

Jelena Kuvač Kraljević  
Gordana Hržica  
Vanja Štefanec

---

Čestotni rječnik hrvatskoga dječjeg jezika

Natuknice

© NAKLADA SLAP

Izdavač

»NAKLADA SLAP«  
Dr. Franje Tuđmana 33, 10450 Jastrebarsko  
www.nakladaslap.com

Direktor

Biserka Matešić

Glavni urednik

dr. sc. Krunoslav Matešić

© NAKLADA SLAP

Lektor

Žarko Taraš  
Petra Trumbetić Prša

Recenzenti

prof. dr. sc. Jasmina Ivšac Pavliša  
prof. dr. sc. Melita Kovačević  
prof. dr. sc. Nikola Ljubešić

Jelena Kuvač Kraljević  
Gordana Hržica  
Vanja Štefanec

Čestotni rječnik  
hrvatskoga dječjeg jezika

Natuknice



© NAKLADA SLAP



NAKLADA SLAP

Copyright © 2022. Naklada Slap

Sva prava pridržana.

Nijedan dio ove knjige ne smije se reproducirati ni prenositi ni u kakvom obliku niti ikakvim sredstvima, elektroničkim ili mehaničkim, fotokopiranjem, snimanjem ili umnožavanjem u bilo kojem informatičkom sustavu za pohranjivanje i korištenje bez prethodne suglasnosti vlasnika prava.

CIP zapis dostupan u računalnom katalogu  
Nacionalne i sveučilišne knjižnice u Zagrebu  
pod brojem 001117330.

ISBN 978-953-191-958-6

ISBN 978-953-191-959-3

Likovno i grafičko oblikovanje: Naklada Slap, veljača, 2022.

## O autorima

**Jelena Kuvač Kraljević** profesorica je na Odsjeku za logopediju Edukacijsko-rehabilitacijskog fakulteta Sveučilišta u Zagrebu. Diplomirala je na istom odsjeku, a magistrirala i doktorirala u području lingvistike na Filozofskom fakultetu Sveučilišta u Zagrebu. Znanstveni interesi su joj jezični razvoj te razvojni i stečeni jezični poremećaji kao što su razvojni jezični poremećaj, disleksija i afazija. Objavila je jednu samostalnu autorsku knjigu, jednu knjigu u suautorstvu te je urednica jedne stručnog priručnika. Objavila je 87 znanstvenih i stručnih radova. Suautorica je sedam testova za procjenu jezičnih sposobnosti kod djece i odraslih.

**Gordana Hržica** izvanredna je profesorica na Odsjeku za logopediju Edukacijsko-rehabilitacijskog fakulteta Sveučilišta u Zagrebu, a diplomirala je i doktorirala na Filozofskom fakultetu Sveučilišta u Zagrebu (Odsjek za hrvatski jezik i književnost i Odsjek za komparativnu književnost; Odsjek za lingvistiku). U znanstvenom radu najviše se bavi jezičnim usvajanjem jednojezičnih i dvojezičnih govornika uporabom niza metoda među kojima se ističe analiza korpusa. Objavila je 54 znanstvena rada u časopisima i knjigama. Suautorica je četiriju testova za procjenu jezičnih sposobnosti kod djece i odraslih.

**Vanja Štefanec** asistent je na Zavodu za lingvistiku Filozofskog fakulteta Sveučilišta u Zagrebu i magistar je informacijskih znanosti, informatičkog usmjerenja. Već desetak godina radi na brojnim europskim i domaćim projektima kao računalni programer, savjetnik za pitanja iz domene informacijsko-komunikacijske tehnologije te stručnjak za prikupljanje i obradu jezične građe.



© NAKLADA SLAP

# Sadržaj

Predgovor	9
Uvod	11
<b>I. dio: natuknice prema čestoti</b>	<b>19</b>
Natuknice razvrstane po čestoti	21
Natuknice razvrstane po abecedi	39
Natuknice razvrstane po vrstama riječi po čestoti	57
Natuknice razvrstane po vrstama riječi abecedno	77
<b>II. dio: natuknice prema dobi usvajanja</b>	<b>97</b>
Natuknice do 1;05 po čestoti	99
Natuknice do 1;05 abecedno	103
Natuknice od 1;06 do 2;00 po čestoti	107
Natuknice od 1;06 do 2;00 abecedno	117
Natuknice od 2;01 do 2;05 po čestoti	127
Natuknice od 2;01 do 2;05 abecedno	141
Natuknice od 2;06 do 3;00 po čestoti	155
Natuknice od 2;06 do 3;00 abecedno	167



© NAKLADA SLAP

© NAKLADA SLAP



# Predgovor

Čestotni rječnik hrvatskoga dječjeg jezika: natuknice i Čestotni rječnik hrvatskoga dječjeg jezika: morfološki i razvojni oblici jedna su cjelina s dva lica, onim definiranim leksičkim opisom riječi i onim određenim njezinim morfološkim oblicima.

Vokabular je temeljni dio jezika. Koliko je važan, govori i podatak da se ponekad količina riječi koju osoba poznaje izjednačava s njezinim jezičnim znanjem. Današnja eksperimentalna i klinička ispitivanja jezika gotovo je nemoguće provesti bez uključivanja i kontroliranja nekih obilježja riječi. Isto tako jezični razvoj nemoguće je proučavati bez obuhvaćanja i opisivanja djetetova receptivnog i ekspresivnog vokabulara, kvantitativno i kvalitativno. Osim u ovladavanju ostalim jezičnim sastavnicama, vokabular ima posebnu ulogu u učenju čitanja i pisanja u školskim danima.

Leksički i morfološki opisi dvaju dijelova čestotnog rječnika proizlaze iz istog izvora jezičnih podataka – Hrvatskog korpusa dječjeg jezika (HKDJ, Kovačević, 2002) – koji pruža uzorke spontane konverzacije troje djece s njihovom neposrednom okolinom. Štoviše, moglo bi se reći da su ovi rječnici prirodan nastavak višegodišnjeg rada na prikupljanju i izradi HKDJ-a. Izrada korpusa dječjeg jezika uključuje redovito i višegodišnje snimanje djetetova jezičnog ponašanja od njegove prve do otprilike treće godine života (dva do tri puta mjesečno), transkripciju i kodiranje podataka, ujednačavanje i provjeru te pripremu korpusa za objavljivanje. Kroz ovaj sažet opis vidljivo je da se rad na čestotnom rječniku nastavlja na rad kolega koji su u različitom opsegu dali svoj doprinos u izradi HKDJ-a u okvirima različitih projekata Laboratorija za psiholingvistička istraživanja Odsjeka za logopediju Edukacijsko-rehabilitacijskog fakulteta Sveučilišta u Zagrebu. To su Maja Anđel, Draženka Blaži, Blaženka Brozović, Antigona Katičić, Petra Makanec, Antonia Ordulj, Marijan Palmović i Jan Šnajder. Zahvaljujemo im na predanom radu, koji je omogućio izradu ovih rječnika. Posebna zahvala Meliti Kovačević na ustupanju HKDJ-a za izradu ovog rječnika te na viziji o važnosti njegova razvoja.

Rječnik je nastao obradom Hrvatskog korpusa dječjeg jezika metodama računalne lingvistike i jezičnim tehnologijama od kojih su neke razvijene u sklopu projekta *Računalni asistent za podršku pri unosu teksta osobama s jezičnim poremećajima* (EU - Strukturni fond; RC.2.2.08-050) u okviru Poziva Jačanje kapaciteta za istraživanje, razvoj i inovacije (2014-2016). Naime, riječi se u konverzaciji pojavljuju u različitim oblicima. Postupak koji se naziva lematizacija omogućava da se različiti oblici vezuju uz zajednički osnovni, tj. kanonski, oblik riječi – lemu ili natuknicu. Za glagole taj je oblik jednak infinitivu, za imenice nominativu jednine, za pridjeve nominativu jednine muškoga roda itd. Na taj se način, primjerice, glagolski oblici «snimam» i «snimamo» povezuju s natuknicom «snimati», imenički oblici «kamionu» i «kamiona» s natuknicom «kamion», a pridjevski oblici «lijepa» i «lijepo» s natuknicom «lijep». Osim toga, morfološkim se oblicima riječi dodjeljuju određene morfosintaktičke kategorije. Ovaj se postupak naziva morfosintaktičko označavanje. Tako

dobiveni podatci usustavljeni su, razvrstani i prikazani na različite načine (primjerice, po abecedi, čestotno, prema određenom razdoblju...). Prikazi su oblikovani tako da mogu što lakše odgovoriti na osnovna pitanja koja bi mogao postaviti istraživač dječjega jezika ili stručna osoba kojoj su u radu nužni podatci o dječjem jeziku.

Motivacija za izradu rječnika proizlazi iz naših istraživanja. Puno smo se puta našli u potrebi za brzim i jednostavnim dohvatom podataka o čestoti riječi u dječjem jeziku. Izrada bilo kakvih eksperimentalnih jezičnih zadataka, dijagnostičkih testova, materijala za terapiju pa i bilo kakvog obrazovnog teksta namijenjenog djeci podrazumijeva kontrolu čestote riječi koje su u njih uključene. Ovi su podatci sadržani u korpusima, ali iz njih nisu lako dohvatljivi, niti su razvrstani na načine prikladne za brzu uporabu. Ubrzo smo uvidjeli da bi od takvog rada mogla prosperirati i šira znanstvena i stručna zajednica. No, pripremiti podatke tako da su prikladni za široku primjenu zahtijevalo je posebna znanja. Stoga su u pripremi rječnika od samih početaka uključeni stručnjaci za računalnu lingvistiku, a metode i alati računalne lingvistike omogućili su prikladno označavanje i razvrstavanje korpusnih podataka. Dakle, ovi su rječnici plod petogodišnje interdisciplinarnе suradnje autora različitih znanja i vještina koji su oba dijela rječnika uspješno sjedinili viziju potrebnu za daljnji objektivni rad u području logopedije i lingvistike. Oba rječnika nastala su u sklopu dvaju projekata: *Računalni asistent za podršku pri unosu teksta osobama s jezičnim poremećajima* (EU - Strukturni fond, RC.2.2.08-050) u okviru Poziva Jačanje kapaciteta za istraživanje, razvoj i inovacije (2014-2016) te *Višerazinski pristup govornom diskursu u jezičnom razvoju* (UIP-2017-05-6603) Hrvatske zaklade za znanost.

U rječnicima smo se trudili predvidjeti sve potrebe koje biste vi, dragi čitatelji i korisnici, mogli imati. Trudili smo se također i te podatke predstaviti na jednostavan, lako čitljiv i pregledan način. Stoga u uvodnom dijelu rječnika možete pronaći teorijske osnove na kojima su rječnici utemeljeni i upute kako ih upotrebljavati.

*Autori*



# Uvod

Ovo je rječnik o čestoti dječjih riječi, odnosno rječnik u kojem je iznesena lista riječi koje su proizvela djeca govornici hrvatskog jezika u svojoj stvarnoj uporabi jezika u dobi do 3. godine. Čestotni ili frekvencijski rječnici ili nekad zvani knjige riječi (Davies, Gardner, 2018) nisu novina u području leksikografije. Međutim, ono što razlikuje ovaj rječnik od ostalih je njegova ponešto drugačija namjera, a time i način njegova nastajanja. Čestotni rječnik hrvatskoga dječjeg jezika: natuknice nastao je na temelju stvarnih uzoraka govornog jezika troje djece govornika hrvatskog jezika u dobi od 13 do 28 mjeseci. Uzorci su sjedinjeni u Hrvatskom korpusu dječjeg jezika (Kovačević, 2002), koji sadrži oko 136 000 pojava. Korpus je dio svjetske baze dječjeg jezika CHILDES, s otvorenim pristupom i vrijedan je izvor lingvističkih podataka o ranom jezičnom razvoju.

## Razvoj i namjena čestotnih rječnika

Frekvencijski, odnosno čestotni rječnici dostupni su gotovo 100 godina. Istina, najčešće su izrađivani i naknadno nadopunjavani za engleski jezik (Thorndike, 1921, 1932) i to u području podučavanja i učenja engleskog kao drugog ili stranog jezika. U ediciji Routledge Frequency Dictionaries 2018. objavljeni su frekvencijski rječnici i za 12 drugih jezika poput španjolskog (Davies i Hayward Davies, 2018), perzijskog (Miller i Aghajanian-Stewart, 2018) ili turskog (Aksan, Aksan, Mersinli i Demirhan, 2018). Cilj te edicije bio je izraditi sržni rječnik od 5 000 najfrekventnijih riječi za nastavnike i učenike tih jezika. Za to je, kao i za sve druge vrste rječnika, bilo potrebno prethodno osigurati korpus, odnosno zbirku tekstova reprezentativnih za dani jezik, dijalekt ili bilo koji jezik namijenjen lingvističkoj analizi (Francis, 1992). U svim su tim jezicima razvijeni korpusi koji su prelazili nekoliko milijuna riječi i uključivali su pisani i govoreni jezik. Preporuka za izradu čestotnih rječnika je da korpus ne smije biti manji od 500 000 riječi jer što je opseg manji, manji je broj najčešćih riječi koje se mogu rangirati tako da odgovaraju stvarnosti (Lončarić, 1977). Korpus Hrvatskog dječjeg jezika (Kovačević, 2002), iz kojeg je izrađen ovaj rječnik, sadrži dječje iskaze, ali i iskaze koje su proizveli njihovi odrasli sugovornici. Dječji iskazi ukupno sadrže nešto manje od 78 000 riječi. To znači da korpus ne zadovoljava danu preporuku. Međutim, ovdje treba naglasiti da je riječ o specijaliziranom dječjem jeziku i ograničenom razdoblju uzorkovanja. Specifični korpusi su uvijek opsegom manji jer pokrivaju jedan uski raspon populacije (vidi više u Kuvač Kraljević i sur., 2016).

Kada se razviju dovoljno veliki korpusi, i to upravo oni koji uključuju različite jezične modalitete, moguće je dobiti pouzdane podatke o čestoti riječi u tom jeziku, odnosno razviti sržni vokabular (eng. *core vocabulary*). Prema Nation (1990), 4 000 do 5 000 najčešćih riječi u engleskom čini čak 95% pisanog tekstova, a 1 000 najčešćih riječi 85% govorenog jezika. To znači da će istraživač ili kliničar, ako se oslanja na čestotni rječnik u procesu poučavanja

novih riječi, potaknuti da osoba nauči one riječi koje će moći rabiti u većini komunikacijskih situacija.

U hrvatskoj leksikologiji čestotni rječnici nisu tako uobičajeni. Prvi je rječnik napravio Ivan Furlan u svojoj disertaciji *Raznolikost rječnika. Struktura govora* (1961), koji se temelji na pisanim neprofesionalnim tekstovima djece u osnovnoj školi. U deset školskih sati djeca su pisala sastavak na deset različitih tema i na taj način je dobio podatke o broju pojavnica koje djeca s obzirom na dob proizvode. Na temelju tog rječnika Furlan je donio zaključke o razvoju vokabulara djece školske dobi, ali što je još bitnije, definirao prvih stotinu najčešćih riječi i usporedio ih s rječnikom odraslih (Kuvač, Palmović, 2007). Nažalost, rječnik od skoro 3 000 riječi nikada nije objavljen kao zasebna publikacija, nego samo kao dio disertacije u djelu *Lista od blizu tri tisuće najčešćih riječi pisanog hrvatskog ili srpskog jezika*. Međutim, kao što to i jest smisao razvoja frekvenzijskih rječnika, i ovaj je imao praktičnu implikaciju. Naime, na temelju informacija o čestoti riječi djece školske dobi Furlan je razvio *Jednominutni ispit čitanja* (1965) za djecu nižih razreda osnovne škole. Test čini lista od 120 riječi, koje su poredane od kraćih i lakših prema težima i duljima. Rezultat se izražava kao broj točno naglas pročitanih riječi u jedinici vremena (jednoj minuti).

Prvo naručeno korpusno istraživanje u hrvatskom jeziku bio je frekvenzijski rječnik tekstova Vjesnika i Večernjega lista u opsegu od 130 000 pojavnica, koji je 1976. obradio Zorislav Šojat (Šojat, 1976; više o pregledu razvoja korpusa i korpusne lingvistike u Hrvatskoj može se pronaći u radovima Klobučar Srbić, 2008).

Prvi sveobuhvatniji i utemeljen na raznolikim testovima bio je Hrvatski čestotni rječnik (Moguš, Bratanić, Tadić, 1999). Rječnik je nastao na milijunskom korpusu, koji je strukturiran u 5 jednakih potkorpusa – proza, poezija, drama, udžbenički tekstovi i novine. Iako je ovaj korpus utemeljen najvećim dijelom na tekstovima profesionalnih pisatelja, on je danas uz hrLex (<http://hdl.handle.net/11356/1232>; Ljubešić, 2019) vrijedan izvor podataka o čestoti riječi u hrvatskom jeziku.

## Psiholingvistička obilježja riječi

Brzina kojom će se neka riječ usvojiti ili naučiti, osim o njezinim gramatičkim obilježjima i značenju, ovisi i o nizu psiholingvističkih obilježja te riječi koja čine njezina inherentna svojstva. Ta obilježja su, primjerice, predočivost, konkretnost/apstraktnost, čestota, dob usvajanja ili poznatost riječi (vidi više u Kuvač Kraljević, Olujić, 2019 i 2018). Psiholingvistička obilježja riječi odnose se na tri njezina svojstva: koncept (primjerice, predočivosti i konkretnosti), lingvističko svojstvo (primjerice, vrsta riječi ili fonološka složenost riječi) te razine njezine uporabe (primjerice, čestota ili dob usvajanja) (Lind i sur., 2015; Soares i sur., 2017).

Čestota riječi odnosi se na kvantifikaciju količine govornikove upotrebe neke riječi u pisanom ili govornom jeziku (Balota i sur., 2001). Oldfield i Wingfield (1965) prvi su uočili utjecaj ovog obilježja u jezičnoj obradi i to na zadatku imenovanja kad su ispitanici brže imenovali visoko nego nisko česte riječi. Ovo (psihol)lingvističko obilježje puno je više