

OBRADA DJEČJIH JEZIČNIH PODATAKA NA PRIMJERU HRVATSKO-TALIJANSKE DVOJEZIČNOSTI

MIA BATINIĆ ANGSTER

STOŠA GRGUREV

Odjel za lingvistiku, Sveučilište u Zadru

mbatinic@unizd.hr

srgurev@unizd.com

UDK: 81*246.2=163.42=131.1:37-053.4

Stručni rad

Primljen: 17.6.2024.

Prihvaćen: 4.10.2024.

SAŽETAK

U radu se daje uvid u jedan aspekt obrade dječjih jezičnih podataka na primjeru izrade prijepisa zvučnoga zapisa interakcije dvojezične djevojčice s majkom. Mnogo toga što danas znamo o ranome jezičnom razvoju poznato je zahvaljujući korpusima dječjega jezika nastalima snimanjem spontane interakcije djece i njihovih roditelja te izradom prijepisa tako prikupljenih jezičnih podataka (usp. Kuvač i Palmović, 2007; Behrens, 2008; Ordulj i Hržica, 2015). Jedan od načina prepisivanja i obrade jezičnih podataka razvijen je za potrebe Svjetske banke dječjega jezika CHILDES (engl. *Child Language Data Exchange System*; MacWhinney, 2000). Cilj je ovoga rada prikazati primjenu toga modela i njegovih norma u izradi prijepisa zvučnih zapisa nastalih na projektu izrade dvojezičnoga hrvatsko-talijanskog dječjeg korpusa MaLi (Batinić Angster i Angster, 2024). Na primjeru spontane govorne proizvodnje dvojezične djevojčice u interakciji s majkom koja se odvija većim dijelom na talijanskome jeziku prikazat će se postupci izrade prijepisa zvučnoga zapisa u trajanju od sat vremena u programu CHAT (engl. *Codes for the Human Analysis of Transcripts*; MacWhinney, 2000). Pritom će se poseban naglasak staviti na označavanje razvojnih pogrešaka vezanih za fonetsko-fonološki razvoj i preopćavanje te posebnosti označavanja dvojezične proizvodnje poput prebacivanja kodova.

KLJUČNE RIJEČI:

dječji jezik, usvajanje prvoga jezika, dvojezičnost, rani jezični razvoj, obrada jezika, korpus

UVOD¹

U ovome radu daje se uvid u jedan aspekt obrade dječjih jezičnih podataka na primjeru izrade i obrade transkripcije ili prijepisa zvučnoga zapisa interakcije dvojezičnoga djeteta s majkom. Budući da je krajnji cilj prikupljanja i obrade dječjih jezičnih podataka analiza različitih aspekata dječjega jezika i ranoga jezičnog razvoja, većina istraživača usredotočena je na interpretaciju podataka i potencijalne znanstvene spoznaje, dok je manje stručnih radova poput rada A. Ordulj i G. Hržica (2015) koji pružaju uvid u postupak obrade podataka u postojećim računalnim programima (npr. Kuvač i Palmović, 2007). Štoviše, uvidom u dostupnu literaturu uočava se nedostatak radova koji bi opisali postupak obrade dvojezične interakcije. Upravo zbog nedovoljne posvećenosti toj temi naglasak ovoga rada stavlja se na postupke izrade prijepisa spontane interakcije u kojoj sudjeluje dvojezično dijete, čime se omogućuje uvid u obradu dječjih jezičnih podataka s naglaskom na posebnosti označavanja dvojezične dječje proizvodnje.

Jezični podatci mogu se prikupiti izravno od ispitanika, npr. zvučnim ili videozapisima interakcije, metodom elicitacije, eksperimentalnim putem itd. (usp. Crystal, 2010: 434–435). Metode prikupljanja i obrade dječjih jezičnih podataka mijenjali su se kroz povijest. Dnevници, koje najčešće vode roditelji, pripadaju metodi bilježenja dječje jezične proizvodnje koja se rabi i danas, međutim napretkom računalnih tehnologija omogućilo se lakše i brže prikupljanje podataka o dječjem govoru (usp. Kuvač i Palmović, 2007: 16–20). Zvučni i videozapisi interakcija između djeteta čiji se jezični razvoj prati i njegovih sugovornika – najčešće roditelja ili istraživača – omogućili su prijenos prikupljenih podataka na računalo i time potaknuli bržu analizu jezičnih podataka i prikupljanje dodatnih informacija (usp. Ordulj i Hržica, 2015). Brojne spoznaje o ranom jezičnom razvoju duguju se korpusima dječjega jezika nastalima snimanjem spontane interakcije djece i njihovih roditelja te izradom prijepisa tako prikupljenih jezičnih podataka (usp. Kuvač i Palmović, 2007; Behrens, 2008; Ordulj i Hržica, 2015). Osim prikupljanja podataka, za istraživanje dječjega jezika važna je i obrada dobivenih podataka. Štoviše, prikupljanje podataka, kako ističe MacWhinney (2000: 5), tek je „početak mnogo većega zadatka“ s obzirom na to da je, između ostalog, „proces izrade prijepisa i analize prirodnih uzoraka iznimno dugotrajan“.

¹ Ovaj je rad izrađen u okviru institucijskoga znanstveno-istraživačkog projekta *Višejezičnost: između teorije i empirije* (ViTE) (IP-01-2023-14), koji financira Sveučilište u Zadru. Kao polazište za rad poslužio je završni rad S. Grgurev (2023), koji je izrađen u okviru prijediplomskoga studija *Language and communication in a multilingual society/Jezik i komunikacija u višejezičnome društvu* pod mentorstvom M. Batinić Angster. Zahvaljujemo recenzentima na korisnim i poticajnim komentarima i sugestijama.

U današnje vrijeme dostupni su brojni programi za transkribiranje ili prepisivanje zapisa dijelom zahvaljujući većoj uključenosti istraživača u područje računalne obrade podataka. Oni znatno smanjuju iscrpljujući zadatak prepisivanja i nude nove funkcije koje potencijalno mogu olakšati cijeli proces (Kreuz i Riordan, 2018: 7). Jedan od načina izrade prijepisa i obrade podataka razvijen je za potrebe Svjetske banke dječjega jezika CHILDES (engl. *Child Language Data Exchange System*; MacWhinney, 2000). Cilj je ovoga rada prikazati primjenu tog modela i njegove norme u izradi prijepisa zvučnih zapisa nastalih na projektu izrade dvojezičnoga hrvatsko-talijanskog dječjeg korpusa MaLi (Batinić Angster i Angster, 2023, 2024). Na primjeru spontane govorne proizvodnje dvojezične djevojčice u slobodnoj interakciji s majkom na talijanskome kao ciljnome jeziku ponudit će se pregled postupaka izrade prijepisa jednoga zvučnog zapisa uporabom programa CHAT (engl. *Codes for the Human Analysis of Transcripts*; MacWhinney, 2000) te će se izdvojiti i opisati oznake rabljene tijekom prepisivanja.

DJEČJI JEZIČNI PODATCI: PRIKUPLJANJE I OBRADA

Izazovi prikupljanja podataka za istraživanje ranoga jezičnog razvoja

Istraživanje ranoga jezičnog razvoja izazovno je područje dijelom zbog činjenice da djeca pripadaju osjetljivoj skupini ispitanika, što podrazumijeva rješavanje etičkih i brojnih drugih organizacijskih pitanja. K tomu je otežano pridobiti ispitanike, razloge čega treba tražiti i u samoj naravi praćenja ranoga jezičnog razvoja. Riječ je naime o longitudinalnome, najčešće višegodišnjem prikupljanju dječje jezične proizvodnje, i to u pravilnim razmacima, zbog čega je velik rizik od pada interesa, kako ispitanika tako i istraživača, te posljedično odustajanja od sudjelovanja u istraživanju, odnosno od njegova provođenja (usp. Lieven i Behrens, 2012; Batinić Angster i Angster, 2024). Razumljivo je stoga što su istraživači najčešće roditelji iz struke, npr. psiholozi, lingvisti, logopedi (Kuvač i Palmović, 2007), koji su svjesni važnosti praćenja ranoga jezičnog razvoja, odnosno spoznaja koje se mogu steći na temelju tako prikupljenih podataka (usp. Batinić Angster i Angster, 2024).

U Hrvatskoj je zabilježeno praćenje jezičnoga razvoja troje djece do treće godine života za potrebe izrade Hrvatskoga korpusa dječjega jezika (Kovačević, 2002), čiji su prijepisi i snimke javno dostupni istraživačima u banci CHILDES i na temelju kojega je nastao veći broj radova o usvajanju hrvatskoga jezika, i to ponajprije na morfološkoj i leksičkoj razini. Istraživači koji promatraju rani jezič-

ni razvoj primjenjuju, između ostaloga, ljestvice i testove među kojima su i hrvatske Komunikacijske razvojne ljestvice – KORALJE (Kovačević i sur., 2007). Na njih se oslanjaju i logopedi pri procjeni jezičnoga statusa djece.

Za dvojezičnu djecu međutim nema razvijenih ljestvica prema kojima bi bilo moguće utvrditi odvija li se razvoj očekivano i skladu s dobi ili pokazuje znakove neuredna razvoja (Hržica i sur., 2015: 36). U Hrvatskoj je zabilježeno jedno praćenje ranoga dvojezičnog razvoja djeteta izloženoga hrvatskomu i engleskomu jeziku (dnevnički zapisi M. Medved Krajnović, 2004), dok se Čamber (2020) bavila dvojezičnim hrvatsko-njemačkim razvojem djece koja odrastaju u Austriji. Premda pojedina obilježja dvojezičnoga razvoja kod djece koja uz hrvatski usvajaju i neki drugi jezik jesu zabilježena, do danas nema sustavno provedenoga longitudinalnog istraživanja niti javno dostupnoga korpusa na kojemu bi bilo moguće provesti sustavniju analizu jezične proizvodnje takve djece.

MaLi – hrvatsko-talijanski dvojezični dječji korpus

MaLi (Batinić Angster i Angster, 2023, 2024) ime je prvoga hrvatsko-talijanskog dvojezičnog dječjeg korpusa u nastajanju.² Prikupljanje jezičnih podataka za korpus započelo je 2020. godine, a sudionici su brat i sestra koji su od rođenja izloženi hrvatskomu i talijanskomu jeziku.³ Djeca odrastaju u Zadru u dvojezičnoj obitelji. Otac je izvorni govornik talijanskoga jezika i služi se hrvatskim na razini B2, a majci je materinski jezik hrvatski te se talijanskim jezikom služi na razini C1/C2. Djeca s obama roditeljima govore talijanskim jezikom, no u razgovoru se s majkom povremeno služe hrvatskim. Izvan užega obiteljskog okruženja (s dadijnom, u vrtiću, u školi, na igralištu s drugom djecom) služe se hrvatskim jezikom. Obitelj godišnje provede prosječno šest tjedana u Italiji, i to četiri odnosno dva tjedna neprekidno, kada su djeca izložena gotovo isključivo talijanskomu jeziku.

U izradi korpusa provodi se kombinacija metoda opisana u cijelosti u radu Batinić Angster i Angster (2024). Od svibnja 2020. godine spontana interakcije djece i troje odraslih sugovornika bilježi se snimanjem zvučnih zapisa profesionalnim diktafonom Zoom H4n u formatu .wav. Svi su odrasli sugovornici djece s kojima je snimana interakcija visokoobrazovani pojedinci, i to u humanističkom znanstvenom području.

² Kako se pojašnjava u Batinić Angster i Angster (2024: 76) ime korpusa s jedne je strane poimeničeni hrvatski pridjev *mali* ili *maleni*, koji upućuje na djecu koja su sudionici istraživanja, a s druge sadrži talijansku riječ *madrelingua* 'izvorni govornik' te aludira na sintagmu *lingua madre* 'materinski jezik'.

³ Metodologija prikupljanja građe za korpus prikazana je u cijelosti u radu Batinić Angster i Angster (2024).

Cilj je izrade prijepisa zvučnih zapisa stvaranje korpusa koji će biti dostupan istraživačima u javnom pristupu u banci CHILDES, stoga se prijepisi zvučnih zapisa izrađuju uz pomoć programa CHAT razvijenoga za potrebe te svjetske banke dječjega jezika.

CHILDES I CHAT

Sustav za razmjenu podataka dječjega jezika CHILDES započeli su razvijati Dan Slobin, Willem Levelt, Susan Ervin-Tripp i Brian MacWhinney 1981. godine (MacWhinney, 2000: 8). Dobitkom dvogodišnje financijske potpore 1984. godine ta se ideja ostvaruje te se u datoteku u kratkom razdoblju postavljaju već postojeći računalni korpusi (Kuvač i Palmović, 2007: 80). Unošenje prikupljenih podataka dječjega jezika bio je jedan od ciljeva unutar projekta CHILDES, dok je drugi cilj, kako dalje navode Kuvač i Palmović (2007: 81), bio unesenim dječjim korpusima osigurati istraživanja koja bi služila u oblikovanju teorija. Programski paket razvijen za potrebe CHILDES-a sastoji od dva glavna programa za obradu podataka, CHAT, koji sadrži format za prepisivanje i označavanje, te CLAN (engl. *Computerized Language Analysis*), programa za analizu na fonološkoj, morfološkoj i sintaktičkoj razini. Zahvaljujući tim alatima za obradu dječjega jezika omogućeno je proučavanje razvojnih obilježja pojedinih jezika, ali i različitih jezičnih pojava karakterističnih za (rani) jezični razvoj, koje bi bilo teško otkriti drugačijim metodama (usp. Kuvač i Palmović, 2007: 81).

Sustav CHAT nudi način izrade prijepisa interakcije. Ima dosljedan format te se može rabiti za različite vrste razgovora, poput onih između djece i roditelja ili učitelja i učenika stranoga jezika (MacWhinney, 2000: 17). Premda se CHAT prvotno razvio za potrebe obrade dječjega jezika, on se danas primjenjuje u prepisivanju zapisa raznovrsnih i raznorodnih vidova govorne interakcije. Oznake koje su mu svojstvene neprestano se osvježavaju prilagođavajući se potrebama novih korpusa. CHILDES je danas, naime, samo jedna od mnogih banaka na repozitoriju pod nazivom Talk-Bank (<https://talkbank.org>), koji okuplja građu namijenjenu istraživanju jezika u autizmu, afaziji, demenciji, u dvojezičnome razvoju, kao i korpusne ovladavanja drugim jezikom ili uporabe jezika u učionici itd. (v. MacWhinney, 2000: 5–6). Tu najveću svjetsku riznicu jezičnih podataka govorenoga jezika povezuje uporaba jedinstvenoga prepisivačkoga formata, odnosno načela na kojima počiva CHAT i s njim povezan sustav programa CLAN, koji je pogodan za daljnju računalnu obradu i analizu jezičnih

podataka. Isto tako, prijepisi izrađeni u CHAT-u mogu se prenijeti u drugi program u kojemu se mogu nastaviti obrađivati, npr. u Praat (praat.org), Phon (phonbank.talkbank.org), ELAN (tla.mpi.nl/tools/elan) itd. (MacWhinney, 2000: 6).

OBRADA DJEČJIH JEZIČNIH PODATAKA NA PRIMJERU INTERAKCIJE DVOJEZIČNE DJEVOJČICE S MAJKOM

Za potrebe stvaranja korpusa MaLi (Batinić Angster i Angster, 2023, 2024) zvučni zapisi spontanij razgovora prepisuju se ručno u računalnome programu CHAT. Prijepisi tako podliježu pravilima i oznakama izrađenima za potrebe CHILDES-a kako bi korpus MaLi jednoga dana postao dvojezični dječji korpus u javnome pristupu s hrvatskim kao jednim od usvajanih jezika. U nastavku će se opisati postupak izrade prijepisa jednoga zvučnog zapisa.

Zvučni zapis

Zvučni zapis *MaLi20220328aita* u trajanju od 58 minuta i 59 sekunda bilježi spontanu interakciju između djevojčice u dobi od tri godine i pet mjeseci (3;5) i njezine majke na talijanskome jeziku. Premda se u razgovoru povremeno pojavljuje hrvatski jezik te vrlo rijetko engleski jezik, talijanski je ciljani jezik zabilježene interakcije, tako da je i najveći dio prijepisa izrađen na talijanskome jeziku.

Naziv zvučnoga zapisa – *MaLi20220328aita* – oblikovan je po shemi koju su ustanovili autori korpusa (v. Batinić Angster i Angster, 2024). *MaLi* označava ime korpusa, zatim slijede godina, mjesec i dan snimanja, slovima *a–z* označavaju se redni brojevi snimke u danu – pri čemu se slovom *a* označava prva snimljena interakcija toga dana – te se na kraju navodi troslovna kratica ciljnoga jezika interakcije, koji je u ovome slučaju talijanski (ovdje kratica *ita*) (Batinić Angster i Angster, 2024). Prema navedenome, zvučni zapis koji je u ovome radu uzet kao primjer nastao je 28. ožujka 2022. godine.

Prijepis: zaglavlje i metajezik

Prijepis čine dva glavna dijela: zaglavlje i prijepis. Jezik zapisa, datum snimanja, dob i spol sudionika i druge informacije poput konteksta interakcije zabilježene po jedinom snimkom nalaze se u zaglavlju datoteke koju slijedi prijepis (v. Sliku 1).

```

@Begin
@Languages:    ita, hrv, eng
@Participants: CH2 Target_Child, MOT Mother
@Options:      CA
@ID:   ita|MaLi|CH2|3;05.19|female|||Target_Child|||
@ID:   ita|MaLi|MOT|37;06.25|female|||Mother|||
@Media: MaLi20220328aita_STE-009, audio
@Date:  28-MAR-2022
@Transcriber:  Mia Batinić Angster, Stoša Grgurev
@Situation:    Mot e Ch2 chiacchierano e giocano a casa.

```

SLIKA 1. Zaglavlje prijepisa zvučnoga zapisa MaLi20220328aita

Sadržaj zaglavlja donosi informacije o varijablama koje istraživačima mogu biti važne za analizu. Zaglavlje se tako sastoji od redaka koji pružaju informacije o dobi, spolu i ulozi sudionika, jezicima koji se pojavljuju u zapisu, datumu nastanka zapisa, nazivu datoteke te situaciji u kojoj je nastala interakcija.

Svaki redak zaglavlja započinje znakom @. U zaglavlju se može nalaziti više redaka, ovisno o potrebi, a među njima su nužni sljedeći: @Begin, @Languages, @Participants, @ID i @End. Svaki prijepis mora početi oznakom za početak prijepisa @Begin ‘početak’, a završiti oznakom za njegov završetak – @End ‘kraj’. Slijedi redak @Languages ‘jezici’ u koji se upisuju jezici na koje se nailazi tijekom prijepisa. Prvi navedeni jezik pritom označava temeljni i dominantan jezik interakcije.

U redak @Participants ‘sudionici’ zapisuju se kratice sudionika interakcije, koje moraju odgovarati podatcima koji se odnose na svakoga pojedinog sudionika, a koji su navedeni u retku @ID. Za zadovoljavanje norma sustava, prema kojemu oznake moraju biti troslovne (MacWhinney, 2000: 22), odabrana troslovna oznaka za dijete jest CH2, a za majku MOT. Oznake su odabrane prema engleskim riječima *child* ‘dijete’ i *mother* ‘majka’ i vrijede za sudionike ove interakcije kao i drugih interakcija u kojima sudjeluju u okviru korpusa MaLi.⁴

U retku @Media navodi se ime zvučnoga zapisa koji se prepisuje. Ime prepisivača navodi se u retku @Transcriber, dok se u retku @Situation opisuje kontekst situacije u kojoj se odvija interakcija, npr. ‘Mot i Ch2 razgovaraju i igraju se u kući’ (prema tal. *Mot e Ch2 chiacchierano e giocano a casa*).

Metajezik zaglavlja dijelom je engleski, a dijelom odgovara dominantnomu jeziku prijepisa. Imena redaka, spol (npr. engl. *female* ‘ženski’), uloga sudionika (*mother* ‘majka’) i kratice jezika propisani su priručnikom o izradi prijepisa u programu CHAT, stoga su na engleskome jeziku. S druge strane, jezik na kojemu se opisuje

⁴ Za kratice kojima se označavaju drugi sudionici v. Batinić Angster i Angster (2024).

situacija odgovara jeziku koji je kao prvi naveden u popisu jezika prijepisa, u ovome slučaju talijanskomu.

Dominantan jezik dvojezične interakcije važan je i u drugim kontekstima tijekom izrade prijepisa. Transkripcija glavnih redaka također mora biti prilagođena ciljnomu jeziku, odnosno njegovu grafemskom sustavu. Ako dijete ne izgovara riječ kako bi je izgovorio odrasli govornik, ona se zapisuje uz pomoć grafema ciljnoga jezika interakcije koji najvjernije prikazuju izgovorene glasove. Tako na primjer ne rabimo dijakritičke znakove prilikom izrade prijepisa interakcije koja se odvija dominantno na talijanske jeziku. Ako dijete nije sposobno izgovoriti *brutto* 'ružan', nego riječ izgovara na način koji bi se fonetski zapisao kao ['blut:o], nećemo zapisati riječ *bluto* jer bi se u talijanske ova riječ čitala drukčije (tj. ['blu:to] naglaskom nalik dugouzlaznome), nego *blutto*, prilikom čijeg se čitanja ostvaruje očekivani izgovor dvostruko pisanih suglasnika koji nalikuje kratkosilaznomu naglasku na samoglasniku *u* u hrvatskome jeziku. Iznimno, u slučaju ostvarenja prebacivanja kodova, prijepis je prilagođen grafemskom sustavu jezika kojemu pripada dio iskaza zahvaćen prebacivanjem kodova.

Tijelo prijepisa i oznake

Kako bi se zadovoljili etički standardi provođenja istraživanja s djecom, odnosno kako bi se zaštitio identitet djevojčice, ona se u prijepisima spominje pod pseudonimom Livia (Batinić Angster i Angster, 2023, 2024).⁵ Prepisuju se izgovorene rečenice ili iskazi. U jezikoslovlju se rečenica rabi kao naziv za apstraktnu jedinicu, dok je iskaz ili izričaj (prema engl. *utterance*) njezin konkretni ostvaraj. MacWhinney (2000) ističe da se uz iskaz može još govoriti o konverzacijskoj jedinici (ili *c-unit* prema engl. *conversational unit*) te je ona zapravo jedinica koja se prepisuje. U slučaju iskaza organiziranoga u rečeničnome nizu, svaka bi se jednostavna rečenica trebala prepisati u zasebnome retku, dok se surečenice zavisnosložene rečenice prepisuju u istome retku. Iskazi se bilježe u cijelosti malim slovima te se velikim početnim slovom pišu samo vlastita imena, a rečenica završava odgovarajućim rečeničnim znakom. Prijepis je slabo zasićen interpunkcijskim znakovima, koji se u načelu izbjegavaju. U nastavku ćemo prikazati primjenu oznaka propisanih priručnikom za uporabu CHAT-a (MacWhinney, 2000)⁶ u izradi korpusa MaLi.

⁵ Njezin brat, koji se spominje i u ovdje prikazanoj interakciji, također ima pseudonim – Paolo.

⁶ Treba istaknuti da je priručnik podložan promjenama, nalazi se u javnome pristupu na mrežnoj stranici CHILDES-a (<https://talkbank.org/manuals/CHAT.pdf>), na kojoj je u trenutku pripreme ovoga članka bila dostupna inačica osvjježena 20. svibnja 2024. godine.

Označavanje prijepisa: opće oznake

Među najučestalijim oznakama koje se rabe prilikom prepisivanja dječjega govora nalaze se *x* i *xxx*. Oznakom *x* prepisuje se nerazgovijetna riječ, dok se troslovna oznaka *xxx* rabi kada se ne razaznaje niz od više riječi ili iskaz u cijelosti. Uporaba obiju oznaka oprimjerena je u isječku prijepisa (1).⁷

- | | | |
|-----|---|--------------------------------|
| (1) | *CH2: <i>x</i> mamma è blutto [: brutto]. | ‘nerazumljivo mama, ružan je.’ |
| | *MOT: ma come mai? | ‘Kako to?’ |
| | *MOT: perché è vestito di nero? | ‘Jer je odjeven u crno?’ |
| | *CH2: <i>xxx</i> . | <i>nerazumljivo</i> |
| | *CH2: quello cattivo. | ‘Onaj zločesti.’ |
| | *MOT: viene dal polo nord? | ‘Dolazi sa Sjevernog pola?’ |

Posebna se pozornost prilikom izrade prijepisa posvećuje označavanju pogrešaka. Jedno od obilježja dječjega jezika jest nemogućnost izgovaranja pojedinih glasova i glasovnih kombinacija, što korelira s djetetovom dobi. Prema istraživanjima fonetsko-fonološkoga razvoja djece koja usvajaju talijanski jezik, predstavljenima u Zmarich i Bonifacio (2005) i Zmarich (2010: 13–14), fonetsko-fonološki sustav postupno raste u dobi između 18 mjeseci i 27 mjeseci života. Djeca tako tek u dobi od 24 mjeseca izgovaraju sve okluzive ili zapornike i frikative ili tjesnačnike koji se razlikuju po načinu i mjestu tvorbe,⁸ a u dobi od 27 mjeseci povećava se broj složenih suglasničkih skupova sastava CVC i CCV, dok do tada prevladava suglasničko-samoglasnička struktura sloga CV. Štoviše, istraživanjem korelacije slogovnoga i fonetskoga sastava prvih riječi i leksičke raznolikosti (Zmarich i sur., 2011), na podacima djece u dobi između 18 i 36 mjeseci utvrdilo se da sve do te dobi u leksičkoj proizvodnji prevladava pojednostavljena slogovna struktura tipa CV te uporaba dvosložnih i trosložnih riječi (usp. Zmarich, 2010: 15). Zanimljivo je pritom što se paralelno s povećanjem dobi i leksičke raznolikosti – unatoč porastu inventara složenijih suglasnika s obzirom na način i mjesto tvorbe – primjećuje i redukcija okluziva ili praskavih suglasnika, koji su inače među prvim suglasnicima koje djeca usvajaju.

⁷ Prijepis je u ovome radu popraćen prijevodom na hrvatski jezik s obzirom na to da je rad pisan na hrvatskome jeziku, no prijevod inače nije sastavnim dijelom prijepisa.

⁸ O hrvatskom nazivlju v. Jelaska (2004).

Pogreške ili koje dijete rabi u odnosu na ciljni oblik karakterističan za proizvodnju odrasloga govornika, a koje su očekivane za dob, smatraju se razvojno uvjetovanim nepravilnostima ili pogreškama. Takve pogreške označavaju se oznakom koja se sastoji od uglate zagrade, dvotočke, razmaka i točnoga, cilj-noga lika riječi, tj. [: lik]. Tako u (1) vidimo da se iza riječi u kojoj djevojčica glas /r/ u suglasničkome skupu u pristupu sloga na početku riječi izgovara kao /l/ nalazi [: brutto]. Budući da za takav kontekst očekujemo svojevrsnu teškoću prilikom izgovora, što je u skladu s njezinom dobi, tu pogrešku vrednujemo kao razvojnu.⁹

U isječcima (2), (3) i (4) donosimo dodatne primjere označavanja razvojnih pogrešaka. Osim već spomenute teškoće s izgovorom glasa /r/ u suglasničkim skupovima, u isječku (2) nalazi se i primjer prijepisa izgovora talijanske riječi *certo* ‘naravno’, pri kojemu se – uz izostavljanje glasa /r/ u suglasničkome skupu /rt/ – izgovor afrikate ili nepčanoga slivenika [tʃ] ostvaruje kao frikativ ili tjesnačnik [s]. Budući da su afrikate ili slivenici među najkasnije usvajanim glasovima koji su nerijetko zahvaćeni govornim, tj. artikulacijskim pogreškama, strategija njihove zamjene drugim glasovima jedna je od očekivanih obilježja karakterističnih za dob djevojčice.

- | | | |
|-----|--|-------------------------|
| (2) | *CH2: noi plendiamo [: prendiamo] chiave. | ‘Mi ćemo uzeti ključ.’ |
| | *CH2: non plendo [: prendo] chiave. | ‘Neću uzeti ključ.’ |
| | *MOT: non prendi la chiave? | ‘Nećeš uzeti ključ?’ |
| | *CH2: no! | ‘Ne!’ |
| | *MOT: la vai a cercare? | ‘Ideš li ga potražiti?’ |
| | *CH2: ma seto [: certo]. | ‘Pa naravno.’ |
| (3) | *CH2: stlega [: strega] è blutta [: brutta]. | ‘Vještica je zločesta.’ |
| | *CH2: collana. | ‘Ogrlica.’ |
| | *CH2: guada [: guarda] colona [: corona]. | ‘Gledaj, ogrlica.’ |
| (4) | *CH2: sì, è stlano [: strano]. | ‘Da, čudno je.’ |
| | *CH2: dome [: dorme]. | ‘Spava.’ |

⁹ To, međutim, ne isključuje mogućnost da će djevojčica u istoj dobi u drugom kontekstu ili čak i prilikom iste interakcije dobro izgovoriti /r/. Nije naime neobično da u djetetovu inventaru istodobno supostoje različiti izlazni oblici, kako na fonološkoj, tako i morfološkoj i sintaktičkoj razini.

U korpusu možemo klasificirati pogreške i s obzirom na to pripadaju li fonološkoj, morfološkoj i morfosintaktičkoj ili semantičkoj razini. Morfološke i morfosintaktičke pogreške koje se tiču, primjerice, vezanih morfema, afiksa, sročnosnih obrazaca ili nenaglasnica bilježe se oznakom [* m]. U isječcima prijepisa (5) i (6) dijete pogrešno tvori množinu imenice (*il) dito* ‘prst’. Umjesto *le mie dita* ‘moji prsti’, dijete rabi oblik *dite*. Zanimljivo je da je riječ o hibridnoj ili dvorodnoj imenici koja je u jednini muškoga roda, a u množini postaje ženskoga.¹⁰ Štoviše, u množini ta imenica traži slaganje u ženskome rodu od člana i pridjeva, no nepravilna je oblika – s dočettim *-a* umjesto očekivanoga dočetta *-e*. U ovome slučaju dijete zaista realizira množinu ženskoga roda tvoreći je analogno drugim imenicama ženskoga roda (npr. *palla* ‘lopta’ – *palle* ‘lopte’), pri čemu dolazi do preopćavanja (prema engl. *overgeneralization*).¹¹

- | | | |
|-----|--|----------------|
| (5) | *CH2: mamma, guada [: guarda]. | ‘Mama, vidi!’ |
| | *MOT: &-eh. | ‘Eh.’ |
| | *CH2: dite [* m]. | ‘Prsti.’ |
| | *MOT: cosa? | ‘Što?’ |
| | *CH2: <le mie dite> [/] le mie dite [* m]. | ‘Moji prsti.’ |
| | *MOT: le tue dita? | ‘Tvoji prsti?’ |
| | *CH2: sì. | ‘Da.’ |

U primjeru (6) djevojčica, umjesto očekivanoga oblika glagola *riuscire* ‘uspjeti’ za prvo lice jednine indikativa prezenta – *riesco* – proizvodi oblik *riescio*. Oblik se asimilira prema obliku za drugo i treće lice jednine *riesci* ‘uspijevamo’.¹² Djevojčica, dakle, i u ovom slučaju poseže za preopćavanjem primjenjujući usvojeno pravilo tvorbe na oblik koji se tvori nepravilno čineći na taj način morfološku pogrešku.

- | | | |
|-----|---|-------------------------|
| (6) | *CH2: +< no(n) riescio [* m] a volale [: volare]. | ‘Ne uspijevam letjeti.’ |
|-----|---|-------------------------|

¹⁰ Dardano i Trifone (1997: 184) ih na talijanskome nazivaju *nomi sovrabbondanti* ‘preobilne imenice’, što bi se moglo razmatrati u okviru pojave morfološkoga preobilja (prema engl. *overabundance*).

¹¹ O preopćavanju kao pojavi i analizi preopćavanja glagolske osnove u usvajanju hrvatskoga jezika pisala je Hržica (2012).

¹² Prvo lice jednine i treće lice množine paradigme glagola *riuscire* ‘uspjeti’ za indikativ prezenta glase *riesco* ‘uspijevam’, *riescono* ‘uspijavaju’, dok ostali oblici glase: *riesci* ‘uspijevaš’, *riesce* ‘uspijeva’, *riusciamo* ‘uspijevamo’, *riuscite* ‘uspijevaju’.

Kada dijete umjesto *un gatto* ‘mačka’ kaže *uno gatto*, kao u (7), pri čemu rabi inače fonološki uvjetovani alomorf neodređenoga člana kakav bismo očekivali u drugom kontekstu, pogrešku označavamo kao fonološku oznakom [* p].

- (7) *CH2: non c’è. ‘Nema je.’
 *MOT: non c’è adesso no. ‘Nema je sada, ne.’
 CH2: non c’è uno [p] gatto. ‘Nema mačke.’

U isječku (8) dijete iz nepoznatoga razloga rabi pogrešan samoglasnik u riječi *caffè* ‘kava’, što je također označeno kao fonološka pogreška.

- (8) *CH2: è cheffè [* p]. ‘Kava.’
 *MOT: no è tisana. ‘Nije nego čaj.’

Među oznakama pogrešaka u govoru valja izdvojiti označavanje ispuštanja jednoga glasa kao u isječku (9). U talijanskome se jeziku niječnica *no* ‘ne’ rabi u niječnom odgovoru, dok se *non* ‘ne’ rabi radi nijekanja glagola. Djevojčica Livia u kontekstu ispred glagola rabi naizmjenice *non* ‘ne’ i *no* ‘ne’, što se u korpusu označava stavljanjem ispuštenoga završnog suglasnika u zagrada: *no(n)*.

- (9) *CH2: no(n) liesco [: riesco]. ‘Ne uspijevam.’

Od paralingvističkih informacija, koje su procijenjene relevantnima za označavanje prijepisa korpusa MaLi, mogu se izdvojiti duljenje glasa (10–11)¹³ ili osobito glasno izgovaranje riječi ili dijela iskaza (12–15). Obje se pojave označavaju tako da se stavi dvotočka iza tako izgovorenoga glasa. Osim toga, označava se i proizvodnja različitih zvučnih radnji koje ne pripadaju govoru, npr. plakanje, kašljanje, vikanje, pjevušenje, zijevanje i sl. (16–19).¹⁴ Za prozodijske aspekte interakcije poput pauza ili tzv. ispunjenih pauza (prema engl. *filled pauses*) na raspolaganju su oznake (MacWhinney, 2000: 92) oprimjerene isječcima (20)–(22).

U primjeru (10) i (11) djevojčica dugo izgovara samoglasnik *e* u prvom slogu participia glagola *seduto* ‘sjedeći’ (dosl. ‘sjednut’), odnosno *o* u *no* ‘ne’.

¹³ Pritom se ne misli na duljinu samoglasnika, nego na iskaze u kojima se neobično dugo izgovara neki glas, npr. nešto što bismo u pismu inače označili ponavljanjem glasa poput *neeeee*.

¹⁴ V. popis u MacWhinney (2000: 66).

- (10) *MOT: ci sta forse. ‘Možda i može stati.’
 *CH2: se:tuto [: seduto]. ‘Sjedeći.’
 *MOT: seduto@e, ecco forse ‘Sjedeći, evo, možda i
 seduto ci sta +... stane sjedeći.’
- (11) *MOT: è forte come te? ‘Je li jak poput tebe?’
 *CH2: sì. ‘Da.’
 *MOT: sì. ‘Da.’
 *MOT: o tu sei più forte di lui? ‘Ili si ti jača od njega?’
 *CH2: no: ‘Neee.’

Isječci (12)–(15) opimjeruju označavanje naglašenih iskaza ili njihovih dijelova. U prepisanome zvučnom zapisu koji ovdje prikazujemo djevojčica katkad uzbuđeno usklikne ili glasno izgovori neki dio iskaza: *no* ‘ne’ u (12), *sì* ‘da’ u (13), *scappa* ‘bježi’ u (14), *è mia gonna* ‘to je moja suknja’ u (15). To se bilježi oznakom [!].¹⁵

- (12) *CH2: qui. ‘Ovdje.’
 *MOT: qui? ‘Ovdje?’
 *CH2: no [!]. ‘Ne.’
- (13)¹⁶ *MOT: è sua mamma? ‘Je li to njezina mama?’
 *CH2: sì [!]. ‘Da.’
 *MOT: &-ah la bimba ha trovato la mamma. ‘Ah, djevojčica je
 pronasla mamu.’
 *CH2: sì [!]. ‘Da.’
- (14) *CH2: colle [: corre]. ‘Trči.’
 *MOT: corre? ‘Trči?’
 *CH2: scappa [!]. ‘Bježi.’

¹⁵ Kada se oznaka odnosi na dio iskaza koji se sastoji od dviju ili više riječi, taj se dio koji prethodi oznaci stavlja u izlomljene zagrade: <dio iskaza>.

¹⁶ Oznakom &- u primjeru prepisana je tzv. ispunjena pauza (ovdje &-ah). V. još primjere (20)–(22).

- (15) *MOT: è una gonna tutta lunga. 'To je neka duga suknja.'
 *CH2: <è mia gonna> [!]. 'To je moja suknja.'

Proizvodnja zvučnih radnji poput kašljanja ili dubokog disanja označena je kao u (16) i (17) oznakom &=, koja uvodi glagol u trećem licu jednine prezenta kojim se opisuje zvučna radnja. Ako se prilikom izgovaranja nekog iskaza, makar i nerazumljivog, izvodi neka zvučna radnja, to se označava unutar uglatih zagrada po načelu [=! glagol] kao u (18). U isječku (19) oprimjerena je uporaba obiju oznaka.

- (16) *MOT: &=tossisce *kašlje*
- (17) *CH2: &=respira profondamente. *duboko diše*
- (18) *CH2: xxx [=! urla]. 'nerazumljivo' *viče*
- (19) *MOT: canticchia. *pjevuši*
 *MOT: che freddo [=! canticchia]. 'Kako li je hladno!' *pjevuši*
 *MOT: <che freddo> [/] che freddo 'Kako li je hladno, kako li je hladno!' *pjevuši*

Unutar iskaza ili konverzacijske jedinice označavamo kraće pauze u govoru oznakom (.) kao u (20), dok pauze ispunjene glasom označene kao &-mhm u (21) i &-ah u (22).

- (20) *CH2: sì (.) ha paula [: paura] qui. 'Da, boji se ovdje.'
- (21) *CH2: mamma s'è lotto [: rotto]. 'Mama, razbio se.'
 *MOT: &-mhm. 'Mhm.'
 *MOT: sì è rotto? 'Razbio se?'
 *CH2: sì. 'Da.'
- (22) *MOT: ma chi te l'ha detto quello? 'Tko ti je to rekao?'
 *CH2: [- hrv] tata. 'Tata.'
 *MOT: &-ah sì, &-mhm (.) va bene. 'Ah, da, mhm, dobro.'

Kada govornik ponovi sugovornikov iskaz, to se navodi oznakom za ponavljanje @e koja se vezuje za ponovljenu riječ (MacWhinney, 2000: 45), kao u ovom isječku prijepisa u kojemu djevojčica ponavlja ono što je netom prije rekla njezina majka:

- (23) *MOT: scappa [/] scappa. ‘Bježi, bježi.’
 *CH2: scappa@e. ‘Bježi.’

U korpusu MaLi rabe se još oznake za: brbljanje – @b (24–25), onomatopeje – @o (26–28), pa čak i novotvorenice – @n, koje međutim nisu potvrđene u zapisu čiji prijepis prikazujemo u ovome radu, kao i za riječi koje dijete rabi u značenju neke druge riječi a koje odrasli sugovornici prepoznaju i prihvaćaju s tim značenjem – @f (29).

- (24) *CH2: pa@b pan@b kan@b pan@b pank@b ‘Pa, pan, kan, pan, pank.’
- (25) *CH2: so@b sou@b sao@b [=! canticchia]. ‘So, sou, sao.’ *pjevusi*
- (26) *CH2: am_am@o è finita. ‘Am, am, gotova je.’
- (27) *CH2: miau@o. ‘Mijau.’
- (28) *MOT: psss@o. ‘Pst.’
- (29) *CH2: è mia ballelina@f [: ballerina]. ‘To je moja haljina.’

U primjeru (29), naime, oprimjeruje se uporaba riječi *ballerina* ‘balerina’ u značenju ‘suknja’ ili ‘haljina’, kojima pak u talijanskome jeziku odgovaraju riječi *gonna* ‘suknja’, odnosno *vestito* ‘haljina’. U korpusu MaLi djevojčica Livia proširila je značenje imenice *ballerina* ‘balerina’ na novi kontekst te su je drugi članovi obitelji prihvatili kao takvu, što se označava dodavanjem oznake @f na riječ.

U usmenoj interakciji među sugovornicima često dolazi do preklapanja iskaza u cijelosti ili njihovih dijelova. Preklapanja se označavaju na način prikazan u isječku (30), u kojemu djevojčica i majka istodobno izgovaraju dijelove iskaza – dok jedna govori *dvije*, druga izgovara *ecco* ‘evo’.

- (30) *CH2: [- hrv] <dvije> [>]. ‘Dvije.’
 *MOT: <ecco> [<] adesso sono assieme. ‘Evo sada su zajedno.’
 *CH2: sono dvije@s. ‘Dvije su.’

U tom se isječku pritom preklapaju iskazi na dvama jezicima – iskaz djevojčice na hrvatskome i majčin iskaz na talijanskome jeziku. Posebnosti označavanja dvojezične interakcije pojasnit ćemo u sljedećem odjeljku.

Učestala su ponavljanja riječi ili skupina riječi. Ako ih govornik ponovi jednom, između riječi i njezina ponavljanja stavlja se oznaka [/] kao u (31). U slučaju više od dvaju ponavljanja istoga dijela iskaza, kako se vidi u primjeru (31), u uglatim se zagradama označava broj ponavljanja iza znaka x i razmaka: [x broj]. Ako govornik prilikom ponavljanja ispravlja ono što je prethodno rekao, to se bilježi oznakom [//] kao u (32).

- (31) *MOT: no adesso <non c'è> [/] non c'è. ‘Ne, sada ga nema, nema ga.’
 (32) *CH2: <vai vi> [//] vai via. ‘Idi, idi.’

Prilikom usmene interakcije govornici povremeno ne dovršavaju vlastite iskaze, a nisu prekinuti od sugovornika, što se događa i u građi koja je predmetom ovoga rada. Nerijetko do toga dolazi u trenutku kada je govornik zaokupljen nečim, kada izvodi neku radnju dok govori ili mu je pažnja skrenuta na što drugo. Nedovršeni iskaz, kao što je to slučaj u isječku (33), označava se oznakom +... na mjestu na kojemu govornik samostalno prestaje govoriti ne dovršavajući iskaz. Ako pak govornik nastavlja iskaz u drugom trenutku, ovom se oznakom na početku iskaza ističe da je riječ o nastavljanju prethodno započeta iskaza istoga govornika: ++.

- (33) *MOT: ma chi è felicissimo? ‘Ma tko je presretan?’
 *MOT: sta +... ‘Ova...’
 *CH2: ka@b. ‘Ka.’
 *MOT: ++ sta mucca? ‘Ova krava?’

Označavanje prijepisa: posebnosti dvojezične interakcije

Govornici koji tečno govore dva ili više jezika prilikom izražavanja pribjegavaju

„stilu“ (MacSwan, 2005: 55) koji podrazumijeva prelaženje s jednoga jezika na drugi i koji se u literaturi naziva miješanjem jezika (prema engl. *language-mixing*) ili prebacivanjem kodova odnosno preključivanjem (prema engl. *code-switching*).¹⁷ Ta se posebnost dvojezične interakcije odražava na prepisivanje. MacWhinneyev (2000: 102–103) priručnik predlaže način označavanja prebacivanja kodova koji se, uz sitne prilagodbe, primjenjuje u korpusu MaLi. Važno je istaknuti da se prilikom prepisivanja interakcije u korpusu MaLi poštuje grafemski sustav onoga jezika kojemu pripada iskaz ili dio iskaza koji je zahvaćen tom pojavom.

Prebacivanje kodova u korpusu MaLi ostvaruje se na različitim jezičnim razinama: među rečenicama, tj. iskazima, dakle na međurečeničnoj ili međuiskaznoj razini, unutar rečenice ili iskaza, kao i na razini same riječi, tj. među morfemima jedne riječi. Prebacivanje kodova na razini iskaza ostvaruje se na način da se u iskaz na jednome jeziku umeće riječ na drugome jeziku ili pak tako da unutar jedne riječi prepoznamo morfeme koji pripadaju dvama jezicima. Najčešće dolazi do umetanja riječi na jednome jeziku u iskaz na drugome. Takav slučaj oprijetuju isječci iz prijepisa (34)–(37), u kojima se na riječ iz jezika koji je manje zastupljen u prijepisu, a što saznamo iz poretka jezika navedenih u zaglavlju prijepisa, vezuje oznaka @s.

- | | | |
|------|---|------------------------|
| (34) | *CH2: sono dvije@s. | ‘Dvije su.’ |
| (35) | *CH2: no. | ‘Ne.’ |
| | *MOT: come no? | ‘Kako ne?’ |
| | *CH2: mamma dammi maca@s. | ‘Mama, daj mi macu.’ |
| (36) | *CH2: volim@s treno [: treno]. | ‘Volim vlak.’ |
| | *CH2: ti piace il treno? | ‘Sviđa li ti se vlak?’ |
| | *CH2: sì. | ‘Da.’ |
| (37) | *CH2: &-oh l’altla [: altra] [* m] policajac@s. | ‘Oh, drugi policajac.’ |

Prebacivanje kodova na međurečeničnoj ili međuiskaznoj razini označava se na drugi način, i to stavljanjem oznake [- hrv] ili [- ita] na početak iskaza koji započinje

¹⁷ O prebacivanju kodova uopće i njihovoj analizi na temelju iskaza zabilježenih u dnevničkim zapisima korpusa MaLi v. Angster i Batinić Angster (2022).

na jeziku koji nije naveden kao prvi jezik interakcije u zaglavlju. Budući da je talijanski prvi i temeljni jezik interakcije zabilježene na zvučnome zapisu koji se prepisuje u ovome slučaju, ova se oznaka odnosi ponajprije na preključivanje s talijanskoga na hrvatski jezik te se navodi kada jedna od sudionica interakcije kazuje cijeli iskaz na hrvatskome jeziku. Uporabu ovih oznaka možemo vidjeti u primjerima (38) i (39).

- | | | |
|------|---------------------------------|------------------|
| (38) | *MOT: ecco prego. | ‘Evo, izvoli.’ |
| | *CH2: grazie. | ‘Hvala.’ |
| | *CH2: [- hrv] to je ruka. | ‘To je ruka.’ |
| (39) | *CH2: quella è mia. | ‘Ona je moja.’ |
| | *CH2: [- hrv] to je klaun. | ‘To je klaun.’ |
| | *MOT: [- hrv] klaun? | ‘Klaun?’ |
| | *CH2: [- hrv] klaun [x 7]. | ‘Klaun.’ |
| | *MOT: questo è un ponte invece. | ‘A ovo je most.’ |
| | *CH2: un ponte@e. | ‘Most.’ |

Iskaz u kojemu govornik rabi dominantan jezik interakcije nakon što je netko od sudionika prešao na drugi jezik ne označava se oznakom za prebacivanje kodova jer se u takvim situacijama nastoji održati interakcija na temeljnom jeziku. Tako u isječku (39) nema oznake [- ita] prije iskaza *questo è un ponte invece* ‘A ovo je most.’ kojim bi se označio povratak na talijanski jezik. Međutim, ima slučajeva u kojima dolazi do naizmjeničnog prebacivanja s jednoga jezika na drugi te s drugoga na prvi unutar istog iskaza. Kada iskaz započinje prebacivanjem na jezik koji nije prvi naveden u zaglavlju prijepisa, a zatim se iskaz nastavlja na dominantnom jeziku, u trenutku ponovnoga povratka na dominantan jezik, navodi se i oznaka za taj jezik. Zato je u isječku (40) dvaput označeno prebacivanje kodova.

- | | | |
|------|------------------------------------|-----------------------------------|
| (40) | *MOT: si è svegliata in un bosco. | ‘Probudila se u šumi.’ |
| | *CH2: si è (s)vegliata. | ‘Da, probudila se.’ |
| | *MOT: sì, con tanti amici attorno. | ‘Da, s puno prijatelja oko sebe.’ |
| | *CH2: [- hrv] to je [- ita] lupo. | ‘To je vuk.’ |

Prebacivanje kodova na razini riječi, odnosno na granicama morfema, označava se na način da se na riječ u kojoj se ostvaruju morfemi dvaju jezika navodi oznaka @s koju slijede dvotočka i kratice jezika na kojima su morfemi, tj. u slučaju isječka (41) oznaka @s:hrv+ita.¹⁸

- (41) *MOT: e quello lì fuori si è un albero.
 ‘Da, ono tamo vani je stablo.’
 *CH2: [- hrv] <ja ću> [/] <ja ću penjele@s:hrv+ita> [/] ja ću penjele@s:hrv+ita [- ita] a +//.
 ‘Ja ću se penjati.’
 *CH2: [- ita] di là [- hrv] ja ću penjele@s:hrv+ita [- ita] di là.
 ‘Tamo ću se penjati.’
 *MOT: cosa vuoi, spegnere?
 ‘Što želiš, ugastiti?’¹⁹
 *CH2: no di là penj@s +//.
 ‘Ne, tamo penjati.’
 *CH2: penjele@s:hrv+ita.
 ‘Penjati.’
 *MOT: ti vuoi arrampicare sull’albero?
 ‘Želiš se penjati (popeti) na stablo?’
 *CH2: sì.
 ‘Da.’

U isječku (41) djevojčica izgovara riječ *penjele* kojom se oprimjeruje prebacivanje kodova unutar riječi. Naime, korijenski se morfem hrvatskoga glagola *penjati* (*penj-*) kombinira s talijanskim infinitivnim sufiksom ili dometkom *-ere*, na što ukazuje oznaka na kraju prepisane riječi.

Uz prebacivanje kodova, u proizvodnji dvojezičnih pojedinaca, pa tako i u slučaju djece čija se proizvodnja bilježi u korpusu MaLi, učestala je pojava i transfer ili prijenos značajki iz jednoga jezika u drugi (v. Angster i Batinić Angster, 2022).

¹⁸ Prijevod ovoga isječka, čiji su iskazi dulji i opterećeni oznakama, iznimno se navodi ispod iskaza na koji se odnosi radi bolje čitljivosti.

¹⁹ Majka u odgovoru rabi glagol *spegnere* ‘ugastiti’ nastojeći razumjeti što djevojčica govori. Majci, naime, ‘penjele’ prvotno zvuči kao oblik talijanskoga glagola *spegnere* ‘ugastiti’ koji se, uslijed činjenja razvojne pogreške, odlikuje drukčijim oblikom – označenim ispuštanjem prvog suglasnika u suglasničkom skupu /sp/ i izgovorom glasa /t/ glasom /l/. Majčina pretpostavka pokazuje se pogrešnom u nastavku razgovora.

Razmatranje označavanja te pojave trebat će se tek uzeti u obzir u daljnjem razvoju korpusa MaLi.²⁰

Završne napomene

Osim glavnih redaka, program CHAT omogućuje prepisivaču da rabi i zavisne retke. Zavisni redci odnose se na glavne, u njima se ne prepisuju iskazi, nego se opisuju različite vrste okolnosti koje istraživaču mogu biti relevantne. Oni započinju znakom % iza kojeg se nalazi troslovna oznaka, npr. %sit (sit prema engl. *situation* ‘situacija’), %add (add prema engl. *addressee* ‘primatelj’) ili %exp (exp prema engl. *explanation* ‘objašnjenje’).

U retku %sit: unutar korpusa MaLi prepisivač navodi bilješku o situaciji koja je potencijalno relevantna za analizu podataka, kao što je dolazak osobe koja inače s djetetom govori drugi jezik, što može biti relevantno pri analizi u slučaju da se ostvaruje prebacivanje kodova. Prilikom prepisivanja zapisa kojim je zabilježena interakcija većeg broja sudionika, što je u kontekstu korpusa MaLi često slučaj u talijanskim zvučnim zapisima interakcije dvoje djece s roditeljima (Batinić Angster i Angster, 2024), pri čemu povremeno dolazi i do razvoja unakrsne interakcije, važno je istaknuti komu se obraća osoba čiji je iskaz prepisan (%add). Zavisni redak %exp u korpusu MaLi služi prepisivaču i istraživaču da zabilježiti opasku o uporabi jezičnog sredstva ili oblika, sintaktičke konstrukcije i sl. na specifičan način, koji može biti neočekivan za odrasloga govornika danoga jezika, ali karakterističan za dijete, za što nije predviđena nijedna druga oznaka.

Kako bi se osiguralo da prijepis bude usklađen s normama CHAT-a popisanim u MacWhinney (2000), potrebno je provoditi redovite provjere prijepisa. Odabirom tipke *escape* ESC i slova L na tipkovnici računala daje se naredba računalnom programu da provede provjeru. Ako program ukaže na problem u prijepisu, što se može očekivati osobito pri prvoj uporabi programa, potrebno je prepraviti mjesta na kojima su prepoznate neusklađenosti. Ova funkcija služi kako bi cijeli zapis bio u skladu s pravilima sustava, a postupak treba ponavljati dokle god na dnu stranice ne bude napisano da je sve u redu. Nerijetko su posrijedi pogreške poput navođenja razmaka u oznaci u kojoj on nije predviđen ili pak njegovo nenavođenje ondje gdje je pretpostavljen i sl. Koristeći funkciju provjere usklađenosti s normama, jednom kada se završi izrada prijepisa, može se prionuti daljnjoj analizi bez brige da dio

²⁰ U priručniku za uporabu CHAT-a (MacWhinney, 2000) nisu istaknute specifične oznake kojima bi se u prijepisu uputilo na tu pojavu.

prijepisa nije proveden poštujući norme i pravila sustava CHAT.

Uporabom oznaka koje su u skladu sa zahtjevima CHAT-a program prihvaća prijepis. Jednom spremljeni prijepis može biti nadopunjen novim informacijama, odnosno podložan je izmjenama ovisno o eventualno naknadno donesenim odlukama o označavanju korpusa.

ZAKLJUČAK

U ovome smo radu, polazeći od jezičnih podataka kao izvora jezikoslovnih istraživanja, razmotrili obradu dječjih jezičnih podataka. Opažanje tijeka i obilježja (ranoga) jezičnog razvoja često se temelji na analizi spontane jezične proizvodnje djece zabilježene u korpusima dječjega jezika. U radu smo prikazali primjenu široko dostupnoga modela prepisivanja na primjeru izrade prijepisa zvučnoga zapisa spontane interakcije između jedne dvojezične djevojčice i njezine majke.

Prijepis izrađen slijedeći norme CHAT-a služiti će za izradu dvojezičnoga hrvatsko-talijanskoga korpusa dječjega jezika MaLi (Batinić Angster i Angster, 2024) i bit će prikladan za daljnju obradu i analizu pomoću programa CLAN. Izrada javno dostupnoga korpusa MaLi omogućit će uvid u različite aspekte ranoga hrvatsko-talijanskog dvojezičnog razvoja i doprinijeti vidljivosti hrvatskoga kao jednoga od dvaju usvajanih jezika u dječjoj dobi. Unatoč sve češćoj uporabi eksperimentalnih metoda, korpusi i dnevnički zapisi ostaju nezamjenjivi među istraživačima ranoga jezičnog razvoja jer omogućuju opetovanu analizu podataka i dugotrajno praćenje svih aspekata djetetova jezika prilikom čega se bilježi čitav niz podataka koji se tiču složenosti djetetove interakcije s okolinom (usp. Caselli i sur., 2015: 59–61).

Priprema prijepisa u skladu s normama CHAT-a prvi je korak ka omogućavanju istraživačima, a i cjelokupnoj javnosti, uvid u temeljna obilježja hrvatsko-talijanskoga dvojezičnog razvoja i razvojnih obilježja hrvatskoga jezika kod dvojezične djece. Konačni primatelji tako obrađenih dječjih jezičnih podataka mogu biti jezikoslovci i stručnjaci iz drugih polja (npr. logopedi, odgojitelji, učitelji, studenti relevantnih struka).

Ovdje opisan prikaz oznaka koji se rabe za označavanje različitih aspekata dječje jezične proizvodnje, ali i uopće govora i usmene interakcije, može se smatrati i svojevrsnim uvodom u uporabu CHAT-a te može biti koristan ne samo budućim istraživačima dječjega jezika i dvojezičnoga razvoja, nego i bilo komu tko bi poželio razviti korpus govorenoga jezika primjenom računalnoga alata prikladnoga za daljnju analizu jezika.

LITERATURA

- Angster, M. i Batinić Angster, M. (2022). Gli effetti del contatto croato-italiano nelle produzioni di un bambino bilingue. *Studia Romanica et Anglica Zagabiensia: Revue publiée par les Sections romane, italienne et anglaise de la Faculté des Lettres de l'Université de Zagreb*, 67, str. 107–122. <https://doi.org/10.17234/SRAZ.67.8>
- Batinić Angster, M. i Angster, M. (2023). *MaLi – kućni laboratorij*. Prilog sa XXXVII. međunarodnoga znanstvenog skupa Hrvatskoga društva za primijenjenu lingvistiku *Jezik i migracije*, Osijek, 15–17. 6. 2023.
- Batinić Angster, M. i Angster, M. (2024). Building MaLi, a Croatian-Italian bilingual child corpus. *Strani jezici*, 53, str. 69–88. <https://doi.org/10.22210/strjez/53-1/4>
- Behrens, H. (2008). Corpora in Language Acquisition Research. History, methods, perspectives. U: H. Behrens (ur.), *Corpora in language acquisition research: History, methods, perspectives* (str. xi–xxx). John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.6.03be>
- Caselli, M. C., Bello, A., Rinaldi, P., Stefanin, S. i Pasqualetti, P. (2015). *Il primo vocabolario del bambino: gesti, parole e frasi. Valori di riferimento fra 8 e 36 mesi delle Forme complete e delle Forme brevi del questionario MacArthur-Bates CDI*. Franco Angeli.
- Crystal, D. (2010). *The Cambridge Encyclopedia of Language. Third edition*. Cambridge University Press.
- Čamber, M. (2020). *Simultaneous acquisition of Austrian German and Croatian at home and in preschool*. Doktorski rad. Sveučilište u Beču.
- Dardano, M. i Trifone, P. (1997). *La nuova grammatica della lingua italiana*. Zanichelli.
- Grgurev, S. (2023). *Jezični podatak i njegova obrada na primjeru dječjeg govora*. Završni rad. Sveučilište u Zadru.
- Hržica, G. (2012). Daj mi to napisaj: preopćavanja glagolske osnove u usvajanju hrvatskoga jezika. *Suvremena lingvistika*, 38 (74), str. 189–208.
- Hržica, G., Brdarić, B., Tadić, E., Goleš, A. i Roch, M. (2015). Dominantnost jezika dvojezičnih govornika talijanskog i hrvatskog jezika. *Logopedija*, 5(2), str. 34–40. <https://hrcak.srce.hr/150404>
- Jelaska, Z. (2004). *Fonološki opisi hrvatskoga jezika. Glasovi, slogovi, naglasci*. Hrvatska sveučilišna naklada.

- Kovačević, M. (2002). *Hrvatski korpus dječjeg jezika*. <https://childes.talkbank.org/access/Slavic/Croatian/Kovacevic.html>.
- Kovačević, M., Jelaska, Z., Kuvač Kraljević, J. i Cepanec, M. (2007). *Komunikacijske razvojne ljestvice (KORALJE)*. Naklada Slap.
- Kreuz, R. i Riordan, M. (2018). 3. The art of transcription: Systems and methodological issues. U: A. H. Jucker, K. P. Schneider, W. Bublitz (ur.) *Methods in Pragmatics* (str. 95–120). De Gruyter Mouton. 10.1515/9783110424928-003.
- Kuvač, J. i Palmović, M. (2007). *Metodologija istraživanja dječjeg jezika*. Naklada Slap.
- Lieven, E. i Behrens, H. (2012). Dense sampling. U: E. Hoff (ur.), *Research methods in child language: A practical guide* (str. 226–239). Wiley-Blackwell. <https://doi.org/10.1002/9781444344035.ch15>
- MacSwan, J. (2005). Précis of a minimalist approach to intrasentential code switching. *Italian Journal of Linguistics*, 17 (1).
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. 3rd Edition*. Lawrence Erlbaum Associates. <https://doi.org/10.21415/3mhn-0z89>
- Medved Krajnović, M. (2004). *Razvoj hrvatsko-engleske dvojezičnosti u dječjoj dobi*. Doktorski rad. Filozofski fakultet Sveučilišta u Zagrebu.
- Ordulj, A. i Hržica, G. (2015). Obnavljanje Hrvatskog korpusa dječjeg jezika. *Logopedija*, 5, 1, 2015, str. 25–31.
- Zmarich, C. (2010). Lo sviluppo fonetico/fonologico da 0 a 3 anni. U: S. Bonifacio i L. Hvastja Stefani (ur.). *L'intervento precoce nel ritardo di linguaggio. Il modello INTERACT per il bambino parlatore tardivo* (str. 17–39). Franco Angeli.
- Zmarich, C. i Bonifacio, S. (2005). Phonetic inventories in Italian children aged 18–27 months: a longitudinal study. Prilog sa skupa INTERSPEECH – Eurospeech, 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, September 4–8. 2005. 10.21437/Interspeech.2005-351
- Zmarich, C., Dispaldro, M., Rinaldi, P. i Caselli, M. C. (2011). Caratteristiche fonetiche del primo vocabolario del bambino. *Psicologia clinica dello sviluppo*, 15 (1), str. 235–256.

PROCESSING CHILD LANGUAGE DATA USING CROATIAN-ITALIAN BILINGUALISM AS A CASE STUDY

ABSTRACT

This paper provides insight into one aspect of processing child language data, using the transcription of an audio recording of the interaction between a bilingual girl and her mother as an example. Much of what we know today about early language development stems from child language corpora, created through recordings of spontaneous interactions between children and their parents, and the subsequent transcription of such collected language data (cf. Kuvač and Palmović, 2007; Behrens, 2008; Ordulj and Hržica, 2015). One method of transcribing and processing language data was developed for the needs of the CHILDES (Child Language Data Exchange System; MacWhinney, 2000). The aim of this paper is to present the application of that model and its standards in the transcription of audio recordings created within the project of constructing the bilingual Croatian-Italian child language corpus, MaLi (Batinić Angster and Angster, 2024). Using an example of the spontaneous speech production of a bilingual girl interacting with her mother, which takes place mostly in Italian, the paper will illustrate the transcription procedures of a one-hour audio recording using the CHAT (Codes for the Human Analysis of Transcripts; MacWhinney, 2000) program. Special attention will be given to marking developmental errors related to phonetic-phonological development and overgeneralization, as well as specific features of bilingual production, such as code-switching.

KEYWORDS:

child language, first language acquisition, bilingualism, early language development, language processing, corpus