

# Slovarski zgledi

*Iztok Kosem*

## **Abstract**

In this paper, the role of examples in dictionary entries is presented, and an overview provided of relevant studies into the use and usefulness of examples. We put forward the different ways of presenting examples in general monolingual dictionaries, list the characteristics of a good dictionary example, and discuss the different methods of finding good examples. The focus then turns to the role and characteristics of examples in the proposal for a dictionary of contemporary Slovenian, the methods for their extraction, and the procedures to be followed for saving examples to the dictionary database and archiving them, before concluding with the different visualisation options for the (on-line) dictionary.

**Keywords:** dictionary examples, good examples, automatic extraction, visualisation, dictionary database

**Ključne besede:** slovarski zgledi, dobri zgledi, avtomatsko luščenje, vizualizacija, slovarska baza

## 1 UVOD

Zgledi so eden pomembnejših elementov slovarskega gesla, saj z njimi prikažemo rabo besed, njihovih kolokacij, stalnih zvez, frazeologije ipd. v kontekstu, torej kot se pojavljajo v dejanski jezikovni rabi. Upoštevajoč dejstvo, da večina jezikovnega opisa v slovarju dekontekstualiziranega, so zgledi kot kontekstualizirani primeri rabe besed za slovarske uporabnike ključnega pomena.

V prispevku najprej opredelimo vlogo zgledov v slovarju in podamo nekaj ugotovitev raziskav o pomembnosti slovarskih zgledov. Sledi pregled tujih in slovenskih slovarskih praks v enojezičnih slovarjih pri podajanju zgledov. Opredelimo tudi lastnosti dobrih slovarskih zgledov in predstavimo različne metode njihovega iskanja. Nato se osredotočimo na vlogo in lastnosti zgledov v predlaganem slovarju sodobnega slovenskega jezika (SSSJ), njihovemu pridobivanju iz korpusnega gradiva in beleženju v slovarski bazi, nekaj besed pa namenimo tudi vizualizaciji zgledov v slovarju. Zaključek je namenjen strnitvi razmišljanj in razmislekom o prihodnosti vloge ter beleženja slovarskih zgledov v slovarjih na splošno.

## 2 VLOGA ZGLEDA V SLOVARJU

Zgledi v slovarju opravljajo dve vlogi: receptivno in produktivno. Glavna receptivna vloga zgledov, ki je sicer tudi njihova temeljna vloga v slovarjih, je dopolnjevati slovarske razlage, zato morajo zgledi vsebovati predvsem s pomenom povezane informacije. Kot poudarjata Atkins in Rundell (2008: 454), uporabnik včasih brez zgledov težko razume razlago. Zgledi so lahko zelo koristni tudi pri navigaciji po (daljših) geslih, saj uporabniki lahko »prepoznajo pomen, ki ga iščejo, tako da poiščejo zglede z ubeseditvijo, podobno tisti, ki so jo prebrali ali jo želijo producirati« (Fox 1987: 137).

Produktivna vloga zgledov se nanaša predvsem na ponazarjanje skladenjskih vzorcev, vezljivosti, kolokacij in podobnih značilnosti iztočnice (Humble 2001), ki naj bi uporabnikom pomagali pri pisanju ali (redkeje) govorjenju. Produktivno naravnane zglede najdemo predvsem v slovarjih za tiste, ki se jezika učijo, npr. slovarjih za tuje govorce ali slovarjih za mlajše matere govorce (šolskih slovarjih ipd.).

Raziskave, ki se ukvarjajo s slovarskimi zgledi, se osredotočajo predvsem na njihovo korist pri produkciji za nematerne govorce. Pri tem največkrat uporabijo metodo, pri kateri morajo udeleženci raziskave oblikovati stavke z (neznanimi) besedami, pomagajo pa si lahko s slovarskimi informacijami. Udeleženci so razdeljeni v več skupin: nekateri imajo na voljo samo definicije, drugi definicije in zglede,

v nekaterih raziskavah (npr. Frankenberg-Garcia 2012, 2014) je uporabljena še skupina udeležencev, ki ima na voljo samo zglede. Ugotovitve večine raziskav (Summers 1988; Laufer 1993; Nesi 1996; Al-Ajmi 2008) niso preveč spodbudne, saj kažejo, da zgledi ne prinesejo veliko dodane vrednosti uporabnikom pri produkciji. Vendar pa Frankenberg-Garcia (2012) poudarja, da imajo omenjene raziskave dve ključni metodološki pomanjkljivosti: kot prvo, čeprav merijo produktivno vrednost zgledov, so naloge zastavljene tako, da morajo udeleženci študije najprej razvozlati pomen neznane besede in jo nato takoj uporabiti v stavku, torej sta združeni receptivna in produktivna raba, kar je pri uporabi slovarjev redko. In drugič, uporaba neznanih besed pri preverjanju produkcije ne odraža realne rabe slovarjev in jezikovne produkcije nasploh, saj »ljudje pri pisanju redko uporabljajo njim povsem nove besede« (Laufer 1993: 138).

Frankenberg-Garcia (2012, 2014) je izboljšala metodologijo prejšnjih raziskav tako, da je jasno ločila preverjanje receptivne in produktivne vloge zgledov ter obenem ločila zglede na receptivno in produktivno naravnane. Udeležence raziskave je razdelila v štiri skupine: kontrolno skupino (brez slovarja), skupino, ki je dobila samo definicije, skupino, ki je dobila en korpusni zglede, in skupino, ki je dobila več korpusnih zgledov. Ugotovitve so pokazale, da je pri razumevanju pomena besede več korpusnih zgledov skoraj enako učinkovitih kot razlaga, pri produkciji pa ima več korpusnih zgledov precej večjo korist kot samo eden, zgledi pa so nasploh precej bolj koristni od razlag.

Raziskave, ki ponujajo informacije o tem, kako pogosto uporabniki pogledajo zglede, so redke. Béjoint (1981) je v svoji študiji s študenti ugotovil, da pri uporabi slovarja zglede pogledajo precej pogosto. O podobnih rezultatih poroča tudi raziskava med 620 študenti (449 maternimi in 171 nematernimi govorcji angleščine) na univerzi Aston (Kosem 2010); zgledi so bili četrtri najpogosteje uporabljeni del slovarskega gesla (za definicijami, izgovorjavo in sinonimi), pri nematernih govorcjih pa celo drugi najpogosteje uporabljeni del gesla (za definicijami).

### 3 ZGLEDI V SPLOŠNIH ENOJEZIČNIH SLOVARJIH

Analiza splošnih enojezičnih slovarjev<sup>1</sup> kaže, da obstajajo glede na obravnavo in prikaz zgledov tri skupine slovarskih virov. V prvo skupino sodijo slovarji, ki zglede podajajo v obliki iztržkov (iz gradiva prevzetih in primerno abstrahiranih zvez), občasno poleg iztržkov ponudijo tudi celostavčne zglede (španski slovar)

1 Analiza je vključevala samo tuje slovarje, ki obstajajo v spletni obliki. Analizirani spletni slovarji imajo lahko tiskane različice oz. so izpeljani iz tiskanih različic, najdemo pa tudi slovarje, ki obstajajo samo v spletni obliki (npr. poljski, nizozemski). Slovenski slovarji so obravnavani ločeno po opredelitvi vseh treh skupin.

in/ali njihov vir (npr. estonski slovar), pri čemer nekateri slovarji zglede ponudijo samo pri določenih (pod)pomenih ali frazah. Gre predvsem za (nekorpusne) slovarje, ki so bili prvotno izdelani za tisk in potem ponujeni tudi na spletu, ali pa za nedavno objavljene slovarje, ki so bili konceptualizirani na osnovi leksikografskih praks prejšnjega stoletja. Med omenjene slovarje spadajo Slovar češkega knjižnega jezika<sup>2</sup> (Slovník spisovného jazyka českého, 1989), slovar španskega jezika Kraljeve španske akademije<sup>3</sup> (Diccionario de la lengua Española, 2014), Razlagalni slovar estonskega jezika<sup>4</sup> (Eesti keele seletav sõnaraamat, 2007) in Hrvaški enciklopedični slovar<sup>5</sup> (Hrvatski enciklopedijski rječnik, 2003).

Drugo skupino predstavljajo slovarji, v katerih prevladujejo celostavčni (korpusni) zglede, iztržki so uporabljeni precej redkeje ali pa sploh ne. Primeri takšnih slovarjev so slovarji Oxford (Oxford Dictionaries<sup>6</sup>), Macmillan (Macmillan English Dictionary<sup>7</sup>) in Merriam-Webster (The Merriam-Webster Online Dictionary<sup>8</sup>) za angleščino, Slovar sodobnega danskega jezika (Den Danske Ordbog<sup>9</sup>), Veliki slovar poljskega jezika<sup>10</sup> (Wielki Słownik języka Polskiego) in Splošni nizozemski slovar<sup>11</sup> (Algemeen Nederlands Woordenboek). V omenjenih slovarjih zasledimo precej različne prakse navajanja zgledov: v Macmillanu in Slovarju sodobnega danskega jezika so celostavčni zglede del prikazanega gesla, podobno velja za Merriam-Webster, kjer pa so celostavčni zglede združeni pod ločeno rubriko proti koncu gesla, medtem ko so pri pomenih, podpomenih ipd. podani iztržki. Tudi slovar Oxford prvotno ponuja iztržke, celostavčni zglede so pod vsakim pomenom oz. podpomenom na voljo s klikom gumba *More example sentences*. Poljski in nizozemski slovar zglede sploh ne ponujata na prvi ravni prikaza rezultatov, ampak si jih uporabnik lahko ogleda s klikom na povezavo (nizozemski slovar) oz. zavihka (poljski slovar). Omenjena slovarja ter tudi Slovar sodobnega danskega jezika za vsak zgled navajajo tudi vir.

V tretji skupini najdemo portale, kot je nemški DWDS<sup>12</sup> (Das Digitale Wörterbuch der deutschen Sprache), ki uporabnikom na enem mestu ponuja zadetke iz slovarjev in korpusov ter druge relevantne informacije o iskani besedi.<sup>13</sup> Za to skupino virov je z vidika zgledov zlasti pomembna povezava med slovarji in korpusi, sploh z vidika ugotovitev Frankenberg-Garcie o koristih ponujanja

2 <http://ssjc.ujc.cas.cz>, (spletna verzija na voljo od 2011).

3 <http://lema.rae.es/drae> (dostop 8. 8. 2015).

4 <http://en.eki.ee/dict/ekss> (dostop 8. 8. 2015).

5 Dostopen prek Hrvaškega jezikovnega portala, <http://hjp.novi-liber.hr> (dostop 8. 8. 2015).

6 <http://www.oxforddictionaries.com/> (dostop 8. 8. 2015).

7 <http://www.macmillandictionary.com/> (dostop 8. 8. 2015).

8 <http://www.merriam-webster.com> (dostop 8. 8. 2015).

9 <http://ordnet.dk/ddo> (dostop 8. 8. 2015).

10 <http://wsjp.pl> (dostop 8. 8. 2015).

11 <http://anw.inl.nl> (dostop 8. 8. 2015).

12 <http://www.dwds.de/> (dostop 8. 8. 2015).

13 Tudi nekateri ostali slovarji, kot je na primer danski slovar, na spletni strani ponujajo dostop do korpusa, vendar pa ne ponujajo enotnega iskanja po vseh virih in hkratnega prikaza zadetkov.

večjega števila zgledov uporabnikom. Slaba plat takšnih portalov je, da je uporabniku predloženo veliko različnih informacij,<sup>14</sup> kar otežuje njihovo interpretacijo in rabo.

Če se ozremo po slovenskih slovarjih, Slovar slovenskega knjižnega jezika (SSKJ) in njegov naslednik SSKJ2 sodita v prvo od zgoraj omenjenih skupin slovarjev, saj ponujata zglede v obliki iztržkov (temelječih na virih in tudi izmišljenih<sup>15</sup>). To za SSKJ ni presenetljiv podatek, saj je nastajal v času predkorpusne leksikografije. Dejansko je količina zgledov v geslih SSKJ precej obsežna in odstopa od marsikaterega podobnega tujega slovarja, tudi takšnih, ki so bili izdelani nedavno, kot je slovar španskega jezika Španske kraljeve akademije. Zgledi so tudi eden od slovarskih elementov v SSKJ, ki je pri pripravi SSKJ2 doživel večje spremembe, saj so jih avtorji popravljali in nadomeščali (zaradi sprememb družbenega sistema ipd.) ali pa so dodajali povsem nove zglede. A kot ugotavlja Krek (2014: 146), spremembe obstoječih zgledov pogosto niso ustrezne oz. potrebne ali pa novo dodani zglede ne prinašajo neke nove dodane informacije za razumevanje pomena. Nadomeščanje oz. popravljanje obstoječih zgledov v SSKJ2, predvsem z razlogom sprememb družbenega sistema, se sploh zdi redundantno glede na to, da avtorji slovar predstavljajo kot vir, ki odseva 150 let slovenskega jezika.<sup>16</sup> To potrjuje tudi Krekov zaključek, da je bil v ponazarjalnem gradivu »izbrisan precejšen del resničnosti dobe pred letom 1991« (ibid: 147).

Še pred SSKJ2 je izšel Slovar novejšega besedja slovenskega jezika (2012; SNB), katerega avtorji so deloma upoštevali sodobne leksikografske trende in v gesla poleg iztržkov vključili (celostavčne) korpusne zglede. Kot je navedeno v Uvodu (SNB: 9), je bil glavni besedilni vir pri pripravi slovarja 318-milijonski korpus Nova beseda:<sup>17</sup>

Na podlagi avtentične besedilne rabe, izpričane v tristomilijonskem besedilnem korpusu Nova beseda, smo v 5384 slovarskih sestavkih interpretirali 6512 pomenov in podpomenov aktualnih besed in besednih zvez, ki imajo svoj izvor v raznolikih področjih družbene dejavnosti.

Analiza zgledov pokaže, da so avtorji slovarja v odsotnosti (dobrih) zgledov črpali zglede tudi iz drugih korpusov, zlasti iz 1,2-milijardnega korpusa Gigafida. To samo na sebi ni sporno, postavlja pa pod vprašaj zgoraj citirano metodologijo izdelave geslovnika, sploh pri geslih, kot je npr. *bandži skok*:

14 DWDS sicer omogoča omejevanje zadetkov samo na določene vire.

15 Kot piše v Uvodu v SSKJ (1991: XXII), »k/adar je bilo slovarsko gradivo pomanjkljivo, so naredili uredniki iztržke po drugih virih ali po spominu«.

16 Marko Snoj 2. novembra 2013 za STA: <http://www.rtvsllo.si/kultura/knjige/akademaska-vojna-okrog-novega-slovarja/321592> (dostop 8. 8. 2015).

17 [http://bos.zrc-sazu.si/s\\_beseda3.html](http://bos.zrc-sazu.si/s_beseda3.html) (dostop 8. 8. 2015).

**bándži skòk** -- skôka in skóka m (ô, ô ó; ɔ)

skok v globino, pri katerem je skakalec pripet z dolgo elastično vrvjo; skok z elastiko: Obnaša se kot frkolin, ki se pred tovarišijo postavi z bandži skokom, ko se privezan na elastično vrv vrže z mostu v globel **E ↑bungee (jumping) in (↑)skòk**

Navedeni zgled je (rahlo) modificiran stavek iz korpusa Gigafida,<sup>18</sup> problematično pa je, da korpus Nova beseda ne vsebuje niti enega zadetka za *bandži skok* (v Gigafidi jih je 5). V tem primeru torej zgled ponazarja iztočnico, za katero sploh ne vemo, kako je prišla v slovar. Poleg tega so ravno zaradi osredotočenosti slovarja na novejšo besedje, ki se praviloma v korpusih pojavlja redko, zgledi velikokrat omejeni zgolj na ponazarjalno vlogo, saj ne prinašajo dodane vrednosti k razumevanju pomena.

Bolj sistematičen (in popolni) korpusni pristop je bil uporabljen pri izdelavi Leksikalne baze za slovenščino (Gantar et al. 2012; LBS), ki vsebuje 2.500 gesel, v katerih najdemo 152.996 zgledov oz. povprečno več kot 61 zgledov na geslo. Vsi zgledi so celostavčni in izhajajo iz korpusa Gigafida (Logar Berginc et al. 2012). Zgledi v LBS niso modificirani, saj se izbira zgledov z vidika oblikovanja slovarske baze in izdelave slovarja razlikuje; v slovarski bazi namreč težimo k izbiri večje količine korpusnih zgledov, ki so tipično celi stavki, vzeti iz korpusa, in predstavljajo kandidate za dobre slovarske zglede (Gantar 2015). LBS je za slovenski, pa tudi širši prostor pomemben vir zaradi uporabljene metodologije. Med izdelavo LBS so bile namreč razvite in preizkušene številne metode, ki kombinirajo leksikografsko delo z avtomatskim luščenjem podatkov (tudi zgledov) ter predstavljajo temelje izdelave slovarja sodobnega slovenskega jezika in z njim povezane slovarske baze (gl. razdelek 5).

#### 4 KAKŠEN JE DOBER SLOVARSKI ZGLED?

Med največkrat omenjenimi lastnostmi dobrih zgledov so naravnost oz. pristnost, tipičnost, informativnost in razumljivost. O pristnosti zglede govorimo, ko deluje naravno, tj. ko je tak, kakršnega bi v jezikovni rabi dejansko srečali. Ravno zato se pristnost velikokrat povezuje tudi z avtentičnostjo, ki naj bi jo zagotavljalo izbiranje zgledov iz zbranih avtentičnih besedil oziroma korpusov, kar je v sodobni leksikografiji tako rekoč standard. Ob tem velja omeniti, da so že slovarji, ki so nastajali v času predkorpusne leksikografije, vsebovali zglede iz avtentičnih besedil (npr. the Oxford English Dictionary) ali vsaj na besedilih vsebovane dele zgledov oz. iztržke (npr. SSKJ). Vendar pa so takrat leksikografi zglede velikokrat oblikovali kar sami oz. so si jih izmislili na podlagi lastne intuicije; omenjeno prakso so korpusne študije (npr. Sinclair 1991; Hunston in

<sup>18</sup> <http://www.gigafida.net/> (dostop 8. 8. 2015).

Laviosa 2001) postavile pod velik vprašaj, zlasti za namene vključevanja zgledov v splošne enojezične slovarje.<sup>19</sup>

Sorodno načelu pristnosti je načelo tipičnosti, ki veleva, da morajo slovarski zgledi pokazati tipično rabo iztočnice z vidika konteksta, skladnje, frazeologije in kolokacij. Sodobna orodja za analizo besedil leksikografom pri iskanju tipičnih zgledov precej pomagajo, saj z njihovo pomočjo lahko poiščejo slovnične strukture, kolokacije in koligacijske lastnosti iztočnice (npr. v kateri obliki ali sklonu se beseda z določenim kolokatorjem najpogosteje pojavlja).

Informativen zglede daje slovarskemu geslu dodano vrednost, največkrat v povezavi z definicijo, ki naj bi jo uporabnik zaradi zglede lažje razumel. Zgled hkrati potrjuje informacije, podane v definiciji, in kontekstualizira rabo iztočnice v določenem pomenu ali podpomenu. Informativnost je deloma povezana tudi s količino zgledov v posameznem geslu; elektronski mediji resda ponujajo možnost vključitve zelo velikega števila zgledov, a dejstvo, da se pri vsakem dodatnem zgledu pojavlja vprašanje, kaj res novega prispeva h geslu, kliče po preudarnosti njihove uporabe, če želimo ohraniti informativnost in uporabniško prijaznost slovarja. Po drugi strani pa ne gre pozabiti ugotovitev Frankenberg-Garcie (2012; 2014) o tem, da je več korpusnih zgledov včasih celo bolj koristnih od same definicije.

Razumljivost zglede dosežemo z izogibanjem kompleksnim strukturam, redki in zahtevni leksiki ter pretirani dolžini. Na ta način se bodo uporabniki lahko osredotočili na iztočnico in od njih ne bomo zahtevali pretiranega miselnega napora za procesiranje informacij v sobesedilu. Seveda se določenim elementom težko izognemo za vsako ceno; npr. redkejša in »zahtevnejša« besede se rade pojavljajo skupaj z drugimi zahtevnimi besedami ali besednimi zvezami, ki jih moramo vključiti v zglede, če se želimo držati načela naravnosti in tipičnosti. Zgledi ne smejo biti predolgi, prekratki pa tudi ne, sploh če je slovar namenjen tudi za produktivno rabo, saj uporabniku v slednjem primeru ponudijo premalo informacij za uspešno produkcijo, pogosto pa tudi recepcijo.

Vse večji pomen ima tudi oblika zglede, saj se je v sodobni leksikografski praksi uveljavila uporaba celostavnih zgledov, tudi v splošnih slovarjih za materne govorce, ki so sicer še pred nekaj desetletji uporabljali skoraj izključno iztržke ali kratke zglede. Temu je nedvomno botrovala prevlada elektronskih medijev, zlasti spletnega, kjer ni potrebno pretirano varčevanje s prostorom, hkrati pa tudi spoznanja, da iztržki in podobni kratki zglede, iztrgani iz stavkov oz. povedi, delujejo abstraktno in nenaravno (gl. Williams 1996).

<sup>19</sup> Nekoliko v manjši meri to velja za slovarje za tuje govorce, saj kot navajata Atkins in Rundell (2008: 456), je veliko predkorpusnih angleških slovarjev za tuje govorce vsebovalo veliko dobrih slovarskih zgledov, ki so delovali avtentično, a niso bili vzeti iz avtentičnih besedil.

Posebna in nadvse pomembna tema pri izbiri zgledov je ideološki vidik, saj se predvsem v zgledih izkazuje slovarska ideološkost, torej resničnost, kot jo vidijo leksikografi. Z zgledi namreč leksikografi velikokrat povedo tisto, česar ne morejo v razlagi zaradi omejitev v definicijskem jeziku in tudi veliko večje eksplicitnosti glede ideološkosti (prim. Meschonnic 1991; Béjoint 2000; Epple 2000; Schutz 2002; pri nas Gorjanc 2004; 2005; 2012). Tako so zgledi dejansko mikrostrukturni element, v katerem se lahko najbolj odsevajo družbene vrednote, preko tega pa tudi izpostavljajo vrednostni sistem slovarske ekipe (Gorjanc 2014). Gorjanc (2014) na primeru izrazja v SSKJ, povezanega s homoseksualnostjo, ponazori, kako se lahko družbeni predsodki v slovarju zabeležijo kot sprejemljivi oz. del norme. Težave z ideološkimi spremembami pri zgledih rabe najdemo tudi v SSKJ2 (Krek 2014: 145–147). Leksikografi se torej morajo zavedati svoje nenevtralne vloge in pri izbiri zgledov (in seveda pri pisanju slovarja nasploh) ravnati družbeno občutljivo in odgovorno (Béjoint 2000: 124).

Najti zgled, ki bi ustrezal vsem naštetim kriterijem, še zdaleč ni enostavno. Čeprav imajo danes leksikografi na voljo zelo velike korpuse in posledično veliko potencialnih zgledov za posamezno iztočnico, se velikokrat zgodi, da najdejo stavke, ki izpolnjujejo dva kriterija, tudi tri, zelo redko pa stavke, ki izpolnjujejo vse kriterije dobrega slovarskega zgeda. Dejansko bi lahko kandidate za zglede razvrstili v stopnje od slabih, bolj slabih kot dobrih, dokaj oz. potencialno dobrih in dobrih, pri čemer naj bi bili dobri tisti, ki jih lahko iz korpusa prenesemo neposredno v slovar. A takšnih neposredno uporabnih dobrih zgledov je malo, precej več je potencialno dobrih, torej takšnih, ki bi ob malenkostnih popravkih postali dobri zgledi. Ampak če se odločimo v slovarju uporabljati modificirane zglede, kaj to pomeni za načelo avtentičnosti? Je naš slovar še korpusni? Kot pravita Atkins in Rundell (2008: 458), je pogosto omenjena izbira med izmišljenimi in avtentičnimi zgledi zavajajoča, saj ne odraža realne leksikografske prakse. Tudi slovarji, kot je COBUILD, izdelani s popolnim korpusnim pristopom, vključujejo modificirane zglede, čeprav velja poudariti, da so se avtorji slovarja COBUILD skušali tej praksi, če je bilo le mogoče, izogniti (Fox 1987).

Najpogostejše oblike modifikacije zgledov so krajšanje oz. izpuščanje nerelevantnih delov, kot so odvisniki ali vrinjeni stavki, poenostavljanje kompleksnih struktur in zamenjava besed ali besednih zvez s pogostejšimi ali ustrežnejšimi poimenovanji. Krajšanje se zdi še najmanj sporno in je dejansko velikokrat povsem legitimno z vidika informativnosti, saj stavki večkrat vsebujejo dele, ki so nepotrebni oz. nerelevantni, če niso podani v širšem kontekstu besedila, kot velja recimo za *na primer* v stavku za iztočnico *anonimnost*:

Jane Austen, na primer, je živela v popolni **anonimnosti**.

Po drugi strani lahko poenostavljanje kompleksnih struktur in zamenjava besed v precej večji meri vplivata na naravnost ali tipičnost zgledov, zato se jima je treba čim bolj izogibati. V nekaterih primerih je sicer zamenjava besed ali besednih zvez



potrebna, npr. pri lastnih imenih, ki jih nadomestimo z zaimki ali generičnimi imeni tipa Janez Novak, ali zaradi izogibanja razžalitvi določene družbene skupine, a tudi tu zadeve niso enoznačne, sploh če gre za javne osebnosti ali če je konkretna oseba tesno povezana z določenim kontekstom rabe iztočnice. Spodnji zgled za *mojstrsko* tako nikakor ne bi imel informativne vrednosti, pa tudi ne bi zvenel naravno, če bi Cristiano Ronaldo zamenjali z generičnim imenom tipa Janez Novak:

Izid polčasa in tudi končni izid je z **mojstrsko** izvedenim prostim strelom postavil Cristiano Ronaldo.

Pogostost in korenitost spreminjanja zgledov sta odvisni tudi od ciljnega uporabnika slovarja. Pri izdelavi slovarja za nematerne govorce ali mlajše materne govorce, ki še razvijajo jezikovno kompetenco in poznajo manjši nabor besedišča, bo posegov v zglede več,<sup>20</sup> medtem ko lahko pri slovarjih za odrasle materne govorce v zglede posegamo precej manj. Pri odločitvah nas mora voditi tudi namen slovarja: če naj bi uporabnikom pomagal pri jezikovni produkciji in ne samo recepciji, potem je sploh pomembno, da zgledi ostanejo čim bolj naravni in tipični.

Posebna oblika modificiranja zgledov je jezikovno popravljanje. Če najdemo dober korpusni stavek, v katerem manjka vejica, lahko vstavimo vejico in stavek uporabimo v slovarju? Kaj pa, če stavek vsebuje narobe črkovano besedo, besedo v napačnem sklonu ali napačen besedni red? Čeprav se zdi popravek napake črkovanja res malenkosten in včasih nujen poseg, moramo pri popravljanju paziti, saj se lahko hitro znajdemo v drugi skrajnosti, ko vidimo kot napako tudi izbiro ubeseditve ipd. in postane popravljanje zgleda skoraj identično temu, kot če bi si zgled izmislili. Dobro se je držati načela, da skušamo na vsak način najti dober korpusni zgled, ki ne potrebuje popravkov, če pa takšnih zgledov ni na voljo in najdemo potencialno dobre zglede z določenimi (manjšimi) jezikovnimi napakami, jih lahko s popravki vključimo v slovar. Na vsak način pa se je priporočljivo izogniti vnašanju novih napak v slovarske zglede.<sup>21</sup>

## 5 METODE PRIDOBIVANJA DOBRIH SLOVARSKIH ZGLEDOV

Iskanje (dobrih) slovarskih zgledov je zelo zahteven in potencialno tudi zamuden ter posledično drag proces. Prvič zato, ker je dober slovarski zgled ob upoštevanju vseh v prejšnjem razdelku omenjenih kriterijev zelo težko najti. Drugič, korpusi

20 Dejansko tudi Atkins in Rundell (2008) svoje odobranje modificiranja zgledov omejeta skoraj izključno na slovarje za nematerne govorce.

21 S tega vidika je recimo vprašljiva nerazumljiva odločitev avtorjev SNB, ki so se odločili na koncu (celostavnih) zgledov opustiti ločila, največkrat piko, kar ob dejstvu, da se zgledi začenjajo z veliko začetnico, ni jezikovno pravilno in tudi na določen način zgledom zmanjšuje avtentičnost.

postajajo vse večji, kar leksikografu sicer ponudi večjo izbiro kandidatov za dobre zglede, a hkrati tudi večjo količino stavkov za analizo. In tretjič, zgledi so ključni element slovarske mikrostrukture, ne samo na ravni gesla, ampak tudi pomenov, podpomenov, kolokacij, stalnih zvez, frazeologije itd. Skratka, leksikograf mora v veliki količini podatkov poiskati veliko dobrih zgledov za vsako slovarsko geslo.

Govorimo lahko o dveh metodah pridobivanja dobrih slovarskih zgledov: ročni in polavtomatski. Pri ročni metodi leksikograf med korpusnimi zadetki oz. konkordancami za določeno iztočnico izbere zglede, pri čemer si lahko pomaga s sortiranjem, filtriranjem in podobnimi funkcijami za urejanje zadetkov. V pomoč mu je lahko tudi vnaprejšnje razvrščanje zgledov glede na kolokacije in slovnične relacije, kar npr. v korpusnem orodju Sketch Engine<sup>22</sup> (Kilgarriff et al. 2004) omogoča funkcija Besedne skice.

Pri polavtomatski metodi leksikografu pomaga orodje za prepoznavo dobrih zgledov, kot je npr. GDEX (*Good Dictionary Examples*; Kilgarriff et al. 2008), ki mu ponudi nabor kandidatov za dobre slovarske zglede, med katerimi leksikograf potem izbere ustrezne. GDEX (gl. tudi razdelek 5.1) razvršča zglede glede na njihovo kakovost pri značilnostih, kot so dolžina zgeda, celostavčna oblika, preprosta ali manj kompleksna skladijska zgradba povedi, prisotnost ali odsotnost redkih besed, spletnih in elektronskih naslovov ipd. Mnoge od teh značilnosti so posredno povezane s tipičnostjo, informativnostjo in razumljivostjo, torej lastnostmi dobrega zgeda. Značilnosti lahko razdelimo v dve skupini: v prvi so tiste, ki jih zged mora vsebovati, npr. celostavčnost, odsotnost spletnih naslovov, odsotnost izredno dolgih ali redkih besed ipd. Če zged ne ustreza vsaj eni od teh značilnosti, dobi toliko kazenskih točk, da se takoj znajde na dnu vseh zadetkov. V drugi skupini so značilnosti, ki so bodisi zaželene bodisi nezaželene (stopnjo (ne)zaželenosti določimo s težo, ki jo pripišemo posamezni značilnosti, in višino dodatnih/odbitih točk), a je pomemben predvsem kumulativni seštevek vrednosti vseh značilnosti v konfiguraciji.

Ključna razlika med metodama je trajanje, saj je polavtomatska metoda precej hitrejša od ročne, a nič manj zanesljiva (Kosem et al. 2012; 2013). V sodobni korpusni leksikografiji tako polavtomatska metoda vse pogosteje nadomešča ročno, zlasti pri projektih, kjer se izdelujejo slovarske baze, ki vsebujejo več zgledov kot na njih temelječi slovarji (npr. nizozemski ANW).

## 6 ZGLEDI V SLOVARJU SODOBNEGA SLOVENSKEGA JEZIKA

V tem razdelku se posvetimo vlogi zgledov v predlaganem slovarju sodobnega slovenskega jezika, in sicer načinu njihovega pridobivanja in beleženja, razliki

<sup>22</sup> <http://www.sketchengine.co.uk/> (dostop 8. 8. 2015).

med zgledi v slovarski bazi in slovarju, nekaj besed pa namenimo tudi različnim možnostim vizualizacije zgledov, ki jih ponuja digitalni medij.

## 6.1 Način pridobivanja in beleženja slovarskih zgledov

Pridobivanje zgledov je, prek orodja GDEX, sestavni del različice polavtomatske metode, imenovane avtomatsko luščenje leksikalnih podatkov (ALLP; Kosem et al. 2012), pri kateri se podatki (skladenjske strukture, kolokacije in zgledi, pa tudi določene informacije o iztočnici in predlogi za slovnične oznake) prek orodja Sketch Engine oz. aplikacije Besedne skice z uporabo API-skripte (angl. *Application Programming Interface*) avtomatsko izvozijo iz korpusa neposredno v program za izdelavo slovarjev, kjer jih leksikograf pregleda, selekcionira in uredi.<sup>23</sup> Metoda leksikografu še vedno ponudi dovolj podatkov za temeljito analizo in izdelavo gesla. Izkušnje pri izdelavi LBS kažejo, da z uporabo takšne metode leksikograf pri izdelavi gesla ne pregleda nič manj zgledov (pogosto jih celo več!), kot bi jih z uporabo ročne ali polavtomatske metode znotraj korpusnega orodja. Prednosti metode so odprava nepotrebnega kopiranja podatkov iz korpusnega orodja in njihovega vnašanja v program za izdelavo slovarjev ter posledično hitrejša, pa tudi bolj razpršena in s tem zanesljivejša analiza.

Za zglede ključni del ALLP je priprava ustrezne konfiguracije oz. konfiguracij za orodje GDEX. Že od leta 2011 je na voljo različica konfiguracije GDEX za slovenščino (Kosem et al. 2011), ki smo jo izdelali za namene izdelave LBS in je dokaj uspešno dosegala cilj vsaj treh dobrih zgledov od desetih ponujenih na kolokator, kar je bila privzeta nastavitvev pri izdelavi LBS. Vendar pa omejena različica ne zadosti potrebam ALLP, kjer izvažamo **prvih** X (največkrat od tri do pet) ponujenih zgledov na kolokator in morajo biti dejansko vsi potencialno dobri. Poleg tega so se že pri uporabi prvotne različice pri analizi v orodju Sketch Engine za potrebe izdelave LBS opazile precejšnje razlike v kvaliteti ponujenih zgledov po posameznih besednih vrstah. Zato smo pri testiranju ALLP pri izdelavi LBS za vsako besedno vrsto izdelali samostojno konfiguracijo GDEX. Izdelava konfiguracij je potekala v dveh korakih: najprej smo na podlagi analize dobrih zgledov v LBS oblikovali izhodiščno konfiguracijo za vsako besedno vrsto, potem pa smo s prilagajanjem nastavitvev izdelali nove različice, katerih evalvacija je vedno potekala tako, da smo rezultate nove različice primerjali z rezultati prejšnje. Postopek smo ponavljali, dokler nismo izoblikovali optimalne končne verzije konfiguracije GDEX za postopek ALLP. Pomemben rezultat tega dela analize je oblikovanje več novih klasifikatorjev, ki jih prvotna verzija GDEX ni vključevala. Klasifikatorji so tako sledeči:

<sup>23</sup> O podobni metodi sta razmišljala že Rundell in Kilgarriff (2011).

- Cela poved. Na ta način prioritiziramo zglede, ki ustrezajo načelu celostavnosti.
- Ne vsebuje pojavnic s frekvenco manj kot 3. Iščemo zglede, ki ne vsebujejo zelo redkih besed, napak in korpusnega šuma.
- Poved mora biti daljša od 7 pojavnic. Nočemo prekratkih zgledov, saj jim velikokrat manjka kontekst. Lažje je krajšati daljše zglede kot iskati nove.
- Poved mora biti krajša od 60 pojavnic. Izločamo samo res dolge povedi, daljše povedi vedno lahko okrajšamo.
- Poved ne sme vsebovati ponovitve iztočnice. Gre za pomemben klasifikator, kajti večkratno ponavljanje iztočnice zgledu jemlje razumljivost in informativnost.
- Vsebuje elektronski ali spletni naslov. Zgledi, ki ustrezajo temu kriteriju, prejmejo visok kazenski pribitek.
- Optimalna dolžina (med X in Y pojavnic). Medtem ko s klasifikatorjema za minimalno in maksimalno dolžino povedi na nek način izločamo prekratke in predolge povedi (jih potiskamo na dno seznama), z optimalno dolžino nagrajujemo povedi z dolžino znotraj danega razpona. Najpogosteje je optimalna dolžina zгледа med 15 in 40 pojavnic, odvisno od besedne vrste. Analiza dobrih zgledov v LBS pri izdelavi prvotne različice (Kosem 2012) je npr. pokazala, da je povprečna dolžina zgledov za pridevniške iztočnice 28,64 pojavnice, za prislove 27,03 pojavnice in za prislov 27,39 pojavnice.
- Vsebuje redke leme. Klasifikator dodeli točkovni odbitek povedi za vsako redko lemo, ki jo vsebuje. Frekvenčna meja, ki opredeljuje redkost, je odvisna od velikosti korpusa.
- Vsebuje pojavnice, daljše od 12 znakov. Klasifikator kaznuje vsako pojavnico, ki izpolnjuje omenjeni kriterij. Analiza je namreč pokazala, da so daljše pojavnice največkrat nebesede ali korpusni šum.
- Število ločil v zgledu (brez vejic). Klasifikator točkovno kaznuje poved, v kolikor je preseženo določeno število ločil v njej, pri čemer se vejice ne upoštevajo.
- Število vejic v povedi. Klasifikator točkovno kaznuje povedi z več kot tremi vejicami, saj je bilo ugotovljeno, da so takšne povedi pogosto kompleksnejše in posledično slabši kandidati za dobre zglede.
- Pojavnice z velikimi začetnicami. Klasifikator točkovno kaznuje povedi, ki vsebujejo pojavnice z velikimi začetnicami, in je namenjen predvsem kot dopolnilo klasifikatorju za lastna imena.

- Pojavnice z mešanimi simboli (npr. črke in številke). Klasifikator točkovno kaznuje nebesede in korpusni šum.
- Lastna imena. Klasifikator točkovno kaznuje povedi s pojavnicami, ki so v korpusu označene kot lastna imena. Če je takšnih pojavnic v povedi več, je kaznovana vsaka posamezna pojavitev.
- Zaimki. Klasifikator z odbitkom kaznuje vsako pojavitev zaimka v povedi. Klasifikator je koristen predvsem, ko je zaimkov v povedi več, saj so takšne povedi ponavadi manj razumljive oz. potrebujejo dodaten kontekst.
- Položaj leme v povedi. Klasifikator točkovno kaznuje povedi, kjer se lema pojavlja izven določenega razpona v povedi. Tako je bilo za glagolske leme ugotovljeno, da so boljši kandidati za dobre zglede tiste povedi, v katerih se glagol ne pojavlja na začetku, tj. v prvih 40 ostopkih pojavnicah povedi.
- Seznam prepovedanih besed na začetku povedi. Pri izdelavi konfiguracij se je izkazalo, da so določene besede na začetku povedi že dober indikator, da ne gre za dobrega kandidata za slovarski zglede. Gre za besede, kot so *sledi*, *tovrsten*, *oboji* ipd., ki nakazujejo, da poved zahteva dodaten kontekst. Za namene klasifikatorja je bil na podlagi analize in opaznan pri evalvaciji konfiguracij izdelan seznam takšnih besed. Klasifikator tako točkovno kaznuje povedi, ki se začenjajo s katero od besed na seznamu.
- Seznam prepovedanih besednih zvez na začetku povedi. Podoben klasifikator kot klasifikator za prepovedane besede, s tem da kaznuje pojavitev določenih večbesednih nizov na začetku povedi.
- Drugi kolokator. Eden najpomembnejših klasifikatorjev, ki točkovno nagrajuje zglede, ki vsebujejo najbolj tipične kolokatorje določene kolokacije, in s tem posredno upošteva merilo koligacijske tipičnosti. Npr. pri kolokaciji *klavrn + podoba* klasifikator dodeli dodatne točke zgledom s statistično pomembnim drugim kolokatorjem *kazati*, pri čemer se izkaže, da tako identificirani zgledi pa vsebujejo tudi tipično širšo strukturo kolokabilne okolice: *kazati klavrnno podobo česa*.
- Levenshteinova razdalja. Gre za algoritem,<sup>24</sup> ki meri podobnost med nizi, v našem primeru povedmi. Če klasifikator najde dve podobni ali celo enaki povedi, tisto z nižjim točkovanjem vrže na dno seznama zadetkov.

Večina razlik med konfiguracijami za različne besedne vrste se pojavlja v nastavitvah posameznih klasifikatorjev, čeprav najdemo tudi razlike med klasifikatorji (npr. samo konfiguracija za glagol vsebuje dodaten klasifikator položaja leme v

<sup>24</sup> [http://en.wikipedia.org/wiki/Levenshtein\\_distance](http://en.wikipedia.org/wiki/Levenshtein_distance) (dostop 8. 8. 2015).

povedi). Vsaki povedi je pripisana določena vrednost med 0 in 1, ki predstavlja seštevek vseh vrednosti klasifikatorjev (pri čemer vsakemu klasifikatorju določimo določeno težo pri končnem izračunu) in služi kot osnova orodju GDEX pri razvrščanju kandidatov za dobre zglede za vsako kolokacijo ter posledično odloča o zgledih, ki jih izvozimo z metodo ALLP.

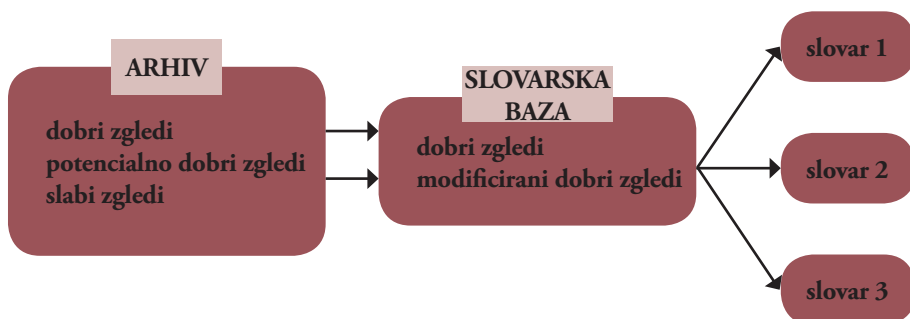
Za vsak izluščeni zglede je koristno izvoziti tudi metapodatke o besedilu, iz katerega zglede izvira, npr. leto, vir, avtor, naslov ipd. To nam zagotavlja sledljivost zglede, ponuja pa tudi možnost prikaza tovrstnih informacij v slovarju. Kot pri vseh ostalih delih slovarja se namreč nikoli ni dobro omejevati samo na potrebe konkretnega slovarja, saj je kasnejše iskanje informacij o zgledih v korpusu precej zamudnejše in dražje. Dober primer koristi metapodatkov je povezan s posodabljanjem slovarja: če namreč želimo enkrat zamenjati stare zglede s sodobnejšimi, s pomočjo podatka o letu nastanka besedila lahko takoj poiščemo vse zglede, ki so nastali pred določenim letom. Metapodatki so lahko koristni tudi pri preprečevanju vključevanja izrazito ideoloških zglede v slovarska gesla. Pri zgledih, ki jih izluščimo z orodjem GDEX, smo lahko pozorni na primere, ko je večina zglede v geslu oz. pri posameznem pomenu iz enega samega vira ali samo nekaj virov (prim. analizo zglede za *pederastija* v Gorjanc 2014).

## 6.2 Zgledi v slovarski bazi in slovarski zglede

Pri diskusiji o iskanju in beleženju zglede je treba upoštevati tudi vlogo odnosa med slovarjem in slovarsko bazo, pa tudi obeh virov do arhiviranih podatkov (Slika 1). Postopki, opisani v tem prispevku, so najbolj aktualni za izdelavo slovarja sodobnega slovenskega jezika, ker pa se bo jezikovni opis delal povsem na novo, bo marsikateri del podatkov (tudi zglede), zabeleženih pri analizi korpusa oz. korpusov, uporaben pri izdelavi drugih slovarjev. Iz enega zglede lahko izpeljemo več različic za različne uporabnike oz. slovarje; za odrasle matere govorce lahko določen dober zglede ostane nespremenjen, medtem ko ga za mlajše govorce po potrebi lahko malenkostno spremenimo, npr. ga skrajšamo ali nadomestimo redko besedišče s pogostejšim, ki ga mlajši govorci poznajo. Zaradi takšne večnamenske rabe zglede moramo izvorno izluščene zglede in vse z njimi povezane metapodatke arhivirati.

Arhiv izvornih zglede omogoča tudi analizo modifikacij zglede in njihov obseg ter posledično izboljšavo konfiguracij za njihovo luščenje. Tudi slabe oz. nerelevantne zglede, ki so sestavni del avtomatsko izluščenih podatkov in jih v bazi nočemo, je treba arhivirati, saj njihova analiza lahko razkrije nove lastnosti slabih zglede, ki jih potem integriramo v konfiguracije za luščenje zglede. Podoben postopek je bil že uporabljen pri izdelavi prve različice konfiguracije GDEX za

slovenščino (Kosem et al. 2011), kjer so se parametri klasifikatorjev testne konfiguracije, izdelane na podlagi analize obstoječih, ročno izbranih zgledov v LBS, izboljšali na podlagi analize izbranih (dobrih) in neizbranih (slabih) zgledov pri pregledovanju testnih rezultatov. Poleg tega ne smemo pozabiti na vlogo slovarskih podatkov pri izdelavi ostalih jezikovnih tehnologij za slovenščino. Skratka, že v samem načrtovanju slovarja je treba posvetiti veliko pozornosti tipom podatkov v slovarski bazi in njihovemu beleženju ter dejansko jemati slovar, tudi splošnega, kot eno izmed njenih izpeljank.



**Slika 1: Zgledi z vidika razmerja med arhivom, slovarsko bazo in slovarji.**

Ker slovarska baza vsebuje več podatkov kot sam slovar, se poraja vprašanje, koliko dodatnega časa vzamejo leksikografski ekipi, ki je osredotočena na izdelavo slovarja, vsi postopki beleženja dodatnih podatkov. Zato mora načrtovanje slovarske baze slediti dvema načeloma: avtomatizirati čim več (rutinskih) leksikografskih postopkov in poskrbeti, da nobena leksikografska odločitev ne ostane neizkoriščena. S tega vidika je uporaba metod, kot je ALLP, pravzaprav nujna, saj si drugače ne moremo predstavljati izdelave baze in slovarja v roku, ki bi zadovoljil tako uporabnike kot financerje, kar ugotavljajo tudi pri podobnih leksikografskih projektih v tujini (npr. nizozemski slovar ANW). Za primer lahko vzamemo povsem osnovno nalogo: vtipkavanje iztočnice in njene besedne vrste v geselski članek v slovarski bazi. Recimo, da za vpis teh dveh podatkov porabimo povprečno 5 sekund. Pri 100.000 iztočnicah to znese 500.000 sekund oz. malo manj kot 139 ur. Pri ALLP se ta dva podatka pripišeta avtomatično – torej prihranimo skoraj en človek-mesec. Precej podobno velja za leksikografske odločitve. Pri ročni analizi oz. analizi v korpusnem orodju, tudi z uporabo orodja GDEX, leksikograf pregleda veliko zgledov in pri vsakem sprejme načelno odločitev, ali je dober ali ne. Ker pa leksikograf iz korpusnega orodja v bazo prenese samo dobre oz. kandidate za dobre zglede, so arhivirane samo tovrstne odločitve. Pri metodi ALLP lahko v bazi beležimo čisto vsako odločitev: identifikacijo dobrega zglada (zglede ostane v bazi v nespremenjeni obliki), potencialno dobrega zglada (zglede je modificiran) in slabega zglada (brisanje zglada).

V postopek izbire zgledov lahko vpeljemo tudi množičenje in tako še dodatno razbremenimo leksikografsko ekipo. A ker mora dober slovarski zgled vsebovati kombinacijo različnih lastnosti, ki se jih mora leksikograf naučiti prepoznati, si je težko predstavljati, da bi takšno delo lahko prepustili neleksikografom oz. procesu množičenja. Da to postane izvedljivo, se moramo zavedati značilnosti in omejitev množičenja (gl. Čibej et al. 2015a in Fišer in Čibej 2015). Kot prvo, naloge morajo biti enostavne, množičarji izbirajo samo med odgovori *Da*, *Ne* in *Ne vem*. Poleg tega naloge ne smejo temeljiti na določanju težko določljivih elementov (kot so npr. lastnosti dobrega zgleda) ali stopenjskosti; vprašanja tipa *Ali je to dober slovarski zgled?* in *Kako dober je ta zgled?* torej ne pridejo v poštev. Kot so pokazala testiranja množičenja na zgledih v LBS (Kossem et al. 2013), je najbolj smiselno uporabiti zglede pri odkrivanju napačnih podatkov (npr. ko raba iztočnice in kolokatorja v zgledu ne ustreza skladijski strukturi, v kateri naj bi se kolokacija pojavljala) ali razporejanju kolokacij in z njimi povezanih zgledov pod relevantne pomene in podpomene.

### 6.3 Vizualizacija zgledov

Leksikografsko delo se z vidika zgledov ne konča z zabeleženjem dobrih zgledov, saj ima svojo vlogo pri uspešnosti izpolnjevanja namena zgledov tudi način predstavitve zgledov slovarskemu uporabniku. Zato je ključno sodelovanje med leksikografi in oblikovalci oz. ekipo, ki skrbi za vizualizacijo slovarskih podatkov. Še tako dober zgled je precej odvisen od ustrezne vizualizacije, kot je na primer izbira ustrezne pisave, ki vpliva na berljivost besedila in tudi na to, kako dobro si uporabnik zapomni prebrano informacijo (Nesi 2011). Če upoštevamo, da zgledi predstavljajo precej velik, če ne celo največji, odstotek besedila v slovarju, je izbira načina vizualizacije še kako pomembna.

Uporabniku lahko pri branju zgledov pomagamo tudi z označevanjem iztočnice, njenih stalnih zvez ali podobnih delov gesla. Sploh v sodobni leksikografiji, ko se prikazujejo celostavčni, daljši zgledi, se zdi smiselno uporabnika opozoriti na iztočnico in s tem na del zgleda, ki naj bi mu posvetil večjo pozornost. Največkrat se za tako označevanje uporabi krepki tisk, v elektronskih slovarjih najdemo tudi rabo drugačne barve (Slika 2), precej redkeje pa ležeči tisk, kar je tudi posledica dejstva, da je ležeči tisk oblika pisave, ki jo za prikazovanje zgledov uporablja večina slovarjev. Slednja možnost se zdi tudi manj učinkovita (Slika 3). Druga vloga krepkega tiska oz. uporabe pisave, ki se razlikuje od pisave ostalega dela zgleda, je izpostavitve določenih zelo tipičnih kolokacij, stalnih zvez in fraz (Slika 4). Odločitev o uporabi oz. obsegu in načinu označevanja iztočnice in ostalih delov zgleda mora biti v vsakem primeru podprta z uporabniško študijo.



## fach

*Prawie 60 lat temu zaczął się uczyć fryzjerskiego **fachu** i nadal pracuje w zawodzie.*

źródło: NKJP: Katarzyna Skrzypek: Cyrkiel za uszami, Dziennik Zachodni, 2005-03-31

*Miał dobry **fach** - przez kilkanaście lat pracował w dużej warszawskiej fabryce jako spawacz, był też ślusarzem, szlifierzem i monterem.*

źródło: NKJP: Monika Mikołajczuk: Między Kantem a Wolterem, Polityka, 2001-07-14

### Slika 2: Rdeče obarvana iztočnica v zglelih v Velikem slovarju poljskega jezika.

#### Examples of CLICK

He *clicked* his heels together and saluted the officer.

Her heels *clicked* on the marble floor.

Press the door until you hear the latch *click*.

To open the program, point at the icon and *click* the left mouse button.

*Click* here to check spelling in the document.

I know him fairly well, but we've never really *clicked*.

### Slika 3: Iztočnica v ležčem tisku v zglelih v slovarju Merriam-Webster.

**3 SURE ABOUT SOMETHING** feeling certain that you know or understand something [↔ clearly]

**clear about/on**

☞ *Are you all clear now about what you have to do?*

**clear whether/what/how etc**

☞ *I'm still not really clear how this machine works.*

☞ *Let me **get this clear** - you hadn't seen her in three days?*

☞ *a clearer understanding of the issues*

**4 THINKING** able to think sensibly and quickly [↔ clarity, clearly]:

☞ *She felt that her thinking was clearer now.*

☞ *In the morning, with a **clear head**, she'd tackle the problem.*

### Slika 4: Izpostavljene kolokacije in fraze v zglelih v slovarju Longman.<sup>25</sup>

<sup>25</sup> <http://www.ldoceonline.com/> (dostop 8. 8. 2015).

Omenili smo že, da je o vsakem zgledu v slovarski bazi dobro imeti čim več metapodatkov, in tudi takšne podatke lahko ponudimo uporabniku. Prikazovanje tovrstnih podatkov, zlasti v splošnih enojezičnih slovarjih, je sicer redkost – izjema je npr. Veliki slovar poljskega jezika (Slika 2) –, iz preprostega razloga: metapodatki, kot so vir, avtor in naslov, imajo referenčno vlogo in sugerirajo, da je bil zgled vzet neposredno iz določenega vira, kar leksikografom odvzame možnost vnašanja kakršnih koli sprememb. Drugi razlog proti prikazovanju metapodatkov je, da predstavljajo za uporabnike dokaj nenujne podatke, ki jemljejo dragoceni prostor na zaslonu in največkrat odvrtačajo njihovo pozornost od osrednje vloge zgleda, namreč prikazovanja rabe iztočnice v določenem (pod)pomenu.

Načelo informativnosti slovarskih zgledov med drugim leksikografom narekuje omejevanje glede števila ponujenih zgledov na geslo ali posamezen pomen v njem. Pa vendar se hitro zgodi, da imamo večje število zgledov na (pod)pomen, skladiščno strukturo ali kolokacijo, kar lahko postane vizualizacijska težava zlasti pri pomensko zelo razčlenjenih geslih. Ena od dobrih rešitev je privzeti prikaz samo določenega števila zgledov, medtem ko se dodatni zgledi uporabniku razkrijejo po kliku na gumb (Slika 5 in Slika 6). Vse več spletnih slovarjev ponuja tudi neposredno povezavo na korpusne zadetke, kar je za zahtevnejše uporabnike vsekakor dobrodošla funkcija, saj si lahko ogledajo številne primere realne rabe iztočnice, njene kolokacije, stalne zveze ipd.

1.1 Pursuing a commercial activity on a significant scale:  
*'many large investors are likely to take a different view'*

MORE EXAMPLE SENTENCES

**Slika 5: Povezava za prikaz dodatnih zgledov (More example sentences) v slovarju Oxford.**

1.1 Pursuing a commercial activity on a significant scale:  
*'many large investors are likely to take a different view'*

MORE EXAMPLE SENTENCES

*'The basic cause of the changed activities of large businesses is a matter of debate.'*

*'The fate of rival bids for NatWest rest in the hands of the faceless large investors.'*

*'Being a large economy, the euro zone is much less open than individual member states.'*

GET MORE EXAMPLES

**Slika 6: Prikazani dodatni zgledi v slovarju Oxford.**

## 7 ZAKLJUČEK

Zaradi svoje nepogrešljivosti v slovarskem geslu na eni strani in zahtevnega načina pridobivanja na drugi bomo zgledom pri snovanju slovarja sodobnega slovenskega jezika posvetili veliko pozornosti. V navodilih leksikografom bomo jasno opredelili kriterije dobrih zgledov skupaj s konkretnimi primeri dobre in slabe prakse, pa tudi z vidika ideološkosti, ter izdelali oz. uporabili orodja, ki zagotavljajo čim večjo doslednost pri upoštevanju teh kriterijev. Hkrati bomo pripravili primere dovoljenih modifikacij in posledično tudi ustrezen sistem arhiviranja izvorno izluščenih zgledov. Ker naj bi SSSJ predstavljal pomemben vir tudi jezikovnotehnološki skupnosti, je nujno zagotoviti čim večje število zgledov in jih opremiti s čim več metapodatki.

Ravno zaradi teženj po vključevanju čim večjega števila zgledov v slovar oz. slovarsko bazo je nujna uporaba polavtomatskih ali avtomatskih metod pri pridobivanju zgledov. V nasprotnem primeru se namreč izdelava slovarja lahko zavleče do takšne mere, da so zgledi že pred dokončanjem slovarja potrebni zamenjave. Iz takšnega razmišljanja izhaja metoda ALLP, ki jo predlagamo za iskanje in beleženje zgledov v slovarju sodobnega slovenskega jezika ter predstavlja nov način leksikografske analize in iskanja zgledov ne samo v slovenskem, ampak tudi v mednarodnem prostoru. Na podlagi slovenskih izkušenj, pridobljenih pri izdelavi LBS, so podobno metodo že začeli uporabljati v Estoniji, in sicer pri izdelavi kolokacijskega slovarja za nematerne govorce estonščine (Kallas et al. 2015).

Pomembna naloga leksikografske skupnosti v zvezi z zgledi pa je tudi izvajanje (nadaljnjih) raziskav o tem, kako, kdaj in na kakšen način slovarski uporabniki uporabljajo zglede ter kakšni zgledi uporabnikom najbolj koristijo. Na podlagi izsledkov bo mogoče v prihodnje še izboljšati postopke izbiranja zgledov, kot tudi način in količino njihovega podajanja v slovarjih.