>IPA⁶</th>
<th>Opis glasu</th>
<th>Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>a</td>
<td>kratki a</td>
<td>riba</td>
</tr>
<tr>
<td>A1</td>
<td>&quot;a&quot;</td>
<td>'a'</td>
<td>naglašeni kratki a</td>
<td>brát</td>
</tr>
<tr>
<td>A2</td>
<td>&quot;a:&quot;</td>
<td>'a:'</td>
<td>dolgi naglašeni a</td>
<td>sám</td>
</tr>
<tr>
<td>B</td>
<td>b</td>
<td>b</td>
<td>objem</td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>b_n</td>
<td>ŉ</td>
<td>favkalni b</td>
<td>óhe</td>
</tr>
<tr>
<td>B2</td>
<td>b_f</td>
<td>b'</td>
<td>zobnoustnični b</td>
<td>ob figi</td>
</tr>
<tr>
<td>C</td>
<td>ts</td>
<td>ts</td>
<td>c</td>
<td>strie</td>
</tr>
<tr>
<td>C1</td>
<td>dz</td>
<td>dz</td>
<td>c pred nezvočnikom</td>
<td>Kocbek</td>
</tr>
<tr>
<td>Č</td>
<td>tS</td>
<td>t'</td>
<td>č</td>
<td>človek</td>
</tr>
<tr>
<td>Č1</td>
<td>dZ</td>
<td>dʒ</td>
<td>džungla</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>d</td>
<td>d</td>
<td>děd</td>
<td></td>
</tr>
<tr>
<td>D1</td>
<td>d_l</td>
<td>d'_</td>
<td>obstranski d</td>
<td>dléto</td>
</tr>
<tr>
<td>D2</td>
<td>d_n</td>
<td>d</td>
<td>favkalni d</td>
<td>dno</td>
</tr>
<tr>
<td>E</td>
<td>&quot;e:&quot;</td>
<td>'e:'</td>
<td>dolgi ozki naglašeni e</td>
<td>Péter</td>
</tr>
<tr>
<td>E1</td>
<td>E</td>
<td>e</td>
<td>kratki široki e</td>
<td>deréza</td>
</tr>
<tr>
<td>E2</td>
<td>&quot;E&quot;</td>
<td>'e'</td>
<td>naglašeni široki e</td>
<td>kméť</td>
</tr>
<tr>
<td>E3</td>
<td>&quot;E:&quot;</td>
<td>'e:'</td>
<td>naglašeni široki dolgi e</td>
<td>těta</td>
</tr>
<tr>
<td>E4</td>
<td>@</td>
<td>ə</td>
<td>polglosnik</td>
<td>maglá</td>
</tr>
<tr>
<td>E5</td>
<td>&quot;@&quot;</td>
<td>'ə'</td>
<td>naglašeni polglosnik</td>
<td>magla</td>
</tr>
<tr>
<td>F</td>
<td>f</td>
<td>f</td>
<td>fant</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>g</td>
<td>g</td>
<td>glábol</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>x</td>
<td>x</td>
<td>měhko</td>
<td></td>
</tr>
<tr>
<td>H1</td>
<td>G</td>
<td>ɣ</td>
<td>h pred zvenecím nezvočnikom</td>
<td>súh grím</td>
</tr>
<tr>
<td>I</td>
<td>i</td>
<td>i</td>
<td>kratki i</td>
<td>mámi</td>
</tr>
<tr>
<td>I1</td>
<td>&quot;i&quot;</td>
<td>'i'</td>
<td>kratki naglašeni i</td>
<td>šit</td>
</tr>
<tr>
<td>I2</td>
<td>'i:'</td>
<td>'i:'</td>
<td>dolgi naglašeni i</td>
<td>píla</td>
</tr>
<tr>
<td>J</td>
<td>j</td>
<td>j</td>
<td>jájce</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>k</td>
<td>k</td>
<td>krik</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>l</td>
<td>l</td>
<td>labilen</td>
<td></td>
</tr>
<tr>
<td>L1</td>
<td>l'</td>
<td>l'</td>
<td>mehčani l</td>
<td>póljski</td>
</tr>
<tr>
<td>M</td>
<td>m</td>
<td>m</td>
<td>mímám</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>F</td>
<td>m</td>
<td>zobnoustnični m</td>
<td>simfonija</td>
</tr>
<tr>
<td>N</td>
<td>n</td>
<td>n</td>
<td>Nina</td>
<td></td>
</tr>
<tr>
<td>N1</td>
<td>N</td>
<td>ň</td>
<td>mehkonobni n</td>
<td>Anhovo</td>
</tr>
</tbody>
</table>


⁶ Notacija IPA je povzeta po članku Računalniški simbolni fonetični zapis slovenskega govora avtorjev Zemljak idr.
Želja ustvariti stroj, ki bi bil zmožen govora in prostega glasovnega odzivanja, vznemirja znanstvenike že od leta 1930, ko je Homer Dudley razvil sintetizator z imenom VODER (angl. Voice Operating Demonstration); upravljal se je z roko, nogo in prsti. Sicer je že v drugi polovici 18. stoletja Christianu Kratzensteinu uspelo z mehanično napravo s pomočjo cevi, oblikovanih po govornem traktu, izgovarjati samoglasnike. Konu 19. stoletja so razvili naprave, namenjene snemanju citatov, z njimi so v poslovnem svetu zmanjšali stroške najema stenografov. Snemalniki so delovali na principu mikrofona, zvočni pritisk je na valj, oblit z voskom, z iglo, podobno gramofonski, vtisal zvočni signal, ki se je lahko naknadno večkrat predvajal. Leta 1952 so Davis, Biddulph in Balashek razvili sistem, ki je s formantnimi in kvantitetnimi analizami že imel sposobnost prepoznave govora posameznega govorca. V šestdesetih letih so tako sistemi poznali besedišče, veliko od 10 do 100 besed, ki

Tabela 1: Seznam glasov

<table>
<thead>
<tr>
<th>Znak</th>
<th>Glavni glas</th>
<th>Zunanji glas</th>
<th>Dvoglasniški glas</th>
<th>Sensibilni glas</th>
<th>Oznaka</th>
</tr>
</thead>
<tbody>
<tr>
<td>N2</td>
<td>n'</td>
<td>n'</td>
<td>mehčani n</td>
<td>sánjski</td>
<td></td>
</tr>
<tr>
<td>N3</td>
<td>F</td>
<td>mj</td>
<td>zobnoustnični n</td>
<td>informácia</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>'o:</td>
<td>o:</td>
<td>ozki dolgi naglašeni o</td>
<td>nós</td>
<td></td>
</tr>
<tr>
<td>O1</td>
<td>0</td>
<td>0</td>
<td>kratki široki o</td>
<td>vodnár</td>
<td></td>
</tr>
<tr>
<td>O2</td>
<td>'O</td>
<td>'O</td>
<td>kratki naglašeni o</td>
<td>snôp</td>
<td></td>
</tr>
<tr>
<td>O3</td>
<td>'O:</td>
<td>o:</td>
<td>široki dolgi naglašeni o</td>
<td>vóda</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>p</td>
<td>p</td>
<td>favkalni p</td>
<td>pipa</td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>p_n</td>
<td>p</td>
<td>favkalni p</td>
<td>rôb mize</td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>p_f</td>
<td>p'</td>
<td>zobnoustnični p</td>
<td>Sápio</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>r</td>
<td>r</td>
<td>tráva</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>silos</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>S</td>
<td>ŵ</td>
<td>ŵ</td>
<td>šumiš</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>tát</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>t_l</td>
<td>t'</td>
<td>obstranski t</td>
<td>tlá</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>t_n</td>
<td>t</td>
<td>favkalni t</td>
<td>tnálo</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>u</td>
<td>u</td>
<td>kratki u</td>
<td>brátu</td>
<td></td>
</tr>
<tr>
<td>U1</td>
<td>'u'</td>
<td>'u'</td>
<td>kratki naglašeni u</td>
<td>krúh</td>
<td></td>
</tr>
<tr>
<td>U2</td>
<td>'u:</td>
<td>'u:</td>
<td>dolgi naglašeni u</td>
<td>krúha</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>v</td>
<td>v</td>
<td>siva</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V3</td>
<td>U</td>
<td>U</td>
<td>dvoglasniški u</td>
<td>siv</td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>U</td>
<td>U</td>
<td>dvoglasniški u</td>
<td>pól</td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>z</td>
<td>z</td>
<td>niz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ž</td>
<td>Z</td>
<td>Ž</td>
<td>žába</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3 Zgodovina sintetizatorjev govora
so ga prepoznavali z enostavnimi akustičnofonetičnimi lastnostmi posnetega govora. V sedemdesetih letih se je besedišče sintetizatorjev povečalo in je tako obsegalo od 100 do 1000 besed. Uvedle so se metode LPC, ena izmed njih se uporablja pri Govorcih 1 in 2. V osemdesetih letih je bilo poznavanje besed že neomejeno, prepoznavava govora pa je temeljila na statističnih metodah predvidevanja. Tu že govorimo o metodi PMM (prikriti modeli Markova), ki jo bomo podrobneje razložili v poglavju Sinteza na strani 9. V devetdesetih letih se je prepoznavava govora dvignila na raven razumevanja in prepoznavanja skladnje v prostem govoru. Zdaj smo že na stopnji, ko je sinteza govora postala uporabna v vsakodnevni in poslovni rabi. 

V Sloveniji se je zanimanje za sintezo slovenščine pojavilo veliko pozneje. Vzrok je verjetno v politični situaciji v preteklosti in samem (ne)dojemanju slovenščine kot govorjenega knjižnega jezika v Jugoslaviji. Prvi z izkušnjami s sintetizatorji govora slovenskega jezika je Hribar. Razvil je postopek za samodejno pretvarjanje slovenskega besedila v fonetični prepis besedila. Razvoj sintetizatorja govora se je nadaljeval na Institutu „Jožef Stefan“ v Ljubljani. Weilguny je, podobno kot pred njim Hribar, besedilo pretvoril v fonetični prepis, mu na zelo preprost način priredil prozodične parametre in ga posredoval dvema formantnima sintetizatorjema, razvitima za angleški jezik (Weilguny 93). Raziskave na omenjenem inštitutu sta nadaljevala Dobnikar in Šef (Šef 96). Razvili so difonski sintetizator govora (Šef 98), ki so ga pred nedavnim poklonili v uporabo slepim in slabovidnim osebam (Žganec Gros 2000). Ravno ta Šefov difonski sintetizator (Govorec 1) je tisti, ki ga zdaj uporablja tudi trenutni Govorec 2. „Program s/m/o razvili na Odseku za inteligentne sisteme (E9) na Institutu „Jožef Stefan“ v sodelovanju s podjetjem Amebis in MMC RTV Slovenija“ (http://dis.ijs.si/rtv-govorec/, dostop 29. 4. 2013). Razvoj in raziskave s področja govornih tehnologij za slovenski jezik so še vedno dejavne. Izvajajo jih na Fakulteti za elekrotehniko Univerze v Ljubljani (Laboratorij za umetno zaznavanje, sisteme in kibernetiko), na Fakulteti za elekrotehniko, računalništvo in informatiko Univerze v Mariboru, na Fakulteti za računalništvo in informatiko Univerze v Ljubljani, na Institutu „Jožef Stefan“, na Naravoslovnotehniški fakulteti Univerze v Ljubljani in v podjetjih

7 Metoda LPC se uporablja pri sintetizatorjih, ki temeljijo na lepljenju govornih enot (npr. dvoglasov). Analiza in sinteza pri tej metodi potekata na podlagi linearne predikcije (angl. linear predictive coding).
9 Pri navajanju celega imena Instituta „Jožef Stefan“ v Ljubljani bomo uporabljali obliko Institut, kot je tudi uradni naziv tega inštituta, drugače pa bomo uporabljali inštitut, saj je to tudi primarna izbira v SP. V Gigafidi smo preverili tudi pogostost rabe oblike institut, ki se uporablja skoraj petkrat manj kot inštitut.
Masterpoint, d. o. o., Hermes Softlab, d. d., in Alpineon razvoj in raziskave, d. o. o. (Žganec Gros idr. 2003: 47). Na seznam lahko dodamo še podjetje Amebis, d. o. o.¹⁰

4 Sintetizator govora

Sinteza je umetno poustvarjanje govora, današnji cilj je, da bi bil čim podobnejši naravnemu govoru. Tu se pojavijo največji izzivi tako za jezikoslovce kot za informatike. Govoreči stroji so različni. Lahko so samodejni odzivniki, ki imajo vse besede posnete že vnaprej, predvsem so uporabni v položajih, ko gre za skladiško manj zahtevna besedila (v dvigalih, na železniških postajah …). Pri obsežnejših besedilih se uporablja samodejna sinteza posameznega zapisanega besedila, kar obravnavamo tudi v diplomskem delu. Za dober sintetizator so pomembne predvsem tri lastnosti:

- kakovost sinteze,
- obseg možnih izgovorjenih besed, stavkov, intonacij,
- nezapletenost izdelave in nezahtevnost računskih ter drugih operacij.¹¹

Postopki ustvarjanja oziroma delovanja sintetizatorja govora so različni. Pri prvih je potekala sinteza govora v naslednjih fazah:

1. faza: grafemsko-fonemska pretvorba besedila,
2. faza: nastavljanje prozodičnih parametrov,
3. faza: izgradnja govorne zbirke,
4. faza: preizkus govorca.

¹⁰ Amebis, d. o. o., je podjetje, ki se ukvarja z jezikovnimi tehnologijami, cilj razvoja je predvsem čim boljše razumevanje besedila in s tem boljši samodejni lektorji, sintetizatorji govora, virtualni asistenti. Več podatkov o podjetju in njihovih projektih je dosegljivih na njihovi spletni strani (http://www.amebis.si/).

Pri novejših postopkih sinteze je 2. faza (nastavljanje prozodičnih lastnosti) samodejna s pomočjo programske analize govornih signalov, ki transkripcijo in govorni signal vzporedno statistično primerja. Rezultati analize se pozneje ročno popravijo in izpopolnijo. Izgradnja zbirke je sicer vedno potrebna, a so njeni postopki in elementi za različne vrste potvorbe prilagojeni.

S podjetjem Amebis, d. o. o., sodelujem slabi dve leti, na začetku sem se ukvarjal predvsem z naglaševanjem in popravljanjem glagolov v slovarju izgovarjav, nato pa s preverjanjem transkripcije posnetkov iz njihove govorne baze. To delo je tudi vir in navdih diplomskega dela. Temo sem izbral, ker sem želel raziskati aplikativnost ob praktičnih izkušnjah s področja fonetike, notacije, pravorečja (in s tem tudi srečevanje s sociolingvistiko). Pri delu s sintetizatorjem govora sem počasi spoznal tudi obsežnost področja raziskovanja svojega dela v vseh korakov, ki so za dober sintetizator potrebni:

1. korak: slovnična analiza vhodnega besedila,
2. korak: nadomeščanje simbolov, števk in kратic z besedami,
3. korak: grafemsko-fonemska pretvorba (s slovničnimi pravili in bazo znanih besed),
4. korak: določanje prozodičnih parametrov in
5. korak: sintetiziranje govornega signala (Rozman 2005: 3–4).

Zaradi obsežnosti teme in predvsem slovenističnega vidika na temo raziskovanja v diplomskem delu bomo delovanje sintetizatorja govora razdelili na dva dela: na obdelavo vhodnega besedila (slovnična analiza, grafemsko-fonemska pretvorba, težave ročnega popravljanja transkripcije, določanje prozodiče …) in na samo sintezo govornega signala ter lastnosti govornih zbirk, ki jih ta potrebuje.

Najprej se bomo posvetili sintezi – vrstam sinteze (čeprav je to zadnja stopnja potvorbe), saj jih lahko predstavimo na kratko, ker je to za jezikoslovca korak, na katerega nima večjega vpliva. Nato bomo v poglavju *Vhodno besedilo* (stran 8) obravnavali še prve štiri točke pretvorbe, pri katerih bomo sproti opisali tudi vse prepreke, ki so se pojavljale pri delu z Govorcem 3.
5 Sinteza

Z besedo sinteza, kot smo že povedali v uvodu, označujemo v tem diplomskem delu sam proces umetnega poustvarjanja človeškega govora, to je zadnja stopnja delovanja sintetizatorja govora (peti del po Rozmanovi delitvi korakov pretvorbe v umetni govor). Ker je teh več vrst, so tudi tipizacije sintez različne. „Po eni izmed delitev se postopki za sintezo govornega signala delijo v naslednje tri glavne skupine [d'Alessandro 95]: artikulatorni sintetizatorji govora, formantni sintetizatorji govora in sintetizatorji govora, ki temeljijo na združevanju osnovnih enot“ (Žganec Gros 2000: 98). Prvi so fizični modeli, zasnovani na podlagi človeškega govornega trakta, pretoka zraka in nastavljanja parametrov za tlak, napetost glasilk, položaj jezika ... Formantni sintetizator temelji na domnevi, da se lahko vpliv in obnašanje govornega trakta opišeta s formantnimi frekvencami in amplitudami, pri sintezi pa pride do umetne rekonstrukcije formantnih značilnosti. Pri drugih delitvah se znotraj tretjih sintetizatorjev po d'Alessandru pojavijo različne podvrste, razlikujejo se predvsem po osnovnih enotah, ki jih uporabljajo za združevanje (difoni, besede, stavki). K novejšim metodam spada metoda PMM (prikriti modeli Markova) oziroma v angleščini HMM (angl. hidden Markov models). Tu je že govora o korpusnih sintetizatorjih. Njihova posebnost je, da imajo v primerjavi s starejšimi več posnetkov iste govorne enote, zato lahko izbirajo najprimernejšo in s tem zmanjšajo agresivno prilagajanje govornih signalov, kot se to dogaja pri Govorcu 1, ki ima v bazi samo po en primer enega difona, zato mora za sestavitev vseh v eno besedo veliko bolj zgraditi signale elementov (osnovni ton in moč), kot bi jih bilo treba, če bi imeli pri postopku sinteze možnost izбирati različne posnetke istih difonov. Primernost bi se določala po podobnosti končnega signala enega difona z začetkom drugega. „Postopek sinteze govora z uporabo PMM-jev se razlikuje od bolj razširjenih postopkov v tem, da ogrodje PMM-jev ne uporablja zgolj za segmentacijo in označevanje govorne zbirke, pač pa gre še korak naprej in ga uporablja tudi kot model za tvorjenje govora“ (Vesnicer idr. 2004). Slika 1 iz tega članka prikazuje proces sinteze po metodi PMM.

12 Difon oziroma dvoglasnik je kombinacija dveh glasov, ki sta izgovorjena eden za drugim. Kot primer so na strani 14 navedeni difoni besede lipa, ti so: _l, l”i:, ”ip, pa, a_ (podčrtaj nakazuje tišino).
koeficiente iz posnetka človeškega govora, torej vse podatke, ki so potrebni za poznejšo uresničitev umetnega govora. Ker pri izgovoru glasu u ne prihaja do turbulenc v govorni cevi, je šumni del glasu zanemarljiv, pri glasu z pa zaradi zvenečnosti pričakovano vidimo tudi harmonski del, a večino energije pri signalu predstavlja šumni del.

Slika 2: Dekompozicija glasu u na harmonski in šumni del (Rozman 2005: 10)
Po seznanitvi z večino sintez lahko zaključimo, da je razdelitev sintetizatorjev po d'Alessandru preozka; tako lahko sintetizatorje razdelimo na:

a) artikulatorne,

b) formantne,

c) sintetizatorje, ki združujejo enote (npr. difonski),

č) sinusne (varianta sinusnih je tudi metoda HNM),

d) tiste, ki temeljijo na metodi HMM oz. modelih Markova.

Druga možna delitev, ki bi bila slovenistom tudi veliko bližje, saj je glavni del našega dela pri vhodu v sintetizator govora, bi se osredotočala na različne vrste vhodnega besedila. Če gre za omejene možnosti branja, ponavadi le za lepljenje vnaprej posnetih stavkov, bi govorili o zelo enostavnih in precej omejenih, specializiranih sintetizatorjih. Kadar pa gre za t. i. „text to speech“, ali poslovenjeno „prenos besedila v govor“ sintetizatorje, poimenujejo jih tudi korpusni sintetizatorji, bi govorili o splošnih ali samodejnih sintetizatorjih, ki
besedišča ne morejo predvideti in zato za delovanje potrebujejo velike slovarje in govorne zbirke.

5.2 Sintetizator govora Govorec

Govorec je ime sintetizatorja govora, ki so ga sprva razvili na Institutu „Jožef Stefan“ v Ljubljani. Trenutno obstajajo tri verzije Govorca. Drugo je razvil Amebis, d. o. o., sam pa sem se pri delu v podjetju ukvarjal z Govorcem 3, ki bo deloval z novo vrsto sinteze (metoda HNM namesto TD-PSOLE) in novo govorno zbirko (več ur posnetkov govorjenega besedila namesto baze difonov); prav tako ga razvija Amebis, d. o. o. Govorca 1 in 2 pa se razlikujeta v tem, da Govorec 2 uporablja novo bazo besed oziroma slovar izgovarjavi, ki je veliko obširnejši in bolj izpopolnjen. Ker se pri pojasnjevanju temeljev delovanja Govorcev 1, 2 in 3 ne moremo izogniti terminu govorna zbirka, bomo v tem poglavju opisali še tega.

5.2.1 Sintetizatorja Govorca 1 in 2

pa frekvenčno in fazno ne obdelujeta trajajočih nezvenečih glasov, saj ti nimajo periode in s tem osnovne frekvence. Ti se po mejah usklajujejo le po energiji. Večje težave se pojavijo pri glasu r, saj ima ta zelo nestalo periodo. Sintetizatorji govora večinoma lepijo posnetke iz govornih baz. Če je npr. vsak difon besede lipa (\texttt{\_l}, \texttt{l"i:}, \texttt{\_ip}, \texttt{\_pa}) element v govorni bazi (glasove morajo predhodno posneti posamezniki), je shema sinteze besede lipa takšna:

![Shema sinteze besede lipa po postopku TD-PSOLA](image)

**Slika 4:** Shema sinteze besede lipa po postopku TD-PSOLA

Pri lepljenju elementov (npr. difonov) se morata usklajati ton in energija, saj je nemogoče vse glasove posneti z isto močjo in višino tona. Ton (\(f_0\) oz. osnovna frekvenca) dveh elementov (npr. \(l\) in \(l"i:\)) bi se izenačil zvezno in počasi, kot se tudi pri človeškem glasu. Pomembni besedi pri lepljenju elementov sta prehod in prelijanje. V difonih se namreč prehodi med glasovoma (npr. glasovoma \(l\) in \("i:"\)) ne spremijajo, energija in frekvenca se ne prilagajata. Zlivanje pa se zgodi med končnim delom prvega difona in začetnim delom drugega difona, v primeru lepljenja difonov \(l\) in \(l"i:"\) se v procesu prelijeta glasova \(l\) iz obeh difonov, notranji prehodi v difonih pa se ne spremijajo. Postopek lepljenja je računanje matematične utežene vsote govornih signalov, ki je viden na Sliki 5.

---

13 Osnovno frekvenco glasu smo iskali s programom Praat in \(F_0\).
14 Pri besedi prelijanje gre za slovenski prevod angleške besede „crossfade“, pogosta v krogih avdioinženiringa.
Slika 5: Prelivanje difonov

Naglas besede in osnovni ton besede prikazuje analiza besede *zavóhala* (Slika 6).

Slika 6: Naglas in analiza besede *zavóhala*

5.2.3 Sintetizator Govorec 3

Za Govorca 3 so pri Amebisu, d. o. o., ustvarili govorno zbirko, ki sta jo posnela profesionalna napovedovalca z nacionalnega radia, oba naj bi predstavljala normo oziroma standard današnjega pravorečja. Gre za dvajset ur posnetkov, ki so razdeljeni na posamezne odseke, ponavadi so to posamezne povedi. Te povedi so vzete iz korpusa besedila in so izbrane tako, da je končni rezultat čim širši izbor kombinacij glasov, možnih intonacij, podredij, besednih vrst (npr. medmeti) … Teh dvajset ur posnetkov (deset ur na enega bralca) je razdeljenih na okoli 4000 delov pri vsakem bralcu (del je večinoma poved). Poleg posnetka govora je posnet še signal utripanja glasilka, kar olajša analizo (predvsem zvenečnosti in nezvenečnosti); govorimo o laringografskem posnetku utripanja glasilka. Ker so se posnetki delali večkrat, so med njimi slišne razlike. Napovedovalca se proti koncu posamezne seje bolj motita, bereta hitreje ali počasneje, spreminjata besedilo in požirata glasove. Na posnetkih se pojavlja tudi šum listov, kar lahko poznejšo računalniško analizo moti, saj je šum papirja zelo podoben šumu človeškega govora. Ta šum pa je pri analizi HNM, ki jo Govorec 3 uporablja, izredno pomemben, zato je tu nevarnost napačne samodejne analize (šuštenje listov je lahko dojeto kot glas). Prednost Govorca 3 bo predvsem široka možna izbira različnih glasov, besed ali pa celo kar celih stavkov. Govorec 3 bo v govorni zbirki, ki so jo posneli, iskal največje možne dele posnetkov, ki jih bo izbral iz baze. Če bo teh možnosti več, bo izbral najprimernejše, takšne, ki bodo potrebovali na prehodih najmanj korekcije. Ker se je pri transkripcijah označevala tudi makroprozodija, bo tako naravnost govora še večja. Pričakuje
se (saj še ni bil testiran), da se bo npr. izboljšala tudi intonacija (vprašalna, povedna, vzklična). Naravnejši bi morali biti tudi prehodi pri različnih stavčnih podredjih ali priredjih. Naravnejši bodo odmori. Na posnetkih so v transkripciji označeni in ohranjeni tudi kašlj, brundanje, razna mašila, ki se bodo lahko nato naključno pojavljala med sintezo, da bo ta še bolj človeška.

5.2.4 Govorne zbirke

Govorne zbirke so obvezni del tako pri današnjih jezikoslovnih raziskavah kot pri ustvarjanju umetnega govora. Zbirke posnetkov so zelo različne. Predvsem pa se njihove lastnosti oblikujejo glede na namen. Pri Govorcu 3 je bilo nujno dobiti čim več možnih kombinacij besed in intonacij dveh bralcov, saj bo lahko uporabnik izbiral med dvema bralcema, in sicer moškim in ženskim glasom. Pomembno je bilo tudi, da obvladata zborni in knjižni pogovorni jezik, ki naj bi bil trenutno standarden. Čeprav se pri tem med slovenisti občasno lomijo različna mnenja, kako bi moralo biti in ni, sta izbrana bralca trenutno reprezentativna predstavnika norme.15 V zbirki so posnete tudi besede, kot so pika, vejica in evro, saj je Govorec uporaben predvsem za slabovidne, ki potrebujejo popolno informacijo o branem besedilu. Govorno zbirko opredeljujejo posnetki govornih signalov, priložene oznake in transkripcije. Pod oznake spadajo tudi podatki o govorcih (spol, starost, poklic, izobrazba ...). „Najpogosteji so mikrofonski posnetki zvočnih signalov, ki se pojavljajo med govorjenjem, in laringografski posnetki vibriranja glasilk v grlu. Uporabljajo se tudi vizualni in rentgenski posnetki govoril. Prvi se nanašajo zgolj na slikovne posnetke obraza, predvsem ustnic, drugi pa na globinske slike prereza celotnega sistema govoril. V zadnjem času se pogosto analizirajo tudi posnetki biomedicinskih signalov, ki se pojavljajo v govorilih in tudi človeških možganih“ (Žganec Gros idr. 2003: 49).

„Gradnja /govorne/ korpusa mora biti skrbno načrtovana, določajo pa jo poleg namena korpusa tudi finančne zmožnosti in razpoložljivti človeški viri (raziskovalci, zbiralci gradiva, transkriptorji, označevalci itd.) …/. Pomembno je, da se načrtovalci korpusa zavedajo, da je gradnja korpusa ciklični proces, ki ob sprotnem evalvirjanju lahko zahteva spreminjanje začetnih izhodišč, pa tudi proces, ki se v idealnih okoliščinah ne bi nikoli

15 Več o standardnem in knjižnem jeziku v drugem poglavju na strani 2.

5.2.5 Slovenske govorne zbirke

V Sloveniji je o govornih zbirkah izdala monografijo Jana Zemljarič Miklavčič, iz katere povzemamo skrajšan seznam in opis nekaterih slovenskih govornih zbirk, ki že obstajajo. Seznam je dopolnjen še z novejšim korpusom GOS (korpus GOvorjene Slovenščine).

- Baza izgovarjav SNABI
  - Izdelana na Fakulteti za elektrotehniko, računalništvo in informatiko Univerze v Mariboru;
  - 128 govorcev;
  - vsak govorec izgovori povprečno 200 stavkov, 80 izoliranih besed, številne nize in abecedo;
  - členjena in fonetično transkribirana.

- Podatkovna zbirka GOPOLIS
  - Izdelana na Fakulteti za elektrotehniko Univerze v Ljubljani;
  - narejena za razvoj sistema za razpoznavo govora in krmilnika dialoga pri govornih poizvedbah o letalskih informacijah;
  - posnetih 25 moških in 25 žensk;
o nima urejenih avtorskih pravic.

- SPEECH-DAT
  o Evropski projekt;
  o za razvoj sistemov telefonskega govornega dialoga za delo v realnem okolju;
  o 1000 govorcev.

- POLIDAT
  o Tudi razvit za avtomatski telefonski sistem GOVORNEGA DIALOĞA;
  o izbranih 1000 govorcev;
  o brana vnaprej pripravljena besedila;
  o majhen delež spontanega govora (Žemljarič Miklavčič 2008: 52–53).

- Govorni korpus GOS
  o Označujejo ga kot korpus govorjene slovenščine;
  o 120 ur posnetkov (po)govora v najrazličnejših situacijah (radijske in televizijske oddaje, zasebni pogovori …);
  o zapis govora na posnetkih je narejen v dveh različicah, standardizirani in pogovorni, ter obsega več kot milijon besed;
  o namenjen vsem, ki želijo raziskovati govorjeno slovenščino (http://www.korpus-gos.net/Support/About, dostop 4. 5. 2013).

Govorne baze se ne zbirajo le za sintetizatorje, ampak tudi razpoznavalnike govora. Govorna baza SNABI je bila prva dostopna baza za razpoznavanje govora. „S stališča kompleksnosti razpoznavalnikov govora so najobsežnejši jezikovni viri potrebni za razpoznavalnike tekočega govora. Če želimo pravilno oceniti parametre akustičnih in jezikovnih modelov, potrebujemo za izpeljavo postopka učenja velike količine učnih podatkov“ (Žgank idr. 2006). O slovenskem jezikovnem viru SloParl, ki ga sestavljata govorni in besedilni korpus, pišejo Žgank in drugi v članku z naslovom Slovenska govorna in tekstovna baza parlamentarnih razprav za avtomatsko razpoznavanje govora.
6 Vhodno besedilo

Vhodno besedilo je katero koli besedilo, ki mora biti strojno prebrano. Najprej je treba to besedilo obdelati in prilagoditi. Ker se v nadaljevanju dela velikokrat sklicujemo na slovar izgovarjav, ga bomo opisali pred postopki analize samega besedila.

6.1 Slovar izgovarjav kot del Govorca


<table>
<thead>
<tr>
<th>ljubeški+p</th>
<th>ljubeških</th>
<th>ljubežen</th>
<th>ljubeženljubljev</th>
<th>ljubeženskajedelcev</th>
<th>ljubeženskačakaljev</th>
<th>ljubeženskajedelcev</th>
<th>ljubeženskačakaljev</th>
<th>ljubeženskajedelcev</th>
<th>ljubeženskačakaljev</th>
</tr>
</thead>
<tbody>
<tr>
<td>ljubeški+p</td>
<td>ljubeških</td>
<td>ljubežen</td>
<td>ljubeženljubljev</td>
<td>ljubeženskajedelcev</td>
<td>ljubeženskačakaljev</td>
<td>ljubeženskajedelcev</td>
<td>ljubeženskačakaljev</td>
<td>ljubeženskajedelcev</td>
<td>ljubeženskačakaljev</td>
</tr>
<tr>
<td>ljubeški+p</td>
<td>ljubeških</td>
<td>ljubežen</td>
<td>ljubeženljubljev</td>
<td>ljubeženskajedelcev</td>
<td>ljubeženskačakaljev</td>
<td>ljubeženskajedelcev</td>
<td>ljubeženskačakaljev</td>
<td>ljubeženskajedelcev</td>
<td>ljubeženskačakaljev</td>
</tr>
<tr>
<td>ljubeški+p</td>
<td>ljubeških</td>
<td>ljubežen</td>
<td>ljubeženljubljev</td>
<td>ljubeženskajedelcev</td>
<td>ljubeženskačakaljev</td>
<td>ljubeženskajedelcev</td>
<td>ljubeženskačakaljev</td>
<td>ljubeženskajedelcev</td>
<td>ljubeženskačakaljev</td>
</tr>
<tr>
<td>ljubeški+p</td>
<td>ljubeških</td>
<td>ljubežen</td>
<td>ljubeženljubljev</td>
<td>ljubeženskajedelcev</td>
<td>ljubeženskačakaljev</td>
<td>ljubeženskajedelcev</td>
<td>ljubeženskačakaljev</td>
<td>ljubeženskajedelcev</td>
<td>ljubeženskačakaljev</td>
</tr>
</tbody>
</table>

Slika 7: Slovar izgovarjav (delovno okno)
V MSD-jih (angl. morpho-lexical descriptor), ki jih vidimo v drugem stolpcu delovnega okna, so besede označene z besednimi vrstami in podrobnimi podklasifikacijami, nekaj jih naštevamo kot primere za lažjo predstavo delovanja slovnične analize:

- samostalnik (vrsta {naselje, občno ime, lastno ime, voda, podjetje ..}, spol, sklon, število, določnost, živost, posebej pomensko število in navadno število);
- glagol (vrsta {polnopomenski, naklonski, vezni}, glagolska oblika {povednik, velelnik, pogojnik, nedoločnik ..}, čas, oseba, število, spol, način {tvornik, trnik}, niknalost, vrsta pomožnika {biti, imeti} in vid);
- pridevnik (stopnje, določnost ..);
- prislov (krajevni {kam, kje}, načinovni, kratnostni, merski, pogostnostni);
- zaimek;
- predlog (določen sklon in število, ki ga predlog zahteva ..);
- veznik (priredni, podredni, notranji in vprašalnice);
- členek;
- člen;
- medmet;
- števnik;
- okrajšava;
- ločilo;
- napake (napačen vrstni red, sumljiva oblika, premena odveč, sporna -lec in -vec, zatipkana oblika, množina namesto dvojine ..).

V slovarju izgovarjajo se uporablja pet stolpcev. Kot primer izpostavljamo iz slovarja besedo solata v imenovalniku množine.

| solata+Soz* | Sozmi---m------- | solate | soláte |

V prvem je osnovna oblika, v drugem oznaka MSD, v tretjem je beseda z vsemi oblikoslovnimi lastnostmi, v četrtem pa je beseda še naglašena. Če besede grafemsko-fonemska pretvorba ne pretvori pravilno ali da gre za kakšno glasovno izjemo znotraj pravil, je treba to zapisati v peti stolpec, vanj se zapiše pravilni fonetični zapis v pisavi MRPA, tako Govorec uporabi za branje ta zapis namesto uporabe pravil. Težave četrtega in petega stolpca so predvsem v dolgotrajnem in ročnem delu vnašanja.
6.1.2 Slovar izgovarjav in Govorec 3

Peti stolpec (fonetični zapis) v slovarju izgovarjav se je kot pomemben in nujno potreben korektiv izkazal predvsem pri nenaglašenih besedah. Prihodnjik biti bo je nenaglašen. Ker je nenaglašen, ga grafemsko-fonemska pretvorba zapiše v [bO]. Tako nenaglašeni bo, čeprav naj bi bil ozek (tako ga bereta tudi napovedovalca), sintetizator bere s širokim ɔ.16 Vse te oblike smo morali vnašati v slovar posebej v peti stolpec. Slovar ima tudi možnost vnašanja dvojnic. Besede se zapišijo ena za drugo, a se vseeno vedno razvrstijo hierarhično, saj slovar primarno izbere obliko na prvem mestu. Takšen je npr. glagol plesati, ki potrebuje dvojnice pri velelniku in deležniku.

Slika 8: Slovar izgovarjav (primer dvojnic)

Seveda se pojavljajo tudi trojnice in četrtnice, a so, če besede niso pogosto v uporabi, nesmiselne, saj bo Govorec še vedno izbral prvo obliko. Do zdaj se je v slovarju izgovarjav naglaševalo po pravopisu, razen če raba v širšem deležu populacije ne odstopa od

16Neskladnosti pravil in drugih rešitev v Slovenskem pravopisu z rabo so podrobneje opisane v članku Podoba govorjenega slovenskega knjižnega jezika v Slovenskem pravopisu 2001, ki sta ga napisala Hotimir Tivadar in Peter Jurgec.
norme. To se je dobro opazilo, ko smo po SP zamenjali vrstni red naglašenih oblik pri številih od dvanajst do dvajset. Po posvetu s sodelavci na Amebisu, d. o. o., smo dvojnice *dvánajst* in *dvanájst* pustili, saj jih ima tudi SSKJ. Tako smo naredili kompromis med SP in SSKJ, obdržali smo obe obliki in na prvo mesto zapisali *dvanájst*. Ko je to novo pravilo začel uporabljati Govorec 2, so se uporabniki pritožili. Oblika *dvanájst* se je zdela večini napačna in tuja. Vrstni red oblik smo zamenjali in dali kot primarno izbiro zopet *dvánajst*, a te v SP sploh ni. Velikokrat sta tudi bralca naglaševala besede drugače, kot so naglašene v SP. Ta težava se je pojavljala predvsem pri tujkah, ki so v SP največkrat normirane z ozkima é in ó, kar pa se v rabi ne izkazuje kot najboljše in se izgovor tujk (sploh angleških besed) v večini primerov bliža izgovoru v originalnem jeziku (zopet angleščini). Takšen primer je angleška beseda *software*, ki ima v SP osnovno obliko poslovenjeno v *softver* in je naglašena kot [s"o:ftv"E:r]. Prvi samoglasnik je pričakovano naglašen ozko, drugi pa v nasprotju z navado široko. Ker sta naša bralca vedno brala prvi o široko, saj je to tudi širša raba, smo izgovor zapisali kot [s"O:ftv"E:r], in tako odstopili od norme Slovenskega pravopisa. V Slovarju novejšega besedja slovenskega jezika je izgovor te besede enak našemu (sōftvēr).

V slovarju izgovarjav je še vedno veliko neoznačenih besed, zato je za njegovo dopolnitev potrebno še veliko dela. Pomembna bo sistematika naglasnih tipov, ki so zapisani v prvi knjigi Slovarja slovenskega knjižnega jezika, a bo ta pravila treba preveriti še na testnih besedilih in pa seveda uskladiti z informatiki, ali jih je sploh možno zapisati v obliki pravil, ki bi jih potem uporabili v slovarju izgovarjav. Slovar, ki je za zdaj še pomanjkljiv, ima za slovenščino veliko vrednost. Jezikoslovci se bodo morali v popravke, če slovar ne bo več namenjen samo za rabo Govorca, resneje vključiti. S tem seveda apeliramo tudi na poenotenje pravorečja in posledično normiranje, saj smo se morali pri delu velikokrat usklajevati med slovarji, rabo in svojim občutkom za jezik.

6. 2 Slovnična analiza pri Govorcih 2 in 3

Sintetizator mora v besedilu najprej opraviti slovnično analizo. Pri Govorcih 2 in 3 se vsem besedam določijo leme, besedna vrsta, sklon, spol, število idr. iz slovarja izgovarjav. Določijo se tudi stavčna vloga besed in skladenjske lastnosti vsake povedi. Analiza temelji na
podlagi glagolskih predlog. „Predloga opiše, kako se določen glagol uporablja. Pove, s katerimi predmeti se veže, katere omejitve so pri tem, katera prislovna določila so lahko ob glagolu, katere predložne zveze so tipično povezane z glagolom in podobno. Predloge so potem povezane v pomene, pri čemer so pripisane tudi potrebne stilne oznake /…/ Predloge določijo tudi stavčne člene, tako da je zaradi tega mogoče povezovati pomene glagolov s pomeni samostalnikov glede na vlogo, v kateri se pojavljajo. Tako je mogoče določiti, da sta glagol (oz. glagolski pomen) »skuhati« in samostalnik »kosilo« povezana tako, da je kosilo tipično predmet v tožilniku“ (Holozan 2004). Tako je Peter Holozan pojasnjeval del analize pri strojnem prevajalniku Presis, ki uporablja isto analizo kot Govorca 2 in 3.

6.2.1 Slovnična analiza in transkripcije pri Govorcu 3

Slovnična analiza je zelo obširna tema, zato navajamo le nekaj primerov, ki smo jih obravnavali. Pri popravljanju transkripcije smo lahko opazovali, kakšne napake se pojavljajo pri analizi in katere besede so napačno naglašene v slovarju, saj so vsa besedila prej obdelali, kot se obdelajo za samo sintezo. Iz napake smo morali najprej ugotoviti, ali je napaka v naglasu ali je analizator npr. izbral napačno besedno vrsto. V razpredelniči je navedenih nekaj napačnih analiz Govorca in napaka, ki smo jo predvideli in bi bila možna. Ker se te napake še preverjajo, popravki niso nujno pravilni.
Najemi sobo, govori s točajem in profesorjem Knuthom.

Albert je bil netakten, žaljiv, porogljjiv in celo grob.

Boleznī rastlin je seveda več, od napada plesni do šičnik, uši, mušic in pršic do rje in koreninske gnilobe.

Pri Naurju je Kitajska ravnala podobno kot zdaj, saj ima tudi ta diplomatske odnose s Tajvanom.

Meni prinese sin Metod, s katerim živiva skupaj v Radovljici, največ česminovih vej iz doline Save v okolici Bodešč in Ribnega.

Po drugi strani velja, da oblast, ki želi postopati preudarno, ve, da se mora samoomejevati.

<table>
<thead>
<tr>
<th>Poved</th>
<th>MRPA</th>
<th>Verjetna napaka</th>
</tr>
</thead>
<tbody>
<tr>
<td>Najemi sobo, govori s točajem in profesorjem Knuthom.</td>
<td>nalj&quot;E:mi s&quot;O:bO gO[lO]r&quot;i: s tO[S]a:&quot;j&quot;m in' prO[l]e:&quot;sO[a]m kn&quot;u:žtOm</td>
<td>Namesto velelne oblike glagola govôrì uporabi povedno obliko govorì.</td>
</tr>
<tr>
<td>Albert je bil netakten, žaljiv, porogljjiv in celo grob.</td>
<td>&quot;a:l@b@r t&quot;jE b&quot;i:U n&quot;E:t&quot;a:k@t@n _ Zal&quot;ji:U _ pO[rO]g&quot;i:U _ in' tsE]l&quot;o: gr&quot;OP</td>
<td>Namesto pridevnika gròb je besedo prepoznal kot samostalnik gròb.</td>
</tr>
<tr>
<td>Nastopil bo lahko vsak, ki ima drsalke in ki ve, kaj je pingpong.</td>
<td>na]st&quot;o:piU bo lax[k]&quot;o: Us&quot;a:k(STA:rp) ki i&quot;m&quot;a: d@r[s]a:&quot;lE&quot;n&quot;i&quot;m in' ki v&quot;e: k&quot;aj j&quot;e: p&quot;i:NkO:nK</td>
<td>Glagol biti je zamenj z glagolom jestì jè.</td>
</tr>
</tbody>
</table>

Tabela 2: Primeri napačnih slovničnih analiz

Slovnična analiza postavi v besedilo tudi stavčne oznake. Te sta bralca včasih pravilno intonirala, včasih pa sta na intonacijo pozabila. Lahko sta tudi pretirvala. Najmanj pa sta spremenila ton in način govora, čeprav bi se to pričakovalo pri velelniku. Označbe velelnosti bo treba verjetno še dodatno pregledati, ali so označene pravilno glede na bralčevo intonacijo.

- samostojni stavčni člen,
- glavni,
- vprašalni,
- pogojni,
- polstavek,
- osebkov,
- predmetni,
- prilastkov,
- prislovni.

Na mestu oznake y pa:

- povedni,
- velelni,
- pogojni,
- želelni,
- občudovalni in
- vezni.

Za še izboljšano analizo ali za podrobnejše jezikoslovne raziskave bo treba dodati še več oznak in transkripcijo podrobneje razdelati, a se je treba zavedati, da je pri našem delu cilj čim naravnejša sinteza, ki pa ne bo pretirano zapletena. S tega vidika je besedilo dobro označeno, Govorec 3 se bo lahko naučil in uporabljal različne intonacije. Najvažnejše bodo seveda vprašalna, velelna, povedna in pogojna. S stavčnimi oznakami se bo veliko lažje naučil tudi uporabljati primerno dolge premore pri različnih stavčnih oblikah in različnih stavčnih situacijah. Bolj kot razdelanost besedila je pomembna pravilnost vnesenih označb. To je bil tudi cilj naših popravkov v transkripciji.

17 Veliko oznak v transkripcijah ni enotnih s slovenistično stroko, saj so namenjene le sintetizatorju in so temu tudi prilagojene. Besedo naklon moramo tu razumeti kot intonacijo.
6.3 Grafemska pretvorba

Grafemska pretvorba je predelava besedila v računalniku razumljivo podobo. Treba je prepoznati vse znake (denarne enote ipd.), ki jih sintetizator nato tudi dobesedno prebere (znak @ pretvori in prebere kot [afna]). Številke pretvori v besede, prepoznati mora, ali gre za vrstilne ali glavne števnike, kratice – te lahko potem bere po slovarju izgovarjav skupaj ali pa črkovno, če so krajše od štirih črk (CD prebere kot [C"e:D"e:]). Skrajšane besede, kot je g., pretvori v cele (gospod). Ločila mora prepoznati kot skladenjska ali neskladenjska, to prepoznava po vnaprej določenih pravilih. Skladenjska pika se npr. določi, če ji sledita presledek in velika začetnica, a le, če ji ne sledi števka ali če je na koncu odstavka in ji sledi prazna vrstica. Ločila (predvsem pika, klicaj in vprašaj) določajo tudi intonacijo in premore sintetiziranega glasu (a sintetizator to upošteva šele pozneje pri določanju prozodije). Samodejno se morajo prilagoditi obrazila iz pretvorjenega števnika v besedo, tako mora analizator poznati spol, sklon in število odnosnic. Z vsemi temi postopki nastane grafemski zapis, ki ga moramo nato pretvoriti v fonetičnega.

6.3.2 Grafemska pretvorba in Govorca 2 in 3

Stavčna analiza še ne ločuje med vezaji in pomišljaji, takšna ločila trenutno še prekinejo analizo, razen če se vezaj pojavi pri kratici, pri kateri je dojet kot del besede. Te težave se pojavljajo, ker ločila še nimajo oznak MSD (oz. oblikoskladenjskih oznak) v slovarju izgovarjav, saj so vsi ti grafemi določeni le kot ločila brez oblikoskladenjskih vplivov. Pri več ločilih (npr. !!!!, ??!!) prepozna samo prvo ločilo in naslednje obravnava kot ločene stavke. Prepoznava krajevnih imen pa poteka le po slovarju izgovarjav in ne tudi po prepoznavi po velikih začetnicah, mestniških predlogih ali stavčnih funkcijah besed. Ne razpoznava tudi premega govora, kar bo v prihodnje povzročalo še nekaj težav, saj se v sodobnejšem leposlovanju pojavlja tudi označitev samogovora, dialoga s kurzivo in ne dosledno z narekovanjem in dvopičjem. Tu se lahko seveda spet vnemo k pomisleku o pomišljalu in vezaju. Njuna uporaba pri latičnih piscih je večinoma nepravilna. Težave z njima imajo v časopisih. Tudi v resnejših s precej naklade namesto pomišljala uporabljajo vezaj. Natančna pravila o stičnosti in nestičnosti vezaja oz. pomišljala bi tako lahko pri Besani (samodejnem
lektorju, ki deluje na istem sistemu kot Govorca 2 in 3 ter Presis) pri popravljанju povprečnega besedila pripeljala do več napak kot popravkov.

6.4 Transkripcija


- širok glas se piše z velikimi črkami,
- oseč glas z majhnimi,
- naglas se označi pred soglasnikom naglašenega samoglasnika z narekovajem 
- dolžina z dvopičjem [:],
- polglasnik z afno [@],
- glas črke h z znakom [x],
- mehčanost z znakom [’],
- dvoglasniški u se zapisuje z velikim ū [U],
- m pred f ali v z [F],
- n pred k, g, h z [N],
- alofon fonema j z [I].

Vsi drugi glasovi, ki jih tu ne omenjamo, so zapisani že v Tabeli 1. V tabeli so tudi zapisi v mednarodni pisavi IPA. Transkribiranje pri Govorcu 3 smo z rabo in izkušnjami prilagajali. „Enako kot vsa načela glede gradnje korpusa so tudi načela za transkribiranje korpusa odvisna od namembnosti korpusa /.../ Fonetik potrebuje manjšo količino podatkov, vendar morajo biti ti mnogo bolje detajlno transkribirani v smislu prozodije in akustične


V Zemljarič Miklavčič (2008) najdemo stopnje označevanja govornih zbirk. Te so:
- ortografska raven (standardni zapis),
- citatno-fonemska raven (zapis s fonemi, zapis iz zapisa besede v izolaciji),
- širša fonetična raven (upošteva se izgovor besed v kontekstu),
- ožja fonetična raven (natančen zapis govora z variantami fonemov),
- akustičnofonetična raven (ločuje vse zaznavne elemente na spektogramu).
6.4.1 Transkribiranje in Govorec 3

Pri transkribiranju se srečamo z mnogimi izzivi. To je najpomembnejši del za jezikoslovca, ki sodeluje pri ustvarjanju umetnega govora, še posebej ko gre za vrste sintetizatorjev, kot je Govorec 3, ki ima po Zemljarič Miklavčič četrto stopnjo označevanja (ožja fonetična raven).


18 Več o različnih transkripcijskih orodjih in njihovih prednostih in slabostih v monografiji Govorni korpusi (121–128).
Slika 9: Delovno okolje v Transcriberju

Pri Govorcu 3 so bile fonetične transkripcije samodejno tvorjene iz grafemske transkripcije po pravilih, ki so jih sestavili na Amebisu, d. o. o., po priporočilih Instituta „Jožef Stefan“. Ta pravila smo že precej popravili in dopolnili, saj so vsebovala nekaj napak. S tem smo se tudi izognili večjemu in obsežnejšemu poseganju v slovar izgovarj. Pri preverjanju rezultatov se je hitro ugotovilo, da je lahko teorija v jeziku (predvsem govorenem) precej relativna in nezanesljiva. V začetnih fonetičnih transkripcijah so se uporabljale vse variante fonemov, upoštevali so se premene po zvenci in vplivi sosednjih besed na sam i zgovor konca in začetka besede. Te vplive (premene po zvenčnosti) na medbesedni ravni je bilo treba velikokrat popravljati, saj sta bralca delala med besedami krajše neslišne premore in se tako oddaljevala od teorije. Vsi premori se v transkripcijah označujejo s tremi znaki, s podčrtanjem [ ] čiste tišine, z znakom [<] vzdihi in z [>] izdihi. Dobro bi bilo različno označevati krajše in daljše premore čiste tišine, kot so na primer označeni v korpusu London-Lund (z dolgim in kratkim podčrtanjem). Težave smo imeli pri označevanju odmorov z različnimi oznakami, saj so se lahko med tišinami pojavljali cmoki in drugi nenamerni zvoki bralcev ali pa se je hkrati pojavila kombinacija več govornih pojavov (npr. tišina, vdih in cmok). Predvsem smo gledali na grafični prikaz govornega signala, na
katerem se je videlo, ali je med besedama daljša tišina ali dih. Z odmori so se na Amebisu, d. o. o., že ukvarjali, a smo pri delu s transkripcijami govornih zbirk naredili še nekaj sprememb. Odmore sta napovedovalca uporabljala precej različno. Moški napovedovalce ima krajše in samo stavčno strukturo bolj kot z odmori izraža z intonacijo, napovedovalka pa uporabla daljše in jasnejše tišine med stavki.

Polglasnik se je pojavljal veliko pogosteje, kot smo pričakovali in kot so jih pri grafemsko-fonemski pretvorbi predvidevala pravila. Za natančnejšo transkripcijo bi potrebovali akustično analizo, saj je pri hitrem govoru težko slišno razločiti, ali gre res za polglasnik. Bralca sta ga vedno izgovarjala, če je bil široki nenaglašeni e na koncu besede v zaprtem zlogu. Navajamo nekaj primerov: [pOtrEb"u:j@S], [pos"e:b@j], [m"ajSp@rk], [d"O:b@r], [v"e:t@r], [s"i:ts@r], [m"E:nEdZ@r], [x"i:t@r] … Pojavljava so se tudi neskladja s samodejno pretvorbo in govorom bralcev pri nenaglašenih besedah, nekaj primerov smo že navedli. Za primer analize smo vzeli členek že. Pri popravljanju transkripcije smo si sproti zapisovali, kolikokrat kdo od napovedovalcev izreče že s širokim in kolikokrat z ozkim e. Pri štetju nismo izključevali nenaglašenih ali naglašenih, obravnavali smo jih enako.

<table>
<thead>
<tr>
<th>Moški napovedovalce</th>
<th>Ženska napovedovalca</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Ze]</td>
<td>[ZE]</td>
</tr>
<tr>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

*Tabela 3:* Primerjava izgovora členka že pri bralcih

Bralca bereta členek zelo različno, na izgovor pa glasovna okolica ne vpliva. Pri bralcih nenaglašenost ne pomeni nujno širine glasu. Naglašenost ni vključena v analizo, saj v programu Transcriber akustična analiza ni mogoča.

Precej dilem je povzročal tudi glas j, ki ga pišemo z i in sledi samoglasniku (npr. *celuloid, pacient*). Po predlogu Instituta „Jožef Stefan“ se je najprej posebej označeval z [I], a smo ta grafem oz. glas ukinili. Namesto tega smo glas zapisovali tako, kot sta ga napovedovalca dejansko izgovarjala, z [i], [ij] ali [j]. Tako smo zmanjšali število glasov in olajšali analizo. Te težave pa bo treba premostiti še pri samodejni pretvorbi grafemskega

Opravili smo tudi nekaj akustičnih analiz. Sicer je testnih vzorcev premalo, da bi bili reprezentativni za celotno govorno zbirko, a rezultati kažejo, kje bo za boljše delovanje Govorca potrebna podrobnjejša analiza. V primerih, v katerih se pričakuje izgovor [ij], je velikokrat prišlo do redukcije glasu i ali pa se i sicer sliši, a se pri akustični analizi pokaže, da nima osnovnega tona in je ustvarjen le s turbulencami v ustni votlini brez pomoči glasilk. Ali gre v tem primeru za en glas, ki ga slišimo kot dva [ts], ali za dva glasova ali pa za kakšen poseben nezveneči i, je treba še posebej raziskati. Na Sliki 10 vidimo analizo besede cianid, ki jo je prebrala bralka, označen pa je glas [tsi]. Modra črta označuje osnovni ton, tako ugotovimo, da ta glas ni zveneč.

Slika 10: Analiza besede cianid (bralka)

Enako je tudi pri podrednem delu besede socialnoekonomski. Pri analizi smo analizirali samo prvi del besede socialno-. Spet se pojavi glas [tsi], ki je sicer manj izrazit, i se sliši le, če ga poslušamo izolirano. Gre za isto bralko kot v prejšnjem primeru.

19 Govorec je namenjen predvsem slabovidnim osebam, zato smo pri delu velikokrat sledili njihovim potrebam. Predlagamo tudi, da sintetizator besede, kot je cianid, bere tako, kot so zapisane, in ne po SP 2001, ki narekuje izgovor [tsjian"i:t], saj je za slabovidne izredno pomembno, da slišijo besedilo tako, kot je zapisano.
Tudi pri bralcu se pojavi isti glas pri istih besedah. Prilagamo dva primera, in sicer cianid in socialno-. Tako lahko trdimo, da pojav ni povezan s posameznim bralcem. Ali je ta pojav posledica istega šolanja govora oz. pravorečja, bi morali preveriti na več in različnejših bralcih.

Slika 11: Analiza besede *socialnoekonomski* (bralka)

Slika 12: Analiza besede *cianid* (bralec)
Slika 13: Analiza besede socialnoekonomski (bralec)

V Tabeli 4 podajamo primere izgovora glasu i pred samoglasnikom. Primeri glasu [tsi], ki smo jih akustično analizirali, so kot posebnosti zapisani krepko. V teh primerih je lahko i bolj ali manj reduciran ali pa bolj ali manj izraziteje zrasel z glasom c.

<table>
<thead>
<tr>
<th>Beseda</th>
<th>Izgovor bralcev v bazi ()</th>
</tr>
</thead>
<tbody>
<tr>
<td>kvocient</td>
<td>k&quot;o:tsijEnt</td>
</tr>
<tr>
<td>hieroglifi</td>
<td>xjErOgl&quot;ifi</td>
</tr>
<tr>
<td>ioni</td>
<td>i&quot;o:ni</td>
</tr>
<tr>
<td>koefficienti</td>
<td>kOEfitsij&quot;E:nti</td>
</tr>
<tr>
<td>postsocialistični</td>
<td>p&quot;o:stsOsja&quot;l&quot;:stSnj</td>
</tr>
<tr>
<td>genialna</td>
<td>gEnij&quot;a:lna</td>
</tr>
<tr>
<td>variira</td>
<td>varij&quot;i:ra</td>
</tr>
<tr>
<td>cianid</td>
<td>tsjan&quot;i:t</td>
</tr>
<tr>
<td>radio</td>
<td>r&quot;a:diJO</td>
</tr>
<tr>
<td>ambiciozen</td>
<td>ambitsj&quot;o:z@n</td>
</tr>
<tr>
<td>fizioterapevt</td>
<td>f&quot;i:ziOterap&quot;e:Ut</td>
</tr>
<tr>
<td>tiare</td>
<td>tij&quot;a:Re</td>
</tr>
<tr>
<td>ciankalij</td>
<td>tsjaNk&quot;a:lji</td>
</tr>
<tr>
<td>komediograf</td>
<td>kOmEdiOgr&quot;a:f</td>
</tr>
<tr>
<td>heliosa</td>
<td>x&quot;e:liOsJa</td>
</tr>
<tr>
<td>niansa</td>
<td>nij&quot;a:nsa</td>
</tr>
<tr>
<td>pionirska</td>
<td>piOn&quot;i:rska</td>
</tr>
<tr>
<td>nacionalizma</td>
<td>natsjOnal&quot;i:zma</td>
</tr>
<tr>
<td>aksiom</td>
<td>aksj&quot;o:m</td>
</tr>
<tr>
<td>varianta</td>
<td>varj&quot;a:nta</td>
</tr>
</tbody>
</table>

Tabela 4: Primeri izgovora i pred samoglasnikom

<table>
<thead>
<tr>
<th>Primeri</th>
</tr>
</thead>
<tbody>
<tr>
<td>v uradu</td>
</tr>
<tr>
<td>s stigmo</td>
</tr>
<tr>
<td>razstavi</td>
</tr>
<tr>
<td>čas za</td>
</tr>
<tr>
<td>razsaja</td>
</tr>
<tr>
<td>v vlogo</td>
</tr>
<tr>
<td>medtem</td>
</tr>
<tr>
<td>izzvali</td>
</tr>
<tr>
<td>vmes z zobmi</td>
</tr>
</tbody>
</table>

**Tabela 5:** Primeri besed, v katerih si v verigi sledita dva soglasnika

Rezultati krajše analize bralcih dolžin so v Tabeli 6. V oklepaju je število primerov, s katerimi smo analizirali glas in izračunali povprečje. Meritve smo opravili za glasove s, z, in d. Dolžina je podana v sekundah.

<table>
<thead>
<tr>
<th>Glasovi</th>
<th>s</th>
<th>z</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>zliti</td>
<td>0,119 (1)</td>
<td>0,140 (1)</td>
<td>0,075 (1)</td>
</tr>
<tr>
<td>enojni</td>
<td>0,115 (3)</td>
<td>0,070 (3)</td>
<td>0,036 (2)</td>
</tr>
</tbody>
</table>

**Tabela 6:** Časovna analiza zlitih glasov (bralca)
Zlita dvojna glasova $d$ in $z$ sta približno dvakrat daljša kot enojni glas, zato bi bila zanju potrebna podrobnejša analiza na večjem vzorcu, da bi se domneve potrdile. Medtem ko pri glasu $s$ ni zaznati razlik. Pri bralcu navajamo rezultate iste analize v Tabeli 7. V oklepaju je poleg rezultatov zopet zapisano število primerov, na katerih je izračunano povprečje.

<table>
<thead>
<tr>
<th>Glasovi</th>
<th>$s$</th>
<th>$z$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>zliti</td>
<td>0,140 (2)</td>
<td>0,132 (3)</td>
<td>0,108 (1)</td>
</tr>
<tr>
<td>enojni</td>
<td>0,121 (3)</td>
<td>0,073 (2)</td>
<td>0,052 (3)</td>
</tr>
</tbody>
</table>

_Tabela 7: Časovna analiza zlitih glasov (bralec)_

Rezultati so podobni kot pri bralki. Pri $z$ in $d$ je zliti glas približno dvakrat daljši, medtem ko lahko pri glasu $s$ govorimo le o 14-odstotni razliki, kar je posledica nižjega povprečja dolžine le enega, izjemoma krajšega izmerjenega glasu. Ker je pri obeh bralcih rezultat podoben, bi bilo dobro o zlitih soglasnikih narediti obširnejšo raziskavo in ugotoviti, ali je res, da so zliti glasovi približno dvakrat daljši in da je pri tem glas $s$ izjema. Tako kot smo naredili kraško analizo zlitih glasov $s$, $z$ in $d$, smo pregledali tudi zveneči $ɣ$, ki smo ga v transkripciji označevali kot $[G]$. Prva bralkina besedna zveza, analizirana v Praatu, je _porodnih bolečin_. Na Sliki 14 je analiza dela [nix bOEtS”i:n’]. Glas $[G]$ je označen z rdečim senčenjem.

_Slika 14: Analiza zvenečega $ɣ$ (bralka)_
Slika 15 prikazuje besedno zvezo parketnih deščic, in sicer del [niG d"@StSits].


Slika 16: Analiza zvenečega ţ (bralec)
Dodajamo še analizo izseka [s@m na t"e:Zjix Et"a:paG b"i:U]. Poleg rdečega osenčenega glasu [G] je označen še nezvenec [x]. Tu se dobro vidi, da prvi glas x nima osnovnega tona, drugi ţ pa ga ima. Gre torej za dve varianti glasu x.

Slika 17: Analiza zvenečega ţ s primerom nezvenečega x (brlec)

Slika 18: Analiza primera [sl"a:bix tr"u:Sp]

S temi akustičnimi analizami smo nakazali nekaj praktičnih izkušenj, s katerimi smo se srečevali pri popravljanju transkripcij. Analize so sicer kratke, a jasno nakazujejo, kaj vse bo še treba postoriti za izboljšavo sintetizatorja. Nekaj teh je:

- podrobnejši pregled polglasnikov,
- analiza dolžine zlitih soglasnikov,
- podrobnejši pregled variant glasov (primer sta zveneči ţ oz. [G] in glas [tsi]),
- vnos novih pravil (npr. premena po zvenečnosti za več kot en glas nazaj).
7 Anketa in rezultati


<table>
<thead>
<tr>
<th>Ocena</th>
<th>Ocenjevalci</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Povprečje</th>
</tr>
</thead>
<tbody>
<tr>
<td>glasovna ustreznost</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>naglasna ustreznost</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>barva glasu</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>intonacija</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>register</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>poudarki</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>dolžina in mesto premorov</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>uporabnost uravnavanja hitrosti govora</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

**Tabela 8:** Številčni rezultati ankete

Na začetku so anketiranci s števili od 1 do 5 ocenjevali kakovost govora. V Tabeli 8 so izpisani ocene in povprečja vseh ocenjevalcev. Najboljše so ocenjeni glasovna in naglasna ustreznost, barva glasu in uporabnost nastavljanja hitrosti govora. Težave v kakovosti se pričakovalo pojavijo pri intonaciji in registru glasu, premorih in poudarkih (premori so tudi v transkripcijah pri Govorcu 3 povzročali največ preglavic). Pri podrobnejših vprašanjih so
anketiranci kot najbolj moteče in najmanj kakovostne elemente navajali predvsem intonacijo in uporabo premorov. Motila sta jih še izgovor glasu v v različnih glasovnih položajih (pred soglasniki in kot dvoglasniški ů) in naglaševanje zloženek ([l"E:Odaj"alts@U]). Kot težave so omenjali tudi barvo, register in hripavost glasu. Samo tretjina anketirancev bi Govorca 2 uporabljala kot bralnik ali kot del aplikacij tehničnih pripomočkov.

Rezultati ankete so pričakovani, vzrok nekaterih motečih elementov je predvsem v naravi delovanja difonskega sintetizatorja, ki lahko dvoglasniški ů zaradi fonetičnih pravil (preverjena bolj v teoriji kot v praksi) nenaravno vsiljuje v jezik, to se bo z Govorcem 3 izboljšalo, saj je primerov izgovora glasu v v bazi veliko več in so izgovorjeni v najrazličnejših glasovnih okoljih, natančneje pa smo uredili tudi pravila grafemsko-fonemske pretvorbe.
8 Zaključek

Sintetizatorji govora imajo v slovenščini še veliko možnosti za razvoj in napredek, predvsem bi se morali v raziskave vključiti tudi slovenisti. Medtem ko sta slovnica in pravopis že dodelana, pa je govorjeni (standardni) knjižni jezik še v večini nedoločen in neraziskan. Seznam glasov v Slovenskem pravopisu je teoretično dodelan, a prihaja v praksi do razhajanj. Novejše ugotovitve o zvenečih in nezvenečih ustničnoustničnih šumnih glasovih ([w] in [W]) so premalo poznane, tako so avtomatika in starejša grafemsko-fonemska pravila in predvsem zanašanje neslovenistov na razčiščeno normo v Slovenskem pravopisu povzročili v transkripcijah nekaj težav, ki jih bo treba ročno popraviti, saj bralca v pred glasovoma l in r bereta večinoma kot zobnoustnični v. Glede na zadnje rezultate raziskav med jezikoslovci smo pri Amebisu, d. o. o., ustničnoustnični šumni (zveneči in nezveneči) v ukinili, to smo izpeljali tudi že v transkripcijah. Težava ni, da uporabniki slovenistične stroke ne bi upoštevali, pač pa da slovenisti rahlo zamujamo in zaostajamo za potrebami hitro spreminjajočega se sveta. Analize, kot smo se jih lotili v diplomskem delu, niso namenjene le slovenistični strokovni javnosti, njihova dodana vrednost je, da so posredno uporabne v informatiki, poslovnem svetu in zabavní industriji. Takšni sintetizatorji, kot je Govorec, se lahko uporabljajo v mobilnih aplikacijah, e-bralnikih in e-asistentih, s takšnimi pripomočki se lahko zmanjšajo stroški poslovanja podjetij, olajša vsakdan slabovidnim in obogati izkušnja z zabavnimi aplikacijami. Glede na to, da so tudi slovenska narečja temelj humorja, bi bili razvoj mobilne aplikacije z osnovami narečnosti ali posnetki narečnih govorjev za uporabnike izredno zanimivi in lahko tudi dobičkonosni, poleg tega pa še seveda priložnost slovenistične stroke, da javnosti približa jezik in znanje, ki ga premore in je že zapisano ter v večini raziskano.

V diplomskem delu smo na kratko povzeli zgodovino in osnovne lastnosti sintetizatorjev govora. Prvi mehanični sintetizatorji so se pojavili že v drugi polovici 18. stoletja, pravi sintetizator govora z imenom VODER pa leta 1930. V šestdesetih letih so razvili formantne sintetizatorje, ki so obsegali do 100 enot v besedišču, število pa je že v sedemdesetih letih zraslo do 1000. Potem se je že začel razvoj sintetizatorjev, ki temeljijo na lepljenju enot (metoda LPC), ta metoda je tudi temelj Govorca 2. Nazadnje so se pojavili še

sintetizatorji, ki temeljijo na metodi PMM, in sinusni sintetizatorji, varianta teh pa je Govorec 3. Pojasnili smo njihov način delovanja in vse potrebne elemente, ki so nujni za uspešno sintezo (govorna zbirka, slovnična analiza, transkripcije in grafemsko-phonemska pretvorba, slovar izgovarjav ter dobra sinteza glasu). Podrobneje smo opisali svoje izkušnje z delom pri transkripciji in naglasih besed, z akustično analizo pa pokazali na nekaj pojavov, ki potrebujejo konkretnjejšo in obsežnejšo analizo. Prihodnost jezikoslovja (predvsem slovenščine kot jezika z manjšim številom govorcev) je v digitalni humanistiki. A bodo za uspešno izpeljavo takšnih projektov, kot je Govorec, zopet potrebne nekatere osnovne raziskave jezika (tako fonetične kot oblikoslovne) z novimi temelji in novimi cilji glede na potrebe informacijskih tehnologij. Anketa je pokazala, da je potreben novejši sintetizator z naravnejšim govorom, to je seveda tudi cilj dela sodelavcev na Amebisu, d. o. o., in moje jezikoslovne pomoči pri Govorcu 3.
9 Povzetek

Pri delu z Govorcem 3 in raziskovanju za diplomsko nalogo smo opazili, da je v slovenskem jeziku še veliko nedorečenega, da bo treba opraviti nove fonetične raziskave, dozdajšnja spoznanja, ki se razlikujejo od SP 2001, pa na novo normirati. Vprašljivo je uvajanje obvezne širine nenaglašenih glasov e in o, pojavljajo se težave pri ozkem izgovoru naglašenih samoglasnikov v tujkah, izgovori s širino, ki so pogosti, tako odstopajo od norme. Neenotnost se pojavlja v SP 2001 tudi pri naglaševanju (npr. dvánajst in dvanájst), predvsem zaradi nedoločenosti osnove za knjižni jezik.

Slovenistična stroka se premalo posveča digitalni humanistiki in s tem tudi sintezi govora. Informatiki tako sami rešujejo vprašanja, ki bi jih morali reševati jezikoslovci, popravki napak in izboljšanja jezikovnih tehnologij so zato dolgotrajnejši, razvoj pa se upočasnjuje. V jeziku bi bilo treba najti sistem, ki bi ga lahko razumel tudi računalnik. Tako smo z določenimi novimi grafemsko-fonetičnimi pravili na Amebisu rešili veliko težav pri prevajanju besedila v fonetični zapis. Dobro nam je uspelo določiti na primer pravila za izgovor črke v, zanj nam ni treba več vnašati posebnih fonetičnih zapisov v slovar izgovarjav. V zadnjem mesecu smo v transkripcijah ukinili večino alofonov, saj ti z naravo delovanja metode HNM niso več potrebni. Fonetična pravila so nujna in morajo biti jasna, a je treba izpostaviti, da naravni govor od teh zakonitosti velikokrat odstopa. V naši govorni zbirki smo našli tudi nezveneči glas i, ki do zdaj ni bil poznan, pojav bo treba podrobnejše raziskati. Raziskave so bile do zdaj opravljene večinoma na naglašenih in čistih samoglasnikih, pozornost bo treba preusmeriti tudi na soglasnike in pa predvsem na prosta govorjena besedila, ki niso studijsko posnete. Naravni govor bo treba raziskati podrobneje.

Govorna zbirka, ki so jo ustvarili na Amebisu, d. o. o., je lahko zelo pomembna za vnaprejšnje raziskave. Je verjetno ena največjih govornih zbirk, transkribirana na ožji akustični ravni (raven po Zemljarič Miklavčič). Ko bomo imeli dokončane še samodejno označene posnetke, se bodo lahko dolžine določenih glasov preprosto statistično primerjale, temeljile bodo na podlagi več tisoč primerov (žal le dveh govorcev).
V prihodnosti bo predvsem pomembno razumevanje sistema jezika z vidika računalništva in informatike, kjer so potrebna drugačna in bolj sistemična pravila. Ta nam bodo omogočala pravilno digitalizacijo jezika, ki bo slovenščino še bolj približala uporabniku, nove izboljšane jezikoslovne tehnologije pa bodo jezikoslovcem omogočile hitrejše, natančnejše analize in lažje ter boljše opravljanje dela (npr. lekture).
10 Viri in literatura


11 Priloge

11.1 Anketa o kakovosti in uporabnosti Govorca 2

Spoštovani,

sem Nejc Robida, študent Filozofske fakultete Univerze v Ljubljani. V diplomski nalogi raziskujem sintezo govora, pri raziskovanju pa bi mi bila v veliko pomoč tudi vaša ocena govora Govorca 2.


<table>
<thead>
<tr>
<th>OCENA</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>fonemská ustreznost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>naglasna ustreznost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>barva glasu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intonacija</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>register</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>poudarki</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dolžina in mesto premorov</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>uporabnost uravnava hitrosti govora</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Kaj je bilo napačno izgovorjeno (naglas, intonacija, barva, register, premori)?

Kaj vas je pri Govorcu 2 najbolj motilo?

Bi Govorca 2 uporabljal kot bralnik ali kot del aplikacij tehničnih pripomočkov (dlančniki, mobilni telefoni, zvočne knjige)?
11.1.1 Besedilo 1

Vlada Republike Slovenije je na 15. redni seji seznanila državni zbor s paketom protikriznih ukrepov, ki jih bo sprejela še letos. Razprava je trajala le 2 uri. Vse poslanske skupine so se do načrta opredelile in podale tudi svoje predloge. Med možnimi elementi racionalizacije delovanja države so se tako znašli še nujna reforma zdravstva, pretirana in neuravnovešena regionalizacija in plačni sistem v javnem sektorju. Omenjala se je tudi vnovična reforma dela, saj naj bi bilo pri zdajšnji premalo popuščanja tako s strani delodajalcev kot delojemalcev. Vsi so se strinjali, da je zdaj pravi čas za korenite in pogumne spremembe.

11.1.2 Besedilo 2

Ah, nič ni, samo spi se mi.
A je to tvoj pulover rdeč?
Kva pa teb dons tok dogaja, čist si fuknen.
Kiseu krompir, to je pa krampirjeva župa, nter se pa da mal jisha.
11.2 Zgoščenka s posnetki sintetiziranega govora
IZJAVA O AVTORSTVU

Izjavljam, da je diplomsko delo Sinteza govora in Govorec 3 v celoti moje avtorsko delo in da so uporabljeni viri in literatura navedeni v skladu z mednarodnimi standardi in veljavno zakonodajo.

Podgora, 2. september 2013

Nejc Robida