

Kristina Štrkalj Despot
 Institut za hrvatski jezik i jezikoslovlje
 Ulica Republike Austrije, HR-10000 Zagreb
kdespot@ihjj.hr

Christine Möhrs
 Institut für Deutsche Sprache
 Postfach 10 16 21, D-68016 Mannheim
moehrs@ids-mannheim.de

POGLED U E-LEKSIKOGRAFIJU

U radu se daje pregled temeljnih pojmova i klasifikacija u području e-leksikografije. Donosi se klasifikacija e-rječnika, prikazuje se leksikografski proces izrade e-rječnika te pregled najraširenijih sustava za izradu rječnika (*DWS*) i sustava za pretragu korpusa (*CQS*). Kao primjer dobre prakse detaljnije se opisuje mrežni rječnik *ellexiko* (Institut za njemački jezik u Mannheimu): prikazuju se njegovi ciljevi i namjena, teorijske i metodološke postavke, moduli te mogućnosti uporabe. Kao moguća osnova za izradu korpusno utemeljenoga e-rječnika hrvatskoga jezika koji bi bio u skladu s najrecentnijim leksikografskim (i uopće lingvističkim) teorijama i praksama prikazuje se rad na mrežnome leksičko-semantičkome repozitoriju hrvatskoga jezika (baza semantičkih okvira, predodžbenih shema, kognitivnih primitiva i leksičkih jedinica) u okviru projekta *Repozitorij metafora hrvatskoga jezika*.

1. Era e-leksikografije¹

Hrvatska leksikografija², čija je povijest dulja od petsto godina, uvijek je bila dio europske leksikografije izrasle na tradiciji latinskih rječnika i uvijek je bila na razini europskih leksikografskih dosega. No danas je, nakon velike i

¹ U radu dajemo prikaz jednojezične e-leksikografije (usmjereni smo samo na jednojezične e-rječnike), dok su dvojezični i višejezični e-rječnici te enciklopedijska leksikografija izvan opsega ovoga rada. Ne ulazimo ni u složenu problematiku odnosa i granica između e-rječnika, e-leksikona i e-enciklopedije.

² Za pregled hrvatske leksikografije kroz povijest v. npr. Tafra i Fink Arsovski (2013).

važne promjene leksikografske paradigme koju je donijela era e-rječnika, a u koju se nismo na prikladan način uključili, vrlo vidljiv zaostatak u broju i kvaliteti mrežnih rječnika u odnosu na druge europske, pa i samo slavenske jezike.³ Naime, područje e-leksikografije danas je već toliko razvijeno da se smatra zasebnim istraživačkim područjem, a u hrvatskoj leksikografiji gotovo je potpuno izostao i prikaz razvoja i temeljnih pojmova toga područja, i što je još važnije, praktična primjena novih istraživačkih uvida i ugledanje u primjere dobre prakse.

U radu se stoga daje pregled temeljnih pojmova i klasifikacija u području e-leksikografije. Donosi se klasifikacija e-rječnika, prikazuje se leksikografski proces izrade e-rječnika te pregled najraširenijih sustava za izradu rječnika (*DWS*) i sustava za pretragu korpusa (*CQS*). Osim na relevantnoj literaturi iz područja e-leksikografije te na detaljnu uvidu u jedan od primjera dobre prakse (*elexiko* Instituta za njemački jezik), rad se temelji i na uvidima i rezultatima nastalima u okviru COST akcije *European Network of e-Lexicography*, dalje *ENeL* (<http://www.elexicography.eu/>)⁴.

Elektronička leksikografija počela se razvijati sredinom prošloga stoljeća uglavnom pod nazivom računalna leksikografija (*computer/computational lexicography*), a rječnici su se nazivali strojno čitljivima, a ne elektroničkim rječnicima ili e-rječnicima⁵, što je danas uobičajen naziv. Prema Granger (2012), u to je vrijeme obzor istraživačkoga područja bio ograničen na „uporabu računala” u leksikografskome poslu izrade rječnika (v. npr. Knowles 1990). U začetcima elektroničke leksikografije ni leksikografi ni korisnici rječnika nisu se služili računalom: leksikografi su nastavili s „papirnatom” praksom leksikografske obrade (rabeći doduše sve strukturiranije i strukturiranije forme), a računalni su stručnjaci onda obrađivali podatke i pretvarali ih u digitalni oblik. Prvi rječnik koji je načinio ključan iskorak od strojno čitljiva rječnika prema leksičkoj bazi podataka, koja je značila i golemu pomoć leksikografima u izradi rječnika, bio je *Longman Dictionary of Contemporary English* (1978.), a podjednako je re-

³ Eri e-rječnika i e-leksikografije prethodila je „korpusna revolucija”, kojoj su se hrvatski jezikoslovci priključili u samim začetcima (tako je npr. 1999. nastao potpuno korpusno utemeljen *Hrvatski čestotni rječnik* Milana Mogaša, Maje Bratanić i Marka Tadića). I korpusno su se i računalno jezikoslovlje nastavili zadovoljavajuće razvijati, no uglavnom je izostala sustavna leksikografska obrada dostupnih korpusa (izuzetak je *Hrvatski školski rječnik* koji se temelji na *Hrvatskome jezičnom korpusu*, računalnome korpusu tekstova hrvatskoga jezika koji je sastavnica institutskoga programa *Hrvatska jezična riznica*; pri obradbi obrađivači su provjeravali značenja i uporabu riječi u korpusu, ali primjeri nisu izravno preuzimani iz korpusa).

⁴ Obje su autorice članice Upravljačkoga odbora navedene akcije.

⁵ E-rječnikom smatramo rječnik u digitalnom obliku. Za podvrste e-rječnika v. sljedeće poglavlje.

volucionaran bio i projekt *COBUILD* Johna Sinclaira⁶ s početka osamdesetih godina prošloga stoljeća, čije je najveće postignuće bilo izrada i analiza elektroničkoga korpusa suvremenih tekstova, što je rezultiralo rječnikom objavljenim 1987. (*Collins COBUILD English Language Dictionary*).

Prošlo je zatim nekoliko godina do pojave prvih rječnika u obliku CD-ROM-a koji su konačno omogućili i korisnicima da otkriju prednosti rječnika u elektroničkome obliku, nakon čega se, od devedesetih godina prošloga stoljeća, razvoj znatno ubrzao prateći razvoj tehnologija i doveo do pojave prvih mrežnih rječnika početkom tisućljeća, a zatim, u današnje doba, i do potpunoga izjednačavanja pojmova leksikografije i e-leksikografije te do brojnih predviđanja o skorom potpunom nestanku papirnatih rječničkih izdanja (Granger 2012). Ono što danas područje e-leksikografije čini zasebnim istraživačkim područjem jest golem broj mrežnih (popularno *online*) rječnika te golema bibliografija istraživačkih doprinosa usmjerenih na to područje.⁷ No pitanje postojanja leksikografske teorije ostalo je aktualno unatoč promjeni paradigme i medija. Tako primjerice predstavnici aarhuske škole (s Bergenholtzom kao najistaknutijim predstavnikom, v. ur. Fuertes-Olivera i Bergenholtz 2011) drže da je leksikografija autonomna disciplina u odnosu na lingvistiku i da pripada informacijskim znanostima te, dakako, ima razvijenu vlastitu teoriju (ili niz različitih teorija)⁸ i praksu (v. npr. Tarp 2012). Naslanjaju se pritom, osim na dansku, i na rusku tradiciju (npr. Scerba 1940, Sorokoletov 1978), njemačku (npr. Duda i dr. 1986, Wiegand 1999, ur. Weber 1996 itd.) te na leksikografe poput Gouwsa (2011) ili Zguste (1992). No neki leksikografi koji pripadaju anglosaskoj tradiciji (npr. Atkins i Rundell 2008 te Béjoint 2010), slijedeći primjerice Landaua (1984), drže da se ne može govoriti o teoriji leksikografije, nego samo o praksi izrade rječnika.⁹

⁶ U velikom međunarodnom višejezičnom leksikografskom projektu Vijeća Europe koji je vodio John Sinclair sudjelovala je i Hrvatska (voditeljica je bila Maja Bratanić).

⁷ Iscrpnj popis mrežnih rječnika i bibliografskih jedinica koji pripadaju području e-leksikografije jesu npr. OBELEX (<http://www.owid.de>), bibliografski projekt (baza podataka) Instituta za njemački jezik u Mannheimu te Hartmann (2007).

⁸ Aarhuska škola zastupa funkcijsku teoriju leksikografije.

⁹ Usp. Béjoint (2010: 381): „I simply do not believe that there exists a theory of lexicography, and I very much doubt that there can be one. Those who have proposed a general theory have not been found convincing by the community, and for good reasons. A theory is a system of ideas put forward to explain phenomena that are not otherwise explainable. A science has a theory, a craft does not. All natural phenomena need a theory, but how can there be a theory of the production of artefacts? There are theories of language, there may be theories of lexicology, but there is no theory of lexicography. Lexicography is about all a craft, the craft of preparing dictionaries, as well as an art, as Landau (2001) says. It may be becoming more scientific, but it has not become a science.”

2. Klasifikacija e-rječnika

U de Schryver (2003: 146) ujedinjeno je nekoliko različitih klasifikacija elektroničkih rječnika te je ponuđena klasifikacija koja se vrlo često citira i široko je prihvaćena. Ondje se primarno razlikuju: 1.) rječnik na samostojećemu računalu kojim se ne može koristiti više korisnika istodobno, nego jedan po jedan korisnik i 2.) rječnik na mrežnome računalu kojim se može koristiti više korisnika istodobno. Sekundarna podjela drugoga tipa rječnika nadalje uključuje: 1.) unutarmrežni (intranetski) i 2.) mrežni (ili internetski) rječnik (v. tablicu 1).

ELEKTRONIČKI RJEČNIK			
Rječnik na samostojećemu računalu (jedan po jedan korisnik)		Rječnik na mrežnome računalu (više korisnika istodobno)	
Rječnik na dlanovniku	Rječnik na CD-ROM-u	Unutarmrežni (intranetski rječnik)	Mrežni (internetski) rječnik
Jedan korisnik koristi se dlanovnikom kako bi pristupio rječniku pohranjenu na malenome disku.	Jedan korisnik koristi se prijenosnim/stolnim računalom kako bi pristupio rječniku pohranjenu na CD-ROM-u.	Grupa korisnika koristi se prijenosnim/stolnim računalom kako bi pristupila rječniku pohranjenu na lokalnome poslužitelju.	Korisnici širom svijeta koriste se prijenosnim/stolnim računalima, mobilnim uređajima itd. kako bi pristupili rječniku pohranjenu na mrežnome poslužitelju.

Tablica 1. Klasifikacija elektroničkih rječnika (prema de Schryver 2003 i Klosa 2013)

Klasifikacija mrežnih rječnika	
ZNAČAJKE	TIP RJEČNIKA
Izvoran oblik	Izvorno objavljen u tiskanu obliku Izvorno objavljen kao elektronički rječnik na samostojećemu računalu Izvorno objavljen kao mrežni rječnik ¹⁰

¹⁰ Takvi se rječnici u literaturi na engleskome jeziku nazivaju *born digital dictionaries*.

Dovršenost	Statični rječnik Dinamični rječnik
Hipertekstualizacija	Hipertekstualizirani rječnik Rječnik bez hipertekstualizacije
Interakcija s korisnicima	Rječnik s mogućnošću interakcije Rječnik bez mogućnosti interakcije
Multimedijalnost	Rječnik s tekstom, ilustracijama i dijagramima Rječnik s tekstnim i zvučnim zapisima Rječnik s tekstnim, (zvučnim zapisima) i videozapisima Rječnik bez multimedijjskih sadržaja
Pristup rječniku	Rječnik kojemu se pristupa s pomoću pomicanja prema dolje u nizu rječničkih članaka Rječnik kojemu se pristupa s pomoću popisa natuknica s hiperpoveznicama Rječnik kojemu se pristupa s pomoću tražilice Rječnik s kombiniranim pristupom itd.

Tablica 2. Klasifikacija mrežnih rječnika (prema Klosa 2013 te prema Storrer i Freese 1996, Storrer 1998 i Storrer 2001)

Izvoran oblik rječnika ponajviše utječe na izgled mrežnoga rječnika, stupanj i metode hipertekstualizacije, stupanj interakcije s korisnicima, uključivanje multimedijjskih sadržaja i odabira pristupa rječničkomu sadržaju (Klosa 2013). Dovršenost rječnika povezana je dakako ponajviše s računalnoleksikografskim procesom. Tako primjerice mrežni rječnici u izradi nisu statični, nego su uvijek dinamični objekti, baze podataka koje se neprestano mijenjaju.¹¹ Dinamični mrežni rječnici mogu nastati ili iz izvorno tiskanoga rječnika koji je prvo pretvoren u digitalni oblik, a zatim se digitalna inačica konstantno širi i obogaćuje (npr. *Oxford English Dictionary*), ili mogu biti izvorno digitalni (*born digital*), npr. *elexiko* Instituta za njemački jezik, koji ćemo u ovo me radu detaljno opisati.

¹¹ S obzirom na dovršenost rabe se dva para naziva: *dovršeni rječnik* vs. *rječnik u izradi*, prema nazivima koje je uveo Schröder (1997: 60): *Abschlusswörterbuch* (*completed dictionary*) vs. *Ausbauwörterbuch* (*dictionary under construction*) te *statični rječnik* vs. *dinamični rječnik*, prema nazivima koje je uveo Lemberg (2001: 81): *statisches Wörterbuch* (*static dictionary*) vs. *dynamisches Wörterbuch* (*dynamic dictionary*). Mi kao primarne rabimo nazive *statični* vs. *dinamični* jer nam se čini da bolje odlikavaju najvažnije karakteristike tih dviju vrsta rječnika.

Danas je vrlo često udruživanje više mrežnih rječnika u *rječničke portale*, koji po broju mrežnih rječnika koje sadržavaju mogu znatno varirati.¹² *Rječničke mreže*, u kojima su rječnici visokopovezani na mikrostrukturnoj razini, obično sadržavaju manji broj rječnika zbog zahtjevnosti njihova mikrostrukturnoga povezivanja, dok *rječničke zbirke* i *tražilice* obično sadržavaju velik broj rječnika¹³ zbog manje računalne zahtjevnosti njihove izrade (v. Engelberg i Müller-Spitzer 2013).

S obzirom na suvremene hrvatske leksikografske dosege, može se predložiti i drukčija klasifikacija primjerenija specifičnostima postignutih rezultata. U tome se smislu mogu razlikovati:¹⁴

1. rječnici vođeni korpusom: npr. *Hrvatski čestotni rječnik* (Moguš, Bratanić i Tadić 1999), *Rječnik Marulićeve Judite* (Moguš 2001) itd.

2. rječnici utemeljeni na korpusu: npr. *Školski rječnik hrvatskoga jezika* (Birtić i dr. 2012) itd.

3. rječnik je dostupan u digitalnom obliku (zatvoreni pristup: CD/DVD, kod): *Prvi školski rječnik hrvatskoga jezika* (Čilaš Šimpraga, Jojić i Lewis 2008), *Veliki rječnik hrvatskoga jezika* (Jojić i dr. 2015) itd.

4. rječnik je dostupan na mreži (otvoreni pristup): npr. *Hrvatski jezični portal* (<http://hjp.novi-liber.hr/>) itd.

5. leksičke baze otvorenoga pristupa (uključujući i terminološke): npr. CroWN (<http://meta-share.ffzg.hr/repository/browse/croatian-wordnet>), Meta-Net.HR (<http://ihjj.hr/metafore>), Struna (<http://struna.ihjj.hr>), HRANA (<http://hrana.ffzg.hr/>) itd.¹⁵

6. rječnički portali (uključujući i terminološke):

Portal hrvatske rječničke baštine (<http://cro dip.ffzg.hr/>)¹⁶, s pomoću kojega se može pretraživati i pregledavati digitalizirani tekst te pregledavati slika izvornika dvadesetak rječnika hrvatskoga jezika iz razdoblja od 16. do konca 19. stoljeća.¹⁷

¹² O rječničkim portalima v. npr. Klosa, Lemnitzer i Neumann (2008).

¹³ Tako npr. jedna od najpoznatijih rječničkih zbirki na svijetu *YOURDICTIONARY.COM* sadržava više od 2500 rječnika i gramatika na više od 300 jezika (v. Engelberg i Müller-Spitzer 2013).

¹⁴ Zahvaljujemo recenzentu rada na prijedlogu ove klasifikacije.

¹⁵ Hrvatski je jezik uključen i u mnoge međunarodne terminološke baze koje su javno dostupne: npr. EMITEL (<http://www.emitel2.eu>) i *Multilingual Archival Terminology* (<http://icar-chives.webbler.co>).

¹⁶ Ovaj je portal rezultat znanstvenoga projekta 130464 *Hrvatska rječnička baština i prikaz rječničkog znanja*. Glavni je istraživač projekta dr. sc. Damir Boras, izv. prof.

¹⁷ Cjelovitu popisu rječnika može se pristupiti na poveznici: <http://cro dip.ffzg.hr/rjecnici.pdf>.

Hrvatski jezični portal (<http://hjp.novi-liber.hr/>) nije rječnički portal u užemu smislu, nego je to dinamičan mrežni rječnik hrvatskoga jezika nastao digitalizacijom nekoliko izvorno tiskanih leksikografskih djela.¹⁸

Hrvatski terminološki portal (<http://nazivlje.hr/>) Instituta za hrvatski jezik i jezikoslovlje u jedinstvenoj portalskoj tražilici zajednički pretražuje građu iz hrvatske terminološke baze Struna, baze podataka s već objavljenim rječnicima, glosarima i leksikonima, nazive iz resursa Leksikografskoga zavoda Miroslav Krleža te nazive iz norma Hrvatskoga zavoda za norme.

Nažalost, nemamo e-rječnika hrvatskoga jezika izrađenoga prema najsvremenijim leksikografskim standardima, dakle takvoga koji bi uključivao uporabu suvremenih alata za pretragu i ekstrakciju podataka iz korpusa.

3. Leksikografski proces izrade e-rječnika

Leksikografski proces izrade tiskanih rječnika zadovoljavajuće je opisan u okvirima leksikografske teorije (npr. Dubois 1990, Landau 1984, Riedel i Wille 1979, Zgusta 1971). S obzirom na to da leksikografski proces izrade e-rječnika zahtijeva drukčiji pristup, to je istraživačko pitanje postalo vrlo važno u recentnoj leksikografskoj teoriji (v. npr. Wiegand 1999, de Schryver i Prinsloo 2000a i 2000b, Müller-Spitzer 2003, Engelberg, Klosa i Müller-Spitzer 2009, Klosa 2013, Engelberg 2014).

U leksikografskome procesu izrade tiskanoga rječnika uglavnom se jasno razlikuju tri faze: planiranje, pisanje i izdavanje, koje slijede jedna za drugom (v. npr. Landau 1984: 227). U izradi e-rječnika granice među pojedinim fazama teško je, a katkad i nemoguće jasno odrediti, a faze ne slijede sukcesivno jedna za drugom nego su nerijetko simultane ili čak obrnute od slijeda na koji smo navikli pri izradi tiskanih rječnika¹⁹. Klosa (2013) uvodi detaljniju razradu leksikografskoga procesa za mrežne rječnike prepoznajući šest faza računalnoga leksikografskoga procesa za (dinamične) mrežne rječnike: 1. *priprema*, 2. *prikupljanje podataka*, 3. *računalni okvir*, 4. *procesiranje*

¹⁸ Na gore navedenoj mrežnoj stranici navodi se da je rječnička baza Hrvatskoga jezičnoga portala nastala na temelju rječničkih i leksikografskih izdanja Novoga Libera u proteklih 15 godina: *Rječnik hrvatskoga jezika*, V. Anić (1991); *Pravopis hrvatskoga jezika*, V. Anić i J. Silić (2001); *Veliki rječnik hrvatskoga jezika*, V. Anić (2003); *Kronologija – Hrvatska, Europa, Svijet*, (grupa autora, urednik I. Goldstein); *Rječnik stranih riječi*, Anić i Goldstein (1998); *Hrvatski enciklopedijski rječnik*, uređivački odbor: prof. dr. Vladimir Anić, prof. dr. Ranko Matasović, prof. dr. Ivo Pranjković, dr. Dunja Brozović Rončević, prof. dr. Ivo Goldstein, Slavko Goldstein, mr. Ljiljana Jojić, Ljiljana Cikota (2003).

¹⁹ Mrežni se rječnici nerijetko objavljuju dio po dio, pa tako objavljivanje rječnika može početi prije no što je faza pisanja završila.

podataka, 5. analiza podataka, 6. priprema za mrežnu objavu.²⁰ Za statične (završene) mrežne rječnike može se identificirati i sedma faza, faza *čuvanja i održavanja* (u engl. literaturi *afterlife*), koja uključuje održavanje rječnika nakon završetka projekta.²¹

U Tiberius i Krek (2014) postavlja se niz pitanja na koja suvremena e-leksikografija (ili konkretnije *ENeL*) tek treba odgovoriti: Može li se proces izrade različitih mrežnih rječnika doista podvesti pod navedenih sedam faza? Kako se nositi s problemom kontrole inačica i arhiviranja različitih inačica rječnika? S obzirom na to da za mrežne rječnike abecedni poredak nije nužan, kojim se kriterijima koristiti pri odabiru reda kojim će se natuknice obrađivati i prezentirati? Kakav je odnos vizualizacije i leksikografskoga procesa? Kako i u kojoj fazi u leksikografski proces uključiti korisnike? Kako automatska ekstrakcija leksikografskih podataka iz korpusa utječe na leksikografski proces?

3.1. Sustavi za izradu rječnika (*DWS*) i sustavi za pretragu korpusa (*CQS*)

U istraživanju koje su u okviru *ENeL*-a proveli Simon Krek („Jožef Stefan” Institut, Ljubljana, Slovenija), Andrea Abel (EURAC, Bolzano/Bozen, Italija) i Carole Tiberius (Institute for Dutch Lexicology, Leiden, Nizozemska) i u kojemu su sudjelovali leksikografi iz 25 europskih zemalja²² pokazalo se kako 70 % ispitanika rabi specijalizirane softvere za izradu rječnika (*Dictionary Writing System*) (v. Krek, Abel i Tiberius 2014). Istraživanje je pokazalo da se manji broj institucija koristi gotovim sustavima za izradu rječnika (poput *T-lexa* ili *iLexa*), a najveći broj institucija koristi se novim ili prilagođenim softverskim rješenjima, i to najčešće prilagođenim bazama (uglavnom *MySQL*), prilagođenim *XML editorom* (uglavnom *oxygen*) ili prilagođenim *wikijem*. Većina sustava za izradu rječnika koji su u uporabi uključuje kontrolu i validaciju rječničke strukture ili rječničke sintakse te provjeru konzistentnosti.

Istraživanje je međutim pokazalo kako većina sustava za izradu rječnika ne uključuje integrirani pravopisni provjernik ni bilo kakvu integraciju sa sustavom za pretragu korpusa (*Corpus Query System*), a samo 1 % ispitanika koristi se sustavom koji integrira sustav za izradu rječnika i sustav za pretragu korpusa u jednome softverskome rješenju.

²⁰ Jasno je da je ova podjela vrlo shematične i praktične naravi i da ne uključuje detalje svake od navedenih faza, od kojih je možda najsloženija upravo faza pripreme u kojoj treba osmisliti makrostrukturu i mikrostrukturu rječnika (v. o tome npr. Gouws 2011).

²¹ V. Tiberius i Krek (2014).

²² V. popis zemalja sudionica i ostale detalje istraživanja na http://www.elexicography.eu/wp-content/uploads/2015/04/ENeL_WG3_Vienna_DWS_CQS_final_web.pdf.

Specijaliziranim softverom za pretragu korpusa koristi se 65 % ispitanika, a pritom velika većina njih rabi *Sketch Engine*, odnosno *No Sketch Engine*.

4. Primjer dobre prakse: *elexiko* – mrežni rječnik suvremenoga njemačkoga jezika

Mrežni rječnik *elexiko* (v. ur. Hass 2005a, Klosa, Schnörch i Storjohann 2006, Engelberg, Klosa i Müller-Spitzer 2009, ur. Klosa 2011, Klosa 2014) projekt je Instituta za njemački jezik (*Institut für Deutsche Sprache, IDS*) u Mannheimu i pristup mu je slobodan. Izvorno je digitalan i izrađen upravo za objavljivanje na internetu. Objavljen je u sklopu rječničkoga portala, odnosno leksikografskoga informacijskoga sustava OWID²³. *Elexiko* je najrelevantniji leksikološko-leksikografski projekt koji tumači i dokumentira suvremeni njemački jezik.

4.1. Namjena i ciljevi

Elexiko se može definirati kao mrežni e-rječnik (ili internetski rječnik), ali treba imati na umu da je on mnogo više od rječnika u tradicionalnome smislu riječi, iako sadržava opise značenja i uporabe leksema poput tradicionalnih rječnika. *Elexiko* je i hipertekstni rječnik i informacijski sustav leksičkih podataka (Storjohann 2005b, Klosa, Schnörch i Storjohann 2006).

Rječnik se sastoji od indeksa koji čini oko 300 000 kratkih rječničkih članaka s podacima o zapisu, inačicama zapisa, vrsti riječi i citatima automatski izlučenima iz korpusa *elexiko*.

Korpus *elexiko* služi kao temelj svih leksikografskih podataka koji se u rječniku donose. Korpus je nastao na temelju njemačkoga referentnoga korpusa (*Deutsches Referenzkorpus DEREKO*) Instituta za njemački jezik i čini primarni rječnički izvor. Osim primarnoga izvora, rabe se i sekundarni rječnički izvori (ostali tiskani ili e-rječnici) te tercijarni izvori (gramatički priručnici) (usp. Klosa 2013: 520). Prikupljanje i inkorporiranje multimedijских elemenata (ilustracija i zvučnih zapisa) ostvareno je u 5. fazi leksikografskoga procesa (u fazi analize podataka).

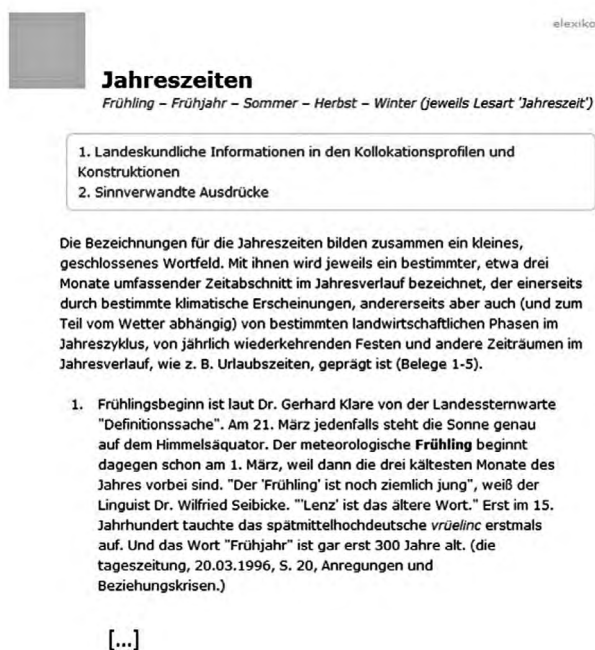
Osim navedenoga, *elexiko* sadržava više od 1000 detaljno razrađenih članaka za visokofrekventne natuknice, u kojima je osobita pozornost usmjerena opširnomu semantičko-pragmatičkomu opisu leksičkih jedinica u stvarnoj jezičnoj uporabi.

²³ www.owid.de

Projektne su leksikografi najdulje radili na leksikografskome modulu s nazivom *Lexikon zum öffentlichen Sprachgebrauch (Rječnik javne jezične uporabe)*, koji sadržava oko 2000 rječničkih članaka (natuknice su također izabrane po načelu čestote u korpusu *elexiko*).

Nakon što je modul *Rječnik javne jezične uporabe* završen, leksikografi su pristupili izradi ostalih rječničkih modula, primjerice modulu koji pristupa fenomenom paronimije, i to iz leksikološke, korpusnolingvističke i leksikografske perspektive (v. npr. Storjohann 2013; v. projektnu mrežnu stranicu: <http://www1.ids-mannheim.de/lexik/paronymwoerterbuch.html>).

Leksikografi *elexika* razvili su novi oblik rječničkoga članka koji upućuje na cijelu grupu rječničkih članaka.²⁴ Analiziraju se kolokacije i paradigmatički partneri dviju ili više već obrađenih natuknica kako bi se opisale sličnosti i razlike među različitim člancima povezanima primjerice u semantičkome ili tematskome polju (v. sliku 1).



The image shows a screenshot of a web page from the 'elexiko' database. The page title is 'Jahreszeiten' with a subtitle 'Frühling – Frühjahr – Sommer – Herbst – Winter (jeweils Lesart 'Jahreszeit')'. Below the title is a list of two items: '1. Landeskundliche Informationen in den Kollokationsprofilen und Konstruktionen' and '2. Sinnverwandte Ausdrücke'. A text block follows, explaining that the terms form a closed word field and are used to denote specific time periods. A numbered list item '1.' provides a detailed definition of 'Frühling' (spring) based on a source from 1996, mentioning its historical and linguistic context. The page ends with an ellipsis [...].

Slika 1. Primjer članka *Jahreszeiten* ('godišnja doba')

²⁴ U projektu takvi se članci nazivaju *Wortgruppenartikel* (<http://www.owid.de/wb/elexiko/gruppen/index.html>).

Osim rada na novim modulima (v. poglavlje 4.3.) postojeći će se moduli održavati i dalje nadopunjavati, a osobita se pozornost posvećuje i dugoročnom spremanju podataka (v. Möhrs 2014).

4.2. Teorijska i metodološka podloga

Opširan korpus njemačkoga jezika nesumnjivo je ne samo primaran izvor nego i temelj leksikografske interpretacije u okviru projekta *lexiko*. Izrada rječnika ne teče abecednim redom, nego se natuknice grupiraju u razrede i polja na temelju čestote i distribucije riječi u *lexiko* korpusu. Popis natuknica nastao je isključivo na temelju *lexiko* korpusa i objavljen je na internetu prije no što je počeo rad na leksikografskim modulima. Kako smo spomenuli, taj popis sadržava 300 000 jednorječnih natuknica za koje su dani automatski ili poluautomatski generirani podatci iz korpusa koji uključuju zapis, inačice zapisa, osnovne gramatičke podatke i automatski odabrane citate iz korpusa (v. sliku 2).

The image shows a screenshot of the 'Briefkasten' entry in the 'lexiko' corpus. The entry is titled 'Briefkasten' and is categorized under 'Lesartenübergreifende Angaben'. It contains several sections: 'Orthografie' (Normgerechte Schreibung: Briefkasten, Worttrennung: Brief|kas|ten), 'Wortbildungsprodukte' (automatisch ermittelt) weiter >, 'Grammatik- und Kookkurrenzprofil' (Grammatische Angaben: canoo.net, Kookkurrenzprofil: CCDB), 'Verteilung im lexiko-Korpus' (Zahl der Quellen: 16 von 31, Zahl der Jahrgänge: 23 von 63, Frequenzschicht: VIII (5.001–10.000 mal belegt)), and 'Belege (automatisch ausgewählt)'. The 'Belege' section includes a text snippet: 'Erst geht es schnurgerade durch den Wald. An jeder Kreuzung mit den Jagen-Wegen wird gebimmelt, der Grenzweg ist (was sonst?) die Trennlinie zwischen Berlin und dem Bundesland Brandenburg. Dann, an der Berliner Straße, wird es kleinstädtisch. Wer als "Wessi" aus der Neuwested-Gegend mit Eichenallee und Ebereschenallee stammt, weiß nun, weshalb manchmal im Jahr Irläufer ohne Postleitzahl im Briefkasten schlummern. Episteln, die aber hierher, nach Schönblick gehören! (L59/SEP.65310 Berliner Morgenpost, 19.09.1999, S. 11, Ressort: BERLIN; Wo der Wald vor lauter Bäumen nicht zu sehen ist)'. Below this is another text snippet: 'Hand auf' s Herz: Hat' s das schon irgendwann einmal gegeben? Der bunte Fiat-Stimmzettel kommt mit der Post, ebenso unsere Wahlbroschüre: eine Extra-Ausgabe der "Kärntner Krone". Alle Kandidaten, alle Informationen. Ja, es könnte Sie recht teuer kommen, am 10. März nicht in Ihren Briefkasten zu schauen... Heute stellen wir Ihnen unseren dritten Bewerber um Ihre Gunst vor; ein bewährter Kandidat mit einem hübschen Namen: Panda. Die "tolle Kiste", die nicht nur praktisch ist, sondern auf die Sie sich auch verlassen können. (094/MAR.18760 Neue Kronen-Zeitung, 02.03.1994, S. 15)'. On the right side of the screenshot, there is a list of 'Sense-independent information' with brackets pointing to the corresponding sections: 'orthography' points to the Orthografie section, 'automatically extracted word formation products' points to the Wortbildungsprodukte section, 'information about grammar and co-occurrence profil' points to the Grammatik- und Kookkurrenzprofil section, 'information about frequency' points to the Verteilung im lexiko-Korpus section, and 'examples from the lexiko-corpus' points to the Belege section.

Slika 2. Primjer automatski generiranoga članka *Briefkasten* (‘poštanski sandučić’)

Za razliku od automatski generiranih članaka, leksikografski detaljno obrađeni članci (poput onih u modulu *Rječnik javne jezične uporabe*) uključuju i detalje o morfologiji i tvorbi riječi, značenjima i njihovim odnosima (v. sliku 3).

The screenshot shows a digital dictionary entry for the word "Dienstleistung". The interface is divided into two main sections: "Lesartenübergreifende Angaben" (Sense-independent information) and "Lesartenbezogene Angaben" (Sense-related information). The word "Dienstleistung" is highlighted in a grey box at the top left. The "Sense-independent information" section includes details on orthography (Normgerechte Schreibung: Dienstleistung; Worttrennung: Dienst|leistung), word formation (Basis: Dienst leisten (Verbalphrase); Suffix: -ung), and etymology (Wandel 1700 bis 1945: -, Wandel seit 1945: -). The "Sense-related information" section includes three sense definitions: "Wirtschaftsgut" (economic good), "Arbeit" (work), and "Bereich" (area), each with a brief description. A final section titled "Zum Zusammenhang der Lesarten" (On the relationship between the senses) explains that "Wirtschaftsgut" is a metonymy of "Arbeit" and "Bereich" is an abstraction of "Arbeit".

Dienstleistung

Lexikon

Lesartenübergreifende Angaben Sense-independent information

- Orthografie**
 Normgerechte Schreibung: Dienstleistung
 Worttrennung: Dienst|leistung
 orthography
- Wortbildung**
 Wortbildungsart/-typ: explizite Derivation
 Basis: Dienst leisten (Verbalphrase)
 Suffix: -ung
 word formation
- Herkunft und Wandel**
 Etymologische Angaben: anzeigen »
 Wandel 1700 bis 1945: -
 Wandel seit 1945: -
 etymology and change of meaning
- Wortbildungsprodukte**
 (automatisch ermittelt) weiter »
 automatically extracted word formation products

Lesartenbezogene Angaben Sense-related information

Lesart **'Wirtschaftsgut'** weiter »
 Mit *Dienstleistung* bezeichnet man ein Wirtschaftsgut besonderer Art, das als Ergebnis menschlicher Tätigkeit entsteht, die nicht unmittelbar mit Herstellung von Sachgütern zusammenhängt, z. B. im Bereich Gastronomie oder Tourismus. Die Erbringung einer *Dienstleistung* geht gleichzeitig mit dem Konsum einher.

Lesart **'Arbeit'** weiter »
 Mit *Dienstleistung* bezeichnet man eine Handlung, bei der eine Person(engruppe) (z. B. ein Unternehmen oder eine Bank) eine gewisse Arbeit verrichtet, die nicht unmittelbar mit der Herstellung von Sachgütern zusammenhängt, wie etwa Beratung, Reparaturen oder Einkaufsservice. Eine *Dienstleistung* wird in der Regel gegen Entgelt oder Vergütung angeboten, kann aber auch freiwillig und unentgeltlich erbracht werden.

Lesart **'Bereich'** weiter »
 Mit *Dienstleistung* beziehen sich Sprecher auf einen Wirtschaftsbereich oder einen Tätigkeitsbereich eines Unternehmens, in dem Personen(gruppen) vergütete Arbeit verrichten, ohne dabei industrielle Produkte herzustellen oder sie zu verkaufen.

Zum Zusammenhang der Lesarten
 Die Lesart 'Wirtschaftsgut' ist eine Metonymisierung der Lesart 'Arbeit'. Die Lesart 'Bereich' ist eine Abstrahierung der Lesart 'Arbeit'. Die Lesarten 'Wirtschaftsgut' und 'Arbeit' liegen semantisch dicht beieinander und werden oft gleichzeitig realisiert.
 on the relationship between the senses

Slika 3. Ulomak početnoga ekrana za članak *Dienstleistung* ('usluga')

Leksikografski detaljno obrađeni članci nude i širok spektar podataka povezanih uz značenja leksema: definiciju značenja, kolokacije, sintagmatske uzorke (konstrukcije), riječi slične po smislu, pragmatičke informacije (osobitosti uporabe) i gramatiku (v. sliku 4).

The screenshot shows a web interface for the word 'Dienstleistung'. At the top, the word is displayed in a large, bold font, with 'Lesart: 'Wirtschaftsgut'' below it. A navigation bar includes 'zur Übersichtseite' and 'Lesarten im Überblick'. A menu of categories is visible: 'Bedeutungs-erläuterung', 'Kollo-kationen', 'Konstruk-tionen', 'Sinnverwandte Wörter', 'Gebrauchs-besonderheiten', and 'Grammatik'. The main content area is titled 'Erläuterung der Bedeutung / Funktion' and contains a definition: 'Mit **Dienstleistung** bezeichnet man ein Wirtschaftsgut besonderer Art, das als Ergebnis menschlicher Tätigkeit entsteht, die nicht unmittelbar mit Herstellung von Sachgütern zusammenhängt, z. B. im Bereich Gastronomie oder Tourismus. Die Erbringung einer **Dienstleistung** geht gleichzeitig mit dem Konsum einher.' Below this is a 'Belege anzeigen' button. Further down, there are sections for 'Sachinformationen' and 'Weitere Informationen', with the latter providing a citation: 'Dienstleistungen können allgemein als ökonomische Güter zur Befriedigung menschlicher Bedürfnisse bestimmt werden. Anders als Sachgüter zeichnen sie sich als an Personen gebundene, Nutzen stiftende Leistungen [...] aus. Brockhaus in Text und Bild Edition (2002). Mannheim (CD-ROM), unter dem Stichwort Dienstleistungsgesellschaft.' At the bottom, it indicates 'Wortklasse: Individuativum'.

Slika 4. Primjer podataka povezanih uz značenje u članku *Dienstleistung* (‘usluga’)

U procesu izrade i objave rječnika na internetu projekt se koristio mnogobrojnim tehnologijama i softverskim alatima poput sustava za pretragu i obradu korpusa (*COSMAS II*)²⁵, koji je također razvijen u IDS-u, te osobito softverskoga paketa sustava za obradu kolokacija (*Statistical collocation analysis and clustering*)²⁶. Osim toga, projekt se koristio i korpusnolingvističkim

²⁵ Za *COSMAS II*, v. <http://www.ids-mannheim.de/cosmas2/>.

²⁶ Za softverski paket *Statistical collocation analysis and clustering*, v. <http://www1.ids-mannheim.de/kl.html>.

istraživačko-razvojnim radnim paketom *CCDB*²⁷ (također razvijenim u IDS-u). Softver za istraživanje korpusa i analizu supojavljivanja (*co-occurrence analysis*) rabi se u mnogobrojnim djelomično ili potpuno korpusno utemeljenim istraživanjima²⁸ u okviru praktične radne procedure.

Za razlučivanje značenja (disambiguaciju) visokofrekventnih polisemnih riječi unutar projekta *ellexiko* razvijena je tehnika utemeljena na empirijskim i teorijskim zasadama. Mnogobrojni su leksikografski preduvjeti nužni za primjenu ovih zahtjevnih tehnika: razrađena teorija, korpusi, softveri za obradu podataka, i, što je osobito važno, lingvistička kompetencija interpretacije podataka (Storjohann 2003: 755).

Rabeći *COSMAS II* i analizu supojavljivanja leksikografi razgraničuju polisemne riječi na trima razinama: kolokacijskoj razini, razini ključne riječi u kontekstu (*KWIC*) i na razini teksta te tako dobivaju detaljan semantički opis (usp. Storjohann 2003: 755, Hass 2005b).

U drugome koraku značenja dobivena s pomoću potpuno korpusno utemeljene metode (*corpus-driven approach*) uspoređuju se sa značenjima u drugim rječnicima. Ako na temelju takva uvida leksikograf zaključi da je opseg značenja i dalje nepotpun, korpus se ponovno pretražuje. Značenja se u mrežnome rječniku *ellexiko* navode po čestoći: u korpusu najpotvrđenije značenje navodi se prvo. Manjkavost je takva pristupa to što su međusobno povezana značenja, poput primjerice doslovnoga i metaforičkoga značenja, katkad razdvojena jer su različita po čestoći. Stoga se uz svaki članak donosi objašnjenje organizacije značenja. Na početnome ekranu nakon unosa tražene riječi pojavljuje se popis svih značenja (usp. sliku 2). U rubrici *O značenjskim odnosima* objašnjava se odnos tražene riječi i navedenoga značenja (u primjeru na slici 2 značenje ‘*Wirtschaftsgut*’ u metonimijskome je odnosu sa značenjem ‘*Arbeit*’).

²⁷ Za *CCDB*, v. <http://corpora.ids-mannheim.de/ccdb/> i Keibel i Belica (2007).

²⁸ Rabimo naziv *djelomično korpusno utemeljen pristup* za engl. naziv *corpus based*, a naziv *potpuno korpusno utemeljen pristup* za engl. naziv *corpus driven*. Tu je distinkciju uvela Tognini-Bonelli (2001) razlikujući pristup koji se korpusom služi kako bi se provjerile unaprijed postavljene hipoteze (engl. *corpus-based approach*) od pristupa koji hipoteze postavlja isključivo na temelju rezultata korpusne analize (engl. *corpus-driven approach*). Vrlo su prikladni i hrvatski nazivi *pristup utemeljen na korpusu vs. pristup vođen korpusom* (v. npr. Mihaljević 1998: 66). Usp. i Hudeček i Mihaljević (2010): „S obzirom na odnos prema korpusu rječnici se dijele na *napametne* i *korpusne*. U korpusnim je rječnicima izbor riječi određen ponajprije izborom korpusa (knjiga, časopisa, novina itd. iz kojih se uzimaju riječi), dok u napametnim rječnicima obrađivač slobodno unosi riječi. Korpusni rječnici dijele se na *rječnike koji se temelje na korpusu* (obrađivač se služi korpusom, ali može slobodno procijeniti što treba unijeti u korpus te korpus može po potrebi dopunjavati) i *rječnike koji su vođeni korpusom* (u njima **mora** biti sve što se nalazi u korpusu i ne smije biti ništa čega u korpusu nema)” (kurzivna isticanja autorica).

Klikom na jedno od ponuđenih značenja korisnik otvara cjelovit opis podataka povezanih sa značenjem (v. sliku 4). Svi su podatci dobiveni iz korpusa, bilo na temelju djelomično (*corpus-based*) ili potpuno (*corpus-driven*) korpusno utemeljena pristupa (usp. Tognini-Bonelli 2001 i Storjohann 2005a).

Podatci o kolokacijama u rječniku *lexiko* primjer su potpuno korpusno utemeljena pristupa: iz korpusa se izdvajaju najčešće riječi koje se supojavljaju s natukničkom riječju (s pomoću softverskoga paketa *statistical collocation analysis and clustering*), a zatim leksikograf analizira njihove semantičke funkcije i na temelju te analize osmišljava članak.

Osim što se u značenjski ovisnim dijelovima rječničkoga članka opisuju sintagmatski semantički odnosi, *lexiko* nudi i nov, temeljit način predstavljanja paradigmatških značenjskih odnosa: sinonimije, antonimije, komplementarnosti, konverzivnosti, reverzivnosti itd. te vertikalnih struktura (poput hiponimije/hiperonimije i meronimije) (Storjohann 2005a: 1, usp. i Storjohann 2005b). Svaka je „partnerska riječ” oprimjerena citatom i sadržava hiperpoveznicu na odgovarajući članak. Kako bi se prikupili ovakvi leksikografski podatci, statistički značajna supojavljivanja riječi utvrđuju se s pomoću softvera *COSMAS II*. Automatska analiza kolokacija prvi je korak na putu prema identifikaciji paradigmatških odnosa kako bi se izbjeglo oslanjanje na intuiciju ili osobnu jezičnu kompetenciju. Drugi je korak leksikografsko vrednovanje automatski prikupljenih podataka: klasifikacija automatski izlučenih značenjskih odnosa i odabir primjera iz korpusa koji određeni značenjski odnos najbolje oprimjeruje. U trećemu se koraku korpusno dobiveni značenjski odnosi upotpunjuju podacima iz ostalih rječnika kako bi se dobio opširan i detaljan paradigmatški opis.

Gramatički je blok u rječniku *lexiko* također korpusno utemeljen, i zasebno se utvrđuje za svako značenje. Stoga su gramatički podatci mnogo pouzdaniji i detaljniji nego u drugim rječnicima (v. Klosa i Müller-Spitzer 2007).

4.3. Leksikografski proces (moduli)

Prvi je projektni modul uključivao objavljivanje cjelokupna popisa od 300 000 natuknica na internetu (2003.) te popunjavanje toga popisa automatski ili poluautomatski generiranim značenjski neovisnim podacima dobivenima iz korpusa izrađenoga za potrebe projekta (v. poglavlje 4.2.).

Drugi je modul, *Demonstrationswortschatz* (*Demonstracijski modul*), uključivao objavljivanje 250 detaljno i potpuno obrađenih natuknica.

Na temelju iskustva stečena u radu na drugome modulu leksikografii su pristupili radu na trećemu modulu: *Lexikon zum öffentlichen Sprachgebrauch (Rječnik javne jezične uporabe)*.

4.4. Korisnička perspektiva

Faza pripreme projekta za mrežno objavljivanje tekla je istodobno sa završnom fazom analize podataka. Ta ja faza uključivala provjeru sadržaja i lekturu obrađenih natuknica, testiranja mrežne aplikacije i prezentacije podataka na internetu te višestruka vrednovanja, provjere i širenje mogućnosti pretrage (v. sliku 5). U toj su fazi izrađene i upute za uporabu te pojmovnik (v. sliku 6).

Hilfe zur Suche

Erweiterte Stichwortsuche in *elexiko*

Stichwort beginnt mit enthält endet auf Dieses Textfeld für die Suche
'Alle Stichwörter mit Merkmal ...' leer lassen...

mit Merkmal Orthografie: -- beliebig --

mit Merkmal (bearbeitete Artikel)

eine/mehrere Lesarten -- beliebig --

Wortart: -- beliebig --

Grammatik: -- beliebig --

Wortbildung: -- beliebig --

semantische Klasse: -- beliebig --

sinverwandte Wörter: -- beliebig --

Gebrauchsbesonderheiten: -- beliebig --

Verwendung als Eigenname: -- beliebig --

Suchen Zurücksetzen

Slika 5. Širenje mogućnosti pretrage

Glossar

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

A

- Abbildung (siehe Illustration)
- Ableitung (siehe Derivation)
- Abstrahierung (siehe Zusammenhang der Lesarten)
- Abtönungspartikel
- Adjektiv
 - adverbial
 - attributiv
 - prädikativ
- Adjunktör
- Adverb
- adverbial (siehe Adjektiv)
- Affix
- Akkusativ (siehe Kasus)
- Aktiv (siehe Genus Verbi)
- Antonymie
- Appellativierung (siehe Zusammenhang der Lesarten)
- Artikel
- Attribut
- attributiv (siehe Adjektiv)
- Aussprache (siehe Gesprochensprachliche Belege)
- Automatisch ausgewählte Belege
- Automatisch ermittelte Wortbildungsprodukte

Slika 6. Pojmovnik

Korisnička perspektiva u okviru projekta *elexiko* bila je temom zasebnoga istraživačkoga projekta *Benutzeradaptive Zugänge und Vernetzungen in „elexiko“* u razdoblju od 2009. do 2011., na kojemu je radio tim leksikografa, računalnih lingvista i sociologa.²⁹ Upravo u okviru toga projekta nastali su mnogi referentni radovi iz područja istraživanja korisničke perspektive u e-leksikografiji uopće (v. npr. Müller-Spitzer, Kopenig i Töpel 2011, Klosa, Kopenig i Töpel 2011, Kopenig 2011, ur. Müller-Spitzer 2014).

²⁹ V. <http://www1.ids-mannheim.de/lexik/bzvelexiko.html>.

5. Mrežni leksičko-semantički repozitorij hrvatskoga jezika kao dio projekta *Repozitorij metafora hrvatskoga jezika*

Izrada potpuno korpusno utemeljenoga e-rječnika hrvatskoga jezika uz uporabu suvremenoga sustava za izradu rječnika i sustava za pretragu korpusa te ugledanjem u najbolje postojeće e-rječnike (poput ovdje opisanoga *alexika*) trebala bi biti prioritet hrvatske leksikografije. Dobra osnova za izradu takva rječnika (primjenom nekih metoda, softverskih rješenja i alata koji su primijenjeni u projektu *alexiko*), koji bi bio u skladu s najrecentnijim leksikografskim (i uopće lingvističkim) teorijama i praksama, mogao bi biti leksičko-semantički repozitorij *MetaNet.HR* koji se izgrađuje u okviru projekta *Repozitorij metafora hrvatskoga jezika*.

U bazi *MetaNet.HR* donosi se hijerarhija semantičkih okvira u širem smislu (koji se dalje razlažu na kognitivne primitive, predodžbene sheme i semantičke okvire u užemu smislu) te se za svaki okvir navode semantičke uloge i odnos s drugim semantičkim okvirima (djelomično po uzoru na *FrameNet*³⁰). U bazi će biti vidljivi³¹ i sintagmatski i paradigmatički značenjski odnosi (kao npr. u *WordNetu*³² ili njegovoj hrvatskoj inačici CroWN-u³³).

MetaNet.HR je primarno baza konceptualnih i jezičnih metafora, metonimija te predodžbenih shema, kognitivnih primitiva i semantičkih okvira hrvatskoga jezika s pripadajućim leksičkim jedinicama (v. Štrkalj Despot 2014). Baza je javno mrežno dostupna na početnoj projektnoj stranici: <http://ihjj.hr/metafore/>.³⁴

Mrežni leksičko-semantički repozitorij hrvatskoga jezika bilo je nužno graditi paralelno uz izgradnju repozitorija konceptualnih i jezičnih metafora kako bi se preslikavanja među pojedinim okvirima (metaforička ili metonimijska) mogla prikladno opisati i prikazati u bazi, ali i kako bi se dobio uvid u strukturu pojmovnoga sustava hrvatskoga jezika, od najsloženijih figurativnih jedinica do najjednostavnijih kognitivnih primitiva. Bez izgradnje takve baze nije moguće pristupiti ni ostvarenju zadaća iz područja obradbe prirodnoga jezika i umjetne inteligencije planiranih u okviru projekta.

Baza *MetaNet.HR* uključuje popis i detaljnu obradu: konceptualnih metafora, semantičkih okvira, predodžbenih shema i kognitivnih primitiva. Izgrađene su i hijerarhija semantičkih okvira (ontologija) i hijerarhija metafora.

³⁰ <https://framenet.icsi.berkeley.edu/>.

³¹ Odnosi se već unose u bazu, a grafički će prikaz odnosa biti realiziran koncem 2016. godine.

³² <https://wordnet.princeton.edu/>.

³³ <http://meta-share.ffzg.hr/repository/browse/croatian-wordnet/32d93d48703d11e28a985ef2e4e6c59e166ec06132a740cbb36e515b093096b2/>.

³⁴ Baza je izgrađena po uzoru na projekt *MetaNet: A Multilingual Metaphor Repository* (International Computer Science Institute i UC Berkeley, SAD) (<https://metanet.icsi.berkeley.edu/metanet/>).

Metafore su podijeljene i pretražive po tipu te po tematskim ili konceptualnim skupinama. Po tipu su metafore razdijeljene u: primarne metafore, složene metafore i izvedene metafore.

Prema tematskim ili konceptualnim skupinama metafore su podijeljene na: metafore strukture događaja, metafore uma, metafore moralnosti, skalarne metafore, metafore vremena, metafore emocija, metafore ekonomije, metafore vladanja, metafore dobrobiti itd. Kognitivnolingvistička analiza metafora provodi se u skladu s postavkama neurokognitivne teorije jezika i mišljenja te neurokognitivne teorije metafore (Feldman 2006, Lakoff 2008).

Metafore, metonimije, okviri, predodžbene sheme te gramatičke i leksičke konstrukcije formalizirano se označuju sustavom označivanja utjelovljene konstrukcijske gramatike (v. Feldman, Dodge i Bryant 2009).

Za svaku se konceptualnu metaforu određuje: tip, hijerarhija, razina, izvorni okvir, ciljni okvir, preslikavanja, odnosi, tipovi odnosa, izvedene metafore te primjeri jezičnih metafora koji odslikavaju konceptualnu metaforu.

Za svaki se semantički okvir određuje: definicija, tip, hijerarhija, semantičke uloge, odnosi s drugim okvirima, veze, inferencije, te se donose u korpusima hrvatskoga jezika najčešće leksičke jedinice za taj okvir, što bazu *MetaNet*. *HR* čini izvrsnom podlogom za daljnju leksikografsku nadgradnju.

6. Zaključak

Primjena dosega suvremene e-leksikografije na hrvatsku leksikografsku teoriju i praksu zasigurno nije zadovoljavajuća. U ovome radu izneseni pregled novih leksikografskih postavka i praksa prinos je izmjeni takvoga stanja. Nužno bi bilo ugledati se u primjere dobre prakse, poput ovdje prikazanoga njemačkoga *lexika*, jednoga od leksikografski najuzornijih e-rječnika danas, i pristupiti izradi korpusno utemeljenoga i leksikografski detaljnoga suvremenoga e-rječnika hrvatskoga jezika. Takav bi pothvat pratila i teorijskoleksikografska istraživanja te razvoj prikladnih jezičnih resursa i alata, čime bi se hrvatska leksikografija približila današnjim europskim standardima.

Zahvala

Zahvaljujemo Upravnomu odboru COST akcije *European Network of e-Lexicography* koja je dr. sc. Kristini Štrkalj Despot dodijelila stipendiju u okviru programa *Short Term Scientific Mission* za jednomjesečni boravak na Institutu za njemački jezik u Mannheimu te time omogućila autoricama suradnju na ovome radu. Zahvaljujemo i potpori Hrvatske zaklade za znanost pod projektom 3624.

Literatura:

- ATKINS, SUE; RUNDELL, MICHAEL. 2008. *The Oxford Guide to Practical Lexicography*. Oxford University Press. Oxford.
- BÉJOINT, HENRI. 2010. *The Lexicography of English*. Oxford University Press. Oxford.
- BIRTIĆ, MATEA I DR. 2012. *Školski rječnik hrvatskoga jezika*. Institut za hrvatski jezik i jezikoslovlje. Zagreb.
- ČILAŠ ŠIMPRAGA, ANKICA; JOJIĆ, LJILJANA; LEWIS, KRISTIAN. 2008. *Prvi školski rječnik hrvatskoga jezika*. Školska knjiga – Institut za hrvatski jezik i jezikoslovlje. Zagreb.
- DUBOIS, CLAUDE. 1990. Considérations générales sur l'organisation du travail lexicographique. *Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Ur. Hausmann, F. J.; Reichmann, O.; Wiegand, H. E.; Zgusta, L. Berlin – New York. 1574–1588.
- DUDA, WALTER I DR. 1986. *Zu einer Theorie der zweisprachigen Lexikographie. Überlegungen zu einen neuen russisch-deutschen Wörterbuch*. AkademieVerlag. Berlin.
- e-Lexicography: The Internet, Digital Initiatives and Lexicography*. 2011. Ur. Fuertes-Olivera, Pedro; Bergenholtz, Henning. Continuum. London – New York.
- ellexiko. Erfahrungsberichte aus der lexikografischen Praxis eines Internetwörterbuchs*. 2011. Ur. Klosa, Annette. Gunter Narr. Tübingen. (= *Studien zur Deutschen Sprache. Forschungen des Instituts für Deutsche Sprache*, svezak 55).
- Elexiko – Online-Wörterbuch zur deutschen Gegenwartssprache*. http://www.owid.de/elexiko_/index.html. (pristupljeno 2. listopada 2015.).
- ENGELBERG, STEFAN. 2014. Gegenwart und Zukunft der Abteilung Lexik am IDS: Plädoyer für eine Lexikographie der Sprachdynamik. *Institut für Deutsche Sprache: Ansichten und Einsichten. 50 Jahre Institut für Deutsche Sprache*. Institut für Deutsche Sprache. Mannheim. 243–253.
- ENGELBERG, STEFAN; KLOSA, ANNETTE; MÜLLER-SPITZER, CAROLIN. 2009. Internet lexicography as a challenge: The Internet dictionary portal at the Institute for German Language. *Kernerman Dictionary News* July 2009. 16–25.
- ENGELBERG, STEFAN; MÜLLER-SPITZER, CAROLIN. 2013. Dictionary Portals. *Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume. Recent developments with focus on electronic and computational lexicography*. Ur. Gouws, R. H. i dr. De Gruyter. Berlin – Boston. 1023–1035.
- FELDMAN, JEROME. 2006. *From Molecule to Metaphor. A Neural Theory of Language*. MIT Press – A Bradford Book. Cambridge, Massachusetts.

- FELDMAN, JEROME; DODGE, ELLEN; BRYANT, JOHN. 2009. Embodied Construction Grammar. *The Oxford Handbook of Linguistic Analysis*. Ur. Heine, B.; Narrog, H. Oxford University Press. Oxford. DOI: 10.1093/oxfordhb/9780199544004.013.0006 (pristupljeno 2. listopada 2015.).
- GOUWS, RUFUS H. 2011. Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. Ur. Fuertes-Olivera, P. A.; Bergenholtz H. Continuum. London – New York. 17–29.
- GRANGER, SYLVIANE. 2012. Electronic lexicography: From challenge to opportunity. *Electronic Lexicography*. Ur. Granger, S.; Paquot, M. Oxford University Press. Oxford. 1–11.
- Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum de Wortschatz. Schriften des Instituts für Deutsche Sprache*. 2005a. Ur. Haß, Ulrike. De Gruyter. Berlin – New York.
- HARTMANN, REINHARD R. K. 2007. *Bibliography of Lexicography*. <http://euralex.pbworks.com> (pristupljeno 2. listopada 2015.).
- HASS, ULRIKE. 2005b. Das Bedeutungsspektrum. *Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz. Schriften des Instituts für Deutsche Sprache*. Ur. Hass, U. De Gruyter. Berlin – New York. 163–181.
- HUDEČEK, LANA; MIHALJEVIĆ, MILICA. 2009. *Hrvatski terminološki priručnik*. Institut za hrvatski jezik i jezikoslovlje. Zagreb.
- JOJIĆ, LJILJANA I DR. *Veliki rječnik hrvatskoga standardnog jezika*. Školska knjiga. Zagreb.
- KEIBEL, HOLGER; BELICA, CYRIL. 2007. CCDB: A Corpus Linguistic Research & Development Workbench. *Proceedings of the 4th Corpus Linguistics conference, Birmingham, 2007*. <http://corpora.ids-mannheim.de/cl2007-134.pdf> (pristupljeno 2. listopada 2015.).
- KLOSA, ANNETTE; SCHNÖRCH, ULRICH; STORJOHANN, PETRA. 2006. ELEXIKO – A lexical and lexicological, corpus-based hypertext information system at the Institut für Deutsche Sprache, Mannheim. *Proceedings of the 12th EURALEX International Congress (Atti del XII Congresso Internazionale di Lessicografia), EURALEX 2006, Turin, Italy, September 6th-9th, 2006. Vol 1*. Ur. Marellò, C. i dr. Edizioni dell’Orso Alessandria. Turin. 425–430.
- KLOSA, ANNETTE; MÜLLER-SPITZER, CAROLIN. 2007. Grammatische Angaben in elexiko und ihre Modellierung. *Dictionary Visions, Research and Practice. Selected Papers from the 12th International Symposium on Lexicography, Copenhagen 2004*. Ur. Gottlieb, H.; Mogensen, J. E. Benjamins. Amsterdam. 13–37.

- KLOSA, ANNETTE; LEMNITZER, LOTHAR; NEUMANN, GERALD. 2008. Wörterbuchportale – Fragen der Benutzerführung. *Lexikografische Portale im Internet*. Ur. Klosa, Annette. Institut für Deutsche Sprache. Mannheim. 5–35. <http://www.ids-mannheim.de/pub/laufend/opal/> (pristupljeno 2. listopada 2015.).
- KLOSA, ANNETTE; KOPLINIG, ALEXANDER; TÖPEL, ANTJE. 2011. Benutzerwünsche und Meinungen zu einer optimierten Wörterbuchpräsentation – Ergebnisse einer Onlinebefragung zu *elexiko*. *OPAL – Online publizierte Arbeiten zur Linguistik* 3/2011. Institut für Deutsche Sprache. Mannheim.
- KLOSA, ANNETTE. 2013. The lexicographical process (with special focus on online dictionaries). *Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume. Recent developments with focus on electronic and computational lexicography*. Ur. Gouws, R. H. i dr. De Gruyter. Berlin – Boston. 517–524.
- KLOSA, ANNETTE (UR.). 2014. *Dateninterpretation und Präsentation in Onlinewörterbüchern am Beispiel von elexiko*. *Deutsche Sprache* 4/14. Berlin.
- KNOWLES, F. E. 1990. The Computer in Lexicography. *Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Ur. Hausmann, F. J. i dr. Walter de Gruyter. Berlin – New York. 1645–1672.
- KOPLINIG, ALEXANDER. 2011. Understanding How Users Evaluate Innovative Features of Online Dictionaries – An Experimental Approach. *Electronic lexicography in the 21st century: New applications for new users. Proceedings of eLex 2011, Bled, 10-12 November 2011*. Ur. Kosem, I.; Kosem, K. 147–150. <http://www.trojina.si/elex2011/Vsebine/proceedings/eLex2011-18.pdf> (pristupljeno 2. listopada 2015.).
- KREK, SIMON; ABEL, ANDREA; TIBERIUS, CAROLE. 2014. Survey – WG3 ENeL Dictionary Writing Systems & Corpus Query Systems. *Proceedings of the 3rd COST ENeL Working Group 3 („Innovative e-dictionaries“) Meeting: „Workflow of Corpus-based Lexicography“*. http://www.elexicography.eu/wp-content/uploads/2015/04/LexicographicalWorkflow_DeliverableWG3BolzanoMeeting2014.pdf (pristupljeno 2. listopada 2015.).
- LAKOFF, GEORGE. 2008. The neural theory of metaphor. *The Cambridge Handbook of Metaphor and Thought*. Ur. Gibbs, R. Cambridge University Press. New York.
- LANDAU, SIDNEY I. 1984. *Dictionaries. The Art and Craft of Lexicography*. Cambridge University Press. New York.
- LEMBERG, INGRID. 2001. Aspekte der Online-Lexikographie für wissenschaftliche Wörterbücher. *Chancen und Perspektiven computergestützter Lexikographie*. Ur. Lemberg, I.; Schröder, B.; Storrer, A. Niemeyer. Tübingen. 71–91.

- MIHALJEVIĆ, MILICA. 1998. *Terminološki priručnik*. Hrvatska sveučilišna naklada. Zagreb.
- MOGUŠ, MILAN; BRATANIĆ, MAJA; TADIĆ, MARKO. 1999. *Hrvatski čestotni rječnik*. Zavod za lingvistiku Filozofskoga fakulteta – Školska knjiga. Zagreb.
- MOGUŠ, MILAN. 2001. *Rječnik Marulićeve Judite*. Institut za hrvatski jezik i jezikoslovlje. Zagreb.
- MÖHRS, CHRISTINE. 2014. The lexicographical process of *ellexiko*. *Proceedings of the 2nd COST ENeL Working Group 3 („Innovative e-dictionaries“)* Meeting: „Workflow of Corpus-based Lexicography“. http://www.elexicography.eu/wp-content/uploads/2014/07/Moehrs_2014_COST_Bolzano.pdf (pristupljeno 2. listopada 2015.).
- MÜLLER-SPITZER, CAROLIN. 2003. Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexicographischen Prozessen. *Lexicographica* 19. 140–168.
- MÜLLER-SPITZER, CAROLIN. 2007. Das *ellexiko*-Portal: Ein neuer Zugang zu lexicografischen Arbeiten am Institut für Deutsche Sprache. *Datenstrukturen für linguistische Ressourcen und ihre Anwendungen: Proceedings of the Biennial GLDV Conference 2007, Eberhard Karls Universität Tübingen*. Ur. Rehm, G.; Witt, A.; Lemnitzer, L. Gunter Narr. Tübingen. 179–188.
- MÜLLER-SPITZER, CAROLIN; KOPLINIG, ALEXANDER; TÖPEL, ANTJE. 2011. What Makes a Good Online Dictionary? – Empirical Insights from an Interdisciplinary Research Project. *Electronic lexicography in the 21st century: New applications for new users. Proceedings of eLex 2011, Bled, 10-12 November 2011*. Ur. Kosem, I.; Kosem, K. 203–208. <http://www.trojina.si/elex2011/Vsebine/proceedings/eLex2011-27.pdf> (pristupljeno 2. listopada 2015.).
- OWID – *Online-Wortschatz-Informationssystem Deutsch* 2008. Institut für Deutsche Sprache. Mannheim. <http://www.owid.de> (pristupljeno 2. listopada 2015.).
- RIEDEL, HANS; WILLE, MARGIT. 1979. *Über die Erarbeitung von Lexika*. Bibliograph. Inst. Leipzig.
- SCERBA, LEV V. 1940. Towards a General Theory of Lexicography. *International Journal of Lexicography* 8/4. 315–350.
- SCHAEDER, BURKHARD. 1987. *Germanistische Lexikographie*. Lexicographica. Tübingen.
- SCHRÖDER, MARTIN. 1997. Brauchen wir ein neues Wörterbuchkartell? Zu den Perspektiven einer computerunterstützten Dialektlexikographie und eines Projektes „Deutsches Dialektwörterbuch“. *Zeitschrift für Dialektologie und Linguistik* 64/1. 57–65.

- DE SCHRYVER, GILLES-MAURICE. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16/2. 143–199.
- DE SCHRYVER, GILLES-MAURICE; PRINSLOO, DANIE J. 2000a. The Concept of „Simultaneous Feedback“: Towards a New Methodology for Compiling Dictionaries. *Lexikos. Journal of the African Association for Lexicography* 10. 1–31.
- DE SCHRYVER, GILLES-MAURICE; PRINSLOO, DANIE J. 2000b. Dictionary-Making Process with ‚Simultaneous Feedback‘ from the Target Users to the Compilers. *Proceedings of the Ninth EURALEX International Congress. Stuttgart*. Ur. Heid, U. i dr. Institut für Maschinelle Sprachverarbeitung. Stuttgart. 197–209.
- SOROKOLETOV, FEDOR P. 1978. Traditionen der sowjetrussischen Lexikographie. *Aspekte der sowjetrussischen Lexikographie. Übersetzungen, Abstracts, bibliographische Angaben*. Ur. Wolski, Werner. Max Niemeyer. Tübingen. 63–88.
- STORJOHANN, PETRA. 2003. The lexicographic use of corpora and computational tool for disambiguation. *Proceedings of Corpus Linguistics 2003 Conference. UCREL technical paper number 16. UCREL*. Ur. Archer, D. i dr. Lancaster University. Lancaster.
- STORJOHANN, PETRA. 2005a. Corpus-driven vs. corpus-based approach to the study of relational patterns. *Proceedings of the Corpus Linguistics Conference 2005 in Birmingham* 1/1. www.corpus.bham.ac.uk/pclcl/ (pristupljeno 2. listopada 2015.).
- STORJOHANN, PETRA. 2005b. *elexiko* – A Corpus-based Monolingual German Dictionary. *Hermes, Journal of Linguistics* 34. 55–83.
- STORJOHANN, PETRA. 2013. Korpuslinguistische und lexikografische Ansätze zur Beschreibung deutscher Paronyme. *Im Dienste des Wortes. Lexikologische und lexikografische Streifzüge. Festschrift für Ioan Lazarescu*. Ur. Scheuringer, H.; Sava, D. Stutz-Verlag. Passau. (= *Forschung zur Deutschen Sprache in Mittel-, Ost- und Südeuropa* 3). 401–418.
- STORRER, ANGELIKA. 1998. Hypermedia-Wörterbücher: Perspektiven für eine neue Generation elektronischer Wörterbücher. *Wörterbücher in der Diskussion III*. Ur. Wiegand, H. E. Max Niemeyer Verlag. Tübingen. 106–131.
- STORRER, ANGELIKA. 2001. Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie. *Chancen und Perspektiven computergestützter Lexikographie*. Ur. Lemberg, I.; Schröder, B.; Storrer, A. Max Niemeyer Verlag. Tübingen. 54–69.
- STORRER, ANGELIKA; FREESE, KATRIN. 1996. Wörterbücher im Internet. *Deutsche Sprache* 24. 97–136.

- ŠTRKALJ DESPOT, KRISTINA. 2014. Croatian Metaphor Repository. *Proceedings of the 2nd COST ENeL Working Group 3 („Innovative e-dictionaries“) Meeting: „Workflow of Corpus-based Lexicography“*. http://www.elexicography.eu/wp-content/uploads/2014/07/Strkalj-Despot_2014_COST_Bolzano.pdf (pristupljeno 2. listopada 2015.).
- TAFRA, BRANKA; FINK ARSOVSKI, ŽELJKA. 2013. Хорватская лексикография. *Славянская лексикография*. Ur. Černyševa, M. I. Azbukovnik. Moskva.
- TARP, SVEN. 2012. Do We Need a (New) Theory of Lexicography? *Lexikos* 22. 321–332.
- The Oxford English Dictionary Online*. <http://www.oed.com/> (pristupljeno 2. listopada 2015.).
- Theorie der Semantik und Theorie der Lexikographie*. 1996. Ur. Weber, Nico. Max Niemeyer. Tübingen.
- TIBERIUS, CAROLE; KREK, SIMON. 2014. Workflow of Corpus-Based Lexicography. *Proceedings of the 2nd COST ENeL Working Group 3 („Innovative e-dictionaries“) Meeting: „Workflow of Corpus-based Lexicography“*. http://www.elexicography.eu/wp-content/uploads/2015/04/Lexicographicalworkflow_Deliverable_WG3_Bolzano_Meeting2014.pdf (pristupljeno 2. listopada 2015.).
- TOGNINI-BONELLI, ELENA. 2001. *Corpus Linguistics at Work*. John Benjamins. Amsterdam – Philadelphia.
- Using Online Dictionaries*. VI/387 S. 2014. Ur. Müller-Spitzer, Carolin. De Gruyter. (*Lexicographica: Series Maior* 145). Berlin – Boston.
- WIEGAND, HERBERT E. 1999. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. I. Teilband*. De Gruyter. Berlin – New York.
- ZGUSTA, LADISLAV. 1971. *Manual of Lexicography*. Academia – Mouton. The Hague – Paris – Prague.
- ZGUSTA, LADISLAV. 1992. The Czech–Chinese Dictionary and the Theory of Lexicography. *International Journal of Lexicography* 5/2. 85–128.

A perspective on e-lexicography

Abstract

This paper provides an overview of basic assumptions and classifications in the discipline of e-lexicography. The classification of e-dictionaries, the description of the lexicographical process for e-dictionaries, and an overview of

the most popular dictionary writing systems (DWS) and corpus query systems (CQS) are provided. The online e-dictionary project *elexiko* (Institute of German Language) is presented as a best-practice example. The *Elexiko* project's purposes and goals, theoretical and methodological background, modules, and its user perspective are described. The workflow of the project of building a lexical and semantic database of the Croatian language (*MetaNet.HR*) is described as a potential basis for a corpus-based and corpus-driven e-dictionary of the Croatian language. *MetaNet.HR* is a database of cognitive primitives, image schemas, semantic frames, metaphors, and metonymies in the Croatian language (with their respective lexical units and/or linguistic examples). The database is being compiled as part of the *Croatian Metaphor Repository* project.

Ključne riječi: e-leksikografija, e-rječnik, leksikografski proces, korisnička perspektiva, *elexiko*

Keywords: e-lexicography, e-dictionary, lexicographic process, *elexiko*