

Nejc Robida

I. gimnazija v Celju, Celje

UDK 808:003.035

UDK 811.163.6'355'374

Sinteza govora in *Govorec 3*

Govorec 3 je sintetizator govora podjetja Amebis, d. o. o., in je primer jezikovnih tehnologij, ki so v Sloveniji trenutno v razvoju in izpopolnjevanju. Pri nastanjanju sintetizatorja sem sodeloval dve leti, moje delo je obsegalo urejanje slovarja izgovorjav (trenutno obsega 6.900.000 oblik besed), urejanje transkripcij govorne zbirke in urejanje izgovornih pravil oziroma pravil grafemsko-fonemske pretvorbe. V članku smo se posvetili načinu delovanja *Govorca 3* in procesom njegovega razvoja. Predstavili smo govorno zbirko, različne vrste sintez govora, transkripcije in transkribiranje govorne zbirke, grafemsko-fonemsko pretvorbo in slovnično analizo. V članku predstavljena anketa in ugotovitve kažejo na potrebo po novem sintetizatorju govora.

sinteza govora, sintetizator, govorna zbirka, transkripcija, *Govorec 3*

Govorec 3 is a speech synthesiser produced by the company Amebis, d. o. o. and is an example of the language technology currently under development and improvement in Slovenia. The author was involved in the development of the synthesiser for two years, dealing with the dictionary of pronunciations (currently 6,900,000 word forms), the transcriptions of the spoken database and the pronunciation rules or graphemic-phonemic conversions. The paper looks at how *Govorec 3* works and how it was developed. The spoken database and its transcription is presented, as well as different kinds of voice synthesis, graphemic-phonemic conversions and grammatical analysis. A survey is also presented, which points to the need for a new speech synthesiser.

speech synthesis, synthesiser, spoken database, transcription, *Govorec 3*

1 Uvod

Jezikovne tehnologije postajo vedno pomembnejši del jezikoslovja, njihove stalne prisotnosti se niti ne zavedamo več. S sintetizatorjem govora se srečujemo v dvigalih, na parkomatih, s strojnimi prevajalniki pri brskanju po spletu, urejevalniki besedil pa nam sproti samodejno lektorirajo besedila. S podjetjem Amebis, d. o. o., sem pri razvoju sintetizatorja govora sodeloval dve leti. V članku se bomo tako posvetili *Govorcu 3*, sintetizatorju govora, in jezikoslovnemu delu ob njegovem ustvarjanju. Povzeli bomo tudi anketo o kakovosti njegovega prednika *Govorca 2* in dokazali, da je razvoj novega sintetizatorja smiseln.¹

2 Sinteza govora

Sinteza govora je dolgotrajen in obširen proces, ki zajema veliko dela različnih strokovnjakov. Proces dela je večsmeren, velikokrat se je treba vračati na že narejeno in temelje na novo izpolnjevati. Predpriprava pred razvojem je izredno pomembna. Proces sinteze govora in delo sintetizatorjev lahko strnemo v nekaj točk. Navajamo pet osnovnih korakov, ki so potrebni za dober sintetizator govora:

¹ Ta članek je nastal na osnovi diplomske naloge *Sinteza govora in Govorec 3* (Robida 2013).

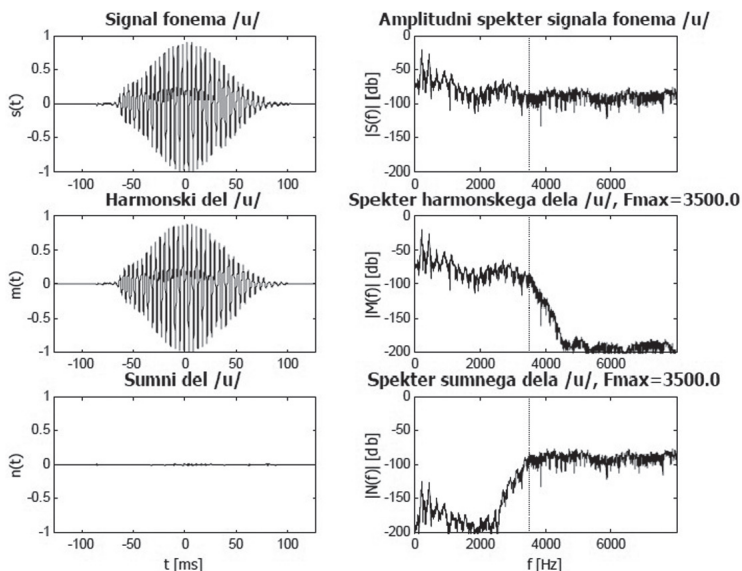
1. korak: slovnična analiza vhodnega besedila,
2. korak: nadomeščanje simbolov, števk in kratic z besedami,
3. korak: grafemsko-fonemska pretvorba (s slovničnimi pravili in bazo znanih besed),
4. korak: določanje prozodičnih parametrov in
5. korak: sintetiziranje govornega signala (Rozman 2005: 3–4).

V članku se bomo posvetili predvsem prvemu, drugemu in tretjemu koraku, ki so tudi najpomembnejši za jezikoslovce. Četrty korak bomo zaradi omejenosti z dolžino članka izpustili. Začeli bomo s petim korakom in navedli različne vrste sintez.

2.1 Vrste sintetizatorjev

Pri načrtovanju sintetizatorja je dobro najprej izbrati vrsto sinteze, ki jo bo uporabljal, in nato najprimernejše postopke pri drugih korakih. Za uresničitev petega koraka obstaja več vrst sintetizatorjev govora,² ti so:

- a) artikulatorni (posnemanje govorne cevi),
- b) formantni (posnemanje glasu s formantnimi frekvencami in amplitudami),
- c) sintetizatorji, ki združujejo enote (npr. difonski),
- č) sinusni (variante sinusnih je metoda HNM),
- d) tisti, ki temeljijo na metodi HMM oziroma modelih Markova.



Slika 1: Dekompozicija glasu *u* na harmonski in šumni del (Rozman 2005: 10)

² Sinteza pomeni umetno poustvarjanje govora.

Govorec 3 deluje na osnovi metode HNM (angl. harmonic plus noise modulation). Na Sliki 1 je prikazano, da dobimo govorni signal tako, »da vzbujevalno funkcijo, ki ponazarja periodične impulze glasilk, filtriramo s filtrom, ki modelira obliko govorne cevi. [...] Bistvo metode je dekompozicija govornega signala na harmonski del, ki ga modeliramo s harmonsko sinusno sintezo, in šumni del, ki ga modeliramo s filtriranjem šuma« (Rozman 2005: 5, 9). Slika 1 tudi jezikoslovcem jasno nakaže bistvo zapletenega teoretičnega opisa. Pri izgovoru glasu *u* ne prihaja do velikih turbulenc v govorni cevi, zato večino energije glasu predstavlja harmonski del, saj je *u* zvoneč glas. Z dekompozicijo tako izračunamo harmonski in šumni del in vse koeficiente s posnetka človeškega govora za poznejše potrebe umetnega poustvarjanja govora.

3 Govorna zbirka

Sama sinteza govornega signala oz. elektronsko oponašanje človeškega govora za delovanje sintetizatorja ni nikoli dovolj. Govorne zbirke so temelj večine vrst sinteze govora.³ Pomembno je, da je zbirka sestavljena iz čim več možnih kombinacij izgovorjav glasov in njihovih intonacij. Tako lahko sintetizator na podlagi okoliških glasov vsakič izbere najprimernejši izsek iz govorne zbirke. Potrebne so za kakovostno posnemanje resničnega človeškega glasu in lahko temeljijo na različnih osnovnih enotah (posamezne besede, difoni,⁴ fonemi). Pri *Govorcu 3* obsega govorna zbirka približno 20 ur posnetkov moškega in ženskega glasu. Izbrana sta bila bralca, napovedovalca nacionalnega radia, ki »sta nedvomno dodatno šolana za govor, obenem pa kot izkušena govorca tudi dovolj neodvisna« (Tivadar 2012: 593). Posnetki vsakega bralca so razdeljeni na približno štiri tisoč enot (večinoma posameznih povedi). Poleg avdiosignala je posnet še laringografski signal, ki prikazuje hitrost utripanja glasilk, z njegovo pomočjo hitro ugotovimo (ne)zvenečnost posameznega glasu.

3.1 Transkribiranje govorne zbirke

Govorno zbirko smo tudi fonetično transkribirali. Najprej je to opravila avtomatska grafemsko-fonemska pretvorba, njene rezultate pa smo še ročno popravili, saj slovenska zborna izreka ponekod ni jasno standardizirana⁵ in sta bralca včasih brala besedilo narobe ali nerazumljivo. Od pravorečnih pravil oziroma naših pričakovanj se je precej razlikovala tudi raba polglasnika. Grafemska pretvorba predela besedilo v obliko, ki je razumljiva računalniku. »Številke pretvori v besede, prepoznati mora, ali gre za vrstilne ali glavne števnike, kratice – te lahko potem bere po slovarju izgovorjav 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« (Robida 2013: 27). Samodejno se morajo prilagoditi tudi obrazila pretvorjenih kratic ali števil, določijo se spol, sklon, število ... Tako razdelan grafemski zapis moramo nato s pomočjo izgovornih pravil predelati v fonetičnega. Zanj smo uporabljali mednarodno uveljavljeno abecedo MRPA. Njene osnovne značilnosti so:

³ Več o govornih zbirkah v monografiji *Govorni korpusi* (Zemljarič Miklavčič 2008).

⁴ Difon oziroma dvoglasnik je kombinacija dveh glasov, ki sta izgovorjena eden za drugim.

⁵ O standardni slovenščini piše podrobneje v več člankih Tivadar (2010a, 2010b, 2012a).

Prihodnost v slovenskem jeziku, literaturi in kulturi

- široka *ε* in *ɔ* se pišeta z velikimi črkami,
- ozka *e* in *o* z majhnimi črkami,
- naglas se označi pred naglašenim samoglasnikom z narekovajem ['],
- dolžina z dvopičjem [:],
- polglasnik se zapisuje z afno [@],
- glas *x* z znakom [x],
- mehčanost z znakom ['],
- mehkonebni glas *ŋ* z znakom [N],
- dvoglasniški *u* z velikim *u* [U].⁶

Pravila grafemsko-fonemske pretvorbe so sestavili na Amebisu, d. o. o., po priporočilih Instituta „Jožef Stefan“. Nato smo jih popravili in dopolnili (npr. pravila izgovora glasov *v*, *n*, *m* ...), nekaj glasov smo zaradi boljše sinteze tudi ukinili (Tabela 2). Za primer v Tabeli 1 navajamo pravila pretvorbe grafemov *nj* znotraj posamezne besede.⁷ Za pravilno pretvorbo je pomemben predvsem vrstni red pravil, tako program od prvega do primernega pravila išče pravo izbiro glede na glasovno okolje aktualne besede (Tabela 1). Na Sliki 2 je izsek iz programske kode C++, v katerem so ta pravila zapisana.

Tabela 1: Pravila grafemsko-fonemske pretvorbe grafemov *nj*

Grafem in glasovno okolje	Glas (zapis MRPA)
1. <i>nj</i> pred <i>g</i>	[N]
2. <i>nj</i> pred <i>h</i>	[N]
3. <i>nj</i> pred <i>k</i>	[N]
4. <i>nj</i> pred soglasnikom	[n']
5. <i>nj</i> na koncu besede	[n']
6. <i>nj</i> pred samoglasnikom	[nj]

```

levo      jedro      desno      novo jedro
L"n",    L"nj",    L"g",      GOVFONEM_N1
L"n",    L"nj",    L"h",      GOVFONEM_N1
L"n",    L"nj",    L"k",      GOVFONEM_N1
L"n",    L"nj",    L"$",      GOVFONEM_N2
L"n",    L"nj",    L" ",      GOVFONEM_N2
L"n",    L"nj",    L"#",      GOVFONEM_N_
L"n",    L"n",     L"g",      GOVFONEM_N1
L"n",    L"n",     L"h",      GOVFONEM_N1
L"n",    L"n",     L"k",      GOVFONEM_N1
L"n",    L"n",     L"f",      GOVFONEM_N3
L"n",    L"n",     L"v",      GOVFONEM_N3
L"n",    L"n",     L" ",      GOVFONEM_N
L"n",    L"\x014b", L"n",     GOVFONEM_N1
L"n",    L"\x0148", L"n",     GOVFONEM_N2

```

Slika 2: Izsek iz programske kode C++

⁶ V oglatih oklepajih uporabljamo fonetični zapis MRPA, v poševnem tisku pa, če ne gre za grafeme, mednarodni zapis IPA.

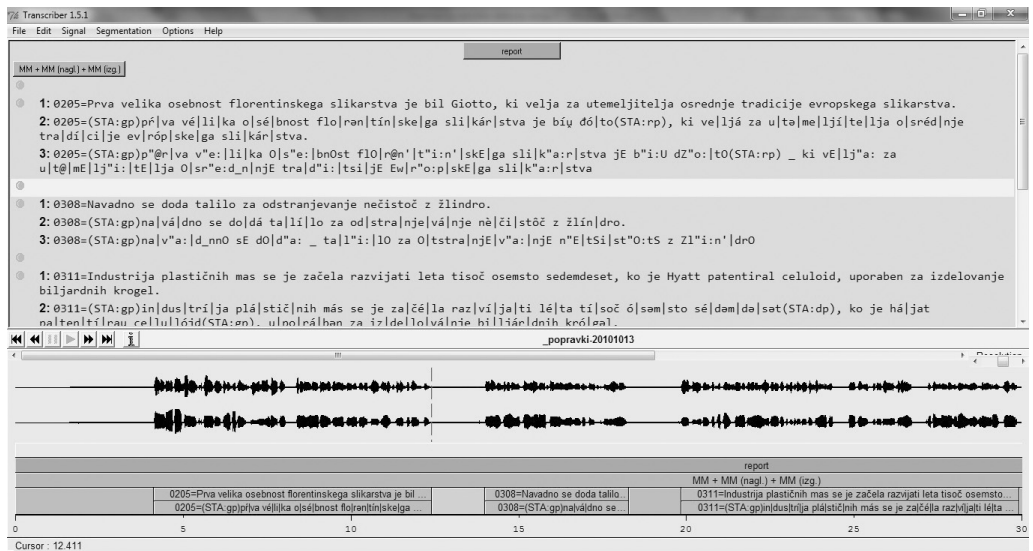
⁷ Grafemsko-fonemska pravila delimo na tista, ki urejajo pretvorbo znotraj besed, in tista, ki urejajo pretvorbo na medbesednih glasovnih stikih.

Tabela 2: Primeri ukinjenih glasov

MRPA	IPA	Opis glasu	Primer
b_n	ɓ	favkalni <i>b</i>	območje
b_f	ɸ	zobnoustnični <i>b</i>	ob fakulteti
d_l	ɖ	obstranski <i>d</i>	dlan
d_f	ɸ̥	favkalni <i>d</i>	odfrleti
w	w	ustničnoustnični šumni zveneči <i>v</i>	vzeti
W	ʍ	ustničnoustnični šumni nezveneči <i>v</i>	vsaj

Dodali smo tudi nove simbole – grafeme za ozka nenaglašena *o* in *e*. Ozkih nenaglašanih glasov prej niso poznali, saj naj bi obstajali samo široki nenaglašeni glasovi. To pravilo se je velikokrat izkazalo kot nepraktično, predvsem pri breznaglasnicah. Ozek izgovor glasu *o* smo uvedli pri prihodnjiku *bo*, čeprav bi moral biti izgovorjen široki *ɔ*. V transkripcijah smo imeli tudi dve varianti izgovora besede *že* (*že* in *žē*), obe smo nato zapisali še v slovar izgovorjav. Poleg znaka za premor (_) smo dodali tudi znaka za vdih (<) in izdih (>).

Transkripcije govorne zbirke smo popravljali v programu Transcriber. »V spodnjem delu zaslona je tudi grafični prikaz govornega in laringografskega signala. Tako smo lahko lažje ločili med zvenečimi in nezvenečimi glasovi ter odmori, in sicer brez natančnejše sonogramske analize« (Robida 2013: 30).



Slika 3: Delovno okolje v programu Transcriber

4 Slovnična analiza in slovar izgovarjav

Sintetizator mora na vhodnem besedilu opraviti tudi slovnično analizo. Besedam določi besedno vrsto, sklon, spol, število idr. Določi tudi nekatere stavčne oznake povedi, ki bodo npr. omo-gočile vprašalno, povedno ali veledno intonacijo. 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 pome-ne glagolov s pomeni samostalnikov glede na vlogo, v kateri se pojavljajo. Tako je mogoče določiti, da sta glagol (oziroma glagolski pomen) »skuhati« in samostalnik »kosilo« povezana tako, da je kosilo tipično predmet v tožilniku. (Holozan 2004)

Pri popravljanju transkripcij govorne zbirke smo morali popravljati tudi napake slovnične analize, nekaj primerov smo opisali v Tabeli 2.

Tabela 3: Primeri napačnih slovničnih analiz (Robida 2013: 25)

Poved	MRPA	Verjetna napaka
Najemi sobo, govori s točajem in profesorjem Knuthom.	na j«E: mi s«O: bO gO vOr”i : s tO tS”a: j@m in prO f”e: sOr j@m kn”u: tOm	Namesto veledne oblike glagola <i>govôri</i> uporabi povedno obliko <i>govorí</i> .
Albert je bil netakten, žaljiv, porogljiv in celo grob.	»a: l b@rt jE b«i:U n«E t«a:k t@n _ Za lj«i:U _ pO rO glj«i:U _ in tsE l«o: gr”Op	Namesto pridevnika <i>grôb</i> je besedo prepoznal kot samostalnik <i>gròb</i> .
Nastopil bo lahko vsak, ki ima drsalke in ki ve, kaj je pingpong.	na st«o: piU bo lax k«o: Us«a:k ki i m«a: d@ r s«a: l kE _ in ki v«e: k«a:j j”e: p”i:Nk p”O:Nk	Glagol <i>biti je</i> zamenja z glagolom <i>jesti jé</i> .

Lastnosti vseh besed so določene v slovarju izgovarjav, ta je pri slovenščini za sintezo govora zaradi nepredvidljivega naglaševanja nujen. Obsega približno 6.900.000 oblik besed z določenimi MSD-ji (to so oblikoskladenjske oznake) in naglasom. V slovarju se besede delijo na leme, ena označuje vse oblike določene besede znotraj iste besedne vrste. Primer sta samostalnik *ljubezen*, ki ima svojo lemo, in pridevnik *ljubezenski* s svojo.

Jezikovne tehnologije, z njimi tudi *Govorec 3*, so prihodnost ohranjanja jezika in priložnost širjenja slovenščine ter njene standardizacije. Kvalitetnejše bodo, bolj se bodo uporabljale, bolj bomo lahko z njimi jezikoslovci vplivali na vsakdanje uporabnike slovenskega jezika in ga tako ohranjali.

Viri in literatura

- HOLOZAN, Peter, 2004: Uporaba glagolskih predlog pri strojnem prevajanju. Tomaž Ejavec, Jerneja Žganec Gros (ur.): *Jezikovne tehnologije: zbornik B 7. mednarodne multikonference Informacijska družba IS 2004, od 9. do 15. oktober 2004*. Ljubljana: Institut „Jožef Stefan“. 128–129. <http://nl.ijs.si/isjt04/zbornik/sdjt04-04holozan.pdf>
- ROBIDA, Nejc, 2013: *Sinteza govora in Govorec 3. Diplomsko delo*. Ljubljana: Filozofska fakulteta.
- ROZMAN, Simon, 2005: *Sinteza govornega signala na osnovi metode HNM. Magistrsko delo*. Ljubljana: Fakulteta za računalništvo in informatiko.
- TIVADAR, Hotimir, 2010a: Gradivna utemeljenost opisa slovenskega govorjenega jezika. Vojko Gorjanc, Andreja Žele (ur.): *Izzivi sodobnega jezikoslovja*. Ljubljana: Znanstvena založba Filozofske fakultete. 53–62.
- TIVADAR, Hotimir, 2010b: Normativni vidik slovenščine v 3. tisočletju – knjižna slovenščina med realnostjo in idealnostjo. *Slavistična revija* 58/1. 105–116. <http://www.dlib.si/details/URN:NBN:SI:DOC-BHCJ5UO8>
- TIVADAR, Hotimir, 2012a: Nevarna razmerja med pisnim in govorjenim jezikom. Nataša Jakop, Helena Dobrovoljc (ur.): *Izzivi sodobnega jezikoslovja*. Ljubljana: Založba ZRC, ZRC SAZU. 203–211.
- TIVADAR, Hotimir, 2012b: Nove usmeritve pri raziskavah govora s pogledom v preteklost. *Slavistična revija* 60/4. 587–601. http://www.srl.si/sql_pdf/SRL_2012_4_02.pdf
- ZEMPLJAK, Melita, KAČIČ, Zdravko, DOBRIŠEK, Simon idr., 2002: Računalniški simbolni fonetični zapis slovenskega govora. *Slavistična revija* 50/2. 159–169.
- ZEMPLJARIČ MIKLAVČIČ, Jana, 2008: *Govorni korpusi*. Ljubljana: Univerza v Ljubljani, Filozofska fakulteta.