

Dubravka Ivšić
Marina Bergovec
Institut za hrvatski jezik i jezikoslovlje
Ulica Republike Austrije 16, HR-10000 Zagreb
divsic@ihjj.hr, mbergov@ihjj.hr

VELIKI ETIMOLOŠKI RJEČNIK HRVATSKOGA JEZIKA I RAČUNALNA STRUKTURA NJEGOVE NATUKNICE

U sklopu projekta »Onomastička i etimologijska istraživanja hrvatskoga jezika« *Instituta za hrvatski jezik i jezikoslovlje* izrađuje se veliki etimološki rječnik hrvatskoga jezika. Rječnik će se oslanjati na opširnu Skokovu građu, a bit će obogaćen novijim etimološkim, onomastičkim i dijalektološkim znanjima. *Institut za hrvatski jezik i jezikoslovlje* nastoji uskladiti svoju leksikografsku djelatnost sa standardima za digitalizaciju leksičke građe. Veliki etimološki rječnik trebao bi također biti izrađen prema istim standardima. Postojeći modeli računalnog kodiranja leksičke građe uglavnom predlažu strukturu natuknice općih rječnika, a za specijalne rječnike (kao što je etimološki) potrebno ih je nadograđivati.

Ovim se radom najavljuje *Veliki etimološki rječnik hrvatskoga jezika* te se predstavlja računalna struktura natuknice toga rječnika. Problemi pri izradbi računalne strukture natuknice uglavnom proizlaze iz činjenice da svaka riječ ima svoju povijest čime je i njezina etimološka natuknica jedinstvena. Model strukture prikazat će se na odabranim primjerima iz novog etimološkog rječnika. Model računalne strukture nije konačan, već se daje znanstvenoj zajednici na razmatranje, ocjenu i nadopunu.

1. Uvod

Za hrvatski jezik postoje dva etimološka rječnika, četverosveščani *Etimologijski rječnik hrvatskoga ili srpskoga jezika* Petra Skoka i jednosveščani *Hrvatski etimološki rječnik* Alemka Gluhaka. U sklopu projekta »Ono-

mastička i etimologijska istraživanja hrvatskoga jezika« pri *Odjelu za onomastiku i etimologiju Instituta za hrvatski jezik i jezikoslovlje* izrađuju se dva nova hrvatska etimološka rječnika, jednosveščani (po uzoru npr. na Snovej slovenski ili Klugeov njemački etimološki rječnik) i višesveščani. Novi jednosveščani opći etimološki rječnik standardnoga jezika izrađuje prof. dr. sc. Ranko Matasović, a rječnik obrađuje naslijeđenu hrvatsku jezičnu građu. Novi višesveščani hrvatski etimološki rječnik izrađuje skupina suradnika pod vodstvom prof. dr. sc. Ranka Matasovića i prof. dr. sc. Dunje Brozović Rončević¹. Taj će rječnik sadržavati opširnu Skokovu građu, osuvremenjenu i obogaćenu novim etimološkim, dijalektološkim i onomastičkim podacima. Nije potrebno isticati da su oba ta rječnika nasušna potreba hrvatske leksikografije jer postojeći (više) ne zadovoljavaju potrebe ni znanstvene ni šire javnosti.

Skokov *Etimologijski rječnik hrvatskoga ili srpskoga jezika* opširan je rječnik visoke znanstvene vrijednosti i predstavlja temelj hrvatske (i srpske) etimologije. Skok je prikupio obilje građe (etimološke, dijalektološke, onomastičke, tvorbenne), koliko iz hrvatskih (i srpskih) govora, toliko i iz okolnih jezika (posebno balkanskih), no nažalost preminuo je prije nego je prikupljenu građu stigao uobličiti u rječnik. Desetak godina poslije njegove smrti Skokov rukopis, 10000 rukom pisanih stranica, za tisak je priredio Valentin Putanec. Ne možemo znati kako bi taj rječnik izgledao da je Skok poživio dovoljno dugo da ga uredi, ali je sigurno da mnogi nedostaci proizlaze upravo iz te nedovršenosti. Danas, više od pola stoljeća nakon Skokove smrti i gotovo četiri desetljeća nakon objavljivanja rječnika, mnogi su dijelovi toga rječnika zastarjeli: od znanstvenih spoznaja (mnoge tadašnje praslavenske i indoeuropske rekonstrukcije nadvladane su) preko građe (u međuvremenu je prikupljeno mnoštvo dijalekatske i onomastičke građe, objavljen je niz dijalektoloških rječnika te Babićeva *Tvorba riječi*) do same koncepcije rječnika. Tih desetak tisuća natuknica složenih po »etimološkim familijama« (Skok 1956:20) zapravo su skup iznimno vrijednih i podacima bogatih lingvističkih eseja, razumljivih uglavnom užem krugu stručnjaka, dok će se širi krug zainteresiranih čitatelja teško snaći u tom mnoštvu podataka. S obzirom na to da rječnik nije oblikovan po leksikografskim načelima, bez četvrte knjige kazala on je gotovo neupotrebljiv.

Gluhakov jednosveščani etimološki rječnik zamišljen je kao »informativni rječnik namijenjen širem čitateljstvu« (Gluhak 1993), međutim mnogo dublja jezična prošlost (nostratička), koja se u njemu donosi, otvara

¹ Uz urednike na leksikografskoj obradbi trenutačno rade Dubravka Ivšić i Tijmen Pronk.

puno više pitanja nego što daje jednoznačnih odgovora (Vinja 1998:6).²

Gotovo istodobno s objavljivanjem sada već klasičnog Skokova rječnika javila se i potreba za njegovom nadgradnjom. Predlagani su različiti tipovi nadgradnje, od objavljivanja »malog Skoka« pa sve do izradbe novog etimološkog rječnika (Vinja 1998:7). U nedostatku znanstvenih i organizacijskih potencijala, a u želji za kontinuitetom etimološke znanosti, pokrenuto je sastavljanje *Dopuna Skokovu rječniku*. Objavljene su romanističke dopune Skokovu rječniku — iznimno vrijedne i iscrpne *Jadranske etimologije* Vojmira Vinje u tri sveska. No, unatoč navedenim dopunama, potreba za izradbom novih etimoloških rječnika hrvatskoga jezika i dalje postoji.

Predradnja velikom višesveščanom etimološkom rječniku hrvatskoga jezika trebalo je biti dopunjeno i temeljito prerađeno izdanje Skokova rječnika. Taj je zadatak osmišljen u *Institutu za hrvatski jezik i jezikoslovlje* u okviru projekta »Onomastička i etimologijska istraživanja hrvatskoga jezika« te je bilo predviđeno da se Skokov rječnik dopuni, osuvremeni i ustroji po strožim leksikografskim načelima. »Novi Skok« trebao je biti etimološki rječnik »srednjojužnoslavenskog dijasistema« s naglaskom na hrvatski jezik. Zamisljeno je bilo da se Skokov rječnik nadopuni novom hrvatskom dijalektološkom i onomastičkom građom, s tim da se samostalne onomastičke natuknice izdvoje, zatim da se Skokove etimologije isprave, a rekonstrukcije osuvremene, da se njegove esejističke natuknice razdvoje na više manjih natuknica te da se sve natuknice oblikuju na isti način. U prvoj fazi rada cijeli je Skokov rječnik skeniran, optički pročitani, korigiran i djelomično kroatiziran. Izdvojene su samostalne onomastičke natuknice, označena su problematična mjesta, neke natuknice razdvojene su na manje. Započelo se i s okvirnom podjelom natuknica po podrijetlu na naslijeđene i posuđene riječi (romanizme, orijentalizme te ostale posuđenice) koje bi u sljedećoj fazi rada stručnjaci za pojedine jezike (romaniستي, orijentalisti, kroatisti, indoeuropeisti i drugi) nadopunili i osuvremenili. Međutim, već u pripremljenoj fazi rada postalo je jasno da je ponovno izdavanje Skokova etimološkog rječnika na opisan način zadatak koji ne bi dao zadovoljavajući rezultat, te se zbog toga započelo raditi na sasvim novom *Velikom etimološkom rječniku hrvatskoga jezika* koji će se, naravno, naslanjati na iznimno vrijednu Skokovu građu.

U sljedećim redcima opisat ćemo u osnovnim crtama kakvim je taj novi rječnik zamišljen, te predložiti moguću računalnu strukturu natuknice tog rječnika.

² Za detaljan prikaz Gluhakova rječnika v. Brozović-Rončević 1994.

2. Općenito o etimologiji i etimološkom rječniku

Etimološki je rječnik, u širem smislu, svaki rječnik koji daje etimološku informaciju o natuknici. U užem smislu, etimološki je onaj rječnik u kojem je etimološka informacija dominantna. Etimološki opisati riječ znači dati informaciju o postanku izraza riječi, o razvoju značenja riječi prema korijenu riječi te otkriti koliko se i na koji način ono razlikuje od današnjega. Promjene je moguće uočiti na svim jezičnim razinama: fonetskoj, gramatičkoj (morfološkoj, tvorbenoj itd.) te semantičkoj. No, etimološki opis riječi često podrazumijeva uključivanje mnoštva drugih informacija osim onih o razvoju izraza riječi. Za etimologiju neke riječi može biti važno, primjerice, kada je prvi put zabilježena, odnosno kada je ušla u leksik, iz kojih izvora, u kakvu tipu teksta, u kojem dijalektu ili društvenoj kategoriji.

Etimologija u svojoj analizi razlikuje načelno dva tipa riječi: naslijeđene i posuđene. Za naslijeđene su riječi važne razvojne promjene mlađeg jezika u odnosu na stariji te usporednice s drugim srodnim jezicima, dok je za posuđenice važan kontakt govornika jezika primaoca i jezika davaoca te fonološka, morfološka i semantička prilagodba jeziku primaocu. Za neke se riječi etimologija može utvrditi prilično pouzdano, za druge postoji nekoliko podjednako vjerojatnih etimologija, od kojih je točna samo jedna ili čak nijedna, a za treće se ne može ponuditi nijedno etimološki i znanstveno korektno rješenje.

Zbog nepredvidljivosti i raznolikosti sadržaja etimološke natuknice neki su dosadašnji etimološki rječnici nerijetko podsjećali na skup raznoliko strukturiranih lingvističkih eseja. Takav je i Skokov rječnik, ali i Gluhakov. Kako bi rječnik bio što jednostavniji za korištenje, bez obzira na tip sadržaja, važno ga je smisljeno i predvidljivo strukturirati, odnosno poštivati barem temeljna leksikografska načela. Tim će se načelima voditi i sastavljači *Velikog etimološkog rječnika*, a dodatni razlog tomu jest i namjera da se *Veliki etimološki rječnik* ostvari i u računalnome obliku te na taj način u budućnosti bude javno dostupan na institutskim mrežnim stranicama.

3. Veliki etimološki rječnik hrvatskoga jezika

Novi, *Veliki etimološki rječnik hrvatskoga jezika* trebao bi biti moderan etimološki rječnik hrvatskoga jezika. To znači da se opisuje suvremeni hrvatski jezik, da su u njega uključene nove spoznaje, da se natuknica prikazuje na moderan i zanimljiv način, da je računalno oblikovan te da postoji internetska verzija. U *Institutu za hrvatski jezik i jezikoslovlje* dogovorene su osnovne smjernice za njegovu izradbu. *Veliki etimološki rječnik hrvatsko-*

ga jezika neće unaprijed imati sustavnih ograničenja u opsegu etimoloških podataka, već će ona proizlaziti iz samih riječi, odnosno znanstvenih spoznaja o pojedinim riječima. Uz etimologiju će, kao i Skokov, sadržavati i prvu potvrdu riječi, podatke o tvorbi riječi, dijalektološke potvrde, onomastičke primjere, podatke o akcentuaciji i bibliografiju. Drugim riječima, *Veliki etimološki rječnik* neće sadržavati manje podataka od Skokova rječnika, ali će svi podatci biti osuvremenjeni, obogaćeni novim spoznajama i strukturirani u skladu sa suvremenom leksikografskom praksom.

3.1. Izbor natuknica

Osnovni popis natuknica činit će sve riječi hrvatskoga standardnog jezika. U popis natuknica ulazit će i riječi potvrđene u dijalektalnoj književnosti te u dijalektalnim rječnicima. Dijalektalne riječi koje su potvrđene samo u terenskim istraživanjima neće se navoditi, osim iznimno ako su osobito bitne za tumačenje pojedine riječi.

3.2. Prva potvrda riječi

Uz svaku lemu navest će se podatak o prvoj potvrdi te riječi. Prva potvrda riječi značajna je jer njome počinje njezina povijest. Sve što je bilo prije pisane potvrde pripada pretpovijesti te riječi i proučava se pomoću metoda poredbenopovijesne lingvistike (usp. Skok 1957:12). Za riječi koje su rano posvjedočene, osim vremena potvrde, relevantan je i podatak u kakvoj su vrsti teksta zabilježene (npr. jesu li zabilježene na natpisu, u književnom ili leksikografskom djelu ili u kakvom pravnom dokumentu). U hrvatskom jeziku, koji je dijalektalno raznolik, važan je i podatak o mjestu, odnosno dijalektu u kojem je riječ posvjedočena. Kod razmjerno novih riječi nemoguće je utvrditi kada su prvi put posvjedočene.

3.3. Tvorba riječi

Nesporna je povezanost etimologije i tvorbe riječi s obzirom na to da je proučavanje nastanka svih postojećih riječi predmet zanimanja etimologije. Još se u 19. stoljeću u hrvatskom jezikoslovlju za *tvorbu riječi* upotrebljavala riječ *etimologija* (Gluhak 2000:8). Danas su to dvije samostalne lingvističke discipline koje, doduše, imaju puno dodirnih točaka. No, valja razmotriti trebaju li podatci o tvorbi riječi biti uvršteni u etimološki rječnik, na koji način i u kojem opsegu.

Skok je smatrao (1957:19) da etimološke natuknice treba pisati prema »etimološkim familijama«, odnosno da tvorba riječi pripada etimološkom rječniku. S druge strane, Babić smatra (1985:68) da etimološkom rječniku pripadaju samo netvorbene riječi, a od tvorbenih riječi one koje su nastale

dijakronijskom tvorbom, zatim riječi kod kojih tvorba nije potpuno jasna (bilo jer glasovne promjene zamućuju tvorbene odnose, npr. *bezdan*, bilo pojedinačne tvorbe, npr. *maćeha*, *maćuha*, *očuh*, *švelja*, *zelembać*), te one u kojima se prepoznaje čisto formalni tvorbeni obrazac, a zapravo su preuzete iz drugih jezika (npr. *princeza*, *ekspertiza*, *bombonijera*). Svemu tome pridodaje i tvorbene formante, sufikse i prefikse.

Sastavljači *Velikog etimološkog rječnika hrvatskoga jezika* smatraju da u rječnik treba uvrstiti tvorbu riječi, s obzirom na to da je hrvatski jezik tvorbeno produktivan. Raspravlja se jedino o principu unošenja tvorbenih riječi. Skokove natuknice složene su po »gnijezdovnom principu«, tj. unutar jedne natuknice nalaze se i sve izvedenice i složenice od osnovne riječi. Takav princip vjerojatno bolje predočava unutarjezične etimološke veze, ali je i manje pregledan od »leksemskog« (Ivić 1998:XI). Leksemski princip pisanja natuknica olakšava korisniku nalaženje pojedine riječi, jer svaka riječ ima vlastitu natuknicu. U *Velikom etimološkom rječniku hrvatskoga jezika* težit će se leksemskom principu, zbog čega će se, u odnosu na Skoka, znatno povećati broj natuknica i uputnica. Ipak, leksemski princip neće biti proveden apsolutno. Tvorbene riječi unosit će se u rječnik na dva načina: neke tvorbene riječi bit će samostalne natuknice, a neke će biti unutar druge natuknice. Koja će od tvorenica ući u popis natuknica, a koje će se navoditi pod natuknicom osnovne riječi u polju tvorbe, ovisit će o tvorbenoj porodici, starosti riječi i njezinu razvoju. Očite izvedenice, odnosno riječi koje su i značenjski i gramatički neposredno vezane uz osnovnu riječ te se iz nje izvode prema pravilima i bez semantičkih i obličnih promjena, navodit će se uz osnovnu riječ u polju tvorbe riječi, a sve ostale tvorenice iz istog korijena bit će zasebne natuknice. Tako će, na primjer, *mačkica* i *mačkin*, izvedenice riječi *mačka*, biti navedene u natuknici osnovne riječi (*mačka*), dok će *maglenka* i *maglušćina*, izvedenice od *magla*, imati samostalne natuknice, s obzirom na to da su izvedene slabo produktivnim ili neproduktivnim sufiksima. U natuknicama osnovnih riječi bit će navedena potpuna etimologija, a etimologija tvorenica koje imaju samostalnu natuknicu upućivat će na osnovnu riječ kako se informacije ne bi ponavljale i tako opterećivale prostor rječnika. Time će se također izbjeći esejistički tip natuknice i postići preglednost i sustavnost rječnika.

3.4. Dijalektologija

S obzirom na to da je hrvatski jezik i dijalektalno razgranat, u natuknice će se unositi i dijalektni primjeri, no ne iscrpno, nego tamo gdje su relevantni za etimologiju. Relevantni su za etimologiju ako nam daju dodatnu informaciju o distribuciji etimona (je li općehrvatski ili samo štokavski; npr.

standardnojezična riječ *kuća* (< psl. **koŭja*) nije rasprostranjena u čakavskim i kajkavskim govorima – tamo ju zamjenjuju posuđenice *dom* i *hiža*) te o rekonstrukciji praslavenskog etimona. Dijalekatski odrazi iste standardne riječi mogu znatno pridonijeti etimološkoj argumentaciji. Oni mogu čuvati važne informacije, primjerice, o akcentuaciji u praslavenskom ili druge podatke važne za etimologiju (npr. istarski oblik *mihunja* ‘mahuna’ upućuje na korijen s jatom **měhuna*). Izvori za dijalektološke podatke bit će dijalektološki rječnici (*Rječnik kajkavskog književnog jezika*, *Čakavisches-deutsches Lexikon*, *Rječnik Gole*, *Rječnik govora otoka Vrgade*, *Rječnik varaždinskog govora*, *Rječnik grobničkog govora...*).

3.5. Onomastika

Za razliku od Skokova rječnika, u novom velikom etimološkom rječniku neće biti samostalnih onomastičkih natuknica jer se u okviru istog projekta u *Institutu* izrađuje i sveobuhvatna onomastička baza podataka, odnosno antroponimijski i toponimijski rječnici. Onomastički podatci u *Velikom etimološkom rječniku* navodit će se u posebnom polju unutar natuknice kao oprimjerenje apelativa (npr. pod natuknicom *brijeg* kao onomastički primjeri navest će se *Bregar*, *Bregeš*, *Bregović*; *Bregana*, *Ludbreg*, *Podbrežje* i sl.). Apelativi koji danas postoje samo kao vlastito ime, a iz dijalekatskih se potvrda zna da su postojali i kao opće riječi, navodit će se kao natuknice sa zvjezdicom (*). Imenica *čret* (‘močvarno mjesto u šumi’) od koje se izvode toponimi *Čret*, *Črečan*, *Začretje* danas postoji samo u dijalektima (kajkavskim), pa će se u rječniku navesti kao **čret*.

3.6. Akcentuacija

Novost je u ovome rječniku sustavno unošenje akcenatskih podataka u obliku akcenatske paradigme. Za riječi naslijeđene iz praslavenskog, kao i za praslavenske rekonstrukte sustavno će se navoditi akcenatska paradigma kojoj pripadaju. Za praslavenski se rekonstruiraju tri akcenatske paradigme koje, uz značajne promjene, svoje odraze imaju i u hrvatskome: akcenatska paradigma (a.p.) *a*, kojoj pripadaju riječi s fiksnim naglaskom na korijenu, zatim a.p. *b*, kojoj pripadaju riječi s fiksnim naglaskom na nastavku te a.p. *c*, kojoj pripadaju riječi s pomičnim naglaskom. Od postojećih slavenskih etimoloških rječnika akcenatska se paradigma navodi u Derksenovom *Etymological Dictionary of the Slavic Inherited Lexicon*. Za novije posuđenice u hrvatskom dat će se standardna akcenatska informacija.

3.7. Etimologija

Svakoj lemi u rječniku nastojat će se pridružiti najdublja pouzdana etimologija. Za naslijeđene riječi to je praindoeuropski oblik, a za posuđene riječi oblik u jeziku koji se može utvrditi kao ishodište.

Za naslijeđene riječi navodit će se i praslavenski i indoeuropski rekonstruirani oblik. Nastojat će se rekonstruirati cijelu riječ, a ako to nije moguće, barem korijen riječi. Kao oprimjerenja praslavenske rekonstrukcije davat će se primjeri iz starocrkvenoslavenskog, te iz jednoga od južnoslavenskih, istočnoslavenskih i zapadnoslavenskih jezika. Dalje će se navoditi usporednice iz baltijskih te, ako su relevantne, i ostalih indoeuropskih jezika, u pravilu iz latinskoga, grčkog i srednjovisokonjemačkog ako postoje. Tako će se npr. za hrvatsku riječ *brada* uz praslavensku rekonstrukciju (psl. **bordá*) navesti slavenske usporednice stosl. *brada*, slv. *bráda*, rus. *borodá*, polj. *broda*, zatim baltijska lit. *barzdà*, i indoeuropske lat. *barba*, stvnj. *bart*, uz indoeuropsku rekonstrukciju (ie. **b^hord^heh₂*).

Za posuđenice će se navoditi neposredan izvor i put kojim je riječ dospjela od ishodišnog jezika do hrvatskoga. Ako je riječ posuđena iz nekog indoeuropskog jezika, navodit će se i indoeuropska rekonstrukcija, odnosno najdublja etimologija do koje se pouzdano može ići. Npr. riječ *asistent* posuđena je (vjerojatno posredništvom njem. *Assistent*) iz lat. *assistentens*, što je participij prezenta glagola *assistere*. Latinski je glagol složen od *ad+sisto*; *sisto* je postalo od ie. **sti-sth₂-*, redupliciranog indoeuropskog korijena **sth₂-*. Od istog je korijena nastao i hrvatski *stajati*.

3.8. Bibliografija

Sustavno će se navoditi reference na Skokov i Gluhakov etimološki rječnik te na etimološke rječnike drugih slavenskih jezika. Za ostale jezike upućivat će se na pripadajuće etimološke rječnike. Neće se navoditi sva postojeća literatura o pojedinim riječima, već samo rasprave nastale nakon pojave Skokova rječnika, i to one koje donose moguću, vjerojatnu ili sigurnu etimologiju. Pogrešne etimologije (znanstveno neutemeljene ili nadvladane) neće se navoditi.

4. Računalni etimološki rječnik

Uporaba računala uvela je mnoge promjene u suvremenu leksikografiju. Mnogi zadatci pojednostavljeni su uporabom računala, otvoren je veći prostor za rječnike, a pretraživanje i unutarrječničko povezivanje višestruko je unaprijeđeno. Kako bi se iskoristile prednosti računalno kodirane

leksikografije, potrebno je rječničku natuknicu računalno strukturirati. S obzirom na to da ne postoji univerzalna računalna struktura rječničke natuknice, pa tako ni etimološke natuknice, u ovom će se članku predložiti računalna struktura u obliku XML-a koja bi bila prikladna za gore opisani hrvatski etimološki rječnik. Struktura modelirana u XML-u može se (uz manje preinake) uklopiti u računalne programe za leksikografsku obradbu kojima je podloga jezik za obilježavanje XML.

Postojećim računalnim leksikografskim standardima i programima zajedničko je to da značajno pojednostavnjuju etimološku informaciju te da računalno kodiraju manji broj podataka. Takav pristup etimološkoj natuknici možda jest najjednostavnije rješenje, ali nije i najbolje. Tako se, primjerice, u preporukama TEI³ (Text Encoding Initiative) za rječnike etimološka informacija daje u obliku proze u kojoj je, od etimološki relevantnih informacija, računalno kodirana samo oznaka jezika i prvi etimon. S obzirom na to da dajemo prijedlog strukture natuknice za rječnik u kojem je etimološka informacija dominantna i najvažnija, željeli bismo da u njemu bude računalno kodirano (tj. računalno čitljivo) što više etimoloških podataka. Osim toga, preporuke TEI u postojećem obliku predviđaju samo jedan stupanj etimologije, a to nikako nije dovoljno za hrvatski etimološki rječnik.

U *Institutu za hrvatski jezik i jezikoslovlje* rječnici su se dosad izrađivali u leksikografskom programu *Softlex* koji znatno olakšava izradbu tiskanih rječnika, ali nije sasvim prikladan za razgranatu strukturu etimološkoga rječnika niti za podlogu ima bazu podataka. Stoga će se *Veliki etimološki rječnik hrvatskoga jezika* izrađivati pomoću programa *Tshwanelex*. *Tshwanelex* je računalni programski paket (*software application suite*) koji služi za sastavljanje rječnika ili terminoloških glosara. Može se koristiti za sve jezike, za sve vrste rječnika, a podloga su mu industrijski standardi XML i Unicode.

Natuknica planiranog etimološkog rječnika sadržavat će veliku količinu podataka, a sve će natuknice biti strukturirane na isti način. Stoga je potreban model etimološke natuknice u kojoj su uravnotežene ograničenost svake rječničke (i računalne) strukture i neograničenost etimološke informacije.

Ovdje će se prikazati računalni model natuknice u koji bi se mogli uvrstiti svi podatci koje će sadržavati novi *Veliki etimološki rječnik hrvatskoga jezika*. Model ćemo prikazati pomoću strukture jezika za obilježavanje XML.

³ Preporuke TEI prvenstveno su predviđene za računalno kodiranje već tiskanih rječnika.

5. Model računalne strukture⁴

Svaka se natuknica u rječniku sastoji od sljedećih elemenata i atributa:⁵

- (I) Lema
LemaNaglašena
[Gramatika]
VrstaRiječi
PrvaPotvrda
Značenje
LeksOdrednica
LatinskiNaziv

- (II) [Riječ]
[Dijalekti]
[Onomastika]
[Tvorba]
[Etimologija]
[References]
[Bibliografija]

Prvi (I) je dio natuknice opći, tj. on može biti dijelom bilo kojeg rječnika. U atribut *Lema* upisuje se lema, nenaglašena riječ koja je nosilac natuknice. Ista ta riječ naglašena standardnim hrvatskim naglaskom upisuje se u atribut *LemaNaglašena*. U element [Gramatika], pod odgovarajuće attribute, upisuju se npr. genitiv i množina imenica, ili 1. lice prezenta glagola i sl. Vrijednost atributa *VrstaRiječi* bira se iz unaprijed određenog popisa mogućih vrsta riječi. U atribut *PrvaPotvrda* upisuje se stoljeće u kojem je lema prvi put posvjedočena. Atribut *LeksOdrednica*, koji daje informaciju o području kojem neka riječ pripada (npr. botanika, kemija, glazba i sl.), bira se iz unaprijed određenog popisa. U općejezičnim rječnicima važno je što detaljnije razraditi i definirati značenje, no u ovom se etimološkom

⁴ Na izlaganju u HAZU potkraj 2009. godine predstavljena je osnovna struktura natuknice. U međuvremenu je započet rad na rječniku, te je struktura razrađena i nadopunjena novim elementima i atributima. Osnovna ideja ostala je ista. Ovdje donosimo noviji, poboljšani model natuknice.

⁵ *Elementi* i *atributi* tehnički su termini koji se koriste u XML-u. *Element* je strukturalna jedinica XML-a. *Atributi* služe da se opiše informacija koja je u određenom smislu deskriptivna za pojavljivanje nekog elementa, ali nije dijelom njegova sadržaja. U praktičnoj primjeni, u Tshwanelexu, *elementi* pod sobom objedinjuju druge elemente, a *atributi* opisuju te elemente. *Atributima* se pridružuje neka vrijednost, bilo tako da se upiše tekst, bilo tako da se izabere vrijednost iz unaprijed određenog popisa vrijednosti. U ovom radu elementi se zapisuju u uglatim zagradama [], a atributi su bez posebne oznake.

rječniku *Značenje* leme ne upisuje sustavno. Značenje se navodi kad riječ nije općepoznata, kad razvoj značenja možda nije očit ili kad je riječ o homonimima. Za biljke i životinje navodi se sistemski *LatinskiNaziv*. Općem dijelu rječnika pripada još i element [References], u kojem se nalaze uputnice na druge leme.

Za preostale elemente (II) možemo reći da su specifični za veliki rječnik hrvatskoga jezika, odnosno za etimološki rječnik. Element [Riječ] može se ponavljati. Tu se navode riječi koje su varijante leme, npr. riječ s uobičajenim naglaskom (koji je različit od standardnog), ili starije potvrde te riječi koje se razlikuju oblikom. U element [Dijalekti] unose se potvrde iz dijalekata, u elementu [Onomastika] navode se imena koja su nastala od leme, a u elementu [Tvorba] trivijalne izvedenice od leme (npr. posvojni pridjev, česte umanjnice, obični mocijski parnjaci i sl.). U elementu [Etimologija] donosi se detaljna etimološka informacija. Element [Bibliografija], premda uobičajen uglavnom u etimološkim rječnicima, teoretski može biti dijelom kojega god rječnika. U njega se upisuju podatci kako je najavljeno u prvo- me dijelu ovoga rada (v. t. 3.8.).

Kao što je već navedeno, natuknica etimološkog rječnika sadržava mnogo riječi koje nisu leme i u različitim su međusobnim odnosima. Najekonomičnije je opisati ih na isti način te ih grupirati u hijerarhijski više elemente. Taj element nazvali smo [Riječ], on je najvažniji dio strukture i može se pojavljivati na svim razinama. Osim na najvišoj razini, neposredno ispod leme, elementi [Riječ] grupiraju se i u elemente [Dijalekti], [Tvorba] i [Onomastika] koji se pojavljuju unutar leme jednom ili nijednom.

Element [Riječ] pojavljuje se i unutar elementa [Etimologija] i to na nekoliko razina. Element [Etimologija] može se ponavljati (u slučaju kad postoji više mogućih etimologija). Element [Etimologija] sadrži sljedeće atribute i hijerarhijski niže elemente:

[Usporednice]
EtimološkaVeza
[Riječ]
[UspOpće]
[Etimologija]
[Bilješka]

Atribut *EtimološkaVeza* bira se iz unaprijed određenog popisa. Riječi nekog jezika mogu se ugrubo podijeliti na naslijeđene, posuđene i izvedene, a to se kodira ovim atributom. Za naslijeđene riječi bira se znak »<«, a

za posuđene i izvedene odgovarajući izraz (npr. »posuđeno iz«, »izvedeno od«, »kalk prema«). Element [Riječ] na toj je razini etimon, riječ od koje je lema postala. U elementima [Usporednice] i [UspOpće] navode se riječi iz drugih jezika. Razlika je između tih dvaju elemenata u tome što se element [Usporednice] odnosi na naslijeđene riječi, a element [UspOpće] odnosi se na sve druge slučajeve kad je potrebno navesti primjere iz drugih jezika. Element [Etimologija] može sadržavati još jedan element [Etimologija], s obzirom na to da se u rječniku ne daje samo jedan stupanj etimologije, nego se daje što je više moguće stupnjeva etimologije. Većina lema imat će više stupnjeva etimologije, dakle ne samo etimon leme, već i etimon etimona i tako dalje. Naslijeđene riječi obično imaju dva stupnja etimologije: na prvome je etimon praslavenska riječ, na drugome indoeuropska. Riječi posuđene iz indoeuropskih jezika imaju barem dva stupnja, a vrlo često i više stupnjeva, jer se navode i riječi iz jezika posrednika i indoeuropska rekonstrukcija.

Element [Bilješka] predviđen je za slobodan unos teksta u kojem se objašnjavaju neke pojedinosti etimologije, zatim sumnjive etimologije, razvoj značenja, i sve ostalo što bi moglo biti zanimljivo korisnicima rječnika.

Element [References] i atribut *Tekst* pojavljuju se unutar svih elemenata.

Preostaje još opisati najvažniji element iz strukture, [Riječ], koji se nalazi na svim razinama strukture. Slijedi redoslijed atributa i elemenata koje sadrži element [Riječ]:

Jezik
Riječ
[Gramatika]
VrstaRiječi
Značenje
Potvrda

U osnovi ti su atributi isti kao i opći dio leme, jer se sve riječi (tvorenice, dijalektalne riječi, usporednice, etimoni...) unutar ovog rječnika obrađuju kao »podleme«. Atribut *Jezik* jedini je atribut koji se nalazi unutar elementa [Riječ], a ne nalazi se i unutar elementa [Lema]. To je zato što je za leme unaprijed definiran jezik hrvatski, pa ga nije potrebno navoditi. Za sve ostale riječi može se iz unaprijed definiranog popisa (koja se po potrebi dopunjuje) izabrati jezik. Za potvrde iz dijalekata to će biti naziv narječja, a za sve ostale riječi, to će biti posvjedočeni ili rekonstruirani jezik.

U elementu [Riječ] samo je jedan atribut, *Riječ*⁶, u koji se upisuje riječ. Element [Gramatika] isti je kao unutar leme. U njega se unosi i akcenatska paradigma za rekonstruirane praslavenske etimone. *Potvrda* je atribut u koji se unosi podatak, primjerice, o mjestu u kojem je neka dijalektalna riječ zabilježena ili o literaturnom izvoru za određenu riječ. Ostali atributi, *Vrsta-Riječi* i *Značenje*, isti su kao unutar leme.

Kao što je već istaknuto, etimološke natuknice sadrže veliku količinu podataka, a svaka riječ ima svoju povijest te je nemoguće unaprijed predvidjeti sve moguće tipove natuknica i sve detalje u njihovoj strukturi. Zato naglašavamo da opisani model etimološke natuknice nije konačan. Prikazani model nastao je uspoređivanjem različitih etimoloških rječnika, novijih i starijih, jednosveščanih i višesveščanih, računalnih i tiskanih. U obzir su uzete i mnogobrojne rasprave o računalnom kodiranju etimološke informacije unutar rječnika, a neki detalji razradbe dolaze iz prakse.⁷

6. Primjeri

Navest ćemo četiri primjera konkretnih natuknica novog etimološkog rječnika. Prvo navodimo »tiskani« oblik, a zatim strukturalni oblik u XML-u⁸.

6.1. Naslijeđena riječ

MED mēd /G mēda/ (*im.*) (13.st.)

♦ *mēt* /G mēda/ (Cres) *mēđ* /G mēda/ (ČDL, Brač) *mēđ* /G mēda/ (Novi, Vrgada)

(stcsl. *medъ* sln. *mēđ* rus. *мѣд* polj. *miód*) < psl. **mēdъ* /G **medú*/ [a.p.c] (lit. *medūs* 'med' skr. *mádhū-* 'med, medovina' stvnj. *metu* 'medovina' stir. *mid* 'medovina' grč. *μέθυ* 'vino') < ie. **medh-u* 'medovina'

Lit.: Skok II, 396 Snog 389; Derksen 306

<Lema Lema="med" LemaNaglašena="mēd" VrstaRiječi="im." PrvaPotvrda="13.st." >

<Gramatika G="mēda"/>

<Dijalekti>

<Riječ Jezik="" Riječ="mēt" Potvrda="Cres">

<Gramatika G="mēda"/>

</Riječ>

<Riječ Jezik="" Riječ="mēđ" Potvrda="ČDL, Brač">

<Gramatika G="mēda"/>

</Riječ>

⁶ Za razliku od dva unutar leme, *Lema* i *LemaNaglašena*.

⁷ Iz dosad obrađenih tristotinjak natuknica.

⁸ Struktura XML-a u ovome je radu nešto pojednostavljena kako bi bila čitljiva, ali i dalje su poštivana sintakсна pravila XML-a.

```
<Riječ Jezik="" Riječ="mêd" Potvrda="Novi, Vrgada">
  <Gramatika G="mêda"/>
  </Riječ>
</Dijalekti>
<Etimologija etimNumber="1" EtimološkaVeza="" &lt;" >
  <Usporednice >
    <Riječ Jezik="stsl." Riječ="medъ"/>
    <Riječ Jezik="sln." Riječ="mêd"/>
    <Riječ Jezik="rus." Riječ="мѣд"/>
    <Riječ Jezik="polj." Riječ="miód"/>
  </Usporednice>
  <Riječ Jezik="psl." Riječ="*mêdъ">
    <Gramatika AkcentParadigma="c" G="*medú"/>
    </Riječ>
  <Etimologija etimNumber="1" EtimološkaVeza="" &lt;" >
    <Usporednice >
      <Riječ Jezik="lit." Riječ="medùs" Značenje="med"/>
      <Riječ Jezik="skr." Riječ="mádhū" Značenje="med,
medovina"/>
      <Riječ Jezik="stvnj." Riječ="metu"
Značenje="medovina"/>
      <Riječ Jezik="stir." Riječ="mid"
Značenje="medovina"/>
      <Riječ Jezik="grč." Riječ="μέθυ" Značenje="vino"/>
    </Usporednice>
    <Riječ Jezik="ie." Riječ="*med%rh%r-u"
Značenje="medovina"/>
  </Etimologija>
  </Etimologija>
  <Bibliografija Skok="II, 396" note="Snoj, 389; Derksen 306"/>
</Lema>
```

6.2. Riječ posuđena iz indoeuropskog jezika; dvije moguće etimologije

MEDALJA mèdalja (17.st.) *medâlja* (Božava, Rab) *medâlja* (ŽK)

♦ čak. *medāja* (ČDL)

Posuđeno iz tal. *medaglia* < vlat. **medalia* (srlat. *medal(l)ia*, *medal(l)a* 'kovanica u vrijednosti pola denara'), što je nastalo disimilacijom od *mediālia* (*monēta*), množina srednjeg roda pridjeva lat. *mediālis* 'središnji' (kasno potvrđen), koji je izveden od pridjeva *mēdius* 'u sredini, središnji, osrednji', v. **medij**, **međa**. Druga je mogućnost da je talijanska riječ posuđena iz nekog galoromanskog idioma u kojem -t- > -d-, a da je izvor vlat. **metallea* (*monēta*) 'kovanica', pridjev izveden od *metallum*, v. **metal**. Riječ se proširila u druge europske jezike: fr. *medaille* njem. *Medaille* eng. *medal*.

Lit.: Skok II, 397; Snj 388; DEI 2401; REW 5451; EWD 1081

```
<Lema Lema="medalja" LemaNaglašena="mèdalja" PrvaPotvrda="17.st." >
  <Riječ Jezik="" Riječ="medâlja" Potvrda="Božava, Rab"/>
  <Riječ Jezik="" Riječ="medâlja" Potvrda="ŽK"/>
  <Dijalekti>
    <Riječ Jezik="čak." Riječ="medāja" Potvrda="ČDL"/>
  </Dijalekti>
```

```
<Etimologija etimNumber="1" EtimološkaVeza="Posuđeno iz">
  <Riječ Jezik="tal." Riječ="medaglia"/>
  <Etimologija etimNumber="1" EtimološkaVeza="&lt;">
    <Riječ Jezik="vlat." Riječ="*medalia"/>
    <UspOpće>
      <Riječ Jezik="srlat." Riječ="medal(l)ia, medal(l)a"
Značenje="kovanica u vrijednosti pola denara"/>
    </UspOpće>
    <Etimologija etimNumber="1" Tekst=", što je nastalo disimilacijom
od" EtimološkaVeza="" >
      <Riječ Jezik="" Riječ="mediālia (monēta)"/>
      <Etimologija etimNumber="1" Tekst=", množina srednjeg
roda pridjeva" EtimološkaVeza="">
        <Riječ Jezik="lat." Riječ="mediālis"
Značenje="središnji" Potvrda="kasno potvrđen"/>
        <Etimologija etimNumber="1" Tekst=", koji je"
EtimološkaVeza="izveden od">
          <Riječ Tekst="pridjeva" Jezik=""
Riječ="mēdius" Značenje="u sredini, središnji, osrednji">
            <References>
              <reftypegroup reftype="v">
                <reference target="medij"/>
                <reference target="međa"/>
              </reftypegroup>
            </References>
          </Riječ>
        </Etimologija>
      </Etimologija>
    </Etimologija>
  </Etimologija>
  <Etimologija etimNumber="2" Tekst="Druga je mogućnost da je tali-
janska riječ" EtimološkaVeza="posuđena iz">
    <Riječ Tekst="nekog galoromanskog idioma u kojem -t- &gt;
-d-" Jezik=""/>
    <Etimologija etimNumber="1" Tekst=", a da je izvor" Eti-
mološkaVeza="">
      <Riječ Jezik="vlat." Riječ="*metallea (monēta)"
Značenje="kovanica"/>
      <Etimologija etimNumber="1" Tekst=", pridjev"
EtimološkaVeza="izveden od">
        <Riječ Jezik="" Riječ="metallum" >
          <References>
            <reftypegroup ref-
type="v">
              <referen-
ce target="metal"/>
            </refty-
pegroup>
          </References>
        </Riječ>
      </Etimologija>
    </Etimologija>
  <Bilješka Tekst="Riječ se proširila u druge europske jezike:">
    <Riječ Jezik="fr." Riječ="medaille"/>
    <Riječ Jezik="njem." Riječ="Medaille"/>
    <Riječ Jezik="eng." Riječ="medal"/>
  </Bilješka>
</Etimologija>
<Bibliografija Skok="II, 397;" note="Snoj 388; DEI 2401; REW 5451; EWD 1081"/>
</Lema>
```

6.3. Riječ posuđena iz neindoeuroskog jezika

MELEZ mèlez (*im.*) (19.st.) 'mješanac, križanac' **mèlez** (Karadžić), **melèz** (Kosovo)

Posuđeno preko tur. *melez* 'mulat, hibrid' iz arap. *malas* 'što je sastavljeno od dvije različite stvari'. Balkanski turcizam, usp. bug. *мелез*, alb. *melez*, rum. *melez*. U dijalektima se riječ odnosi i na druge "miješane" stvari, npr. *melez* 'platno, u kojega je osnova prtena, a potka pamučna' i *meleza* 'šarena tkanina'.

Lit.: Skok II, 404; BER 726

```
<Lema Lema="melez" LemaNaglašena="mèlez" Notes="provjeriti arapski oblik" VrstaRiječi="im."
PrvaPotvrda="19.st." Značenje="mješanac, križanac" >
  <Riječ Jezik="" Riječ="mèlez" Potvrda="Karadžić"/>
  <Riječ Jezik="" Riječ="melèz" Potvrda="Kosovo"/>
  <Etimologija etimNumber="1" EtimološkaVeza="Posuđeno preko">
    <Riječ Jezik="tur." Riječ="melez" Značenje="mulat, hibrid"/>
    <Etimologija etimNumber="1" EtimološkaVeza="iz">
      <Riječ Jezik="arap." Riječ="malas" Značenje="što je sastavljeno od dvi-
je različite stvari"/>
    </Etimologija>
    < Bilješka Tekst="Balkanski turcizam, usp.">
      <Riječ Jezik="bug." Riječ="мелез"/>
      <Riječ Jezik="alb." Riječ="melez"/>
      <Riječ Jezik="rum." Riječ="melez"/>
    </ Bilješka>
    < Bilješka Tekst="U dijalektima se riječ odnosi i na druge &quot;miješane&quot;
stvari, npr. ">
      <Riječ Jezik="" Riječ="melez" Značenje="platno, u kojega je osnova pr-
tena, a potka pamučna"/>
      <Riječ Jezik="i" Jezik="" Riječ="meleza" Značenje="šarena tkanina"/>
    </ Bilješka>
  </Etimologija>
  < Bibliografija Skok="II, 404;" note="BER 726"/>
</Lema>
```

6.4. Izvedenica, s uputnicom na glavnu riječ

MENTOL mèntōl (*im.*)

Izvedeno od **menta** sufiksom **-ol**. Mentol je glavni sastojak ulja *metvice* (*mente*), v. **metvica**.

```
<Lema Lema="mentol" LemaNaglašena="mèntōl" VrstaRiječi="im.">
  <Etimologija etimNumber="1" EtimološkaVeza="Izvedeno od">
    <Riječ Jezik="">
      <References >
        <refypegroup reftype="">
          <reference target="menta"/>
        </refypegroup>
      </References>
    </Riječ>
  <Riječ Jezik="sufiksom" Jezik="">
```



```
<References >
  <refypegroup reftype="">
    <reference target="-ol"/>
  </refypegroup>
  </References>
</Riječ>
<Bilješka Tekst="~[Mentol] je glavni sastojak ulja %imetvice%i (%imente%i)">
  <References >
    <refypegroup reftype=", v.">
      <reference target="metvica"/>
    </refypegroup>
  </References>
</Bilješka>
</Etimologija>
</Lema>
```

7. Zaključak

U ovome radu najavili smo izradbu novog, *Velikoga etimološkog rječnika hrvatskoga jezika* te opisali načela po kojima će se on izrađivati. Predstavljen je također i računalni model natuknice toga rječnika.

Predloženi računalni model nema uzora u hrvatskoj leksikografiji, budući da se postojeći hrvatski etimološki rječnici nisu izrađivali pomoću leksikografskih računalnih programa (Gluhakov rječnik), a neki niti pomoću računala (Skokov rječnik). Strukture natuknica etimoloških rječnika drugih jezika koji postoje u računalnom obliku ne zadovoljavaju specifične zahtjeve *Velikoga etimološkog rječnika hrvatskoga jezika*.

Opisani model natuknice ima nekoliko prednosti. Ponajprije, to je njezova hijerarhičnost koja, između ostalog, odražava i etimološku dubinu vremena. Nadalje, prednost je i mogućnost da se svaka riječ unutar natuknice opiše pomoću istih parametara, što je vrlo važno za rječnik kao što je etimološki. Budući da svaka natuknica unutar sebe sadrži mnogo riječi (etimona, izvedenica, etimoloških usporednica, dijalektalnih varijanti...), takav je pristup opisivanju pregledan, dosljedan i sistematičan. S obzirom na to da svaka riječ ima svoju povijest i da svaka etimološka natuknica može sadržavati mnoštvo podataka, nemoguće je unaprijed predvidjeti sve moguće tipove natuknica i sve pojedinosti njihove strukture. Zato je i predstavljena računalna struktura, kojom bi bila obuhvaćena povijest svih riječi, nužno apstraktna i s neizbježnim pojednostavljenjima. No, ta struktura nije konačna. Dugotrajan i u svakom smislu zahtjevan posao izradbe *Velikoga etimološkog rječnika hrvatskoga jezika* tek je u začetcima, pa se i računalna struktura po potrebi izmjenjuje i dopunjuje.

Literatura

- Alt, Susanne. 2006. Data structures for etymology: Towards an etymological lexical network. *Bulag: Numéro Etymologie* 31. http://www.atilf.fr/perso/salmon-alt/telechargement/Bulag_2006.pdf, [01.12.2009].
- Babić, Stjepan. 1985. Etimologija i tvorba riječi u etimološkim rječnicima. *Zbornik u čast Petru Skoku*, Zagreb : JAZU. 65–69.
- Babić, Stjepan. 1989. Sinkronija i dijakronija u tvorbi riječi. *Jezik* 37 (1989/90):1, 1–9.
- Babić, Stjepan. 2002. *Tvorba riječi u hrvatskome književnome jeziku*. Zagreb : HAZU – Nakladni zavod Globus.
- Brozović-Rončević, Dunja. 1994. O imenima u »Hrvatskom etimološkom rječniku«. *Folia onomastica Croatica* 3, 149–154.
- Crist, Sean. 2005. Toward a formal markup standard for etymological data. http://www.panix.com/~kurisuto/publications/crist_etym_markup.pdf. [01.12.2009].
- Derksen, Rick. 2008. *Etymological Dictionary of the Slavic Inherited Lexicon*. Leiden : Brill.
- Furlan, Metka. 1990. Etimologija in besedotvorje. *Slavistična revija* 38:4, 363–369.
- Gluhak, Alemko. 1993. *Hrvatski etimološki rječnik*. Zagreb : August Cesarec.
- Gluhak, Alemko. 2000. *Tipologija etimoloških rječnika* (doktorska disertacija). Zagreb.
- Hoffmann, Karl, Eva Tichy. Checklist for positing etymologies in Indo – European languages. U Mayrhofer, Manfred. 1980. *Zur Gestaltung des etymologischen Wörterbuches einer Großcorpus – Sprache*. Vienna, 47–52.
- Hraste, Mate, Petar Šimunović, Reinhold Olesch. 1979. *Čakavisch-deutsches Lexikon*. Köln : Böhlau.
- Ivić, Milka (ed.). 1998. *Etimološki rečnik srpskog jezika: Ogledna sveska*. Beograd : Institut za srpski jezik SANU.
- Jojić, Ljiljana, Ranko Matasović (eds.). 2002. *Hrvatski enciklopedijski rječnik: Uvod – Etimologija*. Zagreb : Novi Liber.
- Jurišić, Blaž. 1966. *Rječnik govora otoka Vrgade*. Zagreb: JAZU.
- Kapović, Mate. 2008. Razvoj hrvatske akcentuacije. *Filologija* 51, 1–39.
- Malkiel, Yakov. 1976. *Etymological dictionaries – A Tentative Typology*. Chicago : The University of Chicago Press.
- Matasović, Ranko. Hrvatski etimološki rječnik – Predgovor. <http://www.ihjj.hr/dokumenti/Predgovor.pdf>, [05.01.2010].
- Matasović, Ranko. 2008. *Poredbenopovijesna gramatika hrvatskoga jezika*. Zagreb : Matica hrvatska.

- Salmon-Alt, Susanne, Laurent Romary, Eva Buchi. 2005. *Modelling diachrony in dictionaries*. ACH–ALLC. Vancouver.
- Skok, Petar. 1971–1974. *Etimologijski rječnik hrvatskog ili srpskog jezika*. Zagreb : JAZU.
- Skok, Petar. 1957. O etimološkom rječniku hrvatskoga ili srpskoga jezika. *Filologija* 1, 7–21.
- Burnard, Lou, Syd Bauman (eds.). 2008. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. TEI Consortium.
- Večenaj, Ivan, Mijo Lončarić. 1997. *Rječnik Gole*. Zagreb : IHJJ.
- Vinja, Vojmir. 1998–2004. *Jadranske etimologije. Jadranske dopune Skokovu etimologijskom rječniku I–III*. Zagreb : Školska knjiga.

Comprehensive Etymological Dictionary of Croatian Language and Digital Structure of its Entry

Abstract

As part of the project Onomastic and Etymological Research of Croatian Language at the Institute of Croatian Language and Linguistics, a comprehensive etymological dictionary of Croatian language will be made. The dictionary will follow Skok's extensive lexicographical data, but will also be enriched with more recent etymological, onomastic and dialectal information. The Institute of Croatian Language and Linguistics aims to adjust its lexicographical activities with standardized digital encoding formats of lexicographical data. The comprehensive etymological dictionary will also be made according to the same standards. Existing models of digital encoding formats mainly suggest the structure of general purpose dictionaries, while the structure of special dictionaries (like etymological) needs to be expanded.

This paper will present the new, Comprehensive Etymological Dictionary of Croatian Language and digital structure of its entry. Problems in the process of structuring the entry occur mainly because of the fact that every word has its own history, hence every entry in the dictionary has the unique structure. The model of the structure will be presented on the examples from the new etymological dictionary. The structure is not definite, but it is open for consideration, evaluation and suggestions of the scientific community.

Ključne riječi: etimologija, Veliki etimološki rječnik hrvatskoga jezika, Tshwanelex, računalna struktura

Key words: etymology, Comprehensive Etymological Dictionary of Croatian Language, Tshwanelex, digital structure

