

PISANI DISKURS GOVORNO JEZIČKE BAZE HABIL

Miško Subotić, Silvana Punišić, Nataša Čabarkapa
Institut za eksperimentalnu fonetiku i patologiju govora, Beograd
iefpg@eunet.yu

Sadržaj - Govorno jezičke baze postale su jedan od polaznih elemenata u različitim oblastima izučavanja govora i jezika. Ispostavilo se da je i samo kreiranje ovih baza složen i zahtevan naučno istraživački postupak u kome treba sagledati uticaj velikog broja parametara na pojedine karakteristike govora i jezika. Jedna od osnovnih podela govorno jezičkih baza je podela na: baze koje su nastale čitanjem jezičkog materijala i baze iz spontanog govora. Baza HABIL pripada prvoj grupi i njenu jezičku osnovu čini pisani tekst. Pisani diskurs ima velikog uticaja na prozodijske karakteristike govora nastalog čitanjem istog. Zbog toga smo u radu analizirali pisani diskurs jezičkog korpusa baze HABIL sa ciljem da se dobiju parametri koji će se koristiti za istraživanje funkcionalnih međuzavisnosti prozodijskih karakteristika govornog korpusa i karakteristika pisanog diskursa odabranog jezičkog korpusa.

1. UVOD

Postoje različite strategije i pristupi pri kreiranju govorno jezičkih baza. Zajedničko svim pristupima jeste da namena govorno jezičke baze ima presudan uticaj na njen sadržaj i organizaciju. Prateći razvoj govora i jezika dece, od 0 do 7 godina, sa poremećajima verbalne komunikacije, uočili smo da postoji kritični govorno jezički materijal dovoljan za razvoj govora i jezika. Sama ova činjenica je intrigantna jer nas upućuje na zaključak da konačan skup pokriva dovoljno široko varijaciono polje govora i jezika, te može poslužiti kao osnova za izučavanje govorno jezičkih fenomena [1, 2]. Polazeći od ove činjenice mi smo u kreiranju baze HABIL odabrali knjigu "Dnevnik jednog audiolingviste" [3] jer je u njoj, na osnovu pedesetogodišnjeg iskustva u radu sa decom, hronološkim redom izložen jezički materijal. Posebna vrednost ovog materijala je upravo u tome što su sve jezičke jedinice (fonološke, leksičke, semantičke, sintaksičke, gramatičke) koje mogu uticati na prozodiju govora izbalansirane i ravnomerno zastupljene tako da u korpusu postoji bar po jedan primer koji pokriva osnovni koncept i relaciju od interesa. Iako je nemoguće uključiti sve reči, oblike, sintaksne varijacije, korpus pokriva dovoljan broj jedinica pri čemu značajnu ulogu igra njihova frekvencija u datom kontekstu. Polazeći od diskursnog pristupa izgrađenog na uverenju da samo analiza jezičkih jedinica većih od rečenice u konkretnoj upotrebi, kontekstu ili situaciji objašnjava komunikaciju i ponašanje[4], mi smo iskoristili ponuđeno pravo da proširimo i modifikujemo odabrani pisani diskurs uključivši u njega i jedinice nižeg nivoa od rečenice [5]. Postoji jasan uticaj pisanog jezika na govor pri čitanju jer se neke karakteristike pisanog diskursa (interpunkcija, veliko slovo, rečenica, paragraf, poglavlje ili čitavo delo) jasno realizuju odgovarajućim govornim činovima (intonacija, pauza, jačina, visina itd) [6]. Upravo ove karakteristike govornoj formi diskursa daju visok stepen varijabilnosti koja je posebno ogleđa na prozodijskom planu [7, 8].

Izučavanje uticaja pisanog diskursa na govor u prvom koraku zahteva detaljnu analizu samog jezičkog materijala i njegovo opisivanje jedinica samog diskursa.

Cilj ovog rada je analiza jezičkog materijala baze HABIL jedinicama diskursa kako bi se opisao jezički korpus i omogućilo istraživanje funkcionalnih međuzavisnosti prozodijskih karakteristika govornog korpusa i karakteristika pisanog diskursa odabranog jezičkog korpusa.

2. REZULTATI I DISKUSIJA

Sadržaj knjige "Dnevnik jednog audiolingviste" je organizovan kao skup nezavisnih govorno jezičkih vežbi, čiji redosled prati dinamiku razvoja govora i jezika dece od 0 do 7 godina, pretočen u formu korpusa. Opređeljujući se za gramatičku analizu jezičkog korpusa analizirali smo odabrani materijal u okviru sledećih jedinica: 1. izolovani glasovi; 2. glasovi u nizu; 3. izolovani slogovi; 4. slogovi u nizu; 5. izolovane reči; 6. reči u nizu; 7. izolovane rečenice; 8. rečenice u nizu; 9. tekstovi; 10. recitacije. Odrednica "u nizu", odnosi se na glasove, tj. slogove, reči i rečenice povezane misaono-logičkom komponentom i kontekstualno organizovane. Javljanje rečenica u nizu čini manje celine od paragrafa ili drugih većih diskursnih jedinica.

Struktura jezičkog korpusa "Habil" je data u tabeli 1. čini ukupno 113598 glasova, 35683 reči, 4812 rečenica, 194 recitacije i 141 tekst.

Tabela 1. Struktura jezičkog korpusa

| | |
|--------------------|---------|
| Ukupno glasova | 113.598 |
| Ukupano reči | 35.683 |
| Ukupno rečenica | 4.812 |
| Ukupano recitacija | 194 |
| Ukupno tekstova | 141 |

Ukupan broj glasova čine izolovani glasovi (A, O, R...), glasovi u nizu (PA, OP, BA...) ili glasovi koji su sastavni deo reči i rečenica. Izolovani i glasovi u nizu najčešće predstavljaju onomatopeje ili oponašanje zvukova iz prirode i mogu se pojavljivati samostalno, u okviru rečenica (izolovanih i u nizu) ili rečenica koje su sastavni deo većih diskursnih jedinica. Reči se u datom korpusu javljaju kao izolovane reči (AUTO, MAMA, UVO...), reči u nizu (TATA-TATA...), u okviru rečenica (izolovanih i u nizu) ili rečenica koje su sastavni deo većih diskursnih jedinica. Sve rečenice u korpusu posmatrane su kao izolovane (AVION LETI. BEBA SPAVA...), rečenice u nizu (TO JE BATA. BATA JE PAO. BATU BOLI NOGA), u paragrafu, u kraćem ili dužem tekstu. Tekstovi su dati u vidu kratkih priča, basni i bajki.

Vrednosti prikazane na grafikonu br.1. odnose se na raspodelu izolovanih glasova, slogova, reči i rečenica. U odnosu na ukupan broj pojavljivanja posmatrane kategorije izolovano se javljaju: glasovi 2.6%, reči 5.9% i rečenice 32%. Vidi se da se u odnosu na ukupan broj pojavljivanja pojedinih kategorija kao izolovane najčešće javljaju rečenice. Najveći broj od ukupnog broja izolovanih rečenica i rečenica u nizu (grafikoni br. 2 i br. 3) struktuirano je od 3 reči. Po zastupljenosti slede rečenice od 2 i 4 reči dok su rečenice od 5 i više reči nisko zastupljene. Analizom broja reči u rečenicama koje su organizovane u tekst (graf. br. 4) uočavamo visoku zastupljenost rečenica struktuiranih od 5 i 4

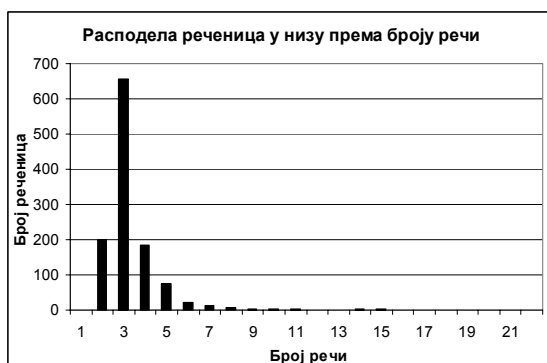
reči; po zastupljenosti slede rečenice od 3, 6, 7, 8 reči dok su rečenice od 9 i više reči nisko zastupljene.



Графикон 1. *Расподела језичких сегмената у корпусу*



Графикон 2. *Расподела изолованих реченица према броју речи*

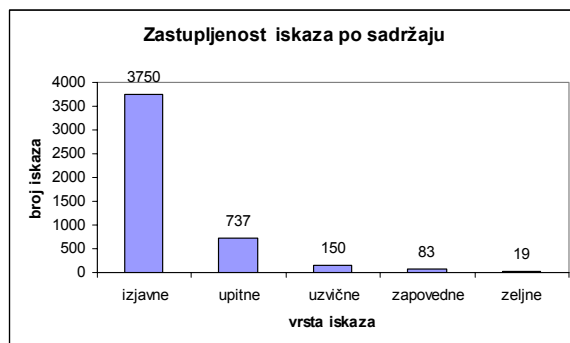


Графикон 3. *Расподела исказа у низу према броју речи*



Графикон 4. *Расподела исказа у тексту према броју речи*

Na osnovu istraživanja, rečenica se kod dece razvija kroz sledeće faze: rečenice od jedne reči, rane rečenice, kratke rečenice, potpune rečenice. Usvajanje gramatike nije samostalan proces već je omogućen ostalim aspektima govornog razvoja. Jedno od merila gramatičkog (morfološkog i sintaksičkog) razvoja kod dece je prosečna dužina iskaza u smislu reči ili morfema. Maksimalna postignuća u sazrevanju sintaksičkog podsistema jezika na bazi biološkog kapaciteta, deca postižu u periodu od 2 do 4 godine starosti. Ovi podaci su u saglasnosti sa distribucijom po zastupljenosti rečenica preme sastavu: proste pa složene (obe sa komunikativnom funkcijom) i prema sadržaju: obaveštajne, upitne, uzvične, zapovedne, željne od kojih svaka može biti potvrdna i odrična [9]. Uzimajući u obzir i podelu iskaza u jeziku na potvrdne i odrične, iskazi u potvrdnom obliku javljaju se 4379 puta dok se u odričnom obliku javljaju 73 puta. Ove vrednosti izražene u procentima: 98,5% i 1,5% pokazuju da je potvrdni način, kao forma iskaza kojom će dete izražavati sebe, prioritetan u odnosu na ostale mada sadržaj korpusa obezbeđuje neophodnu zastupljenost i drugih rečenica po sastavu i sadržaju koje mogu pokriti lingvističke i paralingvističke uticaje na prozodiju srpskog jezika [10]. Iskaze možemo posmatrati i prema tome da li pored jezičke nose i određenu emotivnu poruku (iz konteksta se pretpostavlja da će u toku čitanja one biti izgovorene sa određenom emfatičnom). Posmatrajući iskaze sa aspekta prisustva ili odustva ove karakteristike, u korpusu se nalazi 296 iskaza koji prilikom čitanja (govorenja) imaju emfatični izgovor i 4516 iskaza koji su bez emfatičnog



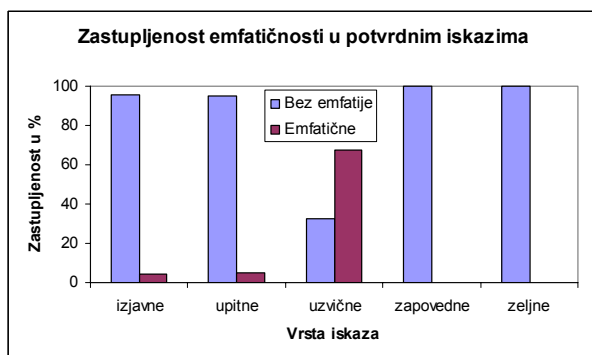
Графикон 5. *Zastupljenost potvrdnih iskaza po sadržaju*

izgovora. Emfatični izgovor bi u okviru govornog korpusa baze HABIL trebalo da bude zastupljen na nivou od 6.15%.

Uvodeći još jednu podelu iskaza prema sadržaju na: izjavne, upitne, zapovedne, uzvične i željne iskaze (graf. 5), vidimo da u korpusu preovlađuje izjavna a zatim sledi upitna, uzvična, zapovedna i željna forma izražavanja. Izjavna forma izražavanja javlja se u 79% od ukupnog broja iskaza i daleko nadmašuje sve ostale načine izražavanja.

Pošto se određena emocija može izraziti različitim vrstama iskaza, od interesa je utvrditi zastupljenost emfatičnosti, tj. pojavu udruživanja intonacije sa drugim prozodijskim i parajezičkim obeležjima u vrstama iskaza koji se javljaju u korpusu. Ako se normalizuje emfatičnost na ukupni broj iskaza date vrste onda se dobija zastupljenost prikazana na grafikonu 6. U odnosu na ukupan broj javljanja pojedine vrste iskaza, emfatičnost se najčešće javlja u formi uzvičnih iskaza. Od njihovog ukupnog broja u jezičkom

korpusu 67.33% je tako kontekstualno određeno da u čitanju (govoru) nosi pored jezičke i emotivnu poruku koja bi u govoru bila prepoznata.



Grafikon 6. Zastupljenost emfatičnosti u različitim vrstama iskaza, jezičkog korpusa baze HABIL

Uočava se da smo u posmatranoj analizi koristili iskaz kao osnovnu jedinicu koju smo posmatrali u okviru jezičkog korpusa. Pitanje izbora fokusa istraživanja i osnovne merne jedinice u analizi određenog sistema po pravilu uvek podleže kritičkoj analizi, odnosno, postoji određeni broj činjenica koje opravdavaju takav izbor dok ga neke činjenice, na drugoj strani, mogu dovesti u pitanje. Mi smo se opredelili za iskaz kao osnovnu mernu jedinicu, imajući u vidu da u govoru i jedan izolovani izgovorni glas može predstavljati celovit komunikacioni iskaz koji u sebi sadrži neophodno značenje socializovano i za govornika i za slušaoca. Na osnovu dobijenih rezultata svaki od grafema (koji će biti pretočen u izgovorni glas) možemo opisati većim brojem parametara koje smatramo relevantnim i utičućim na prozodijske karakteristike u toku čitanja (govora). Polazeći od opštijih ka pojedinačnim obeležjima svaki konstitutivni element iskaza u okvirima jedinica iskaza možemo opisati sledećim atributima: **diskursna celina** (izolovani iskaz, iskazi u nizu, recitacija, tekst [bajka, basna, kratka priča]); **vrsta iskaza prema konstitutivnim elementima** (glas, slog, reč, rečenica); **broj konstitutivnih elemenata u iskazu** (broj konstitutivnih elemenata u iskazu, redni broj konstitutivnog elementa); **vrsta iskaza prema sadržaju** (potvrđni [izjavni, upitni, uzvični, zapovedni i željni], odrični); **emfatičnost u iskazu** (postoji, ne postoji).

3. ZAKLJUČAK

Kreiranje govorno jezičke baze na osnovu odabranog jezičkog materijala uslovljava da se opisivanje karakteristika jezičkog materijala u određenoj meri razlikuje od opisa koji bi se primenio da se kreira samo jezička baza odnosno jezički korpus. U opisivanju jezičkog korpusa rečenica predstavlja polazni element u opisivanju teksta. Čini nam se da je za analizu govornih korpusa iskaz taj polazni element u opisivanju govora. Obzirom da je osnovni cilj formiranja baze HABIL analiza prozodije to smo se i u opisu jezičkog materijala opredelili za iskaz. U ovakvoj analizi uočava se da iskazi nisu identični sa rečenicama odnosno iskaz se može ali i ne mora podudarati sa rečenicom u posmatranom tekstu. Ovo je bitan preduslov za istraživanje uticaja diskursa na karakteristike govora čitanog teksta. Tek će slična analiza govornog materijala (čitanoj teksta) i formiranje relacione baze podataka omogućiti detaljnija istraživanja uticaja teksta na govor. Ovakva analiza bi trebala da ukaže koliko diskurs i

karakteristike pisanog teksta utiču na prozodijske karakteristike srpskog jezika.

Napomena: Ovaj rad je finasiran od strane Ministarstva za nauku i zaštitu životne sredine Republike Srbije, u okviru projekta OI 1784.

LITERATURA

- [1] M. Subotić, S. Punišić, S. Pantelić, N. Čabarkapa, M. Mitić, N. Vasić, S. Milenković: Govorna baza podataka srpskog jezika, Govor i jezik, IEFPG, 202-209, Beograd, 2003
- [2] M. Subotić, S. Punišić, N. Čabarkapa: Govorna baza za rehabilitaciju i rehabilitaciju dece sa poremećajima verbalne komunikacije, poglavlje u monografiji: S. Jovičić, M. Sovilj (editori): Govor i jezik: interdisciplinarna istraživanja srpskog jezika, I, IEFPG, Beograd, 70-97, 2004.
- [3] N. Domski: Dnevnik jednog audiolingviste, IEFPG, Beograd, 2002.
- [4] S. Savić: Diskurs analiza, Filozofski fakultet, Novi Sad, 1993.
- [5] R. Robin: L'analyse du discours entre la linguistique et les sciences humaines, Languages, no 81, str.121-128, Paris, 1986
- [6] D. Kristal: Kembrička enciklopedija jezika. Nolit, Beograd, 1996.
- [7] M. Subotić, M. Sovilj, S. Punišić, S. Pantelić; S. Djoković: Stabilnost prozodijskih elemenata govornika, Najnowsze przemiany fonetyki slowianskiej, Acta Universitatis Nicolai Copernici, studia slavica IX, 37-45, Torun, Poljska, 2004.
- [8] J. Jokanović-Mihajlov: Prozodijska sredstva u organizaciji diskursa, Naučni sastanak slavista u Vukove dane, Sveska 21/II, Beograd, 1991.
- [9] Stanojčić, Lj. Popović: Gramatika srpskoga jezika. Zavod za udžbenike i nastavna sredstva. Beograd, 1999.
- [10] M. Sečujski, V. Delić: A Serbian Speech Database for Speech Synthesis, web site

Abstract – Speech - language basis had become one of the crucial elements in different fields of investigation in the area of speech and language. The creation process of this basis is very complicated scientific process because scientists had to perceive an influence of a great number of parameters on different characteristics of speech and language. One of the basic speech and language basis classification consider: a) basis created by reading language materials and b) basis from spontaneous speech. HABIL basis is in the first group and writing text is her basis element. Writing discourse has a great influence on prosodic characteristics of speech during the process of reading. We analyze writing discourse of language corpus HABIL basis. Our goal was to find parameters we can use in the prosodic characteristics of speech corpus and characteristics of writing discourse of selected language corpus.

WRITING DISCOURSE OF SPEECH - LANGUAGE HABIL BASIS

Miško Subotić, Silvana Punišić, Nataša Čabarkapa