

UDK 81'23:004
81'276-053.2
811.163.42'23
Izvorni znanstveni članak
Prihvaćeno za tisak 2. 10. 2001.

Jelena Kuvač & Marijan Palmović
Laboratorij za psiholingvistička istraživanja POLIN
Odsjeka za logopediju, Edukacijsko-rehabilitacijskog fakulteta Sveučilišta u
Zagrebu

Računalna obrada dječjeg jezika na primjeru usvajanja umanjenica

Članak je pismeni oblik predavanja održanog na Lingvističkom krugu i sadrži kratak pregled baze podataka CHILDES i njegove primjene na hrvatski korpus dječjeg jezika. Drugi dio članka ogledna je analiza, koju programi razvijeni unutar sustava CHILDES olakšavaju, a odnosi se na istaknutu odliku dječjeg jezika u hrvatskome – brojnost umanjenica. Prikazuje se tijek usvajanja umanjenica, a raspravlja se o njihovoj ulozi u procesu usvajanja morfologije u hrvatskome.

Uvod:

Veća ili manja razmimoilaženja velikog broja teorija i pristupa u području psiholingvistike o tome kako djeca usvajaju jezik ipak ujedinjuje ista polazišna osnova odnosno isti pristup u prikupljanju podataka. Naime, sve teorije polaze od toga da svoje analize i objašnjenja trebaju temeljiti na korpusu svakodnevnoga spontanog međudjelovanja djeteta i njegovih sugovornika. Na tako stvorenom korpusu stvaraju se transkripti ili prijepisi koji imaju najvažniju ulogu u istraživanju govornog diskursa jer u vremenu zamrzavaju složene događaje i vidove međudjelovanja (Edwards, J. 1993.). Dvije paralelne linije razvoja utjele su na to kako transkribirati dječji jezik da bi se objasnio jezični razvoj. Jedna od tih linija je konceptualna: različita pitanja koja istraživači postavljaju određuju vrstu podataka koji se prikupljaju. Druga linija, elektroničke inovacije, pridonijela je razvoju razne sofisticirane opreme kao nadopuna papiru i olovci (Bloom, L. 1993.).

Naime, najranija su istraživanja bila ograničena na bilježenje početaka, tj. zabilježavanje prve izgovorene riječi. Ovakva metoda otvorila je put k određivanju razvojnih aspekata jezičnog razvoja, ali je isto tako uvjetovala da istraži-

vači, a još češće roditelji, nastave bilježiti dječji jezik, tj. da pišu dnevnik. Zahvaljujući dnevnicima omogućeno je opisivanje jezičnog razvoja. Najveći broj dnevnika stvoren je kasnih 50-ih godina prošlog stoljeća uvođenjem audioopreme. Ta je oprema s jedne strane omogućila prikupljanje velikog uzorka spontanoga govora i govora na unaprijed osmišljenim ispitivanjima, a s druge strane potpunije zapisivanje jezične proizvodnje, tj. ta je oprema omogućila da zapis jezične proizvodnje bude potpuniji nego što je to mogao biti dnevnik. Prednost je pohranjivačke tehnologije smještanje neograničene količine izravno snimljenih podataka i na taj način stvaranje korpusa prijepisa. Prijepisi pri tome zadovoljavaju dva bitna uvjeta: vjerodostojnosti i praktičnosti. Prednosti koje daje tehnologija odražavaju se na dostupnost, kvalitetu i kvantitetu lingvističkih podataka. Snimke postaju dostupne, prijepis provjerljiv, a cjelokupni jezični korpus izmjerljiv – moguće ga je podvrgnuti analizi i interpretaciji.

Primjena suvremenih tehnologija omogućila je danas stvaranje korpusa dječjeg jezika, tj. nacionalnih korpusa ulazi li se u okvire međujezičnih usporedbi. Danas su putem interneta dostupni prikupljeni podatci za više od četrdeset jezika, što istraživačima dječjeg jezika olakšava međujezične usporedbe budući da su podatci prikupljeni istom ili sličnom metodologijom. Razvijeni su mnogi programi za bilježenje i obradu jezika ili samo nekih jezičnih vidova, a od njih su najpoznatiji:

HIAT (Halbinterpretative Arbeitstranskriptionen), CALL (Computer-aided Language Learning), KCLA (Korean Computerized Language Analysis), SALT (Systematic Analysis of Language Transcript) i CLAN i CHAT u okviru CHILDES-a (Child Language Data Exchange System).

Hrvatski korpus dječjeg jezika razvija se unutar baze podataka CHILDES koji sadrži program za kodiranje zapisa CHAT i skup programa za analizu CLAN. Hrvatski je jezik tek jedan od četrdesetak jezika koji svoj korpus jezičnog razvoja razvija unutar ove baze podataka. Primjena ovih programa toliko je rasprostranjena da banka podataka CHILDES danas sadrži jezične korpusse djece s posebnim jezičnim teškoćama, znakovnoga jezika, dvojezičnoga razvoja te pripovjedačke sposobnosti. Uporaba ovih programa u hrvatskome jeziku ograničena je na praćenje urednoga jezičnog razvoja, dok je u pripremi obrada dječje pripovjedačke sposobnosti u sklopu magistarskoga rada. Velika je prednost CHILDES-a to što uporaba ovih programa nužno ne pretpostavlja teorijske okvire pa se njima služe istraživači koji djeluju unutar raznih teorija i pristupa.

Opis programa CHAT i CLAN

Cijeli su sustav 1983. osmislili Brian MacWhinney sa Sveučilišta Carnegie-Mellon i Catharine Snow sa Sveučilišta Harvard.

Prijepis tonskoga zapisa unosi se u poseban format iz paketa programa CHAT koji ujedno sadrži i standardizirane kodove, što znači jedinstveno obilježavanje iste pojave u svim jezicima. Isto tako program omogućava unošenje novih kodova u datoteku »depfile.cut« za potrebe nekog jezika. Izgled prijepisa

i položaj svih kodova strogo je određen jer se nepridržavanjem pravila kodiranja onemogućava analiza na razini CLAN programa. Svaki prijepis sadrži *retke zaglavlja* koji sadrže opće podatke o djetetu (slika 1.) te *sadržajne* ili *glavne* i *zavisne retke*.

Slika 1. Izgled redaka zaglavlja

```
@Begin
@Participants: ANT Antonija Target_Child, DRA Draženka
                Mother, ANA Grandmother
@sex of ANT:   female
@Birth of ANT: 31-JUL-1993
@Age of ANT:   2;00.2
@Date:        02-AUG-1995
@Coder:       Maja Majić
@Filename:    antbla18.cha
@Coding:      CHAT
@Situation:   Antonija i baka igraju se u sobi, ali
                Antonija primjećuje da majka snima.
```

Sadržajni retci nose sadržaj teksta s kodovima koje ispitivač proizvoljno uzima. Program CHAT za potrebe kodiranja u odnosu na sve jezične sastavnice pruža stotinjak kodova. Koliko i koje kodove će istraživač upotrijebiti, ovisi o interesima i ciljevima istraživanja. Sadržajni redak uvijek započinje zvjezdicom (*) i troslovnom oznakom govornika. Zavisni retci nadopunjavaju sadržajne retke istraživačevim komentarima, objašnjenjima situacija ili označavanjem pogrešaka na fonološkoj i morfološkoj razini.

Slika 2. Izgled sadržajnog i zavisnog retka

```
*ANT:   čekaj # kaj@d(h)oćeš # čekaj .
%com:   dolazi k stolu da vidi kasetofon .
*DRA:   hoæemo se malo snimati ?
```

Nakon što je primijenjen program CHAT prijepis mora proći program CHECK koji provjerava odgovara li zaista prijepis svim uvjetima programa CHAT. Pogreške koje obilježava ovaj program odnose se na pravilnu upotrebu kodova, njihovo smještanje te pravilni izgled prijepisa. Provjeravanje sadržajne odnosno pravopisne točnosti sadržajnih redaka ovim programom nije moguće. Tek nakon što program CHECK propusti prijepis moguće je prijeći na program CLAN i započeti jezičnu analizu. Program CLAN sadrži ukupno 39 programa za obradu različitih vidova dječjeg jezika. S obzirom na to da su ti programi namijenjeni različitim svrhama, možemo ih podijeliti u dvije osnovne skupine:

1. programi za pomoć – programi iz ove skupine ne sudjeluju izravno u analizi, ali olakšavaju neke dodatne postupke. Tako, na primjer, program DATES izračunava djetetovu kronološku dob, program LINES automatski ispisuje broj retka u prijepisu, FREOMERG služi za povezivanje većeg broja frekvencijskih datoteka. Isto tako u ovu skupinu možemo ubro-

jiti i statističke programe kao STATFREQ za izračunavanje riječi ili kodova i WDLLEN za histogramsko prikazivanje duljine riječi.

2. programi za analizu jezika – programi iz ove skupine mogu biti usmjereni prema jednoj od tri jezične sastavnice koje su predviđene za analizu ovim programom ili za interakcijske analize. Tako programi COMBO, COOCCUR, MAXWD služe sintaktičkoj analizi, MOR, MLU morfološkim analizama, PHONFREQ fonološkoj analizi; a KEYMAP interakcijskoj analizi. Uporabom programa FREQ omogućeno je dobivanje frekvencijske pojavnosti riječi u svakom pojedinom prijepisu te omjera natuknica i pojavnica.

Trenutni interesi u proučavanju usvajanja hrvatskoga jezika usmjereni su prema morfološkoj analizi. Osnovni je program za takvu analizu program MOR koji pruža dvije mogućnosti za svoje ostvarenje. Prvi je način unošenje cjelokupnoga hrvatskog leksikona odnosno korijena riječi i afiksa u posebne datoteke. Ovakav način stvaranja elektronske gramatike prikladan je za jezike s malo gramatičkih morfema, npr. za engleski. Za hrvatski jezik ovaj je način teži jer bi bilo potrebno unijeti veliki broj vezanih morfema koji su, uz to, često homofoni (Anđel, Blaži 2000.). Kada bi se i stvorio tako veliki leksikon, razina njegove prikladnosti bila bi niska jer se dječja gramatika uveliko razlikuje od formalne gramatike. Iz tih razloga za morfološku analizu hrvatskoga dječjeg jezika pribjegli su drugom načinu, stvaranju leksikona iz samih prijepisa dječjega govora odnosno leksikona koji sadržavaju samo one riječi koje su se pojavile u prijepisima (slika 3a.).

a@b	{{scat ?}}	ade	{{scat V:1:IMPF:INTR}} "hajde&IMP:2S"
aa@b	{{scat ?}}	ajde	{{scat V:1:IMPF:INTR}} "hajde&IMP:2S"
aaa@b	{{scat ?}}	ajdemo	{{scat V:1:IMPF:INTR}} "hajdemo&IMP:1P"
aaaa	{{scat ?}}	ajmo	{{scat V:1:INTR:IMPF}} "hajdemo&IMP:1P"
au	{{scat ?}}	ajejuja	{{scat INTERJ}} "aleluja"
auuu	{{scat ?}}	ajmete@	{{scat V:1:INTR:IMPF}} "hajde&IMP:2P"
avion	{{scat ?}}	bac	{{scat V:5:PFV:TRANS}} "baciti&IMP:2S"
aviončić	{{scat ?}}	baca	{{scat V:5:IMPF:TRANS}} "bacati&PRES:3S"
batina	{{scat ?}}	bacaj	{{scat V:5:IMPF:TRANS}} "bacati&INF:2S"
bebicu	{{scat ?}}	bacala	{{scat V:5:IMPF:TRANS}} "bacati&PART-FEM:SG"
biće	{{scat ?}}	bacala	{{scat V:5:IMPF:TRANS}} "bacati&PART-NEUT:PL"
bježe	{{scat ?}}	bacali	{{scat V:5:IMPF:TRANS}} "bacati&PART-MASC:PL"
blaženka	{{scat ?}}	baca@	{{scat V:5:IMPF:TRANS}} "bacati&PART-MASC:SG"
blaženka	{{scat ?}}	bacat	{{scat V:5:IMPF:TRANS}} "bacati&INF"
blaženki	{{scat ?}}	bacati	{{scat V:5:IMPF:TRANS}} "bacati&INF"
blaženku	{{scat ?}}	baci	{{scat V:4:PFV:TRANS}} "baciti&IMP:2S"
bolje	{{scat ?}}	bacila	{{scat V:4:PFV:TRANS}} "baciti&PART-FEM:SG"
bonga	{{scat ?}}	bacili	{{scat V:4:PFV:TRANS}} "baciti&PART-MASC:PL"
bočica	{{scat ?}}		
bočice	{{scat ?}}		

Slika 3a. Neobrađeni leksikon

Slika 3b. Obradeni leksikon

Na neobrađenom leksikonu, koji je dobiven automatski, istraživač ručno unosi sve gramatičke osobine unutar vitičaste zagrade i iza nje. Unutar zagrada nalaze se obavijesti o vrsti riječi («scat» – »syntactic category«), dok se izvan zagrada nalaze ostale morfološke obavijesti – lice, broj, padež, vrijeme i sl. (slika 3b.). Nakon što su u leksikon unesene sve gramatičke osobine program MOR automatski će u prijepisu stvoriti nove zavisne retke – morfološke retke (%mor) u kojem se pojavljuje svaka riječ iz sadržajnog retka s njoj pripadajućim gramatičkim obavijestima (slika 4.).

```
*ANT: čekaj # kaj@d(h)oćeš # čekaj.  
%mor: V:5:IMPF:TRANS|čekati&IMP:2S PRO:INT|kaj  
V:1:IMPF:TRANS|htjeti&PRES:2S  
V:5:IMPF:TRANS|čekati&IMP:2S .  
%com: dolazi k stolu da vidi kasetofon .  
*DRA: hoćemo se malo snimati ?  
%mor: V:1:IMPF:TRANS|htjeti&PERS-1P PRO:REFL|sebe&ACC:SG:CLIT  
ADV:QUANT|malo  
V:5:IMPF:TRANS|snimati&INF ?
```

Slika 4. Primjer morfološki obradenog prijepisa

Ovdje je kratko izložen postupak potreban za obradu podataka vezanih za usvajanje hrvatske morfologije. Detaljne upute za upotrebu svih programa objavljene su u knjizi (MacWhinney, 1997.), dok se s najnovijim verzijama programa moguće upoznati putem internetske stranice CHILDES-a (<http://childes.psi.cmu.edu>). Novije verzije programa sadrže dodatne mogućnosti – povezivanje teksta prijepisa sa zvukom ili slikom uz dodatne kodove za, na primjer, geste.

Za oglednu obradu kojom će se predstaviti mogućnosti sustava CHILDES odabrali smo umanjence kao istaknutu pojavu u dječjem jeziku u hrvatskome.

Cilj, metode i ispitani uzorak

Obrada umanjencia u govoru djece u hrvatskome napravljena je za projekt Austrijske akademije znanosti »Pre- and Protomorphology« kojemu je polazište teorija prirodne morfologije (Dressler, Karpf, 1995.). Ukratko, teorijom se ne pretpostavlja urođeni mehanizam za usvajanje jezika, nego se jezični razvoj objašnjava samoorganizacijom jezičnih elemenata što, pak, znači da dijete usvaja jezik korak po korak¹. Redoslijed koraka, tvrdi se, ovisi o strukturi jezika – svaki korak na neki način olakšava savladavanje sljedećeg. Zbog toga se svakom usvojenom jezičnom elementu (u najopćenitijem smislu te riječi) u teoriji prirodne morfologije obično pridaje neka uloga. Tako se tvrdi da umanjence pojednostavljaju morfološki sustav redukcijom broja imenskih klasa, kao na primjer u litavskom (Savickienė, 1986.) ili čine paradigme »prozirnima«, tj. izbjegavaju se glasovne promjene na rubovima morfema, kao u finskom (Laalo, 1986.). Te dvije tvrdnje provjerit ćemo na hrvatskom korpusu dječjeg jezika.

Obrada umanjencia u dječjem govoru u hrvatskome temelji se na zapisima jednog djeteta, Antonije. Ona s roditeljima živi u Zagrebu, govori zagrebačkom štokavštinom s elementima kajkavskog, što je naročito naglašeno u interakciji djevojčice s djedom i bakom iz Sesveta. Antonija je snimana u dobi između 1;3 i 2;8, prosječno tri puta mjesečno po 45 minuta.

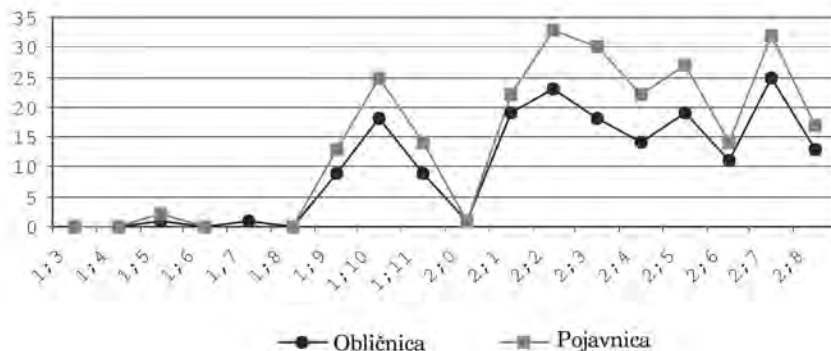
1 O eksperimentima sa samoorganizirajućim neuronskim mrežama koje ne mogu strukturirati jezični materijal ako im se da odjednom, nego samo u malim dozama vidi npr. MacWhinney ed. 1999.

Za ovaj su rad obrađena 42 zapisa Antonijina govora i u njima je izmjerena pojavnost umanjenica, istovremena pojavnost umanjenica i odgovarajućih ne-umanjenih riječi i dječje pogreške. Na taj način moguće je procijeniti primjenjuje li dijete pravilo ili oponaša roditelje.

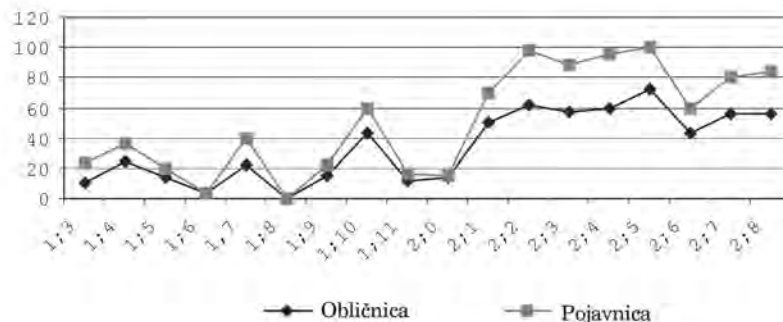
Kvantitativna obrada umanjenica

Obradene su samo imeničke umanjenice budući da se pridjevske i glagolske u cijelom korpusu pojavljuju tek nekoliko puta (*malički, crvenkast, piškiti i smijuljiti se*). Također, od svih vrsta tvorbe umanjenica, zabilježena je samo sufiksalna tvorba. Prefiksalna tvorba nije zabilježena.

Antonija je počela upotrebljavati umanjenice već u dobi od 1;5 (*mamica*), ali značajnija upotreba počinje u dobi od 1;9 kad je za vrijeme jednog snimanja upotrijebila devet različitih umanjenica. Broj umanjenica varira od mjeseca do mjeseca, ali ostaje visok (dvadesetak umanjenica mjesečno) sve do kraja razdoblja snimanja. Kretanje broja umanjenica u Antonijinu govoru prikazano je na slici 5. Broj umanjenica u govoru Antonijinih roditelja također varira i polagano raste kako roditelji prilagođuju svoj govor djetetovu. Kretanje broja umanjenica u govoru Antonijinih roditelja prikazano je na slici 6.



Slika 5. Broj umanjenica u Antonijinu govoru



Slika 6. Broj umanjenica u govoru Antonijinih roditelja

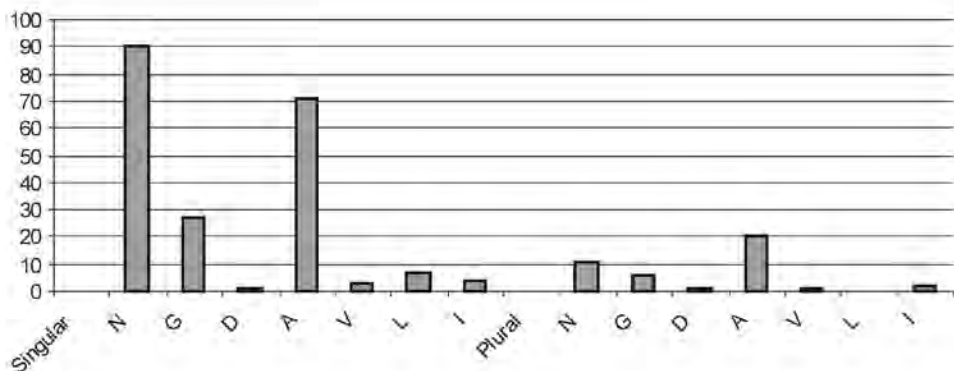
Iz slika 5. i 6. može se primijetiti kako broj umanjjenica varira istovremeno i kod Antonije i kod njezinih roditelja. Taj broj umanjjenica po snimanju najviše ovisi o temi: kad je ona »ozbiljna«, na primjer posjet liječniku, broj je umanjjenica manji. U svakom slučaju, umanjjenice čine do 1/4 ukupnog broja Antonijinih imenica. Prosječna duljina iskaza (PDI) u riječima iznosila je od 1,7 na početku snimanja do 3,2 na kraju, što znači da je Antonija snimana upravo u vrijeme kad se morfologija i sintaksa počela pojavljivati, tj. kad je Antonija izašla iz razdoblja jedne riječi. Pokaže li se kako je Antonija upotrebljavala umanjjenice proizvodno – tj. primjenjujući pravila tvorbe, tvorba umanjjenica bit će jedno od prvih morfoloških pravila koje to dijete usvaja.

U teoriji prirodne morfologije proizvodnost se u dječjem jeziku procjenjuje postojanjem opreka – u ovom slučaju opreka između imenice i odgovarajuće umanjjenice. Upotrebljava li dijete oboje, znak je da upotrebljava pravila, a ne samo imitira roditelje. Mnoge Antonijine umanjjenice jedine su riječi kojima Antonija označava neku stvar, druge se imenice javljaju i kao umanjjenice. Antonija svakako razumije vezu između imenice i njezine umanjjenice – ona zna da je *Helenica* ista osoba kao i *Helena*, da je *olovčica* isti predmet kao i *olovka*, a *narančica* isto voće kao i *naranča*. S druge strane, ona umanjjenice ne upotrebljava da označi malu stvar – *narančica* za Antoniju nije *mala naranča*. Tablica 1. prikazuje odnos imenica i odgovarajućih umanjjenica.

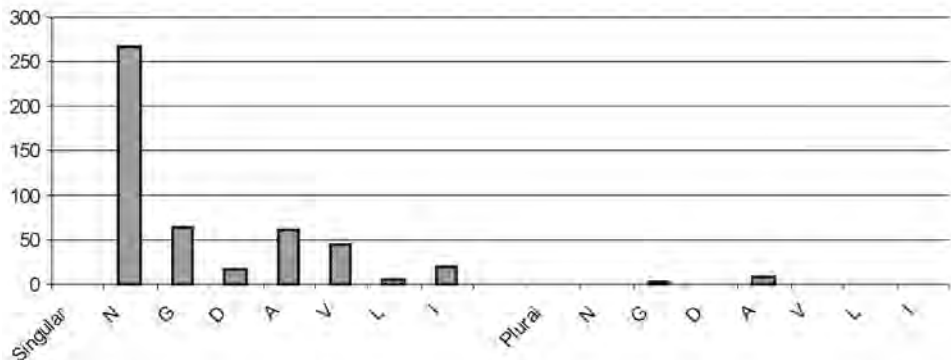
DOB	BROJ UMANJENICA U OPRECI	UKUPAN BROJ UMANJENICA	UKUPAN BROJ IMENICA
1;3	0	0	7
1;4	1	1	9
1;5	0	1	7
1;6	0	0	5
1;7	1	7	47
1;9	0	8	40
1;10	2	16	69
1;11	1	9	79
2;0	1	1	43
2;1	0	16	54
2;2	3	17	69
2;3	2	14	72
2;4	3	13	81
2;5	2	13	60
2;6	2	12	68
2;7	5	21	85
2;8	2	8	44

Tablica 1. Umanjjenice u opreci prema imenicama

Iz tablice 1. vidi se kako postoje opreke između imenica i odgovarajućih umanjjenica, ali je njihov broj malen. Iz tih se podataka ne da zaključiti o tome upotrebljava li Antonija pravila tvorbe ili samo imitira roditelje. U tom razdoblju Antonija ne deklinira većinu imenica, tj. ne upotrebljava mnoge imenice u dva ili više padeža (terminologijom teorije prirodne morfologije – rijetke se imenice pojavljuju u mini paradigmama). Raspodjela Antonijinih umanjjenica po padežima vidi se iz slike 7., dok je na slici 8. prikazana raspodjela svih Antonijinih imenica po padežima.



Slika 7. Raspodjela umanjjenica po padežima

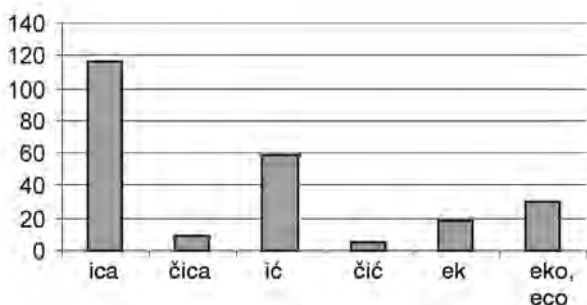


Slika 8. Raspodjela Antonijinih imenica po padežima

Iz te se raspodjele vidi kako u cijelom razdoblju u kojem je Antonija snimana nema mnogo imenica u drugim padežima osim nominativa i akuzativa jednine (a i tu se često radi o istom obliku) ili čak samo nominativa jednine. Genitiv je ograničen na svega nekoliko riječi (*sok*, *mljekeco*, *juhica* i sl.), u množini se javljaju samo one imenice koje označavaju predmete koji dolaze u paru (*cipele*, *rukavice* i sl.). S množinom Antonija ima mnogo problema koji se uočavaju u pogreškama koje Antonija radi naročito kad su u pitanju *pluralia tantum*: »gaću« umjesto *gaće*, »u Fesetu« umjesto *u Sesvetama*. Ako je prema kriteriju postojanja opreka ili mini paradigmi teško govoriti o proizvodnoj upotrebi umanjjenica, isto je tako teško govoriti o proizvodnoj upotrebi padeža. Uspo-

redbom dviju slika, slika 7. i 8. može se uočiti da je Antonija čak spremnija deklinirati umanjenicu od neumanjene imenice.

Od svih sufiksa, kojima se tvore umanjenice u hrvatskome, samo se najproizvodniji mogu naći u Antonijinu govoru (Babić, 1986.). Slika 9. prikazuje raspodjelu sufiksa u Antonijinu govoru.



Slika 9. Raspodjela sufiksa

Iz te se raspodjele može rekonstruirati Antonijin sustav: opći sufiks za ženski rod je *-ica*, njegova je fonološka varijanta *-čica*, za muški je rod opći sufiks *-ić*, a njegova je fonološka varijanta *-čić*. U nekim se riječima javlja i dijalektalni sufiks *-ek* (*pesek*, *zeček*). Za srednji rod takvog općeg sufiksa nema (*-ce*), ali se mogu naći dijalektalni sufiksi *-eko* i *-eco*. Zanimljivo je da su svi kajkavski sufiksi ograničeni na svega nekoliko riječi; kao što je *-ek* ograničen na dvije životinje, tako su i *-eko* i *-eco* ograničeni na *sunčeko*, *jajčeko* i *mljekeco*.

Često se pogreške i gramatička poopćavanja uzimaju kao mjera djetetove proizvodne upotrebe nekog gramatičkog pravila. Antonija nije mnogo griješila kad su u pitanju umanjenice: svega je nekoliko puta pogriješila stavljajući sufiks *-ić* kad bi se očekivao *-ek* kao u sljedećem primjeru:

```
%com: Antonija mazi i ljubi sliku.
*ANT: <ja sam> [/] ja sam pusu dala pesiću.
%com: Antonija ljubi psića.
*DRA: jesi? dobro maco.
%com: Antonija i dalje mazi i ljubi psića na slici
*DRA: u kak je lijepi pesek mali, hm?
*ANT: ja ga ninam.
```

Valja, međutim, imati na umu i to da su ovom metodom snimanja mnoge pogreške manje učestalosti propuštene. S obzirom na to da se od Antonije ne može očekivati poznavanje razlika između standarda i dijalekta, primjer *pesića* bio bi kakav-takav dokaz Antonijine proizvodne upotrebe tvorbenog pravila.

Dječje umanjenice razlikuju se u značenju od umanjenica u govoru odraslih. Prvo, one ni u kojem slučaju ne mogu imati pejorativno značenje kakvo mogu imati u hrvatskom, ali i u drugim jezicima (npr. u ruskom – Voekova, 1998.). Drugo, one općenito ne označuju malu stvar. Kad dijete želi nešto reći o maloj ili velikoj stvari, upotrijebit će pridjev *mali* ili *velik* bez obzira slijedi li umanjenica ili ne. Umanjenica je često jedina riječ kojom Antonija označava neki

predmet. Dječje su umanjenice posljedica njihove česte roditeljske upotrebe. Umanjenice stvaraju ozračje usmjereno prema djetetu (»child-centered atmosphere«, Jurafsky, 1996.), usmjereno prema malim, dječjim stvarima i problemima. S druge strane, umanjenice su fonološki teške, teže od neumanjenih imenica. Kao što je već spomenuto, u teoriji prirodne morfologije nastoji se riješiti ta nelogičnost time što se umanjenicama daje neka uloga u jezičnom usvajanju. Ta se uloga razlikuje od jezika do jezika. U načelu, već spomenuta objašnjenja (redukcija broja imenskih klasa ili paradigma bez glasovnih promjena) vrijede i za hrvatski. Jedina Antonijina riječ imenske klase –i, *stvar* pojavljuje se samo kao umanjenica, *stvarčica*, pa se reducira na zastupljeniju klasu –e. Isto tako, umjesto *ruka*, *ruke*, *ruci* Antonija će upotrebljavati *rukica*, *rukice*, *rukici* što, jezikom prirodne morfologije, čini paradigmu prozirkom (nema morfonoloških promjena kao u *ruka* – *ruci*). Olakšava li to usvajanje padeža ili ne, teško je reći budući da ne možemo provjeriti predviđanje da će dijete, koje manje upotrebljava umanjenice, biti slabije u deklinaciji. Iz same činjenice, da je pravilo tvorbe umanjenice prvo pravilo tvorbe koje dijete usvaja, i da ga vjerojatno usvaja prije pravila fleksije, ne može se odmah zaključiti da prvo na neki način uvjetuje ili olakšava drugo. Može se prihvatiti općenitiji i stoga slabiji zaključak: fonološka »težina« umanjenica nadoknađuje se većom postojanošću kraja riječi, tj. onog dijela koji nosi gramatičke obavijesti što djetetu olakšava otkrivanje gramatičkih elemenata u riječi.

Zaključak

Programi CHAT i CLAN nisu najprikladniji za obradu usvajanja hrvatskoga, iako su zamišljeni da budu prilagodljivi mnogim jezicima (pravila i kodovi pohranjuju se u posebnim datotekama ili se odrede u zaglavljinama). No, za sada ovo je jedini alat kojim se sustavno proučava usvajanje hrvatskoga. Bez obzira na nedostatke, sustav olakšava međujezične usporedbe budući da su podatci iz drugih jezika obrađeni na sličan način. Cijela se baza CHILDES stalno nadopunjava, a autori su dostupni putem elektronske pošte.

Ovaj je skup programa – zajedno s cijelom bankom podataka CHILDES – upotrebljiv za kvantitativnu obradu bilo kojeg svojstva dječjeg jezika, uz uvjet da je to svojstvo na neki način zapisano u korpusu. Na primjeru umanjenica prikazana je kvantitativna obrada kojom se empirijski provjerava pretpostavka o usvajanju pravila tvorbe umanjenica. Takva provjera – čega zaista ima u dječjem jeziku – omogućava potvrdu ili opovrgavanje pretpostavki o jezičnome razvoju.

Stvaranje korpusa hrvatskoga dječjeg jezika i njegov ulazak u bazu podataka CHILDES višestrukog je značenja jer će omogućiti:

1. određivanje normi dječjeg jezičnog razvoja kvalitativno i kvantitativno,
2. lakši pristup objašnjenju usvajanja drugog jezika,
3. razumijevanje jezičnih teškoća te oblikovanje terapijskih postupaka i radnog materijala,

4. jačanje položaja hrvatskog jezika na svjetskoj pozornici lingvistike, poglavito psiholingvistike.

A smjernica za budućnost bila bi izrada nacionalnoga računalnog programa koji bi kao takav ipak bio najprikladniji za hrvatski jezik.

Literatura

- Andel, M., Blaži, D. (2000.) CHAT i CLAN – alati za računalnu obradu jezičnih transkripata. Primijenjena lingvistika u Hrvatskoj – izazovi na početku XXI stoljeća. Hrvatsko društvo za primijenjenu lingvistiku. Opatija. (neobjavljeni rad)
- Babić, S. (1986.) Tvorba riječi u hrvatskom književnom jeziku. Zagreb: Globus.
- Bloom, L. (1993.) Transcription and Coding for Child Language Research: The Parts are More than the Whole. U: Edwards, J. A., Lampert, M. D. (ur.) Talking Data Transcription and Coding in Discourse Research. New Jersey: Lawrence Erlbaum Associates. 149–169.
- Dressler, W., Karpf, A. (1995.) The theoretical relevance of pre- and protomorphology in language acquisition. Yearbook of Morphology. 99–122.
- Edwards, J. A. (1993.) Principles and Constrasting Systems of Discourse Transcription. U: Edwards, J. A., Lampert, M. D. (ur.) Talking Data Transcription and Coding in Discourse Research. New Jersey: Lawrence Erlbaum Associates. 3–33.
- Jurafsky, D. (1996.) Universal tendencies in the semantics of the diminutive. Language 72, 533–578.
- Laalo, K. (1998.) Diminutives in Finnish child-directed and child speech. U: Gillis, S. (ur.): Studies in the Acquisition of Number and Diminutive Marking. Antwerp: Universiteit Antwerpen, Antwerp Papers in Linguistics 95. 137–148.
- MacWhinney, B. (1997.) The CHILDES Project Tools for Analysing Talk. New Jersey: Lawrence Erlbaum Associates.
- MacWhinney, B. (1999.) The Emergence of Language. London: Lawrence Erlbaum Associates.
- Savickienė, I. (1998.) The acquisition of diminutives in Lithuanian. U: Gillis, S. (ur.): Studies in the Acquisition of Number and Diminutive Marking. Antwerp: Universiteit Antwerpen, Antwerp Papers in Linguistics 95. 115–136.
- Voekova, M. (1998.) Acquisition of Diminutives by a Russian Child: Preliminary Observations in Connection with the Early Adjectives. U: Gillis, S. (ur.) Studies in the Acquisition of Number and Diminutive Marking, Antwerp: Universiteit Antwerpen, Antwerp Papers in Linguistics 95. 97–113.

Computer Analysis of Child Language on the Example of the Acquisition of Diminutives

The article is a written form of a lecture given at Linguistic Circle in Zagreb with a purpose to give an overview of the CHILDES system and its implementation on the Croatian corpus of child language. The second part of the article is an example of the analyses that are facilitated with the programs included in CHILDES. Diminutives are a prominent feature of the child language in Croatian and their acquisition is presented. A brief discussion of the role of diminutives in the acquisition of Croatian morphology is given at the end of the article.

Ključne riječi: CHILDES, računalna obrada jezika, dječji jezik, umanjeње, usvajanje jezika, morfologija, hrvatski jezik

Key words: CHILDES, language computer processing, child language, diminutives, language acquisition, morphology, Croatian language