

Gordana Hržica

Edukacijsko-rehabilitacijski fakultet
Borongajska cesta 83f, HR-10000 Zagreb
orcid.org/0000-0001-6067-9148
gordana.hrzica@erf.hr

Ivana Trtanj

Fakultet za odgojne i obrazovne znanosti
Ulica cara Hadrijana 10, HR-31000 Osijek
orcid.org/0000-0002-4700-8443
itrtanj@foozos.hr

MJERE RJEČNIČKE RAZNOLIKOSTI U PRIČAMA DJECE PREDŠKOLSKE I RANE ŠKOLSKE DOBI

Razvoj rječnika u usvajanju prvoga jezika može se pratiti analizom transkripata govorenog jezika djece predškolske i rane školske dobi, dobivenih, primjerice, pričanjem nepoznate priče prema slikovnom predlošku. Pri analizi rječnika takvih jezičnih uzoraka upotrebljava se niz mjera poznatih kao mjere rječničke raznolikosti. Za dvije mjere, broj različitih riječi (BRR) ili omjer različenica i pojavnica (ORP), utvrđeno je da ovisе o duljini transkripta te zbog toga nisu prikladne za analizu. Preporučuje se da se te mjere izračunavaju na ograničenom broju riječi ili da se upotrebljavaju druge mjere, primjerice parametar/mjera D (D) ili pomični prosječni omjer obličnica i pojavnica (MATTR). Dosadašnja istraživanja nisu u potpunosti razriješila dvojbu oko prikladnih mjera za mjerenje rječničke raznolikosti. Validacijska istraživanja u najvećem broju slučajeva temelje se na analizi pisanoga jezika te na analizi engleskoga jezika. Zbog toga je cilj ovog rada bio istražiti koje mjere rječnike raznolikosti mogu razlikovati govorenu jezičnu proizvodnju djece različite dobi u hrvatskom jeziku. S pomoću slikovnice bez riječi *Frog, where are you?* ispitano je ukupno 157 djece u četiri dobne skupine. Podatci pokazuju da rječnička raznolikost mjerena svim četirima mjerama raste s porastom kronološke dobi. Tri mjere (BRR, ORP, D) statistički značajno razlikuju dvije najmlađe skupine (četverogodišnjaci i šestogodišnjaci) od najstarije (desetogodišnjaci). No, samo mjera MATTR razlikuje dvije najmlađe skupine od dviju najstarijih skupina (osmogodišnjaci i desetogodišnjaci).

1. Uvod

Kada govorimo o ranom leksičkom razvoju u literaturi se često susrećemo s pojmom *razvoj rječnika* (engl. *vocabulary development*). Johansson (2008) navodi primjer iz priručnika Fletchera i MacWhinneya (1995) gdje se čitatelja s pojma *leksički razvoj* (engl. *lexical development*) upućuje na *razvoj rječnika* (engl. *vocabulary development*). Isto pronalazimo kod Dromi (1999) i Crystal (1997). Razvoj rječnika može se pratiti mjernim sredstvima kao što su standardizirani jezični testovi (za hrvatski jezik Peabody slikovni test rječnika – PPVT-III-HR, Dunn i Dunn, Kovačević i dr. 2009) ili roditeljske ljestvice (za hrvatski jezik Komunikacijske razvojne ljestvice – Koralje, Kovačević i dr. 2007). No, moguće je rabiti i drugačije, neizravnije metode, to jest proučavati jezične uzorke – pisane radove ili transkripte govorenog jezika. Pri analizi jezičnih uzoraka za procjenu rječnika upotrebljava se niz mjera koje su poznate kao mjere rječničke raznolikosti ili rječničkoga bogatstva. Pokazalo se da takve mjere mogu razlikovati djecu različite kronološke dobi (primjerice, Berman 2000, Fey i dr. 2004) te djecu s razvojnim jezičnim poremećajem od njihovih vršnjaka urednoga jezičnog razvoja (Fey i dr. 2004, Mäkinen 2014, Jacobson i Walde 2013, Kelić i dr. 2012, Hržica i dr. 2019).

Pojam *rječničke raznolikosti* (engl. *lexical diversity*) koja je u fokusu ovoga rada, u literaturi je ponekad sinonim za *rječničko bogatstvo* (engl. *lexical richness*) (Daller, van Hout i Treffers-Daller 2003). Međutim, Malvern i dr. (2004) objašnjavaju kako je mjera rječničke raznolikosti samo jedna dimenzija rječničkoga bogatstva. Moguće je, naime, mjeriti različite aspekte rječničkoga bogatstva: rječničku raznolikost (raspon rječnika u jezičnom uzorku i izbjegavanje ponavljanja), rječničku gustoću (udio leksičkih riječi u jezičnom uzorku), rječničku sofisticiranost (udio naprednih riječi, to jest riječi niže čestote u jezičnom uzorku) te udio pogrešaka (primjerice, u odabiru riječi) (Read 2000). Sve te dimenzije rječničke raznolikosti pružaju raznovrsne informacije o razvoju ili kvaliteti rječnika koje se međusobno nadopunjuju (Ménard 1983).

Mjere rječničke raznolikosti upotrebljavane su u različitim istraživanjima (pregled: Malvern i dr. 2004), od forenzičke lingvistike i analize povijesnih dokumenata do demografskih utjecaja na uporabu rječnika. Najčešće se ipak upotrebljavaju u svrhu istraživanja jezičnoga razvoja prvoga jezika, u istraživanjima

dvojezičnosti (primjerice, usporedba dvaju jezika dvojezičnih govornika) te u istraživanjima jezičnih poremećaja (primjerice, razvojnoga jezičnog poremećaja ili afazije). Unatoč relativno dugoj povijesti uporabe, postoje različiti metodološki problemi, prije svega ovisnost o duljini jezičnoga uzroka, koji onemogućuju pouzdan odabir samo jedne mjere rječničke raznolikosti (Malvern i Richards 1997, Vermeer 2000). Svaka mjera rječničke raznolikosti ovisi i o karakteristikama mjerena jezika (primjerice, Hržica i Roch 2020, Hržica i dr. 2020). Ta je problematika u velikoj mjeri zanemarena. Većina velikih validacijskih istraživanja mjera rječničke raznolikosti provedena je na engleskom jeziku. Istraživanja u kojima se, primjerice, uspoređuju jezični uzorci u dvama jezicima, često ne uzimaju u obzir jezične razlike (vidi Ucelli i Paez 2007, Bedore i dr. 2010; Simon-Cerijido i Gutiérrez-Clellen 2009, ali i Daller i dr. 2003, Treffers-Daller i Korybski 2016, Hržica i Roch 2020 za drugačiji pristup).

1.1. Mjere rječničke raznolikosti

Mjere jezičnoga razvoja temelje se na razlikovanju triju temeljnih pojmova. Pojavnica (engl. *token*) svaka je riječ u jezičnom uzorku, to jest svaki niz znakova okružen bjelinama. Različnica (engl. *type*) svaki je fonološki različit izraz riječi. Natuknica (engl. *lemma*) citatni je oblik riječi kakav se pojavljuje u rječnicima (više o razlikovanju tih pojmova u Jelaska 2005). Tako u rečenici *U grad je krenula, u grad će stići i u gradu će ostati.* ima 13 pojavnica, 9 različenica (jer se ponavlja oblik ‘grad’, ‘u’ i nenaglašeni oblik pomoćnoga glagola ‘htjeti’) te 8 natuknica (‘u’, ‘grad’, ‘biti’, ‘krenuti’, ‘stići’, ‘i’, ‘ostati’, ‘htjeti’ (> će)). Najčešće ispitivane mjere leksičkoga razvoja broj su različitih riječi (različnica) u jezičnom uzorku (često se označava engleskom kraticom NDW) te omjer različenica i pojavnica (Johnsson 1944, Templin 1957) (ORP – prema Kuvač i Palmović 2007, no često se označava engleskom kraticom TTR). Jasno je, naravno, da se mjerom broj različitih riječi mogu mjeriti samo jezični uzorci slične dužine. Bilo koji govornik, čak i onaj s manje razvijenim rječnikom, proizvest će više različenica ako govori duže. Zbog toga se kao pouzdanija mjera dugo rabio omjer različenica i pojavnica. Njegovim izračunom dobivamo podatak o tome koliki je udio različitih riječi u ukupnom broju svih riječi u jezičnom uzorku. Ako je omjer bliži vrijednosti 1, to znači da je ispitanik svaku riječ rekao u manje navrata,

dok omjer različenica i pojavnica bliži vrijednosti 0 znači da je ispitanik svaku riječ upotrijebio više puta. Viši ORP pokazuje veću rječničku raznolikost jer je u jezičnom uzorku više različitih riječi (pretpostavljamo da je govornik imao više riječi na raspolaganju jer ima bogatiji rječnik), a što je taj omjer manji, to je rječnička raznolikost manja jer je veći broj riječi koje se ponavljaju (Kuvač i Palmović 2007) (pretpostavljamo da je govornik imao manje riječi na raspolaganju jer mu je rječnik siromašniji). No, pokazalo se da dužina uzroka ipak utječe i na mjeru omjer različenica i pojavnica. Istraživanja su pokazala (pregled: Malveren i dr. 2004) da ta mjera s duljenjem uzorka opada. Naime, što duže govorimo, to će se više ponavljati, primjerice, funkcionalne riječi te se tako omjer različenica i pojavnica smanjuje iako govornik upotrebljava različite leksičke riječi (Johansson 2008). U istraživanju Gayrauda (2000) koji je usporedio ORP i broj pojavnica pokazalo se da iako broj pojavnica značajno raste s porastom kronološke dobi govornika, ORP pada. Drugim riječima, ORP može izmjeriti rječničku raznolikost na jezičnim uzorcima jednake duljine, ali nemoguće je usporediti jezične uzorke različitih duljina.

Potencijalna rješenja toga problema su dvojaka. Moguće je ujednačiti jezične uzorke koji se uspoređuju po duljini (primjerice, uzeti jednak broj riječi ili iskaza: primjerice Miller 1991, Klee 1992, Thordardottir i Ellis Weismer 2001). No, time se gube dragocjeni podatci jer tekstove je potrebno ujednačiti prema najkraćem jezičnom uzorku od onih koje želimo usporediti. Također, često se skraćivanjem uspoređuju različiti dijelovi nekoga teksta. Primjerice, uvod priče jednoga djeteta može biti jednako dug kao čitava priča drugoga. Alternativno rješenje ponudili su autori koji su osmislili mjere koje neće biti ovisne o duljini jezičnoga uzorka, to jest, njima se mogu pouzdano usporediti tekstovi i transkripti govorenoga jezika koji su različite dužine, a da se pri tome ne moraju kratiti kako bi se ujednačili. Nekoliko je autora (Guiraud 1960, Carroll 1964, Maas 1972) predložilo različite transformacije ORP-a kako bi se nadoknadila varijabilnost duljine uzorka (pregled: Hržica i Roch 2020, Hržica i dr. 2020). Od mnogih mjera dvije su dio programskoga paketa CLAN, koji omogućuje njihov automatiziran izračun te su nešto zastupljenije u istraživanjima jezičnoga razvoja. Jedna je od takvih mjera VOCD (VOCabulary D), indeks rječničke raznolikosti D ili jednostavno parametar / mjera D (engl. *diversity*). Mjera D, koju su predložili Malvern i Richards (1997) i koju su dalje razvili Malvern i dr. (2004), izračunava se u softveru *vocd* (McKee i dr. 2000) primjenom pristu-

pa prilagođavanja krivulji. Temelji se na predviđenom padu ORP-a porastom uzorka. Viša vrijednost D pokazuje višu rječničku raznolikost, a samim time i bogatiji rječnik. Druga mjera za koju se također pokazalo da je otporna (ili barem otpornija) na duljinu jezičnoga uzorka (Covington i McFall 2010) jest prosječni omjer različenica i pojavnica (engl. *moving-average type–token ratio* ili MATTR). Predložili su je Covington i McFall (2010: 96–97), a njihov pristup zagovarali su i Köhler i Galle (1993). MATTR se izračunava tako da se odabere duljina teksta (npr. 500 riječi) i izračunava se ORP za riječi 1–500, zatim za riječi 2–501 pa za 3–502 i tako dalje do kraja teksta. Srednja vrijednost svih tih ORP-ova jest mjera rječničke raznolikosti cijeloga teksta te stoga na nju ne bi trebala utjecati duljina teksta. Covington i McFall (2010) opisuju MATTR kao mjeru neovisnu o duljini teksta, koja može otkriti promjene unutar teksta kao i među tekstovima. Fergadiotis i dr. (2015) utvrdili su da MATTR ima veću valjanost od D.

Važno je napomenuti da se mjere koje su razvijane tako da budu otporne na utjecaj duljine jezičnoga uzorka češće upotrebljavaju u istraživanjima pisanoga jezika, primjerice pri analizi pisanih tekstova osoba koje uče neki strani ili drugi jezik. U području usvajanja prvoga jezika još se uvijek često upotrebljavaju temeljne mjere, posebice broj različitih riječi u omeđenom broju riječi, kako bi se izmjerila rječnička raznolikost. No, s obzirom na pokazane dobre metrijske karakteristike novijih mjera, njihova je primjena u analizi uzoraka govorenoga jezika poželjna kako bi se dobilo podatke o prikladnosti mjera za takva istraživanja.

1.2. Rječnička raznolikost u istraživanjima jezičnog usvajanja

Mjere rječničke raznolikosti često su u području procjene jezičnoga znanja u prvom i drugom jeziku služile kao opće mjere razvoja pisanoga i govorenoga jezika (Malvern i dr. 2004: 23). Zasiurno najopsežnije istraživanje koje je procjenjivalo rječničku raznolikost u prvom jeziku jest istraživanje Ruth Berman *Developing literacy in different contexts and in different languages* (Berman 2000). To međujezično istraživanje koje je obuhvatilo sedam jezika usporedilo je djecu različite dobi i odrasle te proizvodnju jezika u pripovjednom i izlagačkom diskursu u pismu i govoru (Berman i Verhoeven 2002). Istraživanje je potvrdilo utjecaj dobi, žanra i jezika na rječničku raznolikost, dok se razlike u jezičnim

djelatnostima (govorenje i pisanje) nisu pokazale kao značajne u raznolikosti rječnika. Strömquist i dr. (2002), koristeći se mjerom D, usporedili su govoreni i pisani izlagački i pripovjedni diskurs odraslih govornika iz četiriju zemalja. Rezultati su tog istraživanja, za razliku od Bermaničina istraživanja, pokazali velike razlike između govorenja i pisanja pri čemu je viša rječnička raznolikost uočena u pisanju. Jarvis (2002) je u pisanom pripovjednom diskursu adolescenata koji engleski uče kao strani jezik ispitivao i rezultate u znanju rječnika. Njegov je zaključak da rječnička raznolikost, ako je mjerena pouzdanim mjerama, pozitivno korelira s količinom formalne nastave i poznavanjem rječnika u inom jeziku, a u kojoj je mjeri rječnička raznolikost povezana s kvalitetom pisanja ovisi isključivo o prvom jeziku pojedinca.

Važno je napomenuti da se dobar dio validacijskih istraživanja mjera rječničke raznolikosti, u kojima se govori o prednostima određenih mjera nad drugima, temelji na pisanim tekstovima (primjerice, McCharthy i Jarvis 2007, 2010). S obzirom na to da je Jarvis ustanovio razlike između pisanog i govorenog modaliteta u rječničkoj raznolikosti, potrebno ih je s oprezom interpretirati kad je riječ o istraživanjima na govorenom jeziku. Tomu u prilog govore i razlike u validacijskim istraživanjima koja se temelje na jezičnim uzorcima govorenoga naspram pisanoga jezika (primjerice razlike u rezultatima Malverna i dr. 2004 te McCharthyja i Jarvisa 2010). Također, validacijska istraživanja u najvećoj su se mjeri provodila na transkriptima napisanima na engleskome, jeziku čije je morfološko bogatstvo, koje se odražava i mjerama rječničke raznolikosti, manje od mnogih drugih jezika. Istraživanja koja bi produbila spoznaje o primjeni mjera na govorenim jezicima koji nisu engleski nameću se kao potreban nastavak istraživanja rječničke raznolikosti.

1.3. Međujezične razlike u izračunu mjera rječničke raznolikosti

Temeljne mjere rječničke raznolikosti, broj različitih riječi te broj različenica i pojavnica, osmišljene su za engleski jezik. Engleski je jezik morfološki relativno jednostavan, to jest, u odnosu na morfološki bogatije jezike, svaka natuknica ima daleko manje morfoloških oblika (obličnica). Za mjere rječničke raznolikosti to ima presudan utjecaj. Naime, ako se računalnim programima izračuna mjera broj različitih riječi, broje se obličnice, a ne natuknice. U jezicima u koji-

ma postoji manje obličnica po natuknici (kao što je engleski), ta će mjera davati rezultate koji su bliži izračunu broja različitih natuknica. U jezicima u kojima postoji više obličnica po natuknici, rezultati će se udaljavati od izračuna broja različitih natuknica. Do slične razlike u izračunu u različitim jezicima dolazi se i pri izračunu mjera omjer različenica i obličnica. Taj omjer ovisi o brojanju obličnica, a s obzirom na razlike u morfološkom bogatstvu različitih jezika, broj obličnica može biti znatno bliži ili znatno dalji od broja natuknica.

Na to upozoravaju i Kuvač i Palmović (2007), napominjući da bi za hrvatski jezik bolja mjera rječničke raznolikosti bila mjera omjer natuknica i pojavnica. U zanimljivom istraživanju odmaknutom od teme jezičnoga razvoja, u usporedbi 21 jezika Kettunen (2014) pokazuje da su omjer različenica i pojavnica, ali i pomični prosječni omjer različenica i pojavnica, povezani s morfološkom složenošću jezika. Što se jezik određuje morfološki bogatijim, to jest, ima više morfoloških oblika po natuknici, to su mu više mjere rječničke raznolikosti. U longitudinalnom istraživanju o pripovijedanju u finskom jeziku (Mäkinen i dr. 2014), leksička raznolikost, odnosno mjera broja različitih riječi, tumači se i u kontekstu finskoga kao aglutinativnoga jezika. Autori tvrde da u takvim jezicima mjera broj različitih riječi ne mjeri samo semantičke sposobnosti, već odražava i morfosintaktičku složenost. Slična se usporedba može povući i s flektivnim jezicima kao što je hrvatski, naročito takvima koji su morfološki bogatiji.

Dok je jedan metodološki nedostatak mjera rječničke raznolikosti (ovisnost o duljini transkripta) doveo do brojnih istraživanja i razvoja novih mjera, utjecaj jezičnih razlika na izračun rječničke raznolikosti često nije uziman u obzir. Primjerice, u istraživanjima dvojezičnosti izravno se uspoređuju mjere rječničke raznolikosti u različitim jezicima (primjerice, Ucelli i Paez 2007, Bedore i dr. 2010, Simon-Cerijido i Gutiérrez-Clellen 2009). Pri tome samo neki istraživači uočavaju metodološki problem i pokušavaju ga na neki način razriješiti. Ponekad je to prilagodbom jezičnih uzoraka (primjerice, lematizacija transkripata u Treffers-Daller i Korybski 2016) ili prilagodbom analize (primjerice, uporaba z-vrijednosti u Daller i dr. 2003 ili u Hržica i Roch 2020).

S obzirom na razlike u morfološkom bogatstvu među jezicima, ali i u prisutnosti i čestoti slobodnih gramatičkih morfema (primjerice, članovi, pomoćni glagoli...), nameću se pitanja o usporedivosti mjera u različitim jezicima. Takva su pitanja važna za istraživanja dvojezičnosti, ali i za proučavanje jednojezičnoga

razvoja. Naime, mjere u različitim jezicima mogu zahvaćati u različite sastavnice jezičnoga znanja. Primjerice, u morfološki bogatim jezicima mogle bi odražavati ne samo rječničko bogatstvo, već i morfosintaktički razvoj. Zbog toga i njihova snaga u razlikovanju različitih skupina ispitanika (primjerice, djece različite dobi, osoba s jezičnim poremećajima) može biti drugačija u različitim jezicima, posebice onima koji su morfološki bogatiji od engleskoga jezika.

2. Cilj i istraživačka pitanja

Dosadašnja istraživanja nisu u potpunosti razriješila dvojbu oko prikladnih mjera za mjerenje rječničke raznolikosti. Validacijska istraživanja u najvećem broju slučajeva temelje se na analizi pisanoga jezika te na analizi engleskoga jezika. Zbog toga je cilj ovog rada istražiti koje mjere rječnike raznolikosti mogu razlikovati govorenu jezičnu proizvodnju djece različite dobi u hrvatskom jeziku. Postavljamo dva istraživačka pitanja.

1. Koje mjere rječničke raznolikosti odražavaju razlike među skupinama djece različite kronološke dobi?
2. Koje mjere rječničke raznolikosti nisu vezane uz temeljnu jezičnu produktivnost, to jest uz duljinu transkripta?

3. Metode ispitivanja

3.1. Uzorak ispitanika

Uzorak ispitanika činilo je ukupno 157 djece raspoređenih u četiri kronološke skupine: od 3;00 do 5;11, od 6;00 do 7;11, od 8;00 do 9;11 te od 10;00 do 11;03. Najmlađu skupinu čine predškolska djeca šireg raspona kronološke dobi (tri godine). U kasnijim je skupinama raspon dobi manji te obuhvaća dvije godine. Naime, velik je utjecaj formalnog obrazovanja, posebice čitanja i pisanja, na rječničku raznolikost djece. Kako djeca sve bolje ovladavaju tim složenim vještinama, njihova rječnička raznolikost raste (primjerice, Nippold 2002) te su skupine oblikovane tako da se preciznije uzorkuju djeca koja su ušla u formalni školski sustav.

Tablica 1. Demografske karakteristike ispitanika

skupina	n	kronološka dob			spol	
		raspon	M	SD	m	f
četverogodišnjaci	36	3;00 – 5;11	4;05	1,49	19	17
šestogodišnjaci	42	6;00 – 7;11	6;55	0,78	25	17
osmogodišnjaci	41	8;00 – 9;11	8;55	0,78	25	16
desetogodišnjaci	38	10;00 – 11;11	10;55	0,78	24	14

3.2. Istraživački materijal

U ovome su radu za potrebe istraživanja uzorci prikupljeni situacijskim uzorkovanjem (vidi Kelić i dr. 2012, Trtanj i Kuvač Kraljević 2017). U takvom uzorkovanju snimaju se reakcije na različite metode elicitanja željenih jezičnih elemenata ili samih struktura. Uporaba istoga elicitirajućega materijala omogućuje lakšu i precizniju usporedbu jezičnih uzoraka i pojedinačno unutar jedne dobne skupine, ali i među sudionicima različitih dobnih skupina. Budući da isti elicitirajući materijal omogućuje i barem djelomično ujednačavanje veličine datoteka, čini i dobar preduvjet za uspoređivanje različitih mjera.

Slikovnica *Frog, where are you?* (Mayer 1969) poslužila je kao elicitirajući materijal pomoću kojega su prikupljeni uzorci govorenoga jezika za izračun rječničke raznolikosti. Riječ je o materijalu često rabljenom u istraživanjima dječjega jezika. Berman i Slobin (1994) proveli su na temelju slikovnice *Frog, where are you?* opsežno ispitivanje pripovjedne sposobnosti u pet jezika (engleskom, njemačkom, turskom, hebrejskom i španjolskom) nakon čega je ta priča postala glavni materijal međujezičnih usporedbi za različite uzorke ispitanika, od dvojezične djece preko djece oštećena sluha do one koja imaju jezične teškoće. Slikovnica je to s 24 crno-bijele slike bez teksta koje prikazuju klasičnu dječju priču

s junakom, problemom, nizom akcija koje proizlaze iz tog problema i sretnim završetkom.

Svakom djetetu dane su jednake upute prije ispitivanja te je svako dijete ispitano pojedinačno. Djeca su najprije trebala prelistati slikovnicu, a potom gledajući ju ispričati priču ispitivaču koji je djetetu mogao pomoći jedino neutralnim komentarima, ne utječući na oblikovanje iskaza. Dječje je pripovijedanje tonski snimljeno, a dobiveni su zapisi transkribirani u CHAT-u (*Codes for the Human Analysis of Transcripts*) i analizirani u programu CLAN (*Computerized Language Analysis*) (MacWhinney 2000).

3.3. Izračun mjera produktivnosti i rječničke raznolikosti

U ovom je istraživanju izračunata jedna mjera produktivnosti te četiri mjere rječničke raznolikosti. Mjera produktivnosti jest ukupan broj riječi, a označava broj pojavnica koje je određeni govornik izrekao. Dvije mjere rječničke raznolikosti pripadaju starijim mjerama, koje su u validacijskim istraživanjima pretrpjele dosta kritika. Riječ je o mjerama omjer različenica i pojavnica (ORP) te broj različitih riječi (BRR). Obje su mjere, kako bi se kontrolirao utjecaj veličine transkripta, izračunate na 100 prvih pojavnica. Druge dvije mjere pripadaju u novije mjere te se tvrdi da ne bi trebale biti ovisne o broju pojavnica. To su mjera rječničke raznolikosti D (D) te pomični prosječni omjer različenica i pojavnica (MATTR). Zbog toga su izračunate na ukupnom broju pojavnica. Sve su navedene mjere izračunate automatizirano u programu Computerized Language Analysis (CLAN, MacWhinney 2000).

Tablica 2. Mjera produktivnosti i mjere rječničke raznolikosti u istraživanju

naziv mjere	kratica	vrsta mjere
ukupan broj pojavnica	UBP	mjera proizvodnje
broj različenica na 100 pojavnica	BRR	mjera rječničke raznolikosti
omjer različenica i pojavnica na 100 pojavnica	ORP	mjera rječničke raznolikosti
mjera rječničke raznolikosti D	VOCD	mjera rječničke raznolikosti
pomični prosječni omjer različenica i pojavnica	MATTR	mjera rječničke raznolikosti

3.4. Analiza

Za testiranje normalnosti distribucije proveden je Kolmogorov-Smirnov test. Test se pokazao značajnim za varijable ‘ukupan broj riječi’, ‘omjer različenica i pojavaonica’ te ‘pomični prosječni omjer različenica i pojavaonica’. Za ‘ukupan broj riječi’ dobivena je tendencija k značajnosti ($p=.053$). Normalnu je distribuciju pokazala jedino ‘mjera D’. Zbog tih su rezultata rabljene metode neparametrijske analize, MannWhitney U test te Spearmanov koeficijent korelacije. U usporedbi rezultata među skupinama za mjeru D rabljena je ANOVA.

4. Rezultati

4.1. Rezultati prema skupinama kronološke dobi

4.1.1. Ukupan broj riječi

Četiri dobne skupine pokazuju različite rezultate u ukupnom broju riječi rabljenom u zadatku pripovijedanja. Skupina četverogodišnjaka upotrebljava najmanji broj riječi ($M=204,83$), a raspršenje je rezultata unutar te dobne skupine veliko. Šestogodišnjaci proizvode više riječi od četverogodišnjaka ($M=259,98$), uz nešto manje raspršenje rezultata, ali još uvijek uz velike razlike. Osmogodišnjaci proizvode više riječi i od četverogodišnjaka i od šestogodišnjaka ($M=300,2$), a u odnosu na te skupine raspršenje je rezultata najmanje. Desetogodišnjaci proizvode veći broj riječi od četverogodišnjaka, ali manji i od šestogodišnjaka i od osmogodišnjaka ($M=229,55$). Od svih skupina pokazuju najmanje raspršenje rezultata.

Tablica 3. Rezultati na varijabli ukupan broj riječi za četiri skupine sudionika

skupina	<i>n</i>	<i>M</i>	minimum	maksimum	SD
četverogodišnjaci	36	204,83	40	436	100,561
šestogodišnjaci	42	259,98	96	588	96,348
osmogodišnjaci	41	300,20	157	515	83,486
desetogodišnjaci	38	229,55	107	342	58,244

Usporedba testom Kruskal-Wallis pokazala je da se skupine međusobno razlikuju ($p<.001$). *Post-hoc* analiza pokazuje da četverogodišnjaci proizvode značajno

manje riječi od osmogodišnjaka ($p < .001$) te da desetogodišnjaci proizvode značajno manje riječi od osmogodišnjaka ($p < .005$). Razlike među ostalim skupinama nisu se pokazale značajnima.

4.1.2. Broj različitih riječi na 100 pojava

U sve četiri dobne skupine primjetan je trend u povećanja prosječnoga broja različitih riječi u 100 pojava s povećanjem kronološke dobi. Četverogodišnjaci proizvode najmanje različitih riječi ($M=48,92$), šestogodišnjaci proizvode nešto više različitih riječi ($M=54,52$), osmogodišnjaci još više ($M=56,52$), a skupina najstarije kronološke dobi, desetogodišnjaci, proizvodi najviše različitih riječi u 100 pojava ($M=59,71$).

Tablica 4. Rezultati na varijabli broj različitih riječi za četiri skupine sudionika

skupina	<i>n</i>	<i>M</i>	minimum	maksimum	SD
četverogodišnjaci	36	48,92	41	70	16,737
šestogodišnjaci	42	54,52	45	73	10,714
osmogodišnjaci	41	56,52	41	70	5,988
desetogodišnjaci	38	59,71	49	75	5,589

Usporedba testom Kruskal-Wallis pokazala je da se skupine međusobno razlikuju ($p < .001$). *Post-hoc* analiza pokazuje da četverogodišnjaci proizvode značajno manje različitih riječi od desetogodišnjaka ($p < .001$) te da šestogodišnjaci proizvode značajno manje različitih riječi od desetogodišnjaka ($p < .05$). Razlike među ostalim skupinama nisu se pokazale značajnima.

4.1.3. Omjer različenica i pojava na 100 pojava

Omjer različenica i pojava četverogodišnjaka ($M=,41$) niži je od omjera različenica i pojava šestogodišnjaka ($M=,45$) ali jednak onomu osmogodišnjaka ($M=,41$). Desetogodišnjaci imaju najviši omjer različenica i pojava od svih skupina ($M=,49$).

Tablica 5. Rezultati na varijabli omjer različenica i pojava za četiri skupine sudionika

skupina	<i>n</i>	<i>M</i>	minimum	maksimum	SD
četverogodišnjaci	36	0,49	0,41	0,7	0,167
šestogodišnjaci	42	0,54	0,45	0,73	0,107
osmogodišnjaci	41	0,56	0,41	0,7	0,059
desetogodišnjaci	38	0,59	0,49	0,75	0,056

Uspredba testom Kruskal-Wallis pokazala je da se skupine međusobno razlikuju prema omjeru različenica i pojava na 100 pojava ($p < .001$). *Post-hoc* analiza pokazuje da četverogodišnjaci imaju značajno manji ORP od desetogodišnjaka ($p < .001$) te da šestogodišnjaci imaju značajno manji ORP od desetogodišnjaka ($p < .05$). Razlike među ostalim skupinama nisu se pokazale značajnima.

4.1.4. Mjera D

Deskriptivne vrijednosti pokazuju da rezultati rječničke raznolikosti mjereni mjerom D rastu s porastom kronološke dobi. Skupina četverogodišnjaka pokazuje najniže vrijednosti ($M=25,35$), skupina šestogodišnjaka nešto više ($M=37,64$), skupina osmogodišnjaka još više ($M=42,37$), a desetogodišnjaci imaju najviše vrijednosti od svih dobnih skupina ($M=47,91$).

Tablica 6. Rezultati na varijabli mjera D za četiri skupine sudionika

skupina	<i>n</i>	<i>M</i>	minimum	maksimum	SD
četverogodišnjaci	36	35,35	71,24	35,35	12,127
šestogodišnjaci	42	37,64	89,78	37,64	14,202
osmogodišnjaci	41	42,34	87,40	42,34	13,714
desetogodišnjaci	38	47,39	71,91	47,39	11,245

S obzirom na normalnu distribuciju te varijable, rabljena je ANOVA s post-testom koja je pokazala značajne razlike među skupinama ($p < 0,05$). *Post-hoc* testom Bonferrini pokazalo se da četverogodišnjaci na ovoj mjeri imaju značajno niže rezultate od desetogodišnjaka ($p < 0,05$) te da šestogodišnjaci imaju značajno niže rezultate od desetogodišnjaka ($p < 0,05$). Ostale razlike među skupinama nisu se pokazale značajnima.

4.1.5. Pomični prosječni omjer različenica i pojavnica

Deskriptivne vrijednosti pokazuju da rezultati rječničke raznolikosti mjereni pomičnim prosječnim omjerom različenica i pojavnica rastu s porastom kronološke dobi. Skupina četverogodišnjaka pokazuje najniže vrijednosti ($M=,89$), skupina šestogodišnjaka nešto više ($M=,90$), skupina osmogodišnjaka još više ($M=,92$), a desetogodišnjaci imaju najviše vrijednosti od svih dobnih skupina ($M=,93$).

Tablica 7. Rezultati na varijabli pomični prosječni omjer različenica i pojavnica za četiri skupine sudionika

skupina	<i>n</i>	<i>M</i>	minimum	maksimum	SD
četverogodišnjaci	36	,89	,79	,95	,037
šestogodišnjaci	42	,90	,80	,96	,029
osmogodišnjaci	41	,92	,84	,97	,023
desetogodišnjaci	38	,93	,86	,96	,020

Usporedba testom Kruskal-Wallis pokazala je da se skupine međusobno razlikuju ($p<,001$). *Post-hoc* analiza pokazuje da četverogodišnjaci imaju značajno niže vrijednosti od osmogodišnjaka ($p<,005$) i od desetogodišnjaka ($p<,001$). Šestogodišnjaci također imaju značajno niže vrijednosti od osmogodišnjaka ($p<,05$) i desetogodišnjaka ($p<,001$). Značajno se ne razlikuju četverogodišnjaci od šestogodišnjaka, niti osmogodišnjaci od desetogodišnjaka.

Usporedba dobivenih značajnih razlika među skupinama na mjerama rječničke raznolikosti pokazuje da su sve mjere pokazale razlike među nekim dobnim skupinama. No, najveći je broj razlika pokazala mjera pomični prosječni omjer različenica i pojavnica (slika 1).

BRR	4 godine	6 godina	8 godina	10 godina
4 godine	-			*
6 godina		-		*
8 godina			-	
10 godina	*	*		-
ORP	4 godine	6 godina	8 godina	10 godina
4 godine	-			*
6 godina		-		*
8 godina			-	
10 godina	*	*		-
VOCD	4 godine	6 godina	8 godina	10 godina
4 godine	-			*
6 godina		-		*
8 godina			-	
10 godina	*	*		-
MATR	4 godine	6 godina	8 godina	10 godina
4 godine	-		*	*
6 godina		-	*	*
8 godina	*	*	-	
10 godina	*	*		-

Slika 1. Prikaz statistički značajnih razlika među skupinama na različitim mjerama rječničke raznolikosti

4.2. Rezultati korelacija mjera rječničke raznolikosti i proizvodnje

Kako bi se pokazala povezanost mjera rječničke raznolikosti s mjerom proizvodnje (ukupan broj riječi/pojavnica), provedena je Spearmanova korelacija. Pokazalo se da sve mjere značajno koreliraju s mjerom proizvodnje i to na sličan način. Broj različenica na 100 pojava, ORP na 100 pojava, mjera D i pomični prosječni omjer različenica i pojava pokazuju pozitivnu nisku korelaciju s mjerom proizvodnje ($\rho=0,221$, $p<0,05$; $\rho=0,221$, $p<0,05$; $\rho=,158$, $p<0,05$; $\rho=,227$, $p<0,05$). Dakle, s ukupnim brojem proizvedenih riječi, sve mjere rječničke raznolikosti blago rastu i po tome se međusobno malo razlikuju.

5. Rasprava

Analiza rječničke raznolikosti u analizi pisanih i govorenih jezičnih uzoraka ima dugu tradiciju. Mjere koje se pri tome upotrebljavaju pokazale su se pouzdanima u različitim vrstama istraživanjima. Različite mjere rječničkoga bogatstva pokazale su se dobrima u razlikovanju djece različite kronološke dobi (pregled: Malvern i dr. 2004), djece s jezičnim poremećajima (primjerice, Mäkinen i dr. 2014), odraslih osoba s jezičnim poremećajem (primjerice, Fergadiotis i Harris Wright 2011), govornika inoga jezika na različitim stupnjevima jezičnoga znanja te u određivanju dominantnosti jezika dvojezične djece (Hržica i Roch 2020) i odraslih govornika (Daller i dr. 2003).

Od mnogih mjera rječničke raznolikosti, u istraživanjima su se najčešće rabile mjere omjer različnica i pojavnica te broj različnica, kojima se zamjera ovisnost o duljini transkripta. Validacijska su istraživanja mjera rječničke raznolikosti koje su predložene kao alternativa tim mjerama najviše provođena na pisanim tekstovima te na jezicima koji su morfološki siromašniji od hrvatskoga jezika. Također, istraživanja se razlikuju u preporukama mjera (pregled: Fergadiotis i dr. 2015).

U ovom se istraživanju pošlo od pitanja prikladnosti mjera za razlikovanje rječničke raznolikosti djece različite kronološke dobi. Sukladno prijašnjim istraživanjima, pretpostavljalo se da će MATTR te D biti osjetljivije od ORP-a na razlike u rječničkoj raznolikosti među djecom različite dobi. Kada su se ORP i broj različnica mjerili sukladno uputama o ujednačavanju duljine uzorka, to jest na jednakom broju pojavnica, pokazalo se da mogu razlikovati rječničku raznolikost četverogodišnjaka od desetogodišnjaka te šestogodišnjaka od desetogodišnjaka. Dakle, obje su mjere razlikovale dvije najmlađe skupine od najstarije skupine. Sličan je rezultat pokazala i mjera D, koja se u prijašnjim istraživanjima preporučivala kao pouzdanija od prvih dviju mjera. Tek je posljednja mjera, MATTR, pokazala da se prema rječničkoj raznolikosti četverogodišnjaci razlikuju od osmogodišnjaka i desetogodišnjaka te da se šestogodišnjaci razlikuju od osmogodišnjaka i desetogodišnjaka. Dakle, pokazalo se da tri mjere razlikuju na sličan način dobne skupine govornika s obzirom na rječničku raznolikost, a četvrta pokazuje još veću osjetljivost.

Osim o tome koje mjere mogu odraziti razvojne razlike, postavilo se i pitanje koje su mjere povezane s produktivnošću, to jest dužinom transkripta. Prijašnja

su istraživanja kao glavni argument nepouzdanosti mjera kao što su ORP ili broj različenica navodili njihovu negativnu korelaciju s mjerama produktivnosti, prije svega brojem pojavnica te je to bila i početna pretpostavka ovoga istraživanja. No, pokazalo se da niti jedna mjera nije negativno povezana s duljinom transkripta. Štoviše, sve su mjere pokazale značajnu pozitivnu nisku korelaciju. Dakle, pokazalo se da je duljina transkripta u pojavnicama povezana s rječničkom raznolikošću mjerenom sa sve četiri mjere.

Nameće se pitanje zbog čega se pokazalo da i kritizirane mjere mogu pokazati razlike u rječničkoj raznolikosti te zašto je njihova povezanost s duljinom uzorka slična onoj drugih mjera. Nekoliko je razloga. Prvo, istraživanja koja su pokazale veliku ovisnost ORP-a i mjera proizvodnje nisu uvijek ORP izračunavala na ujednačenim uzorcima. Također, ponekad su uzorci bili ujednačavani, ali ne prema broju pojavnica, već prema broju iskaza (primjerice: Thordardottir i Ellis Weismer 2001). Važno je napomenuti da novije mjere rječničke raznolikosti ne teže samo usporedbi jezičnih uzoraka koji su proizvedeni na temelju ujednačenih poticaja (primjerice, sličnih slikovnih predložaka), već teže dovoljnoj robusnosti da mogu uspoređivati transkripte koji nisu skraćivani (ujednačavani po duljini) te koji se tematski i duljinom razlikuju. Ovo je istraživanje ograničeno na jezične uzorke pažljivo vođene slikovnim predloškom, dakle, slične tematski te po duljini. No, potrebno je uzeti u obzir da i za takve ujednačene uzroke neke mjere treba dodatno ograničiti (NDW i ORP), dok neke daju jednake ili bolje rezultate bez potrebe za ujednačavanjem uzoraka. Zbog toga, unatoč sličnim rezultatima u usporedbi mjera, može se zaključiti da je u potrebi automatiziranoga mjerenja mjera rječničke raznolikosti, bolje uzeti u obzir mjere koje ne zahtijevaju kraćenje jezičnih uzoraka. Naime, kraćenjem se gube informacije, što je posebno važno u istraživanjima jezičnoga usvajanja, gdje su uzorci govorenoga jezika često kratki. Također, ako uspoređujemo jezične uzorke govornika različite dobi, morat ćemo ih skratiti prema uzorcima mlađih ispitanika te tako usporediti, primjerice, čitavu priču koju je ispričao mlađi ispitanik s dijelom priče starijega ispitanika. Od dviju mjera koje ne treba ograničavati izračunom na fiksnom broju riječi, MATTR se pokazao boljim u razlikovanju djece različite dobi. Sličan su rezultat u usporedbi osoba urednoga jezičnog razvoja i osoba s razvojnim jezičnim poremećajem dobile Hržica i dr. (2019): iako su i mjera D i MATTR uspješno razlikovali skupine, veličina učinka za MATTR bila je nešto veća.

6. Zaključak

Ovo istraživanje pokazalo je da se rječnička raznolikost u jezičnom razvoju može mjeriti različitim mjerama i da one mogu pokazati razlike među dobnim skupinama, uz praćenje metodoloških preporuka primjene tih mjera. U slučaju potrebe za usporedbom jezičnih uzoraka različitih duljina, iz rezultata proizlazi da je najbolje odabrati mjeru MATTR. Istraživanja kojima se ispitivao odnos različitih mjera rječničke raznolikosti najčešće su provedena na uzorcima pisanoga jezika te u kontekstu istraživanja usvajanja drugoga jezika. U ovom radu naglasak je stavljen na istraživanja prvog jezika i uzorke govorenog jezika. Također, ranija su istraživanja redovito provedena na većim svjetskim jezicima. U kontekstu proučavanja usvajanja hrvatskog jezika, ovaj rad pripada među malo-brojne usredotočene na rječničku raznolikost, posebice uzevši u obzir različite mjere. Ono na što se nije odgovorilo, a svakako bi bila zanimljiva tema budućih istraživanja, jest kako karakteristike hrvatskoga jezika (primjerice, morfološko bogatstvo) utječu na mjere rječničke raznolikosti, točnije, u kojoj mjeri one odražavaju rječničko bogatstvo, a u kojoj i, primjerice, morfosintaktički razvoj.

Zahvala

Ovaj rad sufinancirala je Hrvatska zaklada za znanost projektom Višerazinski pristup govornom diskursu u jezičnom razvoju (UIP-2017-05-6603).

Literatura

BEDORE, LISA M.; PENA, ELIZABETH D.; GILLAM, RONALD B.; HO, TSUNG HAN. 2010. Language sample measures and language ability in Spanish English bilingual kindergarteners. *Journal of Communication Disorders* 43/6. 498–510. doi.org/10.1016/j.jcomdis.2010.05.002.

BERMAN, RUTH; VERHOEVEN, LUDO. 2002. Cross-linguistic perspectives on the development of text-production abilities: speech and writing. *Written Language and Literacy* 5. 1–44.

BERMAN, RUTH; SLOBIN, DAN. J. 1994. *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Lawrence Erlbaum Associates. Hillsdale.

- BERMAN, RUTH. 2000. Thematic perspectives on how children talk about interpersonal conflict. *Developing literacy in different contexts and different languages*, Ur. M. Aparici, N. Argerich, E. Rosado, L. Tolchinsky. Working Papers 3. University of Barcelona. 65–76.
- CARROLL, JOHN B. 1964. *Language and thought*. Prentice-Hall. Englewood Cliffs, NJ.
- COVINGTON, MICHAEL A.; MCFALL, JOE D. 2010. Cutting the Gordian knot: The moving-average type-token ratio MATTR. *Journal of Quantitative Linguistics* 17/2. 94–100. doi.org/10.1080/09296171003643098.
- CRYSTAL, DAVID. 1997. *The Cambridge encyclopedia of language*. Cambridge University Press. Cambridge.
- DALLER, MICHAEL HELMUT; MILTON, JAMES; TREFFERS-DALLER, JEANINE. 2007. *Modelling and Assessing Vocabulary Knowledge*. Cambridge University Press. Cambridge.
- DALLER, HELMUT.; VAN HOUT, ROELAND; TREFFERS-DALLER, JEANINE. 2003. LEXICAL richness in the spontaneous speech of bilinguals. *Applied Linguistics* 24/2. 197–222. doi.org/10.1093/applin/24.2.197.
- DUNN, LEOTA M.; DUNN, L. C.; KOVAČEVIĆ, MELITA; PADOVAN, NEVENA; HRŽICA, GORDANA; KUVAČ KRALJEVIĆ, JELENA; MUSTAPIĆ, MAJA; DOBRAVAC, GORDANA; PALMOVIĆ, MARIJAN. 2009. *Peabody slikovni test rječnik PPVT – III – HR*. Naklada Slap. Zagreb – Jastrebarsko.
- DROMI, ESTHER. 1999. Early lexical development. *The development of language*. Ur. Barrett, Martyn. Psychology Press. Hove. 99–131.
- FERGADIOTIS, GERASIMOS; HARRIS WRIGHT, HEATHER. 2011. Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology* 25/11. 1414–1430. doi.org/10.1080/02687038.2011.603898.
- FERGADIOTIS, GERSIMOS; WRIGHT HARRIS, HEATHER; GREEN, SAMUEL. 2015. Psychometric evaluation of lexical diversity indices: Assessing length effects. *Journal of Speech, Language, and Hearing Research* 58/3. 840–852. doi.org/10.1044/2015_JSLHR-L-14-0280.
- FEY, MARC E.; CATTS, HUGH W.; PROCTOR-WILLIAMS, KERRY; TOMBLIN, BRUCE J.; ZHANG, XUYANG. 2004. Oral and written story composition skills of children with language impairment. *Journal of Speech, Language, and Hearing Research* 47/6. 1301–1318. doi.org/10.1044/1092-4388(2004/098).
- RETCHER, PAUL; MACWHINNEY, BRIAN. 1995. *The handbook of child language*. Blackwell. Oxford.
- GAYRAUD, FREDERIQUE. 2000. *Le développement de la différenciation oral/écrit vu d travers le lexique*. Science du Langage. Université Lumière – Lyon 2.
- GUIRAUD, PIERRE. 1960. *Problèmes et méthodes de la statistique linguistique*. Presses Universitaires de France. Paris.

HRŽICA, GORDANA; ROCH, MAJA. 2020. Lexical diversity in bilingual speakers of Croatian and Italian. *LITMUS in Action: Crosscomparison studies across Europe*. Ur. Armon-Lotem, Sharon; Grohmann, Kleantes. John Benjamins Publishing Company Trends in Language Acquisition Research (TILAR). Amsterdam.

HRŽICA, GORDANA; KOŠUTAR, SARA; KRAMARIĆ, MATEJA. 2019. Rječnička raznolikost pisanih tekstova osoba s razvojnim jezičnim poremećajem. *Hrvatska revija za rehabilitacijska istraživanja* 55/2. 14–30. doi.org/10.31299/hrri.55.2.2.

JACOBSON, PEGGY F.; WALDEN, PATRICK R. 2013. Lexical diversity and omission errors as predictors of language ability in the narratives of sequential Spanish-English bilinguals: A crosslanguage comparison. *American Journal of Speech-Language Pathology* 22. 554–565. doi.org/10.1044/1058-0360(2013/11-0055).

JARVIS, SCOTT. 2002. Short texts, best fitting curves, and new measures of lexical diversity. *Language Testing* 19/1. 57–84. doi.org/10.1191/0265532202lt220oa.

JOHANSSON, VICTORIA. 2008. Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund University Department of Linguistics and Phonetics Working Papers* 53. 61–79.

JOHNSON, WENDELL. 1944. Studies in language behavior: A program of research. *Psychological Monographs* 56. 1–15.

KELIĆ, MAJA; HRŽICA, GORDANA; KUVAČ KRALJEVIĆ, JELENA. 2012. Mjere jezičnog razvoja kao pokazatelji posebnih jezičnih teškoća. *Hrvatska revija za rehabilitacijska istraživanja*, 48/2. 23–40.

KETTUNEN, KIMMO. 2014. Can Type-Token Ratio be Used to Show Morphological Complexity of Languages? *Journal of Quantitative Linguistics* 21/3. 223–245. doi.org/10.1080/09296174.2014.911506.

KLEE, THOMAS. 1992. Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders* 12. 28–41.

KÖHLER, REINHARD; GALLE, MATTHIAS. 1993. Dynamic aspects of text characteristics. *Quantitative text analysis*. Ur. L. Hrebicek i G. Altmann. Wissenschaftlicher Verlag Trier. Trier. 46–53.

KOVAČEVIĆ, MELITA; JELASKA, ZRINKA; KUVAČ KRALJEVIĆ, JELENA; CEPANEC, MAJA. 2007. *Komunikacijske razvojne ljestvice (KORALJE)*. [Test].

KUVAČ, JELENA; PALMOVIĆ, MARIJAN. 2007. *Metodologija istraživanja dječjeg jezika*. Naklada Slap. Jastrebarsko.

MAAS, HEINZ-DIETER. 1972. Zusammenhang zwischen Wortschatzumfang und Länge eines Textes. *Zeitschrift für Literaturwissenschaft und Linguistik* 8. 73–79.

MACWHINNEY, BRIAN. 2000. *The CHILDES Project. Tools for analyzing talk*. Lawrence Erlbaum Associates. Mahwah, NJ.

- MÄKINEN, LEENA; LOUKUSA, SOILE; NIEMINEN, LEA, P., LEINONEN, EEVA; KUNNARI, SARI. 2014. The development of narrative productivity, syntactic complexity, referential cohesion and event content in four- to eight year-old Finnish children. *First Language* 34/1. 24–42. doi.org/10.1177/0142723713511000.
- MALVERN, DAVID; RICHARDS, BRIAN; CHIPERE, NGONI; DURÁN, PILAR. 2004. *Lexical diversity and language development*. Palgrave Macmillan. New York.
- MALVERN, DAVID; RICHARDS, BRIAN. 1997. A new measure of lexical diversity. *Evolving models of language. Papers from the Annual Meeting of the British Association of Applied Linguists*. Ur. A. Ryan; A. Wray. Clevedon: Multilingual Matters. 58–71.
- MAYER, MERCER. 1969. *Frog, where are you?* Dial Books for Young Readers. New York.
- MCKEE, GERARD; MALVERN, DAVID; RICHARDS, BRIAN. 2000. Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing* 15/3. 323–337. doi.org/10.1093/lc/15.3.323.
- MÉNARD, NATHAN. 1983. *Mesure de la richesse lexicale*. Slatkine. Geneva.
- NIPPOLD, MARILYN. 2002. Lexical learning in school-age children, adolescents, and adults: A process where language and literacy converge. *Journal of Child Language* 29/2. 449–488. doi.org/10.1017/S0305000902275340.
- READ, JOHN. 2000. *Assessing vocabulary*. Cambridge University Press. Cambridge.
- SIMON-CEREJIDO, GABRIELA; GUTIÉRREZ-CLELLEN, VERA. 2009. A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics* 30/2. 315–337. doi.org/10.1017/S0142716409090134.
- STRÖMQVIST, SVEN; JOHANSSON, VICTORIA; KRIZ, SARAH; RAGNARSDOTTIR, HRAFNHILDUR; AISENMAN, RAVID; RAVID, DORIT. 2002. Toward a crosslinguistic comparison of lexical quanta in speech and writing. *Written Language and Literacy* 5. 45–67.
- THORDARDOTTIR, ELIN T.; ELLIS WEISMER, S. 2001. High frequency verbs and verb diversity in the spontaneous speech of school-age children with specific language impairment. *International Journal of Language and Communication Disorders* 36. 221–44.
- TREFFERS-DALLER, JEANINE; KORYBSKI, TOMASZ. 2016. Using lexical diversity measures to operationalise language dominance in bilinguals. *Language dominance in bilinguals: Issues of measurement and operationalization*. Ur. C. Silva-Corvalan; J. Treffers-Daller, J. Cambridge University Press. Cambridge. 106–123.
- TRTANJ, IVANA; KUVAC̃ KRALJEVIĆ, JELENA. 2017. Jezična i govorna obilježja dječjega pripovjednog diskursa: analiza na mikrostrukturnoj razini. *Govor: časopis za fonetiku* 34/1. 53–69. doi.org/10.22210/govor.2017.34.03.
- UCCELLI, PAOLA; PAEZ, MARIELA M. 2007. Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associ-

ations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools* 38/3. 225–236. doi.org/10.1044/0161-1461(2007/024).

VERMEER, ANNE. 2000. Coming to grips with lexical richness in spontaneous speech data. *Language Testing* 17. 65–83.

WATKINS, RUTH V.; KELLY, DONNA J.; HARBERS, HEIDI M.; HOLLIS, WENDY. 1995. Measuring children's lexical diversity: differentiating typical and impaired language learners. *Journal of Speech and Hearing Research* 38. 1349–1355. doi.org/10.1044/jshr.3806.1349.

Measures of Lexical Diversity in Narratives of Preschool and Early School Age Children

Abstract

The development of vocabulary in the first language acquisition can be traced by analysing spoken language transcripts. Such language samples are obtained, for example, by eliciting a story by a wordless picture book. Measures of lexical diversity are used to assess vocabulary development. Two measures, namely the number of different words (NDW) and type-token ratio (TTR) have been criticized for their correlation with the length of transcripts. The recommendation is that a fixed number of words should be used to calculate them or that other measures, resistant to differences in length, should be used. Two of such measures are parameter D (D) and moving average type-token ratio (MATTR). However, there is no consensus in the literature about the most applicable measure of lexical diversity. Also, studies aiming to validate these measures are predominantly based on written language analysis of English language. Therefore, the goal of this study was to explore which measures of lexical diversity best reflect spoken language production of Croatian children of different age. In total 157 children were tested using a picture book *Frog, where are you?*. All four lexical diversity measures showed an increase in different age groups of children. Statistically significant differences were obtained for all measures. Three measures (NDW, TTR, D) differentiated two youngest age groups (4-year-olds and 6-year-olds) from the oldest (10-year-olds). However, only measure MATTR differentiated two youngest age groups from two oldest (8-year-olds and 10-year-olds).

Ključne riječi: usvajanje prvog jezika, razvoj rječnika, mjere rječničke raznolikosti

Keywords: first language acquisition, vocabulary development, measures of lexical diversity