

dr. sc. Tomislav Stojanov
Institut za hrvatski jezik i jezikoslovje



21. - 22. travnja 2016.
Nacionalna i sveučilišna knjižnica u Zagrebu

Hrvatska leksikografija i digitalizacija

Od tiskanoga rječnika prema Europskome rječničkom portalu

6. festival hrvatskih digitalizacijskih projekata

Zagreb, 21. – 22. travnja 2016.

Nacionalna i sveučilišna knjižnica

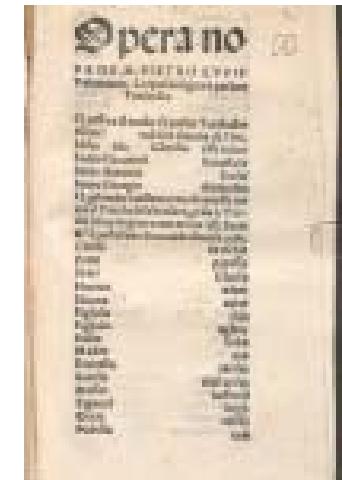
Sažetak

1. ishodi leksikografskoga rada u prošlosti
 - leksikografija i njezina civilizacijska i kulturološka dimenzija
 - veza leksikografije i prvih nacionalnih akademija
2. leksikografska djelatnost za hrvatski jezik danas
 - pretraživi hrvatski leksikografsko-digitalizacijski projekti
 - dostupni mrežni rječnici za hrvatski jezik
3. leksikografija i hrvatski jezik u budućnosti
 - norma digitalizacijske leksikografije za znanstvena istraživanja
 - NSK kao sakupljač tekstne građe na hrvatskome
 - Europski rječnički portal

Ishodi leksikografskoga rada u prošlosti

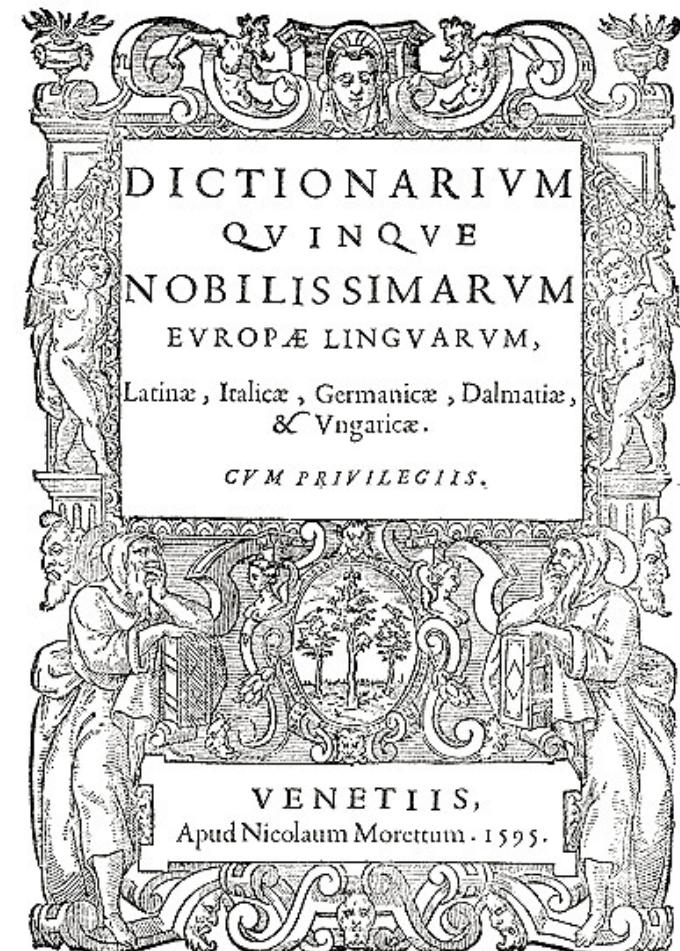
Bogata leksikografska povijest hrvatskoga jezika

- Ako se po jutru dan poznaje...
- 1973. otkriven *Talijansko-hrvatski rječnik* Petra Lupisa Valentiana (*Opera nuova...*) iz 1527., prvi hrvatski tiskani rječnik, 68 godina prije Vrančića, 27 godina nakon ere inkunabula i 25 godina nakon *Dictionarium Amrosiusa Calepinusa*



Bogata leksikografska povijest hrvatskoga jezika

- Ako se po jutru dan poznaje...
- Zora hrvatske leksikografije nije mogla biti reprezentativnija – 1595. *Rječnik pet najuglednijih europskih jezika: latinski, talijanski, njemački, hrvatski i mađarski*



Bogata leksikografska povijest hrvatskoga jezika

- Ako se po jutru dan poznaje...
- Zora hrvatske leksikografije nije mogla biti reprezentativnija – 1595. *Rječnik pet najuglednijih europskih jezika: latinski, talijanski, njemački, hrvatski i mađarski*
- Osnova za sedmerojezični rječnik Petra Lodereckera iz 1605. (s dodatkom češkoga i poljskoga)



Od autorskih rječnika do institucionalne leksikografije

- od svojih početaka leksikografija ima važno kulturno i kasnije, nacionalno obilježje
- „Hrvatska na području leksikografije nije nimalo zaostajala za kulturnom Europom“ (Boras 2008)
- „Hrvatska leksikografija ima četiri stalnice – zapadnoeuropsku orijentaciju, nadregionalnost, zadanost i purizam“ (Znika 1992)

Jezik u središtu srednjovjekovnih akademija

- Motiv za osnivanjem najstarijih svjetskih akademija – talijanske, francuske i španjolske – želja za uspostavljanjem *jezičnoga čistunstva* (tj. jezičnoga standarda)
- Osnivane su u 16. i 17. stoljeću, iako je renesansna svijest o estetsko-umjetničkoj dimenziji jezika postojala i prije (Francesco Petrarca pisao je dva stoljeća prije osnutka talijanske akademije).
- Leksikografija kao prioritetna djelatnost prvotnih akademija
- Prve jezična djela bili su rječnici
- Čekao se – izum tiska

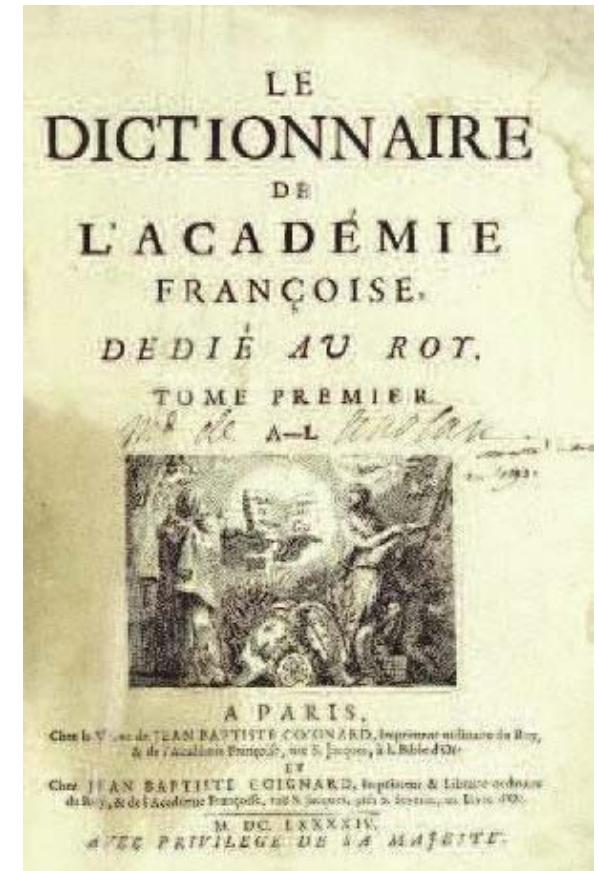
Jezik u središtu srednjovjekovnih akademija

- *Accademia della Crusca* („Akademija mekinja”, „Akademija koja odvaja žito od kukolja”), 1582.
- krilatica: *Il più bel fior ne coglie* („skuplja najljepše cvjetove”)
- rječnik: *Vocabolario degli Accademici della Crusca*, prvo izdanje Venecija 1612.



Jezik u središtu srednjovjekovnih akademija

- *L'Académie française* (Francuska akademija), 1635.
- krilatica: *À l'immortalité* („prema besmrtnosti”)
- rječnik: *Le Dictionnaire de l'Académie française dédié au Roy*, 1694.



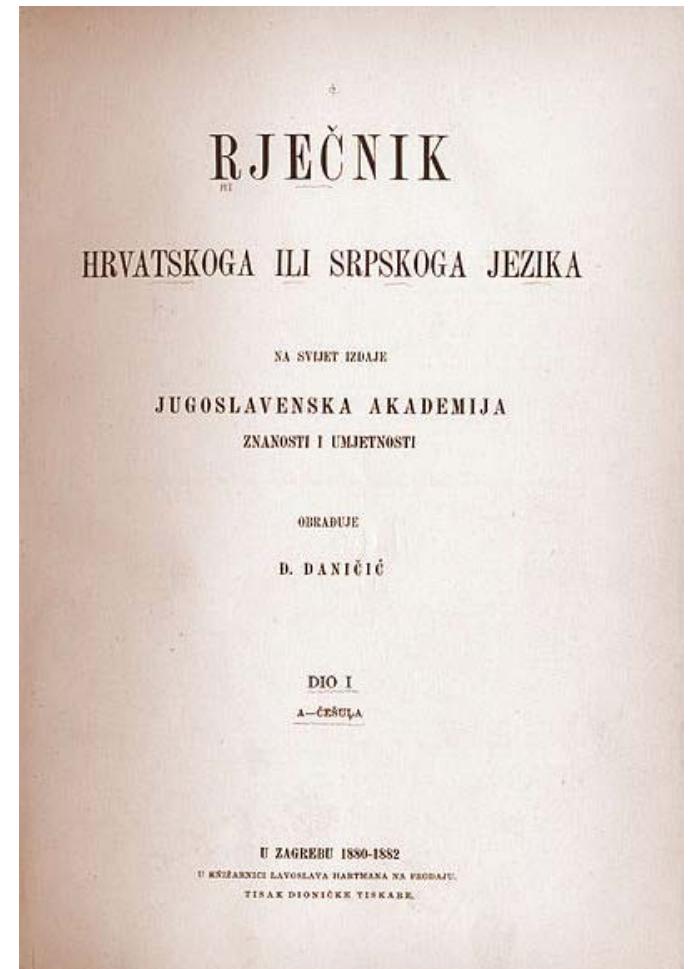
Jezik u središtu srednjovjekovnih akademija

- *Real Academia Española* (Španjolska kraljevska akademija), 1713.
- krilatica: *Limpia, fija y da esplendor* („čisti, uređuje i daje sjaj”)
- rječnik: *Diccionario de autoridades*, 1726. – 1739.



Jezik u središtu JAZU/HAZU

- JAZU/HAZU 1866.
- prva tri razreda: **istoričko-filološki**, filozofičko-juridički, matematičko-prirodoslovni
- *Rječnik hrvatskoga ili srpskoga jezika*, nastajao od 1880. do 1976.
- Za razliku od francuskoga hrvatski jezik ima svoj monumentalni povijesni rječnik (njihov je započeo 1778. i nije odmakao dalje od slova A)



leksikografska djelatnost za hrvatski jezik danas

Pretraživi hrvatski leksikografsko-digitalizacijski projekti

- Ne računajući digitalizirana djela na Google Books i Archive.org

Pretraživi hrvatski leksikografsko-digitalizacijski projekti

1. Digitalizzazione e messa on line del Vocabolario di tre nobilissimi linguaggi, italiano, illirico e latino di Tanzlingher, 2004. – rukopisni rječnik iz 1699.,
<http://tanzlingher.filosofia.sns.it/>



- financirano iz europskih fondova, sudjelovali stručnjaci sa Sveučilišta u Padovi te IHJJ-a
- leksikografska provjera i transkripcija, kodiranje teksta u format XML, navigacija i napredno pretraživanje, višejezično sučelje

Indice alfabetico: A | B | C | D | E | F | G | H | I | J | K | L

U

152 pagine per la lettera U:

1 - da U more utopitti a Ubaahxni ▾ vai

Pagina n. 1 da U more utopitti a Ubaahxni

U more utopitti
U dobbri časc
U glyuubaù navaayati
U kgnighi pijsati
U krijo tischatti
U sadasgne uryme
U suakko dobba
U suakkò uryme
U zaayam datti
U zayam pijtati

Cerca: tutto ▾

Lemma Definizione

Restringi a:

Termini militari (solo italiano) Erbe (solo italiano)

Pretraživi hrvatski leksikografsko-digitalizacijski projekti

2. Portal hrvatske rječničke baštine, 2005., <http://crodip.ffzg.hr/> – pretraživa četiri rječnika: Vrančić 1595., Kašić 1599., Habdelić 1670. i *Libellus alphabeticus*, 1756.

Portal hrvatske rječničke baštine

Unakrsno pretraživanje
Nova pretraga | Impresum | English version | Naslovnica

- znanstveni projekt MZOŠ-a *Hrvatska rječnička baština i hrvatski europski identitet* pod vođenjem prof. D. Borasa s FFZG-a
- transkripcija i napredna tražilica

vlast

Habdelić, Dictionar
 Kašić, talijansko-hrvatski
 Libellus alphabeticus
 Vrančić, Dictionarium quinque nobilissimarum Europeae linguarum

svi jezici
 hrvatski
 latinski
 talijanski
 njemački
 mađarski

zanemari posebne znakove
 koristi transkripciju

Ubaci poseban znak

Traži bilo gdje
 Traži na početku zapisa
 Traži riječ
 Traži početak riječi

Habdelić

Hrvatski	Latinski	Pregledavanje	Slika
Naulaſzt.	Studio, De induſtria, Ex compoſito.	Pregledavanje	Slika
Naulaſztito.	Præſertim, Potiſſimūm.	Pregledavanje	Slika

Kašić

Hrvatski	Talijanski	Pregledavanje	Slika
Vlastelin, na, M.		Pregledavanje	
Vlastiti, a, o		Pregledavanje	

Libellus

Latinski	Hrvatski	Njemački	Kategorija	Pregledavanje	Slika
Præſertim	Navlaſtito	Meiſtens.	ADVERBIA	Pregledavanje	Slika

Vrančić

Latinski	Talijanski	Njemački	Hrvatski	Mađarski	Pregledavanje	Slika
Auctoritas	Autorita	Vvirdigkeit	Oblaszt	Hatalom	Pregledavanje	Slika
Imperium	Imperio	Gebiet	Vladanye	Birodalom	Pregledavanje	Slika
Peculiaris	Particulare	Eigen,befonder	Vlatit	Tulaydon	Pregledavanje	Slika
Priuilegium	Priuilegio	Sunderbare freyheit	Szloboschina	Szabadsag	Pregledavanje	Slika
Proprius,a,m	Proprio	Eigen	Vlatit	Tulaydon	Pregledavanje	Slika

Pretraživi hrvatski leksikografsko-digitalizacijski projekti

3. Rječnik hrvatskoga ili srpskoga jezika JAZU/HAZU, 1880 – 1976., unutar Digitalne zbirke HAZU-a,
<http://dizbi.hazu.hr/>



- Digitalna zbirka pronalazi samo 3 sveska od ukupno 97
- Optičko prepoznavanje i bez ručne provjere ispravnosti teksta
- Tražilica nije prilagođena za leksikografsko pretraživanje nego je zajednička za cijelu Digitalnu zbirku HAZU-a
- Šteta što nije cjelovito mrežno priređen i objavljen (primjer *Svenska Akademiens Ordbok* iz 1874.)

Naslov Rječnik hrvatskoga ili srpskoga jezika : Sv. 56 : 2
trinaestoga dijela : raspuknuti-
1 razbor

Urednik / Pripeđivač /
Prevoditelj Musulin, Stjepan

Mjesto izdavanja Zagreb
Izdavač Jugoslavenska akademija
znanosti i umjetnosti

Godina izdavanja 1953

Serija Rječnik hrvatskoga ili
srpskoga jezika

Jedinica HAZU Knjižnica

Svezak 13/2

rogatičkom kotaru u Bosni.
Popis žit. bos. i herc. 96.
RASPUVATI, pf. vidi
raspuhati. **RASPUZATI**
SE... | ... razputim (t. j.
sredim, svršim) domaće
posle, a biću paka vas u
bogu. 358b. b. refl. razriješiti
se, oslobođiti **se**, razuz-
duti **se**. U rus. pacnyTaBCH
(razmotati **se**, razvezati **se**;
rastumačiti **se**

li ga s konem rastavio. 2,
585. Na nili živu vatru
oboriše, dokle Turke s konma
rastaviše. 3, 331... | ... ovi
jadan stril od gorke jubavi,
kojim mi u sto dil srdace
rastavi. Vetranič 1, 181. Zač
se jur... | pravi i u pas.
službi; rastaviti **se** isto što
razdijeliti **se**, razdvojiti **se**,
odvojiti **se**, raspasti



Hrvatska mrežna leksikografija danas

- „Hrvatska leksikografija prije je bila više uronjena u europske leksikografske tokove“ (Tafra 2005.)
- Gerbrich de Jong (2014), An Inventory of European scholarly dictionaries, raščlanjeno 125 pronađenih *scholarly* rječnika za 83 jezika (s ciljem pronalaska mrežne leksikografske građe)
- Mnoštvo suvremenih hrvatskih rječnika (Anić Novoga Libera, Anićev veliki rječnik Novoga libera, Šonje LZMK j ŠK, HER Novoga Libera, Školski rječnik IHJJ-a, Anićev školski Znanje, VRH Školske knjige), međutim samo je građa Novoga Libera dostupna na mreži (sada Znanje, <http://hjp.znanje.hr/>).
- VRH ima netipičnu leksikografsku politiku: pretraživanje je omogućeno samo za kupce tiskanoga izdanja, i to ne kroz mrežnu stranicu nego kroz aplikaciju.

Hrvatska mrežna leksikografija danas

- Otvoren i besplatan mrežni pristup svim korisnicima standard je za europske akademske rječnike.
- Mrežna leksikografija hrvatskoga jezika danas nije u skladu s bogatom leksikografskom tradicijom hrvatskoga jezika jučer.
- Jedini mrežno dostupan jednojezični rječnik standardnoga jezika ne nadograđuje se (nema sigurnih naznaka o njegovoj budućnosti).
- „Radi razvijanja sustava dostupnosti leksikografskih informacija potrebno je povezati LZMK s drugim srodnim znanstveno-stručnim kompleksima u RH (osobito s NSK) te osigurati globalno leksikografsko-informatičko polje otvorenog pristupa za sve korisnike.” (Vujić 2015.)
- Opći cilj II. Uspostava javno dostupnoga digitalnoga repozitorija leksikografskih sadržaja (Strategija LZMK 2015. – 2020.)

Hrvatska digitalizacijska leksikografija danas

- indeks gospodarske i društvene digitalizacije (DESI) za 2016. – RH ima najveći rast vrijednosti indeksa od svih članica EU-a, ali i dalje na 24. mjestu (od 28), <https://ec.europa.eu/digital-single-market/progress-country>
- dva zaključka konferencije o digitalnom poslovanju *e-biz 2016* u ožujku 2016.: (1) digitalno gospodarstvo i digitalno društvo ne mogu postojati bez digitalnih podataka, (2) digitalni podatci odavno su izvan isključivo tehnoloških krugova
- pojedinačni napori hrvatskih ustanova ili pojedinaca nisu dovoljni (FFZG, NSK, IHJJ itd.)
- nastavi li se ovakvim tempom rječnička digitalizacija i priprema za mrežno objavljivanje hrvatskih leksikografskih djela, teško da u dogledno vrijeme možemo očekivati ispunjavanje ciljeva
- leksikografska digitalizacija ne bi se trebala zaustaviti samo na TIFF-u i PDF-u, potrebno je imati pretraživ vjerodostojan tekst
- za znanstvene potrebe ni to nije dovoljno: traže se strukturirani podatci obogaćeni metapodatcima

leksikografija i hrvatski jezik u budućnosti

Digitalizacija gramatika (i pravopisa)

- pohvalan rad NSK na digitalizaciji hrvatskih gramatika
- potrebno je osigurati mrežnu pretraživost i starih hrvatskih pravopisnih djela (u razdoblju od 1639. do 2015. hrvatsko pravopisanje opisano je u 42 djela, među kojima su 23 grafijska prijestandardna i 19 standardnih, a računaju li se izdanja, pretisci, dotisci 115 od kojih je 46 prijestandardnih i 69 standardnih, <http://ihjj.hr/stranica/pravopisi-hrvatskoga-jezika/30/>)
- (Svih) pet starih slovenskih pravopisa digitalizirano je u jesen 2012., <http://pravopisi.trojina.si/>, a financiralo je slovensko Ministerstvo kulture.

Moguća uloga NSK u digitalizaciji građe na hrvatskome

- ako je temeljna zadaća NSK „priključiti, stručno obrađivati, čuvati i davati na korištenje sveukupnu građu (tiskane i netiskane publikacije) koja dokumentira i iskazuje sveukupno hrvatsko duhovno stvaralaštvo”, zašto NSK ne bi imao ulogu priključivača digitalne jezične građe (prema *Zakonu o knjižnicama o obveznom primjerku*)
- natkorpus i metatražilica kao kapitalni digitalizirani resurs hrvatskoga jezika (nalik na *Hrvatski arhiv weba* NSK – ali za tiskane izvore)
- postojanje kvalitetnoga korpusa preduvjet je za suvremenu leksikografiju

Hrvatska leksikografija sutra

The European Dictionary Portal, curated by experts in the [European Network of e-Lexicography](#), is your guide to good online dictionaries of European languages. More information...

- Europski rječnički portal (<http://dictionaryportal.eu/en/>) – trenutačno u izgradnji (travanj 2016.), rezultat rada radne grupe 1 COST akcije IS1305 *European Network of e-Lexicography* (<http://www.elexicography.eu/>)
- u dogledno vrijeme središnje leksikografsko mjesto za europske jezike
- stvaranje svijesti o zajedničkoj europskoj leksikografskoj tradiciji
- civilizacijska, kulturna i znanstvena suradnja
- standardizacijska nastojanja europske (meta)leksikografije
- ERP trenutačno nije u mogućnosti indeksirati niti jedan hrvatski rječnik!
- prisutnost hrvatskoga jezika na ERP-u trebao bi biti važan cilj

Zaključci

- ljudska potreba za popisivanjem riječi civilizacijski je uvjetovana i nadilazi tehnologiju i medije na kojima se ostvaruje (bitno starija od otkrića tiskarstva i posve evoluirana u digitalnome dobu)
- digitalizacijska, odnosno mrežna leksikografija hrvatskoga jezika nije zadovoljavajuće razvijena u odnosu na europske kriterije
- digitaliziranost u leksikografiji nije dovoljna, traže se strukturirani i opisani podatci
- težiti otvorenome i besplatnome mrežnom pretraživanju leks. građe
- mjerilo uspješnosti: indeksiranost leksikografskoga sadržaja hrvatskoga jezika u Europskome rječničkom portalu
- proaktivan pristup za leksikografsko-digitalizacijske projekte i potreba za hrvatskom međuinstitucionalnom suradnjom u kojoj bi NSK imala važnu ulogu