

Gordana Hržica
Jelena Kuyač Kraljević
Vanja Štefanec

Čestotni rječnik hrvatskoga dječjeg jezika

Morfološki i razvojni oblici

© 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

Gordana Hržica
Jelena Kuvač Kraljević
Vanja Štefanec

Čestotni rječnik
hrvatskoga dječjeg jezika

Morfološki i razvojni oblici



© 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 001117331.

ISBN 978-953-191-958-6

ISBN 978-953-191-960-9

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

O autorima

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.

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.

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
I. dio: natuknice s obličnicama prema čestoti	31
Natuknice s obličnicama razvrstane po abecedi	33
Natuknice s obličnicama i individualnim oblicima razvrstane po abecedi	119
II. dio: natuknice prema dobi usvajanja	335
Natuknice s obličnicama do 1;05	337
Natuknice s obličnicama i individualnim oblicima do 1;05	349
Natuknice s obličnicama od 1;06 do 2;00	355
Natuknice s obličnicama i individualnim oblicima od 1;06 do 2;00	429
Natuknice s obličnicama od 2;01 do 2;05	457
Natuknice s obličnicama i individualnim oblicima od 2;01 do 2;05	567
Natuknice s obličnicama od 2;06 do 3;00	609
Natuknice s obličnicama i individualnim oblicima od 2;06 do 3;00	703



© 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, ujednačivanje 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

Svrha je rječnika *Čestotni rječnik hrvatskoga dječjeg jezika: morfološki i razvojni oblici* omogućiti uvid u morfološke oblike riječi koje upotrebljavaju djeca do treće godine života. Uz oblike i njihove morfološke opise navodi se i čestotnost svakog oblika. Rječnik je nastao 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 (HKDJ – Kovačević, 2002), koji sadrži oko 136 000 pojavnica. Korpus je dio svjetske baze dječjeg jezika CHILDES, s otvorenim pristupom i vrijedan je izvor lingvističkih podataka o ranom jezičnom razvoju.

Za potrebe izrade rječnika, korpus je iz CHAT formata, koji se koristi unutar TalkBank sustava, kojeg je CHILDES dio, pretvoren u standardni vertikalizirani format. Korpus je prvo strojno morfosintaktički označen i lematiziran korištenjem dostupnih jezičnih alata za hrvatski jezik (Agić i sur., 2013). Točnost je takve jezične obrade, zbog specifičnosti ovog registra i činjenice da su ti alati izgrađeni na osnovi pisanoga standardnoga jezika, bila očekivano niska. Zbog toga su cijeli korpus pregledali i ispravili računalni lingvisti specijalizirani za morfosintaktičko označavanje. Pritom, uzimajući u obzir to da se ovdje radi o specifičnom registru govornog jezika i činjenici da se radi o jeziku u procesu usvajanja, lingvisti-annotatori su pri označavanju pazili da morfološka oznaka neke riječi slijedi iz analize komunikacijske situacije te semantičkog i sintaktičkog konteksta riječi, a ne njezina pukog površinskog lika. Oznake koje se koriste za morfosintaktički opis, standardne su oznake unutar MULTEXT-East (Erjavec, 2012) specifikacija za označavanje korpusa. Oznake su slijedile revidiranu inačicu 4 (Ljubešić, 2013), temeljenu na službenoj inačici 4 MULTEXT-East specifikacije za hrvatski jezik (Tadić, 1998). Revidirana inačica 4 u praktički neizmijenjenom obliku postat će službena inačica 5 (Erjavec, Ljubešić, 2016).

Morfosintaktičke oznake unutar MULTEXT-East specifikacije sastavljene su od niza znakova od kojih prvi znak smješta oznaku u jednu od 14 kategorija koje definira MULTEXT standard, a one u nekoj mjeri odgovaraju vrstama riječi (Erjavec, 2012). Svaka od tih kategorija definira popis svojstava koje ima i vrijednosti koje ta svojstva mogu poprimiti. Vrijednosti svojstava prikazane su slovnim znakovima. Primjerice, imenice se opisuju s 5 svojstava, a to su *tip*, *rod*, *broj*, *padež* i *živost*. Svojstvo *tipa* može poprimiti dvije vrijednosti [opća (c) i vlastita (p)], *roda* tri [muški (m), ženski (f) i srednji (n)], *broja* dvije [jednina (s) i množina (p)], *padeža* sedam [nominativ (n), genitiv (g), dativ (d), akuzativ (a), vokativ (v), lokativ (l), instrumental (i)] i *živosti* dvije [ne (n) i da (y)]. To znači da će se obličnica riječi 'kamion' u akuzativu jednine označavati oznakom Ncmsan (imenica, opća, muški rod, jednina, akuzativ, neživo). Oznake su po naravi pozicijske, što znači da je redosljed svojstava određen. Svojstva se s kraja oznake mogu ispustiti ako nisu primjenjiva; na primjer, svojstvo živosti će se određivati samo kod imenica u akuzativu budući da u drugim padežima to svojstvo ne čini nikakvu razliku. Dakle, obličnica riječi 'kamion' u nominativu

jednine označavati će se kao Ncmsn, a ne Ncmsn. Međutim, neprimjenjivost svojstva unutar oznake mora se označiti crticom (-) radi očuvanja pozicije vrijednosti svojstava koja slijede. Glagoli su, primjerice, opisani sa šest svojstava: *tipom, glagolskim oblikom, licem, brojem, rodom i negativnošću*, pa se tako oblik glagolskog pridjeva 'kopao' glagola 'kopati' označava s Vmp-sm (glagol, glavni, particip, -, jednina, muški rod) budući da je svojstvo lica ovdje neprimjenjivo.

Čestotni rječnik hrvatskoga dječjeg jezika: morfološki i razvojni oblici odgovara na dva temeljna pitanja o dječjem jeziku: koje su osnovne morfosintaktičke kategorije usvojene u jeziku djeteta do treće godine te koliko je svaka od njih česta. Osim znanstvenicima, odgovori na ta pitanja potrebni su i velikom broju stručnjaka iz djelatnosti koje se vežu uz rad na dječjem jeziku. Podatci doprinose poznavanju tijeka jezičnoga razvoja, pružaju resurs za razvoj jezičnih testova i zadataka te za uspostavu orijentira u stručnome radu s dječjim jezikom.

Uvodni dio ovog rječnika sastoji se od četiri temeljna dijela. U prvom se dijelu ukratko prikazuje povijest rječnika s morfološkim prikazima te se objašnjavaju temeljni pojmovi: natuknica, obličnica, inačica obličnice, pojavnica. U drugom se dijelu razjašnjava kako raditi rječnik te se prikazuje njegova struktura. U trećem, posljednjem dijelu pokazuje se na koji se način pomoću rječnika može odgovoriti na relevantna znanstvena pitanja, to jest, prikazuje se primjena rječnika u znanstvenom i stručnom radu.

1. Svrha morfoloških rječnika i temeljni pojmovi

Čestotni rječnik hrvatskog dječjeg jezika: morfološki i razvojni oblici spada u rijetke čestotne rječnike s podacima o čestoti oblika riječi, a za razumijevanje svih podataka u njemu nužno je poznavati neke temeljne pojmove. U ovom se poglavlju ovaj rječnik smješta unutar sličnih jezičnih izvora te se objašnjavaju temeljni pojmovi važni za razumijevanje koncepta čestote.

1.1 Povijesni pregled: svrha i namjena morfoloških rječnika

Čestotni rječnici ili liste prikazuju podatke o svakoj riječi u nekom korpusu s brojem njezinog pojavljivanja (čestota). Uz to, uz svaku se riječ prikazuju i morfološki podatci, najčešće vrsta riječi i pripadajuće gramatičke kategorije (primjerice, rod za imenice). No, znatno rjeđe možemo doznati koji su sve oblici određene riječi pronađeni u nekom korpusu te koliko je određeni oblik (primjerice, padež) čest. Štoviše, u izradama čestotnih rječnika, morfološko bogatstvo nekog jezika otežava sjedinjavanje podataka o određenoj riječi. Ako je cilj prikazati ukupnu čestotu riječi, tada je potrebno sve oblike neke riječi pribrojiti ukupnoj čestoti. Iako za izvorne govornike s lingvističkim obrazovanjem to ne mora biti problematično, u vremenima prije računalne obrade tražilo je dodatno vrijeme u prikupljanju podataka (Derocher i sur., 1973). Za automatiziranu računalnu obradu problem je još veći: računalo prepoznaje oblike riječi (nizove istovrsnih odnosno različitih znakova), a svođenje tih oblika