

# OPTIMIZACIJA IZDELAVE SLOVARJA: MOŽNOSTI UPORABE OBSTOJEČIH SLOVARSKIH PODATKOV

Andrej Perdih

Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU, Ljubljana

UDK 811.163.6'374.3:81'271'322.2

Prispevek prikazuje možnosti neposredne uporabe podatkov za novi slovar slovenskega knjižnega jezika iz obstoječih in nastajajočih slovarjev Inštituta za slovenski jezik Frana Ramovša ter drugih zbirk. S tem je mogoče doseči hitrejše nastajanje slovarja. Opozorjeno je na težave pri ponovni uporabi podatkov, zlasti zaradi spremenjenega stanja v jezikovni rabi in zaradi dejstva, da so podatki v slovarjih do določene mere rezultat ene od različnih možnih jezikoslovnih interpretacij gradiva, ki je posledica različnih gradivskih in teoretičnih izhodišč slovarjev.

slovar, avtomatizacija, leksikografija, jezikovni opis

The article explores the possibilities of reusing data from existing or emerging dictionaries for the purpose of compiling a New Dictionary of the Slovene Standard Language. The aim is to speed up the process of dictionary compilation. Several issues regarding data reuse are discussed, especially the language change and the bias caused by various possible linguistic interpretations of data as a result of the different theoretical premises and the use of different language data.

dictionary, automatization, lexicography, language description

## 1 Uvod

31. marca 2015 je dal Inštitut za slovenski jezik Frana Ramovša ZRC SAZU v javno obravnavo Osnutek koncepta novega razlagalnega slovarja slovenskega knjižnega jezika (delovna kratica NSSKJ),<sup>1</sup> po katerem nastaja »temeljni enojezični informativno-normativni razlagalni slovar, ki opisuje jezikovno stanje in pojmovni svet sodobnega slovenskega jezika, kot ju izkazuje jezikovna raba, potrjena v pisnih besedilih, zajetih zlasti v korpusih, v manjšem obsegu pa tudi raba v govorjenih

<sup>1</sup> Pripravili so ga sodelavci Inštituta za slovenski jezik Frana Ramovša ZRC SAZU (Nataša Gliha Komac, Nataša Jakop, Janoš Ježovnik, Simona Klemenčič, Domen Krvina, Nina Ledinek, Tanja Mirtič, Andrej Perdih, Špela Petric, Marko Snoj in Andreja Žele) v sodelovanju s 34-članskim uredniškim odborom, ki ga sestavljajo priznani strokovnjaki z matične ustanove, vseh štirih slovenskih univerz in ene zamejske, člani SAZU, pisatelji, člani relevantnih stanovskih združenj in predstavniki jezikovnotehnoloških podjetij. V enomesečno javno obravnavo je bila dana različica 1.1, v kateri so glede na različico 1.0, ki so jo pripravili avtorji, v največji možni meri upoštevani predlogi in pripombe uredniškega odbora in kolegov Inštituta za slovenski jezik Frana Ramovša ZRC SAZU.

besedilih«. Slovar bo obsegal 100.000 slovarskih sestavkov in bo izdelan povsem na novo, prosto pa bo dostopen na portalu Fran.<sup>2</sup>

Ker slovarji tega tipa in velikosti nastajajo dlje časa, se poraja vprašanje, kako je mogoče zastaviti postopek izdelave slovarja tako, da bi se čas nastajanja skrajšal. Možnosti za to je seveda več, v prispevku bodo prikazane nekatere možnosti uporabe obstoječih slovarskih podatkov v slovarjih Inštituta za slovenski jezik Frana Ramovša ZRC SAZU, ki so že izšli ali pa so trenutno še v nastajanju, ter drugih vsebinsko ustreznih zbirk, ki jih je mogoče uporabiti. Da je to mogoče, potrjuje dejstvo, da je na Inštitutu za slovenski jezik Frana Ramovša ZRC SAZU samo od leta 2000 dalje izšlo več kot 30 slovarjev; številni izmed njih so objavljeni na portalih Fran (glej opombo 2) in Termania (Romih, Krek 2012), seveda pa so slovarji in druge jezikovne zbirke nastajali tudi že pred tem časom. Nekateri izmed teh slovarjev, zlasti specialni, in določene besedne zbirke se v večji ali manjši meri opirajo na prvo izdajo *Slovarja slovenskega knjižnega jezika* (Perdih 2013: 33–34).

Takšno število nastalih slovarjev spodbuja k razmišljanju o tem, kako te že interpretirane jezikovne podatke uporabiti kot pomoč za hitrejšo izdelavo novih slovarjev. Če je pri posameznih nalogah sestavljanja slovarja, na primer pri oblikovanju razlag, mogoče preverjati rešitve v zelo različnih priročnikih (slovarjih, enciklopedijah, leksikonih itn.), pa se poraja vprašanje, katere podatke je iz slovarjev smiselno sistematično uvoziti neposredno v digitalno slovarsko bazo že pred začetkom samega redatorskega dela. Namen takega uvoza je vnaprej pripraviti karseda zanesljive podatke, da bi postopek izdelave slovarja potekal hitreje, ne da bi se ob tem znižala kakovost končnega izdelka zaradi vpliva starejših jezikovnih podatkov in interpretacij.

V pričujočem članku ne govorimo o golem priključevanju obstoječih podatkov oz. celotnih podatkovnih zbirk v slovar, kar je ena od možnosti,<sup>3</sup> ampak se osredotočamo na možnosti uvoza izbranih podatkov iz različnih zbirk, da nastane nabor vnaprej pripravljenih jezikovnih podatkov, ki so v jeziku razmeroma stabilni in jih je zato mogoče dovolj varno uporabiti v nastajajočih slovarskih opisih. Pri takšni uporabi jezikovnih priročnikov se nujno upošteva, da so ti priročniki izdelani po določenih konceptualnih izhodiščih, ki niso nujno enaki kot v nastajajočem slovarju. Zato se upošteva, da se geslovniki med slovarji razlikujejo, da je poglobljenost analize jezikovnih podatkov lahko drugačna in, kar je morda še pomembnejše, da so slovarji izdelani z različnimi vidiki obravnave jezika (s tem je mišljeno konceptualno

<sup>2</sup> Spletni naslov: [www.fran.si](http://www.fran.si). Portal je predstavljen v prispevku Kozme Ahačiča, Nine Ledinek in Andreja Perdiha z naslovom Portal Fran – nastanek in trenutno stanje v: Mojca Smolej (ur.): *Slovnica in slovar – aktualni jezikovni opis. Obdobja 34*.

<sup>3</sup> Tak primer je npr. priključitev paradigem v poljskem slovarju *Wielki słownik języka polskiego* iz slovarja *Słownik gramatycznego języka polskiego* (Żmigrodzki 2010: 159) ali priključitev več podatkovnih zbirk, kar je predvideno v Predlogu za izdelavo *Slovarja sodobnega slovenskega jezika* (Krek, Kosem, Gantar 2013). Tovrstno priključevanje je lahko problematično, saj imajo lahko nekatere podatkovne zbirke povsem drugačna konceptualna izhodišča od nastajajočega slovarja, pogosto pa primarno niti niso namenjena istemu oz. podobnemu namenu oziroma istemu tipu primarnega uporabnika v smislu razmerja človeški : nečloveški uporabnik.

drugačno razumevanje in interpretacija jezikovnih podatkov). Če torej primerjamo splošni slovar knjižnega jezika s specialnim slovarjem, ki obravnava npr. besedotvorje, sinonimijo ali frazeologijo, je razlika v pristopih, načinu obravnave in namenu dela lahko razmeroma velika.

Vsi uvoženi podatki se bodo preverjali. Namreč, uvoz podatkov ne pomeni, da so ti povsem ustrezni glede na koncept novega slovarja in glede na stanje v jeziku, ki ga novi slovar opisuje. Pomeni le, da če se pri analizi gradiva izkaže, da so uvoženi podatki ustrezni, jih ni treba vnašati ročno, kar prinaša časovni prihranek; če pa uvoženi podatki niso ustrezni, jih je pri preverjanju treba odstraniti. Samo po sebi je umevno, da se uvoženim podatkom pripišejo dodatni, ki se pridobijo pri analizi gradiva. Z upoštevanjem navedenega bodo vsi jezikovni podatki skupaj predstavljali konceptualno enoten opis slovenskega jezika z enotno celotno logiko slovarja.

## 2 Tehnični vidik

Za uvoz podatkov iz slovarjev morajo biti ti zapisani v takem računalniškem zapisu, v katerem so ustrezno strukturirani in semantično opredeljeni. Danes se za zapis slovarjev pogosto uporablja oblika XML bodisi primarno ali pa vsaj kot ena od možnosti izvoza iz programov, ki uporabljajo drugačne oblike zapisa podatkov. V programu iLex (Erlandsen 2004) v obliki XML nastaja večina novih slovarjev na Inštitutu za slovenski jezik Frana Ramovša ZRC SAZU. Terminološki slovarji in *Slovenski lingvistični atlas* nastajajo v programu SlovarRed (Košmrlj - Levaičič, Seliškar 2004), etimološki pa je nastal v programu EVA (Jakopin 1995). Ti slovarji (in še kateri drugi) so že bili ali pa še bodo pretvorjeni v obliko XML (Ledinek, Perdih 2012). Slovarji, ki so v okviru tega prispevka obravnavani kot vir za pridobivanje podatkov, so že zapisani v ustrezni obliki.

## 3 Podatki za vključitev in slovarji

Odgovor na vprašanje, katere podatke je smiselno pripraviti za ponovno uporabo v novih slovarjih, je pragmatične narave: glede na zastavljeni način obravnave teh podatkov je primerno uporabiti tiste podatke, ki pospešijo nastajanje novega slovarja, vendar ne vplivajo na leksikografa v smislu neustreznega sprejemanja starejših podatkov in interpretacij jezikovnih dejstev. Če so podatki napačni ali zavajajoči, je namreč njihova uporaba lahko škodljiva, če pa je teh podatkov malo ali jih je mogoče pridobiti približno enako hitro kot jih pripraviti za uvoz, pa je smiselnost avtomatizacije vprašljiva. Kako se na podlagi tega načelnega stališča odločiti v konkretnih primerih, je v nekaterih primerih lažje, v drugih pa težje vnaprej objektivno vrednotiti.

Pri izdelavi NSSKJ je predviden neposreden vnos več skupin podatkov, v tem prispevku pa bo predstavljena problematika uporabe naslednjih skupin podatkov:

- podatki o izgovoru,
- dvojnice,
- sinonimi,
- besedotvorni podatki,

- frazeologija,
- etimologija.

Te podatke je mogoče pridobiti vsaj iz naslednjih slovarjev:

- druga izdaja *Slovarja slovenskega knjižnega jezika* (SSKJ 2),
- *Besedišče slovenskega jezika z oblikoslovnimi podatki* (BSJ),
- *Slovar novejšega besedja slovenskega jezika* (SNB),
- *Slovenski pravopis* (SP),
- *Slovar slovenskih frazemov* (SSF),
- *Slovar sinonimov slovenskega jezika* (v nastajanju),
- *Slovar pravopisnih težav* (SPT, rastoči slovar<sup>4</sup>),
- *Sprotni slovar slovenskega jezika* (SSSJ, rastoči slovar),
- *Slovenski etimološki slovar* (SES).

Najaktualnejši **podatki o izgovoru** so zapisani v SSKJ 2, SNB, SSSJ in SP. Pri pripravi SSKJ 2 so bili podatki iz prve izdaje SSKJ že deloma vsebinsko preverjeni in posodobljeni ter tehnično dopolnjeni, v poštev pa lahko pridejo tudi v SSKJ 2 nezajeti podatki, ki so na voljo v SNB in SP. Ob nastajanju NSSKJ bo opravljena fonetična raziskava izgovora, in kadar se bodo podatki iz navedenih slovarjev ujemali z rezultati raziskave, kar verjetno velja za večino podatkov, bo prihranjenega precej časa pri vnosu osnovnih oblik. V SSKJ 2 je namreč izgovor (zapisan v oglatih oklepajih) naveden v 5.442 slovarskih sestavkih, torej pri približno 5,5 % vseh iztočnic. Večina teh iztočnic bo predvidoma vključena tudi v NSSKJ. Za vse ostale iztočnice, ki nimajo pripisanega podatka o izgovoru, je kot podatek o izgovoru mogoče uporabiti zapis iztočnice. Tako bo večina iztočnic v novem slovarju že opremljena s podatki o jakostnem naglasu in izgovoru. Zapis tonemskega naglasa na osnovni obliki je mogoče v podatkovno zbirko prenesti na podlagi primerjave zapisa tonemskega naglasa v SSKJ 2 (prikazano v navadnih oklepajih) in jakostno onaglašene iztočnice.<sup>5</sup> Ocenjujemo, da bo ročni vnos podatkov o izgovoru za oba tipa naglaševanja potreben le za manjši del slovarja.

V povezavi z **dvojnicami** ločujemo dve vrsti podatkov. Prvi podatek je zapis dvojnice, drugi pa razmerje med dvojnicami. Razmerje med dvojnicami se v jeziku s časom spreminja že samo po sebi, tako kot že sicer kvalificiranje razmeroma hitro zastareva. Poleg tega vsak slovar na svoj način interpretira in zapiše to razmerje. Zaradi navedenega je bila sprejeta odločitev, da prenašanje podatka o razmerju med dvojnicami iz drugih slovarjev v NSSKJ ni smiselno, pač pa je zaradi povezovanja dvojnic med seboj smiselno prenesti podatke o zapisu dvojnic in s tem povezanosti

<sup>4</sup> Z izrazom *rastoči slovar* na Inštitutu označujemo tiste slovarje, ki imajo na portalu Fran objavljene le izdelane slovarske sestavke, ne pa še vseh predvidenih; novi slovarski sestavki se dodajajo v večjem številu, ko so pripravljeni za objavo.

<sup>5</sup> Tako pridobljeni osnovni podatki bodo obenem podlaga za generiranje jakostno in tonemsko onaglašeni paradigem za vsako pregibno iztočnico v slovarju. Mehanizem za generiranje onaglašeni oblik bo pripravljen za tiste pregibnostno-naglasne tipe, ki jim pripada dovolj veliko število iztočnic, ne pa za manjše skupine, kot sta na primer *mati* in *hči*, ki jima bo vzorec pripisan ročno.

med njimi. Te podatke je mogoče pridobiti iz zaglavja polnih slovarskih sestavkov in iz kazalčnih slovarskih sestavkov SSKJ 2.

Podatki o **sinonimiji** so zajeti v različnih slovarjih, kjer opravljajo različne funkcije – od normativne do tezavrske. Smiselno se zdi uporabiti zlasti nabor sinonimov iz SSKJ 2, SNB, nastajajočega *Slovarja sinonimov slovenskega jezika*<sup>6</sup> in drugih priročnikov. Zaradi deloma različne pomenske členitve med slovarji in NSSKJ tako pridobljenega nabora sinonimov ni smiselno vnaprej povezovati h konkretnim pomenom, pravzaprav to pred izdelavo slovarskih sestavkov niti ni mogoče. Zato se vsaki iztočnici pripiše nabor znanih sinonimov, nato pa bo urednik pri pomenski členitvi sinonime ustrezno pripisal posameznim pomenom, dodal morebitne manjkajoče sinonime in izločil za konkretni slovar neustrezne sinonime.

Za prikaz **besedotvorja** so v NSSKJ predvidena naslednja razmerja: manjšalnica, ženska/moška oblika, svojilni pridevnik. Podatke o tem je mogoče dobiti na podlagi sklicevalnih razlagalnih tipov (Napotki 1981: 18; Silvester 1978: 47–55) v SSKJ 2 in SSKJ, in sicer iz SSKJ 2 *manjšalnica od X* (1.427), *svojilni pridevnik od X* (44), iz SSKJ pa *ženska oblika od X* (784).<sup>7</sup> Poleg besedotvornega podatka o tipu tvorjenke se v tej fazi v slovarsko bazo pripiše tudi podatek o povezavi med iztočnicama, ki v elektronski izdaji slovarja deluje kot hiperpovezava.

Nabor **frazeoloških** kandidatov je mogoče pridobiti iz SSF. S tako pridobljenim seznamom bo manj verjetno, da bi se relevantne frazeološke enote spregledale, seveda pa se zaradi meril uvrščanja v NSSKJ ne bodo priključile vse enote iz SSF. Kljub temu da SSF vsebuje še številne druge tipe podatkov, ki bi lahko bili koristni tudi za NSSKJ, npr. razlage, kvalifikatorje, stavčne primere iz korpusa itn., ocenjujemo, da zaradi drugačnih izhodišč SSF teh podatkov ni zaželeno neposredno vključiti v NSSKJ.

Del podatkov o **etimologiji** lahko neposredno pridobimo iz SNB in SES. Časovni prihranek bo največji pri zapisovanju besed in njihovih rekonstrukcij, saj je ročno zapisovanje teh podatkov zaradi številnih posebnih znakov in velikega števila strukturnih elementov razmeroma zamudno, prav tako so možnosti za napake večje. Uporabljeni bodo tudi podatki o pomenu in motiviranosti; na podlagi seznama besed v SES, ki pripadajo isti besedni družini, bodo izdelane tudi sklicevalne etimološke osvetlitve. Do določene mere bo pozornost pri uvozu podatkov posvečena tudi pretvorbi v način ubeseditve, ki je predvidena za NSSKJ.

#### 4 Podatki, ki niso predvideni za vključitev

Na načelni ravni, kot rečeno, ne želimo vključevati podatkov, ki so neskladni z zastavljenimi konceptualnimi izhodišči NSSKJ in zato upočasnjujejo delovni proces ter povečujejo možnost napak. Prav tako ni smiselno vključevati podatkov, ki zaradi

<sup>6</sup> *Slovar sinonimov slovenskega jezika* predstavlja leksikalno sinonimijo slovenskega jezika. Vsebuje tiste slovarske enote, ki najmanj z enim od svojih pomenov vstopajo v sinonimno razmerje z najmanj eno drugo slovarsko enoto (Ahlin idr. 2003: 9).

<sup>7</sup> Tip *ženska oblika od X* v drugi izdaji SSKJ ne obstaja več (Perdih, Snoj 2015), zato bodo podatki o tem pridobljeni iz prve izdaje.

slabe strukturiranosti ali kvalitete slovarja ne prinesejo kvalitetnih podatkov. Še manj primerno je uvažati podatke, ki v veliki meri odražajo starejše jezikovno stanje, saj NSSKJ opisuje stanje v sodobnem jeziku. Za NSSKJ predvidevamo, da med podatke, ki jih zaradi različnih navedenih razlogov ni smiselno vnaprej uvoziti v podatkovno zbirko, sodijo normativni podatki, razlage, nekateri deli podatkov iz zaglavja, kvalifikatorji, nabor frazeologije iz SSKJ in nekateri drugi podatki.

## 5 Zaključek

V prispevku so prikazane nekatere možnosti pridobivanja manj spremenljivih jezikovnih podatkov za NSSKJ iz obstoječih inštitutskih slovarjev, s čimer je mogoče optimizirati nastajanje novega slovarja. Prednosti so zlasti hitrejše nastajanje slovarja (del podatkov je že pred začetkom redaktorskega dela pripravljen v podatkovni zbirki) in zmanjšanje možnosti napak pri ročnem vnosu. Previdnost je potrebna pri izbiri podatkov za avtomatski vnos, saj lahko zaradi drugačnega načina opisa jezikovnih dejstev in sprememb jezika skozi čas ali novejših etimoloških spoznanj rešitve v slovarjih, iz katerih se podatki pridobivajo, odstopajo od rešitev v nastajajočem slovarju. Za NSSKJ bodo uporabljeni podatki o izgovoru, dvojnice, sinonimi, besedotvorni podatki, frazeologija in etimologija. Vsi tako pridobljeni podatki bodo preverjeni in ovrednoteni glede na konceptualna izhodišča NSSKJ in glede na stanje v sodobnem jeziku, ugotovljeno z analizo gradiva.

## Viri in literatura

- AHLIN, Martin, LAZAR, Branka, PRAZNIK, Zvonka, SNOJ, Jerica, 2003: *Slovar sinonimov slovenskega jezika: splošna določila in opis zgradbe slovarskih sestavkov z vzorčno predstavitvijo*. Ljubljana: Založba ZRC, ZRC SAZU.
- ERLANDSEN, Jens, 2004: iLex – New DWS. Third International Workshop on Dictionary Writing Systems (DWS 2004). <http://nlp.fi.muni.cz/dws2004/pres/#15>
- Fran, slovarji Inštituta za slovenski jezik Frana Ramovša ZRC SAZU, 2014–, različica 2.0: [www.fran.si](http://www.fran.si)
- GLIHA KOMAC, Nataša, JAKOP, Nataša, JEŽOVNIK, Janoš, KLEMENČIČ, Simona, KRVINA, Domen, LEDINEK, Nina, MIRTIČ, Tanja, PERDIH, Andrej, PETRIC, Špela, SNOJ, Marko, ŽELE, Andreja, 2015: Osnutek koncepta novega razlagalnega slovarja slovenskega knjižnega jezika. Različica 1.1. Ljubljana: Inštitut za slovenski jezik Frana Ramovša ZRC SAZU.
- JAKOPIN, Primož, 1995: EVA – A Textual Data Processing Tool. Proceedings of the first TELRI seminar: Language Resources for Language Technology. Tihany.
- KEBER, Janez, 2011: *Slovar slovenskih frazemov*. Ljubljana: Založba ZRC, ZRC SAZU.
- KOŠMRLJ - LEVAČIČ, Borislava, SELIŠKAR, Tomaž, 2004: Uporabniški računalniški program SlovarRed 2.0 = SlovarRed 2.0 – a terminological application. Marjeta Humar (ur.): Terminologija v času globalizacije: zbornik prispevkov s simpozija Terminologija v času globalizacije, Ljubljana, 5.–6. junij 2003 = Terminology at the time of globalization: collected papers from the Scientific Conference Terminology at the Time of Globalization. Ljubljana: Založba ZRC, ZRC SAZU. 179–199.
- KREK, Simon, KOSEM, Iztok, GANTAR, Polona, 2013: *Predlog za izdelavo Slovarja sodobnega slovenskega jezika*. [http://trojina.org/slovar-predlog/datoteke/Predlog\\_SSSJ\\_v1.1.pdf](http://trojina.org/slovar-predlog/datoteke/Predlog_SSSJ_v1.1.pdf)
- KRVINA, Domen: *Sprotni slovar slovenskega jezika, različica 1.0*: [www.fran.si](http://www.fran.si)

- LEDINEK, Nina, PERDIH, Andrej, 2012: Izdelava XML-shem za slovarske projekte na primeru nastajajočih tipološko raznovrstnih slovarjev. *Zbornik Osme konference Jezikovne tehnologije*. Ljubljana: Institut »Jožef Stefan«. 123–128.
- LOGAR, Nataša, 2014: Jezikovni viri na klik in dotik. Hotimir Tivadar (ur.): *Prihodnost v slovenskem jeziku, literaturi in kulturi. 50. seminar slovenskega jezika, literature in kulture*. Ljubljana: Filozofska fakulteta. 103–108.
- NAPOTKI = *Napotki za redakcijo samostalnika, pridevnika in glagola v SSKJ. 1. in 2. del*, 1981. Ljubljana: Inštitut za slovenski jezik SAZU, Leksikološka sekcija.
- PERDIH, Andrej, 2013: *Tipologija slovanskih enojezičnih slovarjev. Doktorska disertacija*. Ljubljana: Filozofska fakulteta.
- PERDIH, Andrej, SNOJ, Marko, 2015: SSKJ 2. *Slavia Centralis VIII/1*. 5–15.
- ROMIH, Miro, KREK, Simon, 2012: Termania – prosto dostopni spletni slovarski portal. *Zbornik Osme konference Jezikovne tehnologije*. Ljubljana: Institut »Jožef Stefan«. 163–166. [http://www.trojina.org/slovenscina2.0/arhiv/2014/2/Slo2.0\\_2014\\_2\\_09.pdf](http://www.trojina.org/slovenscina2.0/arhiv/2014/2/Slo2.0_2014_2_09.pdf)
- SILVESTER, Marta, 1978: *Priročnik za tehnično stran redakcije gesel*. Ljubljana: Inštitut za slovenski jezik SAZU, Leksikološka sekcija.
- Slovar novejšega besedja slovenskega jezika*, 2012.
- Slovar pravopisnih težav, različica 1.0*: [www.fran.si](http://www.fran.si)
- Slovar slovenskega knjižnega jezika*, 1971–1991.
- Slovar slovenskega knjižnega jezika, druga, dopolnjena in deloma prenovljena izdaja*, 2014.
- Slovenski pravopis*, 2001.
- ŠIRCELJ - ŽNIDARŠIČ, Ivanka idr., 1998: *Besedišče slovenskega jezika z oblikoslovnimi podatki: A–Ž: po gradivu za slovar sodobnega knjižnega jezika zbrane besede, ki niso bile sprejete v Slovar slovenskega knjižnega jezika*. Ljubljana: Založba ZRC, ZRC SAZU.
- Termania: [www.termania.net](http://www.termania.net)
- ŻMIGRODZKI, Piotr, 2010: Jak powstaje Wielki słownik języka polskiego. Barbara Czopek Kopciuch, Piotr Żmigrodzki (ur.): *Język polski wczoraj – dziś – jutro*. Kraków: Wydawnictwo Lexis. 150–160. [http://wsjp.pl/strony\\_opisowe/Jak\\_powstaje\\_WSJP.pdf](http://wsjp.pl/strony_opisowe/Jak_powstaje_WSJP.pdf)