
LEKSIKALNA BAZA: VSE, KAR STE VEDNO ŽELELI VEDETI O JEZIKU

V prispevku predstavimo izdelavo leksikalne podatkovne baze za slovenščino, ki poteka v okviru projekta *Sporazumevanje v slovenskem jeziku*. Pojem najprej osvetlimo s teoretičnega vidika in ga umestimo v kontekst konkretnih rezultatov po posameznih evropskih jezikih. Dvojni namen leksikalne podatkovne baze za slovenščino, tj. za slovarske aplikacije in računalniško obdelavo naravnega jezika, določa opis leksikalnih enot s treh temeljnih vidikov: pomenskega, skladijskega in kolokacijskega. Na podlagi teh izhodišč predstavimo zgradbo leksikalne baze po vsebinskih nivojih in pojasnimo teoretične razmisleke o vsebinskih rešitvah. Na koncu na kratko predstavimo orodje za izdelavo leksikalno-gramatičnega profila besed *Word Sketch* in opišemo programski vmesnik za izdelavo leksikalne baze.

Ključne besede: leksikalna podatkovna baza, leksikalno-gramatični pristop, leksikografija, leksikologija, besedni pomen, računalniška obdelava naravnega jezika

1 Uvod

1.1 Kaj je leksikalna podatkovna baza

Na splošno je pojem *baze* mogoče razumeti kot osnovo, temelj, iz katerega in na katerem se gradijo veliko večje stvari, kot je baza sama: bolj ko je baza trdna, stabilna, bolj zanesljive so stvari, ki na njej nastajajo. Hkrati je baza zbir vsega pomembnega, osnovnega oziroma organizacija teh stvari na enem mestu.

Splošnima vidikoma je potrebno dodati še jezikoslovni pogled, po katerem je leksikalna podatkovna baza (LPB) organiziran opis besedišča določenega jezika

v elektronski obliki¹ ali na kratko: zbirka podatkov o besedah.² Pri tem se s pojmom besedišče misli tako na zbir leksikalnih enot (LE; leksemov) konkretnega jezika, kot tudi na vse tisto védenje, ki ga obsega besedišče materne govorce, da se lahko sporazumeva v svojem jeziku. Nekoliko bolj praktičen pogled, kot ga zastopa Patrick Hanks,³ opredeljuje LPB kot skup (korpusnih) dejstev, ki predstavljajo vsako besedo v jeziku, pri čemer je poudarek na sintagmاتيki oziroma kolokabilnosti pa tudi na morfologiji in pomenu. Taka LPB velja kot primarni vir za številne aplikacije, zlasti za izdelavo (elektronskih ali spletnih)⁴ slovarjev, jezikovnih učbenikov, popravljanje napak in učenje jezika, računalniško obdelavo naravnega jezika (RONJ) in kodifikacijo besednih pomenov.

LPB imajo – ob skupnem teoretičnem imenovalcu – v praksi več različic glede na namen izrabe. To, kar jim je skupnega, lahko strnemo v naslednje točke:

- predstavljajo inventar vseh v jeziku znanih LE,⁵ jih pomensko razčlenjujejo in pomene pojasnjujejo z zgledi rabe;
- zlasti t. i. ontološke LPB, ki so primarno namenjene RONJ, vzpostavljajo pomenske povezave med LE, kot so: so-, nad- in podpomenskost;
- predvidevajo kategorizacijo LE (zlasti slovnično, glede na področje rabe, besedilne vrste ipd.);
- predvidevajo fonološke, morfološke in skladenjske informacije;
- upoštevajo dejstvo, da materni govorce uporabljajo slovarje (kot rezultate LPB) za preverjanje/konzultiranje glede ustrezne rabe in izbire besed.⁶
- navadno vsebujejo več informacij, kot jih dejansko vključujejo iz njih izhajajoči produkti.

Bistvo jezikovnega opisa v LPB, ki temeljijo na sodobnih leksikalnogramatičnih pristopih in korpusnih podatkih, je v tem, da pomenski opis LE ni strogo ločen od skladenjskega opisa oziroma povedano drugače: pomenske lastnosti LE so neposredno povezane s slovničnimi vzorci, v katerih se pojavljajo. Koncept besedišča, kjer slovnična pravila niso ločena od LE, zahteva torej kombinacijo in integracijo fonoloških, morfoloških, kolokacijskih, skladenjskih, pomenskih in pragmatičnih informacij na različne načine.

¹ *GlossaryOfLinguisticTerms*: <<http://www.sil.org/linguistics/GlossaryOfLinguisticTerms/Index.html>>.

² <<http://www.thefreedictionary.com/lexical+database>>.

³ Patrick Hanks: *Compiling a Monolingual Dictionary for Native Speakers*. Predavanje na posvetu *Slovarji, več kot le besede*. Članek je v slovenskem prevodu objavljen v pričujoči številki *Jezika in slovnstva*.

⁴ Z e-slovarji mislim predvsem na slovarje, izdane na CD-jih, medtem ko so spletni slovarji dostopni prek interneta.

⁵ Obseg LPB je tesno povezan z njeno izrabo, številom sodelavcev, časom izdelave, finančnimi sredstvi in dejstvom, da lahko LPB zajame samo presek besedišča določenega jezika v konkretnem času, saj se besedišče pod vplivom jezikovne in družbene (politične, kulturne, gospodarske itd.) situacije nenehno spreminja.

⁶ Tega segmenta se LPB v praksi lotevajo na različne načine, med katerimi prevladujeta opis (deskripcija), kjer je vključevanje tipičnih segmentov besedila in zgledov dejanske rabe osnovni način podajanja informacij tega tipa, in predpis (preskripcija), ki temelji na stilističnih in jezikoslovnih interpretacijah jezikovne danosti v smislu jezikovne pravilnosti in nepravilnosti.

1.2 Leksikalne podatkovne baze v evropskem okviru

Pojem leksikalne baze je v evropskem kontekstu znan že vsaj dvajset let, dejavnosti v zvezi s tem pa so se začele z evropskimi jezikovnotehnološkimi projekti, kot sta npr. *Genelex I* in *Genelex II* (*GENERIC LEXICON*, 1990–1994),⁸ Od tod je bil način strukturiranja informacij po posameznih ravneh (morfološki, skladenjski in semantični) prenesen v druge evropske projekte, zlasti PAROLE in SIMPLE (izdelava korpusov in leksikonov za 12 jezikov EU) ter v iz njih izvirajoče nasledke po posameznih jezikih.⁹

LPB, ki so nastale v okviru evropskih projektov za posamezne jezike, so namenjene predvsem RONJ ter izdelavi spletnih slovarjev in pripomočkov za učenje jezika. Na drugi strani specifično stanje, v katerem se je znašla slovenščina z zastarelimi, pomanjkljivimi ali neobstoječimi učbeniki in jezikovnimi priručniki, zahteva usmeritev aktivnosti v izdelavo LPB, s katero bo mogoče v čim krajšem času priti do konkretnih za jezikovno skupnost najbolj strateško pomembnih spletnih aplikacij in jezikovnih priručnikov, kot so slovar sodobnega slovenskega jezika, dvojezični slovarji, posamezna poglavja slovnice, specializirani slovarji, npr. šolski, frazeološki, kolokacijski, za učenje slovenščine kot tujega jezika ipd. Pri tem je tako s finančnega vidika kot z vidika sledenja sodobnemu stanju na področju posredovanja jezikovnih podatkov jezikovnim uporabnikom pomembno dejstvo, da so leksikalni podatki strukturirani v elektronski obliki, kar omogoča nenehno dopolnjevanje,¹⁰ nadgradnjo in večkratno izrabo.¹¹

2 Leksikalna podatkovna baza za slovenščino

2.1 Obseg in namen

Oblikovanje leksikalne podatkovne baze za slovenščino (LBS) v okviru projekta *Sporazumevanje v slovenskem jeziku* (SSJ)¹² sledi dvema temeljnima ciljema:

⁷ Teorijo generičnega leksikona je konec osemdesetih let razvil James Pustejovsky s sodelavci (Pustejovsky 1995), njen namen pa je ponuditi potrebno infrastrukturo za v celoti generativno (tj. računalniško obvladljivo), podrobno razčlenjeno in leksikalno orientirano semantično teorijo jezika.

⁸ Podrobneje o tem v člankih Simona Kreka in Špele Arhar v pričujoči številki *Jezika in slovstva*.

⁹ Pri projektu *Sporazumevanje v slovenskem jeziku* smo za namene izdelave LBS pregledali vrsto sorodnih projektov, in sicer: *elexico* za nemščino, CLIPS za italijanščino, CORNETTO za nizozemščino, DAFLES in ALFALEX za francoščino, STO za danščino, ADESSE in GRIAL za španščino, CEGLEX za poljščino, SPRÅKBANKEN za švedščino, PRALED za češčino itd.

¹⁰ Sodobni leksikografiji se, zahvaljujoč spletnim aplikacijam, katerih podlaga so elektronske leksikalne podatkovne baze, ni treba več ukvarjati z zastarelostjo leksikalnih podatkov, ki lahko dosežejo uporabnika tudi več kot deset let kasneje, kot so bili dejansko ustvarjeni, saj omogočajo sprotno objavlanje oziroma dopolnjevanje, ki sledi realnemu jezikovnemu stanju.

¹¹ Pri tem je pomembno poudariti, da so vsi »produkti«, ki bodo nastali v projektu SSJ, javnosti dostopni pod licenco *Creative Commons*, kar omogoča dostop do jezikovnih informacij celotni jezikovni skupnosti na brezplačen in transparenten način.

¹² Projekt delno financira Evropska unija iz Evropskega socialnega sklada ter Ministrstvo za šolstvo in šport. Vse o projektu je mogoče najti na spletni strani: <<http://www.slovenscina.eu/Vsebine/SI/Domov/Domov.aspx>>.

- zadostiti leksikografskim potrebam, konkretno izdelavi za slovenščino najpomembnejših eno- in dvojezičnih slovarjev in specializiranih jezikovnih priročnikov¹³ ter
- zadostiti potrebam RONJ oziroma uporabi v jezikovnotehnoloških aplikacijah za slovenščino.

Na podlagi teh izhodišč bo LBS vsebovala podatke o frekvenčnih, slovničnih, pomenskih, skladijskih, kolokacijskih, frazeoloških in drugih lastnostih predmetnopomenskih LE jedrnega besedišča sodobne slovenščine, ne bo pa vsebovala slovničnih (veznik in predlog) ter funkcijskih (členek in medmet) besednih vrst. Omenjene besedne vrste so vključene v LBS na skladijskem in kolokacijskem nivoju, zato bo mogoče pridobiti tudi informacije o pomenskih in skladijskih lastnostih predlogov in veznikov ter njihovo vlogo v ustaljenih tipičnih segmentih jezika. LBS tudi ne bo vsebovala pomenskih definicij, ker je odločitev za tip definicije odvisna od namena, obsega in ciljnega uporabnika vsakega konkretnega slovarja. Pomenska členitev LE bo prikazana s pomenskimi indikatorji in pomenskimi shemami, ki jih navajamo pri glagolskih (poskusno tudi pri samostalniških in pridevniških) iztočnicah. Pomenske sheme predstavljajo izhodišče za oblikovanje razlag komunikacijskega tipa (Sinclair in Coulthard 1975), realiziranih predvsem v Collins Cobuildovih EFL-slovarjih.¹⁴

Dvojna narava LBS, ki je povezana z namenom njene izrabe v leksikografske in jezikovnotehnološke namene, določa zgradbo celotne LBS po vsebinskih nivojih. Posamezni nivoji so določeni z xml-zapisom DTD-ja (*Document Type Definition*)¹⁵ in vključeni v programski vmesnik DPS (*Dictionary Production System*).

2.2 Organizacija leksikalnih podatkov po nivojih

Vsako LE (iztočnico/lemo) analiziramo s treh osnovnih vidikov: pomenskega, skladijskega in kolokacijskega. Na posameznem nivoju je za vsako LE oziroma njen pomen predvidenih šest različnih vrst jezikovnih podatkov (gl. **Tabelo 1**). Na pomenskem nivoju pripišemo LE pomen(e) oziroma podpomen(e), ki jih opredelimo s pomenskimi indikatorji. Drugi del pomenske informacije predstavlja pomenska shema, ki se teoretično približuje pomenskimi okvirjem, kot jih predvideva *FrameNet*.¹⁶ Skladijski nivo je namenjen formaliziranju skladijskih struktur, v katerih se realizira posamezni pomen ali podpomen obravnavane LE, in je primarno namenjen računalniškemu branju jezikovnih podatkov. Leksikografsko verzijo skladijskih struktur zapisujemo s slovničnimi vzorci. V posebnem razdelku skladijskega nivoja beležimo skladijske zveze,

¹³ Ponovno je treba poudariti, da so v prvi vrsti mišljene elektronske verzije in interaktivne spletne aplikacije, ki so za sodobnega uporabnika veliko bolj praktične. Tiskane verzije so drugotna možnost, ki so glede na hranjenje podatkov v elektronski obliki neproblematične.

¹⁴ EFL/ESL slovarji (*English as Foreign/Second Language*) so namenjeni učenju angleščine kot tujega/dr drugega jezika.

¹⁵ DTD je elektronsko berljiv zapis zgradbe gesla v LBS.

¹⁶ <<http://framenet.icsi.berkeley.edu/>>.

ki predstavljajo bodisi kolokacijsko razširjene segmente jezika bodisi strukturno omejene zveze, ki predstavljajo tipične delčke jezika, vendar niso leksikalizirane do te mere, da bi potrebovale samostojno razlago. Četrti nivo je kolokacijski. Tu navajamo tipične kolokacije, v katere vstopajo obravnavane LE glede na zabeležene pomene in skladijske strukture. Frazeološke enote (FE), v katerih nastopa obravnavana LE, navajamo v samostojnem razdelku. Vsaki FE pripišemo pomen(e) v obliki pomenskega indikatorja, hkrati evidentiramo njeno osnovno in realizacijske oblike ter navedemo korpusne zglede.

NIVO	PODATKI	PRIMER
I. Leksikalna enota	iztočnica besedna vrsta glagol	stisniti
II. Pomenski nivo	pomenski indikator pomenska shema	(1) <i>oprijeti z dlanjo</i> ČLOVEK stisne PREDMET z dlanjo
III. Skladijski nivo	skladijska (mega)struktura slovnični vzorec (1) slovnični vzorec (2)	sbz1 GBZ sbz4 (za sbz4) kdo stisne kaj kdo stisne koga za kaj
IV. Kolokacijski nivo	(1) (2)	stisniti [roko, vrat] stisniti za [ramena, vrat]
V. Zgled	zgled (1) zgled (2)	<i>Ko je uro pozneje Michaels z landroverjem vozil nazaj h koči, je ob menjalniku otipal ženino roko in jo močno stisnil.</i> <i>Stisnil jo je za vrat in jo začel daviti ter jo skušal posiliti, vendar mu zaradi motenj s potenco to ni uspelo.</i>
VI. Frazeologija	frazeološka enota pomenski indikator zgled	stisne pri srcu koga <i>občutek tesnobe; skrb</i> <i>Doma so bila vrata odklenjena.</i> <i>Stisnilo me je prisrcu.</i>

Tabela 1: Razporeditev podatkov v LBS po vsebinskih nivojih (primer glagola stisniti).

2.3 Vsebinski premisleki in teoretična podlaga

LBS predstavlja skladišče podatkov o LE, njihova vsebina pa izhaja iz prepričanja o neločljivi povezanosti pomenskih in skladijskih lastnosti besed. Hkrati sledi LBS zadovoljevanju potreb jezikovne skupnosti in njenega življenjsko relevantnega interesa na področju jezika. Zadnje določa vsebino LBS v tem, da

mora vsebovati spekter podatkov, ki bodo v najširšem smislu večkrat uporabni v jezikovnopriročniških in računalniških aplikacijah različnih vrst. Zato je pomembno na eni strani upoštevati sodobno stanje (v teoriji in praksi) na področju leksikologije, leksikografije in jezikovnih tehnologij, na drugi strani pa izhajati iz obstoječega stanja in potreb glede slovarskopriročniške infrastrukture za slovenščino. V nadaljevanju pojasnimo nekaj temeljnih teoretičnih premislekov, ki smo jih upoštevali pri snovanju LBS, in predstavimo modele, ki so nam služili pri oblikovanju vsebine podatkov, vključenih v LBS.

2.3.1 Pomen ali pomenske tendence?

Poudarili smo, da je primarni namen izrabe LBS izdelava slovarjev in večpredstavnih jezikovnih priročnikov. To, s čimer se slovarji ukvarjajo, pa je pojasnjevanje besed v njihovih različnih pomenih. Ob tem ne moremo mimo na današnji stopnji leksikografske teorije in prakse že uveljavljenega spoznanja (prim. Kilgariff 2006a: 29; Atkins in Rundell 2008: 264), da namreč ni jasno, kaj besedni pomeni dejansko so in celo, če kaj takega, kot je besedni pomen, v resnici sploh obstaja.¹⁷ Bistvo te na prvi pogled kontroverzne ugotovitve je, da večina pomenske informacije konkretne LE izhaja iz njenega bodisi neposrednega bodisi širšega besedilnega okolja. Logična posledica tega je, da vsega, kar se nahaja vsakokrat v sobesedilu vsake konkretne LE, ni mogoče zajeti v njen pomenski opis, kot ga omogoča slovar, in da sta zato postopek pomenskega dekodiranja v vsakdanji jezikovni komunikaciji in postopek razdvoumljanja pomenov v slovarjih med seboj neprimerljiva (Atkins in Rundell 2008: 263). Osnovno izhodišče, ki ga upoštevamo tudi pri pomenskem členjenju LE v LBS torej je, da obstaja bistvena razlika med slovarskimi pomeni (tj. oštevilčenimi pomeni, na katere so navadno razdeljene iztočnice v slovarjih)¹⁸ in pomeni, kot jih dojemajo (slovarski) uporabniki v vsakdanji komunikaciji.

Vendar pa, tako kot je na eni strani jasno, da ne obstaja splošno sprejet in najboljši način pomenskega členjenja večpomenskih LE, obstaja na drugi strani tudi težnja, da se pri organizaciji leksikalnih podatkov uporabljajo notni principi pomenskega členjenja, s čimer se ustvarja potrebna notranja konsistenca. Kot pravita Atkinsova in Rundell (2008: 275), je največ, kar lahko leksikograf doseže pri pomenskem členjenju LE, to, da uporabnik pri pregledovanju večpomenske slovarske enote

¹⁷ Če citiramo van der Meera (2006: 604): »Kljub stoletjem praktičnega ukvarjanja z leksikografijo, še vedno ne obstaja oprijemljiv konsenz glede delitve semantičnega prostora leksikalne enote« in nenazadnje tudi slovenskega jezikoslovca in leksikografa Boža Voduška, ki že v šestdesetih letih 20. stoletja na podlagi takratne evropske leksikografske teorije in prakse ugotavlja, da »/u/gotavljanje in urejevanje besednih pomenov velja na splošno le za nekakšno prakso, nekakšno zgolj empirično delavnost, za katero se ne dajo postavljati nobena sistematična pravila /.../« (Vodušek 1961: 5.)

¹⁸ V slovenskem leksikalnem pomenoslovju (Vidovič Muha 2000: 29 in dalje) je natančneje opredeljen predvsem *slovarski* ali *slovnični* pomen, »ki je določen s kategorialnimi pomenskimi sestavinami, kot so spol, vid idr.« in vključuje obvezni denotativni in neobvezni konotativni pomen.

nima občutka, da gre za nesmiseln in nerazumen način predstavljanja dejstev. Prav to je tudi eno od osnovnih vodil pri pomenskem členjenju besedišča v LBS.

Da podatkovna baza ni že sam slovar, ampak strukturiran zbir podatkov, ki se oblikujejo v analitičnem procesu iz surovega jezikovnega materiala, potrjuje dejstvo, da posamezni slovarji različno členijo besede na pomene. V kako podrobno pomensko členitev se bo spustil posamezni slovar, je odvisno od njegovega namena, obsega, ciljnega uporabnika, specializiranosti itd. Ena izmed pomembnih lastnosti, ki jo izpostavlja sodobna leksikografska literatura v zvezi s tem pa je, da je v fazi oblikovanja podatkovne baze smiselno pomene razdeliti čim bolj podrobno (Atkins in Rundell 2008: 268), česar se držimo tudi pri oblikovanju LBS. S podrobno razdeljenimi pomeni v podatkovni bazi je namreč leksikografom na voljo celoten spekter možnosti, iz katerih lahko izpeljejo več različnih slovarjev. Hkrati upoštevamo tudi »nasprotno« vodilo: razbijanje pomenov mora slediti še obvladljivim in dovolj homogenim enotam, čemur denimo Čermák (2008)¹⁹ pravi »zadostna analogija«.

2.3.1.1 Slovarski vzori

Da bi dosegli ustrezno stopnjo konsistentnosti, smo se pri pomenskem členjenju LE in pri oblikovanju pomenskih indikatorjev naslonili na teoretska in praktična leksikografska spoznanja, ki so se izoblikovala v zvezi z naslednjimi slovarji:²⁰ *New Oxford Dictionary of English* (NODE 1998), *Longman Dictionary of Contemporary English* (LDOCE 2003) in *Macmillan English Dictionary for Advanced Learners* (MEDAL 2007²). Vsi trije so enojezični angleški slovarji, ki se osredotočajo na standardno angleščino ter na to, kar je v jeziku osrednje, pri čemer je NODE namenjen angleškemu govorcem (po obsegu večji), ostala dva slovarja pa sta namenjena tujim govorcem za učenje angleščine in sta po obsegu relativno primerljiva (100.000 gesel). Pri vsakem od slovarjev smo za izhodišče pomenskega členjenja upoštevali lastnosti, ki najbolj ustrezajo naši predstavi o podrobni pomenski členitvi besed, ter lastnosti, ki bi po našem mnenju ponudile čim bolj zanesljivo in glede na sodobni jezik realno informacijo.

2.3.1.1.1 *New Oxford Dictionary of English*

NODE se osredotoča na to, kar je v dejanski jezikovni rabi osrednje in tipično, s čimer sledi Hanksovi teoriji jezikovnih standardov in njihovih priložnostnih izrab oziroma razširitev (Hanks 1994). Rezultat tega je manjše število jedrnih

¹⁹ V prispevku v pričujoči številki *Jezika in slovstva*.

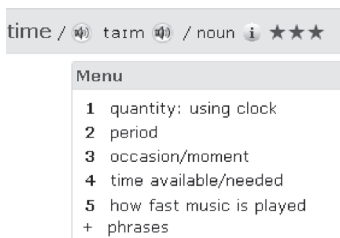
²⁰ Potrebno je poudariti, da je naslonitev na angleško slovarsko teorijo in prakso vezana v tem segmentu LBS zgolj na oblikovanje pomenskih indikatorjev in ne na oblikovanje podatkovne baze v celoti. Vrsta informacij, ki jih vključujemo v LBS, in njihova organizacija po nivojih je produkt in kombinacija izkušenj in teorije različnih leksikalnih baz za evropske jezike in specifičnih potreb slovenske leksikalne baze.

pomenov,²¹ ki so jasno in na kratko razloženi. Pomenska členitev besed temelji na naslednjih izhodiščih:

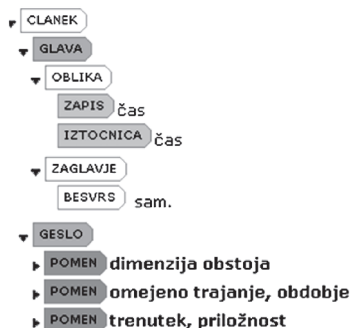
- za vsako besedno vrsto je prva definicija namenjena jedrnemu pomenu;
- vsaka iztočnica ima navadno vsaj en jedrni pomen, na katerega je vezanih več podpomenov, ki so z njim logično povezani;
- jedrni pomen predstavlja tipično, osrednjo rabo besede (kot jo izkazuje BNC),²² kar pomeni, da ga materni govorci prepoznajajo kot najbolj uveljavljenega in osrednjega;
- hkrati je to najbolj »dobesedni« pomen (ne pa nujno tudi najstarejši oziroma izhodiščni);
- jedrni pomen ni nujno najfrekventnejši; najfrekventnejši so lahko tudi figurativni nasledki jedrnega pomena;
- vsak podpomen je jasno izpostavljen;
- logično razmerje med podpomenom in jedrnim pomenom uporabniku omogoča pregled nad iztočnico ter vpogled v strukturiranost in medsebojno povezanost pomenov.

2.3.1.1.2 Macmillan English Dictionary for Advanced Learners

Pri MEDAL smo se zgledovali po principu, po katerem je pri večpomenskih besedah (pet ali več pomenov) na začetku izpostavljen t. i. »meni pomenov«, ki predstavlja povzetek jedrnih pomenov (brez podpomenov), ki so podrobneje predstavljeni v nadaljevanju gesla, in sicer tako, da so v meni pritegnjeni najpomembnejši deli razlage oziroma t. i. mnemonične besede, ki naj bi uporabniku pomagale najti pravi slovarski pomen, ne da bi ga že definirale. Včasih so to kolokatorji, neposredne nadpomenke ali pa oznake področja, na katerem se pomen uporablja (npr. pri nogometu). Tak meni pomenov predstavlja izhodišče za oblikovanje indikatorjev, kot si jih zamišljamo v LBS.



Slika 1: Meni pomenov večpomenskega samostalnika time (čas) v MEDAL.



Slika 2: Jedrni pomeni za samostalnik čas v LBS.

²¹ Po Eleanor Rosch (cit. v Atkins in Rundell 2008: 277 in 279) imenovan tudi *prototipski pomen*, ki pojasnjuje vse tipične in pogoste rabe neke besede: ne opredeljuje vseh možnih (so)besedilnih rab, a je hkrati dovolj širok, da ustrezno upošteva variacije na ravni individualnih jezikovnih situacij.

²² NODE v celoti temelji na BNC (*British National Corpus*): <<http://www.natcorp.ox.ac.uk/>>.

2.3.1.1.3 Longman Dictionary of Contemporary English

LDOCE je namenjen zahtevnejšim učencem angleščine kot tujega jezika. Organizacija pomenov znotraj večpomenskih besed je za nas zanimiva zaradi pomenskih kazalcev (ang. *signpost*), ki so znotraj iztočnice posebej izpostavljeni. Njihova naloga je nakazati pomensko območje posameznega pomena in vzpostaviti jasne pomenske razlike med posameznimi pomeni.

LDOCE: <i>memory</i>	LBS: <i>spomin</i>
1 ability to remember	1 pomnjenje česa iz preteklosti
2 something you remember	2 sposobnost pomnjenja
3 computer	3 nav. množina avtobiografsko literarno delo
	4 zmogljivost računalnika

Zgled: Pomenski indikatorji za jedrne pomene samostalnika *memory* (*spomin*) v LDOCE in LBS.

2.3.1.2 Pomenski indikatorji

Pomenska členitev LE je v LBS prikazana s pomenskimi indikatorji in pomenskimi shemami. Pomenski indikator na kratko določi pomenski obseg ali področje obravnavanega pomena, pri večpomenskih besedah pa vzpostavlja razliko glede na druge pomene.

S kratkostjo in jedrnatostjo se želimo izogniti oblikovanju slovarskih razlag, hkrati pa doseči hitro prepoznavnost pomenskega dosega posamezne LE oziroma njenega pomena. Uporaba sinonima je ustrezna takrat, kadar z njim ne zajamemo področja preširoko in kadar lahko na ta način vzpostavimo ustrezno pomensko razmerje do drugih pomenov in med podpomeni. Z razumljivostjo želimo doseči jasno in preprosto predstavo o pomenskem dosegu LE, zato se izogibamo besed, ki same potrebujejo razlago ali so redko uporabljane v vsakdanjem govoru. Na ta način bo mogoče oblikovati t. i. definicijski ali kontrolni slovar, kot ga poznajo Longmanovi ali Macmillanovi slovarji, v katerem je leksikografom na voljo besedišče, ki je določeno kot primerno za uporabo v pomenskih definicijah. Spodnja tabela prikazuje del pomenske členitve glagola *pasti* s pomenskimi indikatorji in v povezavi s pomenskimi shemami.

Pomen/ podpomen	Pomenski indikator	Pomenska shema
3	znižati se na lestvici	
	3.1 o količini ali vrednosti	KOLIČINA, VREDNOST ali PREDMET, ki ima neko vrednost, pade na lestvici za določeno KOLIČINO ali VREDNOST
	3.2 o intenzivnosti	STANJE, v katerem je ČLOVEK ali USTANOVA, pade tako, da izgubi intenzivnost
	3.3 o statusu ali kakovosti	ČLOVEK ali USTANOVA ali STATUS, KAKOVOST ali POLOŽAJ človeka ali institucije pade na lestvici na določeno TOČKO ali VREDNOST
	3.4 o socialnem položaju	ČLOVEK pade v moralnem ali socialnem smislu tako, da izgubi svoj status na družbeni lestvici
4	zgoditi, dogajati se	NEKAJ, kar pade , se zgodi ali začne obstajati
	4.1 o stanju: doživljati	ČLOVEK, ki pade v določeno STANJE, se v njem nahaja ali ga doživlja
	4.2 o prostoru, situaciji	ČLOVEK, ki pade v PROSTOR, OKOLJE ali SITUACIJO, se nenadoma znajde tam
	4.3 o dejavnosti: ukvarjati se	ČLOVEK, ki pade v neko DEJAVNOST, se z njo ukvarja
	4.4 o dogodku: nastopiti	DOGODEK ali ČAS dogodka pade na določen ČAS
	4.5 o besedah: izreči	ko BESEDE padejo , so izrečene
	4.6 o neprijetnostih: doleteti	na ČLOVEKA ali USTANOVO pade NEKAJ neprijetnega
	4.7 napasti	ČLOVEK pade po drugem ČLOVEKU

Tabela 2: Pomenska členitev in pomenske sheme v LBS za glagol pasti.

2.3.1.3 Pomenska shema in teorija *FrameNet*

Opis pomenske sheme za vsak pomen in podpomen glagola predstavlja v LBS združitev skladenjskih in pomenskih informacij pri opisu glagolov (ponekod tudi samostalnikov in pridevnikov). Pri oblikovanju pomenske sheme smo se približali teoriji pomenskih shem v projektu *FrameNet* (FN) ameriške univerze v Berkeleyju

(Fillmore et al. 2003),²³ hkrati pa pomenska shema, kot jo vidimo v LBS, ohranja glede na izhodiščni projekt nekatere pomembne razlike.

Tako kot pri FN temelji tudi v LBS opis pomenskega scenarija v pomenski shemi LE na analizi velikega števila korpusnih konkordanc.²⁴ Za razliko od FN pa naš namen ni opisati pomena besede na podlagi obstoječega ali vnaprej izdelanega seznama pomenskih shem in pomenskih elementov znotraj njih. V LBS je opis pomenskega scenarija v pomenski shemi prost; edini pogoj je, da redaktor natančno opiše situacijo pomena določene LE in izpostavi vse za pomen nujne udeležence in okoliščine. Za razliko od FN, ki za vsak shemski element predvideva ime, opis in semantično kategorijo, poimenovanje udeležencev in okoliščin v pomenskih shemah v LBS ni fiksirano in vnaprej določeno, pač pa je poimenovanje udeležencev, npr. ČLOVEK, PREDMET, ŽIVAL, SNOV, DEL TELESA itd. podrejeno opisu konkretnega pomenskega scenarija besede oziroma njenega pomena. V nasprotju s FN, kjer formalizirani pomenski okvirji in elementi znotraj njih omogočajo medsebojno povezovanje in vzpostavljanje pomenskih razmerij med LE, namen LBS ni združevati LE na podlagi pomenskih shem, čeprav naknadna analiza omogoča tudi to. Povezovanje pomensko sorodnih besed v FN narekuje obravnavo glede na pomensko povezanost LE, kar pomeni, da v primeru, če seznam v FN ne vsebuje pomenske sheme, ki bi ustrezala opisu določene nove besede, te besede ni mogoče vključiti v sistem, dokler se ne pokaže zadostna potreba oziroma zadostno število besed, ki bi ustrezale novi pomenski shemi. V nasprotju s tem je LE v LBS mogoče obravnavati neodvisno od njihove pomenske (ne)povezanosti, kar omogoča vključitev jedrnega dela besedišča v LBS neodvisno od možnosti pomenskega povezovanja med LE. Iz istega razloga lahko v LBS več pozornosti posvečamo pomenskemu opisu večpomenskih besed, tj. opisu tudi tistih pomenov, ki nimajo v celotni leksikalni bazi izkazane sorodne pomenske povezave.

2.3.1.3.1 Oblikovanje pomenske sheme

V izhodišču je naša naloga ugotoviti mehanizme, po katerih se govorniki slovenščine med seboj razumemo, oziroma mehanizme, na podlagi katerih smo sposobni razločevati med pomenom glagola *stisniti* v zgledih (**a**) od pomena glagola *stisniti* v zgledih (**b**). Osnovni princip, ki ga pri tem upoštevamo, izhaja iz teorije CPA (Hanks 2005) in predpostavlja, da pomeni izhajajo iz slovničnih vzorcev, v katerih besede nastopajo, in ne iz besed samih na sebi.

²³ Teorija in praksa projekta *FrameNet* je podrobno predstavljena v posebni številki revije *International Journal of Lexicography*, Vol. 16, No. 3, september 2003, Special Issue: Framenet and Frame Semantics. Projekt je bil predstavljen tudi z vidika slovenščine (Krek 2008 in <http://videlectures.net/solomon_krek_fn/>). S slovenskim *FrameNetom* se ukvarja tudi Birte Lönneker Rodman, več o tem na spletni strani <<http://www.icsi.berkeley.edu/~loenneke/>>.

²⁴ Analiza besedišča v celoti temelji na referenčnem korpusu *FidaPLUS*. Poleg tega uporabljamo tudi orodje za leksikalno profiliranje besed, kot so *Word Sketches* za slovenščino (Krek in Kilgariff 2006).

- (a) Otital je ženino roko in jo močno **stisnil**.
Anja pa ni Marjana niti enkrat **stisnila** za ramena.
Eleanor je zdravilna zelišča in pismo trdno **stisnila** v *dlan*.
Cezar z *dlanjo stisne* ročaj meča.
- (b) **Stisnite** si LIMONO, POMARANČO, GRENIVKO, MELONO, ANANAS itn.
 Iz GROZDJA SO **stisnili** 70 litrov ledenega vina.
 Počasi segrevajte marelično marmelado, v katero ste **stisnili** malo limoninega soka.
Ulični prodajalci TRS **stisnejo** v *stiskalnici* in soku dodajo nekaj soli in limoninega soka.

Če primerjamo pomene glagola *stisniti* v zgledih (a) in (b) zgoraj, lahko ugotovimo, da ima pomen glagola v zgledih (a) nekatere skupne lastnosti, ki se ločujejo od skupnih lastnosti pomena v zgledih (b). Razlike med pomenoma (a) in (b) v pomenski shemi opišemo na način, s katerim dosežemo najmanjši skupni imenovalec razumevanja konkretnega pomena oziroma skupni imenovalec razlikovanja med obema pomenoma in med drugimi pomeni/podpomeni istega glagola. Za oba pomena glagola *stisniti* v zgledih (a) in (b) smo zapisali pomenski scenarij v obliki pomenskih shem takole:

- (a) ČLOVEK *stisne* PREDMET z *DLANJO*
- (b) ČLOVEK s *PRIPOMOČKOM stisne* SADJE |ali| *stisne* TEKOČINO iz SADJA |ali| *stisne* SADJE v TEKOČINO

Za besede, ki so v pomenski shemi napisane z velikimi črkami, predvidevamo, da so obvezni shemski elementi za konkretni pomen glagola, različno zapisani deli stavka v zgledih pa ustrezajo shemskim elementom v pomenskih shemah.

2.3.1.3.2 (Ne)izraženost elementov pomenske sheme

V opisu pomenskega scenarija zapisujemo skladijsko in pomensko nujne elemente pomenske sheme z velikimi črkami, s čimer nakažemo vezljivostni potencial konkretnega pomena. Na tej stopnji se odločamo tudi o tem, ali so okoliščine imanentna lastnost konkretnega glagolskega pomena. Če je odgovor pritrdilen, je njihova obveznost razvidna iz pomenske sheme z zapisom z velikimi črkami. Z analizo konkordančnega niza lahko ugotovimo, da so obvezni elementi, kot je razvidno tudi iz zgornjih zgledov, v stavku lahko izraženi ali pa ne. Pri tem so (a) izraženi elementi pomenske sheme lahko razvidni (a.1) znotraj ali (a.2) zunaj stavčne zgradbe ali pa so (b) neizraženi. Neizraženost obveznih elementov je povezana s stavčno strukturo, ki omogoča izpuščanje (npr. brezosebna raba, splošni vršilec ipd.) ali pa obstaja splošno vedenje o nujni prisotnosti sicer neizraženega elementa, npr.: *Tomaž jé (hrano)*. Neizraženost elementov na osebkovem mestu v argumentni strukturi je v slovenščini pogosto, kot izkazujejo spodnji zgledi, zajeta

v osebni obliki glagola. V pomenski shemi »**ČLOVEK** s *PRIPOMOČKOM stisne* SADJE |ali| *stisne* TEKOČINO iz SADJA |ali| *stisne* SADJE v TEKOČINO« so:

- (a) izraženi elementi pomenske sheme (SADJE, tekočina)
 - (a.1) znotraj stavka: *Iz GROZDJA so stisnili 70 litrov ledenega vina.*
 - (a.2) zunaj stavka: *JABOLKA je treba najprej umiti, razrezati in stisniti. Jabolčni sok nato pustimo stati.*
- (b) neizraženi elementi pomenske sheme: (tekočina in sredstvo):
Stisnite si LIMONO, POMARANČO, GRENIVKO, MELONO.

2.3.2 Skladenjske informacije

Izhajajoč iz leksikografske teorije in prakse nas na skladenjskem nivoju zanimajo podrobnosti skladenjskega okolja preučevane LE. Če predstavlja pomenska shema opis pomenskega scenarija, v katerem so za glagolske iztočnice jasno izraženi udeleženci, njihova razporeditev v slovničnem vzorcu in okoliščine, beležimo s skladenjskimi strukturami (ang. *constructions*) in slovničnimi vzorci (ang. *syntactic patterns*) tipične skladenjske realizacije konkretnih pomenov preučevane LE oziroma povedano drugače: s skladenjskimi strukturami beležimo skladenjske realizacije obveznih (in tipičnih opcijskih) elementov, izraženih v pomenski shemi. Beleženje skladenjskih struktur je v prvi vrsti namenjeno potrebam RONJ, čemur je v celoti podrejen njihov zapis. V LPB so tovrstni podatki navadno zapisani v obliki kod, v našem primeru v obliki atributov, ki jih pripisujemo posameznim delom struktur ali vzorcev. V ta namen je redaktorjem na voljo zaključen seznam atributov, ki ga vključuje programski vmesnik.

Zapisovanje skladenjskih informacij v LBS se pri glagolskih iztočnicah nekoliko razlikuje od zapisa pri drugih besednih vrstah. Pri glagolih predvidevamo pri vsakem pomenu in podpomenu informacijo o **megastruktur**, ki združuje vse skladenjske strukture,²⁵ ki predstavljajo formalizirano podstavo **slovn**ičnih **vzorcev**. Namen megastruktur je predstaviti vse elemente, tj. besede oziroma besedne zveze, vključno s predlogi, vezniki in odvisnimi stavki znotraj stavčne zgradbe, ki so potrebni za izražanje konkretnega pomena glagola v iztočnici. Poleg obveznih elementov so iz megastrukture razvidni tudi možni (opcijski) elementi in elementi, ki so medsebojno zamenljivi (alternativni elementi strukture). Na primer, pomen glagola *pobirati* 'zaračunavati plačilo' se realizira z megastrukuro: sbz1 GBZ sbz4 (za sbz4| od sbz2), ki jo beremo takole: samostalniku ali samostalniški besedni zvezi v imenovalniku (sbz1) sledi glagol oziroma glagolska besedna zveza (GBZ) *pobirati* in samostalnik/besedna zveza v tožilniku (sbz4). Tej strukturi lahko (opcija) sledi bodisi predlog *za*, ki veže samostalnik/besedno zvezo v tožilniku

²⁵ V fazi izdelave specifikacij za LBS smo oblikovali sezname skladenjskih struktur za posamezne besedne vrste, pri čemer smo strukture razdelili glede na to, ali je obravnavana besedna vrsta kot iztočnica jedro besedne zveze, kar zapišemo z velikimi črkami, npr. **SBZ2** sbz2 – **meseca** [aprila], ali ne, npr. sbz2 **SBZ2** – [meseca] **aprila**.

(za sbz4), bodisi predlog *od*, ki veže samostalnik/besedno zvezo v rodilniku (od sbz2).

Postopek lahko pogledamo tudi z vidika obrnjenega zaporedja, kjer za glagol *stisniti* v pomenu: 'oprijeti z dlanjo' predvidevamo najprej ločene skladijske strukture, s katerimi se konkretni pomen tipično izraža, nato pa jih združimo v megastrukturo, ki poleg osnovne, tj. za izražanje konkretnega pomena minimalne strukture, izraža še opcijske dele (v oklepaju) in alternativne oziroma med seboj zamenljive dele (ločene s poševnico):

sbz1 GBZ sbz4	}	sbz1 GBZ sbz4 (za sbz4 /v sbz4 / z/s sbz6)
sbz1 GBZ sbz4 za sbz4		
sbz1 GBZ sbz4 v sbz4		
sbz1 GBZ sbz4 z/s sbz6		

Medsebojno prepletanje skladnje in pomena pri glagolskih iztočnicah najbolje (poleg pomenskih shem) odražajo **slovnični vzorci**. Predstavljajo namreč tisti del LBS, ki daje informacije o vezljivosti glagola, obenem pa predstavlja posredniško fazo med skladijskimi strukturami (združenimi v megastrukturo) in njihovimi uresničitvami v konkretnem slovničnem vzorcu. Na primer, zapisu pomenske sheme za glagol *roditi* v pomenu 'o rojstvu otroka': ŽENSKA rodi OTROKA ali OTROK se rodi KOMU, sledijo slovnični vzorci, v katerih se ta pomen tipično realizira, zato izražajo tudi okoliščine, najpogosteje s predložnimi zvezami v prislovni funkciji. Konkretna zapolnitev predložnih zvez je zabeležena na kolokacijskem nivoju.

roditi

INDIKATOR: o rojstvu otroka

POMENSKA SCHEMA: ŽENSKA rodi OTROKA ali OTROK se rodi KOMU

VZOREC: kdo rodi koga

KOLOKACIJA: roditi [sina, hčerko, otroka, fantka]

VZOREC: kdo rodi

KOLOKACIJA: [ženska, žena] rodi

VZOREC: kdo se rodi

KOLOKACIJA: [otrok, dete, dojenček, deklca] se rodi

VZOREC: kdo se rodi česa/sbz2 *KDAJ*

KOLOKACIJA: roditi se [aprila, maja]

VZOREC: kdo rodi kaj/sbz4 *KDAJ*

KOLOKACIJA: roditi [vsako leto]

VZOREC: kdo se rodi po čem *KDAJ*

KOLOKACIJA: roditi se po [vojni]

VZOREC: kdo se rodi med čim *KDAJ*

KOLOKACIJA: roditi se med [vojno]

VZOREC: kdo se rodi v čem *KJE*

KOLOKACIJA: roditi se v [Ljubljani, Mariboru, Trstu]

VZOREC: kdo se rodi na čem *KJE*

KOLOKACIJA: roditi se na [Dunaju, Hrvaškem, Ptujju]

VZOREC: kdo se rodi s čim

KOLOKACIJA: roditi se z/s [Downovim sindromom, okvaro, napako]

Zgled: *Slovnični vzorci in kolokacije za pomen glagola roditi.*

2.3.3 Kolokacijski nivo: sintagmatika LE

V tem segmentu LBS beležimo *kolokacije*, ki jih razumemo kot pogoste kombinacije besed, v katerih določena beseda kaže očitno težnjo po sopojavljanju

z drugo besedo, pri čemer je pogostost večja od naključne sopojavitve (Kilgarriff 2006b: 998). Tipične kolokacije, v katere vstopa obravnavana beseda v iztočnici, beležimo glede na registrirane pomene in skladenjske strukture. Z upoštevanjem statističnih izračunov besedne povezovalnosti, kot so razvidni iz programa za izdelavo besednih skic za slovenščino, vključujemo v LBS informacijo o sintagmatskem povezovanju besed, npr. *pasti pod* (*ničlo, vpliv, streli, mejo, kolo ...*); *pasti na* (*pamet, tla, izpitu, bojišču ...*), s podrejanjem kolokacij pomenski členitvi pa upoštevamo tudi paradigmatški vidik besedne povezovalnosti, npr. *pasti pod* (*ničlo, minus, ledišče*) – ‘znižati se’; *pasti na* (*tla, vozišče, pločnik*) – ‘izgubiti ravnotežje’. To pomeni, da ob vsakem pomenu ali/in podpomenu, ki ga za obravnavano besedo registriramo v LBS, beležimo v kolokacijah samo tiste dele struktur in slovničnih vzorcev, ki so kolokabilni:

pogajati se

INDIKATOR: poskušati se dogovoriti

POMENSKA SHEMA: ČLOVEK se pogaja z DRUGIM ČLOVEKOM glede VSEBINE dogovora

SKLADENJSKA (MEGA)STRUKTURA: sbz1 GBZ (s sbz6) (o sbz5 / glede sbz2 / za sbz4)

VZOREC: kdo se pogaja

VZOREC: kdo se pogaja s kom

VZOREC: kdo se pogaja s kom o čem

VZOREC: kdo se pogaja s kom glede česa

VZOREC: kdo se pogaja s kom za kaj

KOLOKACIJA: [vlada, država, vodstvo] se pogaja

KOLOKACIJA: [sindikati, delodajalci, partnerji] se pogajajo

KOLOKACIJA: pogajati se z/s [sindikati, EU, vlado, Brusljem, delodajalci]

KOLOKACIJA: pogajati se z/s [ugrabitelji, uporniki, teroristi]

KOLOKACIJA: pogajati se o [ceni, nakupu, odkupu, prodaji, pogojih, spremembah]

KOLOKACIJA: pogajati se o [članstvu, vstopu, sporazumu]

KOLOKACIJA: pogajati se o [vdaji, izpustitvi]

KOLOKACIJA: pogajati se za [ceno, plačo]

KOLOKACIJA: pogajati se za [vstop, članstvo, podaljšanje, izpustitev]

STRUKTURA: rbz GBZ

KOLOKACIJA: [trdo, tajno, intenzivno, neuspešno, resno] se pogajati

STRUKTURA: Vez-gbz pbz GBZ

KOLOKACIJA: [biti pripravljen] pogajati se

STRUKTURA: gbz GBZ

KOLOKACIJA: [začeti, pričeti, nehati, nameravati, znati] se pogajati

Zgled: *Kolokabilni deli slovničnih vzorcev in struktur za glagol pogajati se.*

Kolokacije predvidevamo tudi na vseh tistih segmentih LBS, kjer se deli predhodno registriranih vzorcev, struktur in skladenjskih zvez kažejo kolokabilni, npr. zvezi: *čas od (česa) do (česa)* in *v času po (čem)* izkazujeta lastno kolokabilnost: *čas od [okužbe] do [pojava bolezni, obolenja]*; *v času po [volitvah, osamosvojitvi, vojni, revoluciji]*, in povsod tam, kjer imajo večbesedne LE, tj. stalne besedne zveze in frazeološke enote (FE), kolokabilno okolje, npr. *glasbena šola*: [*učenec, ravnatelj, gojenec, učenka, ravnateljica*] *glasbene šole*, [*dvorana, prostori*]

glasbene šole, [srednja, nižja] glasbena šola, [obiskovati, vpisati se na] glasbeno šolo, ne izključujemo pa tudi možnosti, da kolokator sam nastopa kot večbesedna enota, npr. *pasti na [pokrov motorja], [berlinski zid, železna zavesa] pade, [minuta molka] v spomin (komu), roditi se z/s [Downovim sindromom]*.

2.3.4 Stalne besedne zveze in frazeološke enote

V LBS predvidevamo dva tipa večbesednih leksikalnih enot: stalne besedne zveze (SBZ) in frazeološke enote (FE).

Kot SBZ obravnavamo tipične besedne kombinacije, ki so oblikovno in pomensko osamosvojene, kar pomeni, da imajo svoj lastni pomen (oziroma konkretnega nanosnika), ki je pogosto vezan na določeno strokovno ali kako drugače omejeno področje rabe (npr. *kisli dež, kislja juha, konj z ročaji*), pri čemer vsaj ena sestavina kaže asociativno povezavo s katerim od pomenov besede v iztočnici (npr. *kisel*: ‘o kemični sestavi’; ‘o hrani’). Tipičen pokazatelj leksikaliziranosti je tudi večpomenskost SBZ oziroma dejstvo, da obstaja tudi dobesedna, navadno neterminološka homonimna besedna zveza, npr. *črna luknja* v primeru: *črne luknje razbitih oken* in *črne luknje so ključ za razumevanje poslednje usode vesolja*. Variantnost sestavin SBZ je omejena na sinonimijo ali skupno pomensko polje (npr. *kratkoročni/kratkotrajni spomin; glasbena skupina/zasedba*), hkrati pa imajo SBZ pogostokrat tudi lastno tipično besedilno okolje, ki ga beležimo v obliki kolokacij, kot smo pokazali zgoraj.

Pomenski indikatorji na kratko opredelijo pomenski obseg SBZ. Z njimi označujemo:

(a) pripadnike določene vrste ali sorte, npr. arabski konj – *pasma*, zlati delišes – *sorta*;

(b) poimenovanja predmetov, stvari, stanj, prepričanj, ustanov, kjer je indikator bodisi širša, npr. iskanje skritega zaklada – *družabna igra*, konj z ročaji – *telovadno orodje*, nilski konj – *žival*, krompir v oblicah – *jed*, ali ožja nadpomenska, npr. črni kos – ptica zelena uš – škodljivec;

(c) poimenovanja stvari, predmetov, pojavov ipd. na določenem strokovnem ali omejenem področju rabe, npr. otrok iz epruvete – *medicina*, zimski čas – *geografija*, kisli dež – *ekologija*;

(d) parafraze ali opise: knežje mesto – *Celje*, mesto pod Pohorjem – *Maribor*.

V samostojnem razdelku LBS (tj. neodvisno od pomenske členitve) beležimo FE. Gre za tiste besedne zveze, ki glede na sestavne elemente izkazujejo samostojni, od sestavnih elementov neodvisni – t. i. *frazeološki pomen*. Ob tem se za FE zahteva tudi prenesenost pomena, ki jo izkazuje zveza kot celota, kar ji daje glede na nefrazeološka večbesedna poimenovanja v besedilu ekspresivno oziroma zaznamovano noto. Prav zadnje iz FE izloča tudi vsa večbesedna terminološka poimenovanja, ne glede na sicer izkazan pomenski prenos, npr. *črna skrinjica* ‘naprava’; *sončni veter* ‘energija’.

Vsako frazeološko enoto v LBS opredelimo s pomenskim indikatorjem, ki ima enako vlogo kot pri enobesednih iztočnicah, kar pomeni, da na kratko opredeli pomen FE, npr. *stisniti zobe* – ‘potrpeti’, *stisniti koga v kot* – ‘spraviti v brezizhoden položaj’, *stisniti glave* – ‘sestati se in skupaj razmisliti’, *stisniti pas* – ‘varčevati’, *stisniti ustnice* – ‘gesta kot znak jeze, nestrinjanja ali kljubovanja’, oziroma vzpostavlja razliko med pomeni večpomenske FE, npr. *stisniti pest/pesti* – 1. ‘gesta kot znak upora, pripravljenosti na boj’, 2. – ‘gesta kot znak podpore, upanja na uspeh’, *dol pasti* – 1. ‘biti presenečen, začuden’, 2. – ‘biti utrujen’, 3. – ‘omagati, izgubiti zavest’.

FE tipično izkazujejo različne variantne možnosti že na ravni posameznih sestavin zveze ali »zunanjih« udeležencev, npr. *luč na koncu tunela/predora*, *stisniti pest/pesti*, *zastor/zavesa pade*; *biti trn v peti komu/koga*, *pasti v kremplje komu/čemu/koga*, pogostokrat v kombinaciji z oblikoslovnimi prilagoditvami, ki jih varianta zahteva, npr. *stisniti skozi zobe* in *stisniti med zobmi*, *stisniti koga v kot* in *stisniti koga ob zid*, *zaiti v slepo ulico* in *znajti se v slepi ulici*. Take variante, vključno s pretvorbenimi možnostmi, ki zajemajo:

(a) različno skladenjsko funkcijo FE, npr. *beraška palica* – *spraviti koga na beraško palico*; *zdrava pamet* – *biti zdrave pameti/pri zdravi pameti* – (*narediti kaj*) *po zdravi pameti*, pri čemer se zavedamo, da pri glagolskih rabah oziroma FE v funkciji povedka na pomenske odtenke vplivajo konkretni glagoli, npr. *princ na belem konju* – *čakati na princa na belem konju*, *iskati*, *najti* (*svoje*) *princa na belem konju* – *sanjati o princu na belem konju* itd.;

(b) različne realizacijske možnosti,²⁶ ki jih omogoča zgradba FE, zlasti posamostaljenje, npr. *prati možgane* – *pranje možganov*, *pihati na dušo* – *pihanje na dušo*, in potrpnjenje, npr. *stresati/zbijati šale* – *šale se stresajo/zbijajo*; *dvigati prah* – *prah se dviga*, *dati besedo* – *beseda je dana* ipd.

(c) število in razporeditev »zunanjih« udeležencev, npr. *prepustiti se na milost in nemilost komu/čemu* – *prepustiti koga na milost in nemilost komu/čemu* – *biti prepuščen na milost in nemilost komu/čemu*;

(č) realizacija v kateri od tipičnih upovedovalnih oziroma naklonskih možnosti, kot so zanikanje, velebnost, vprašalnost ipd., npr. *kaj (ne) pade na pamet/misel/glavo/um komu* in *kaj ne pade niti na kraj pameti komu*, *saj nisem na glavo padel/la!*;

(d) fakultativne sestavine FE: *pasti kot* (*zrela*) *hruška* in *pasti komu* (*v naročje*) *kot* (*zrela*) *hruška*, *podajati si kaj* (*iz rok v roke*) *kot vroč krompir*, *biti* (*vedno*) *korak pred časom*; tudi kadar so opcijski elementi »zunanji« udeleženci, npr. *pasti na limanice* (*komu*), beležimo, če se izkazujejo kot tipične in za posamezno FE relevantne.

Pri tem se držimo pravila, da kot variante (zapis s poševnico) in variantne oblike rabe ter pretvorbe beležimo znotraj enega nadelementa frazeološka enota zaporedoma, če ne prihaja do bistvenih pomenskih sprememb, medtem ko večpomenske FE ter

²⁶ V primeru, ko se katera od pretvorb izkaže kot edina možna, npr. *kopja se lomijo* (*lomiti kopja), navajamo samo to obliko.

sinonimne FE, npr. *pasti na rit* in (*skoraj*) *pasti s stola*, kjer gre v obeh primerih za pomen 'biti presenečen', beležimo kot samostojne FE, torej v samostojnem nadelementu, saj predvidevamo, da lahko tudi homonimne in sinonimne FE izkazujejo lastne variante in pretvorbene možnosti.

Na enak način kot FE beležimo v samostojnem razdelku LBS – tj. neodvisno od pomenske členitve – idiomatske enote. Sem prištevamo FE s stavčno zgradbo, t. i. **pregovore** in **reke** (razen glagolskih FE tipa: *zob časa načne koga/kaj*, *mrak pade na oči komu*, *sekira pade v med komu* ipd., ki se v stavku obnašajo kot povedki). Zanje velja, da so »besedila že sama na sebi«, hkrati pa je zanje značilno, da se nanašajo na konkretno temo ali referenta, izraženega v širšem (so)besedilu, npr. *iz te moke ne bo kruha*, *jabolko ne pade daleč od drevesa*, *čas zaceli vse rane* itd.

2.3.5 Primeri rabe

S primeri rabe, ki v celoti temeljijo na korpusu *FidaPLUS* oziroma bodo v nadaljevanju izdelave LBS temeljili na bodočem v okviru projekta nastajajočem milijardnem korpusu SSJ in jih v programski vmesnik prenašamo s pomočjo funkcije WSE (gl. v nadaljevanju), potrjujemo vse informacije, ki smo jih na višjih nivojih vnesli v LBS: za vsak zabeležen pomen LE morajo primeri, ki odražajo realno jezikovno rabo, potrditi s pomenskim indikatorjem opredeljen pomen, pomensko shemo z obveznimi in opsijskimi udeleženci, skladišne strukture in slovnične vzorce pri glagolskih iztočnicah. Najbolj se primerom rabe približajo kolokacije, ki so tipične potrditve skladišnih struktur ali slovničnih vzorcev. S primeri rabe vedno potrdimo tudi realno obnašanje SBZ, FE v besedilu. V hierarhično urejenih nivojih LBS tako predstavljajo primeri rabe najkonkretnjšo informacijo in skupaj s pomenskimi, skladišnimi in kolokacijskimi podatki o LE zaključujejo celoto, znotraj katere so informacije med seboj neposredno povezane in se medsebojno dopolnjujejo.

pasti

POMENSKI INDIKATOR: 1.2 izgubiti; izpustiti

POMENSKA SHEMA: PREDMET pade ČLOVEKU KAM |ali| MU pade od KOD

MEGASTRUKTURA: sbz1 GBZ sbz3 (iz sbz2) v sbz4/na sbz4/pod sbz4

VZOREC: kaj pade komu v kaj

VZOREC: kaj pade komu na kaj

VZOREC: kaj pade komu pod kaj

VZOREC: kaj pade komu iz česa

KOLOKACIJA: [pištola, cigareta] pade

KOLOKACIJA: pasti v [jarek, jašek]

KOLOKACIJA: pasti na [tla]

KOLOKACIJA: pasti pod [mizo]

KOLOKACIJA: pasti iz [rok, roke, ust]

KOLOKACIJA: pasti iz [žepa]

ZGLED: Vžigalice so ji padle v cestni jarek in so zdaj vse premočene.

- ZGLED: Fantku je plastični tovornjaček padel v jašek.
 ZGLED: Pištola mu je med prerivanjem padla na tla.
 ZGLED: Če vam hrana pade na tla, je lahko okužena v petih sekundah.
 ZGLED: Vse kaže, da mi je padlo pod mizo deset šilingov.
 ZGLED: Poberem knjigo, ki mi je padla iz rok, in nadaljujem branje.
 ZGLED: Sneguljčici zastrupljeni košček jabolka pade iz ust in dekle oživi.
 ZGLED: Med begom mu je ukradena denarnica padla iz žepa.

Zgled: *Leksikalno-gramatične informacije v LBS: od abstraktne pomenske členitve do konkretnih primerov rabe.*

Pri izbiri primerov stremimo k relativni kratkosti, če je mogoče, se izogibamo lastnih imen, predvsem pa smo pozorni na to, da izbiramo primere, ki ne vnašajo dvoma v informacije, ki jih želimo s primeri potrditi. S primeri rabe želimo zajeti tudi tipično besedilno rabo, tip besedila, področje rabe in skladijsko okolje, če ga ne moremo zajeti s skladijskimi strukturami. Pri navajanju števila primerov nismo omejeni, pomembno pa je, da je vsako skladijsko strukturo in kolokacijo mogoče videti v njenem tipičnem življenjskem okolju.

3 Pridobivanje podatkov iz korpusa in postopek izdelave iztočnic

3.1 FidaPLUS in Sketch Engine

Podatki za izdelavo iztočnic, tj. opis LE po posameznih nivojih, v celoti temeljijo na korpusu *FidaPLUS* (Arhar in Gorjanc 2007) in uporabi orodja *Sketch Engine* (SkE),²⁷ zlasti funkcije *Word Sketch* (WSE) (Kilgarriff et. al 2004), katere slovenska verzija (Krek in Kilgarriff 2006) je bila v procesu testiranja nadgrajena za namene projekta. Pri delu s *FidoPLUS* uporabljamo konkordančnik ASP32. Možnosti, ki nam jih ponuja, kot tudi statistične metode in načine iskanja podrobno opisuje spletni priročnik (Arhar 2007), zato jih na tem mestu ne bomo opisovali. Nekoliko več pozornosti zahteva korpusni iskalni sistem SkE, v katerega je vgrajen korpus *FidaPLUS* in katerega beta verzijo uporabljamo za delo s konkordancami, izdelovanje gramatično-kolokacijskih profilov besed in izvažanje primerov rabe. Konkordančnik omogoča iskanje po lemi, frazah, besedni obliki oziroma na podlagi iskalnega pogoja, ki ga določimo s pomočjo kodiranega zapisa v CQL (*Corpus Query Language*). Poleg omenjenega lahko v konkordančniku določimo obseg in besedno vrsto besed v besedilnem okolju iskane besede ter tipe besedil, po katerih želimo iskati, pri čemer lahko izbiramo besedila glede na leto izida, prenosnik in zvrst, če je seveda korpus temu primerno grajen. Posamezno konkordanco lahko razširimo na poljuben obseg odstavka. Konkordančni niz pa lahko urejamo

²⁷ <<http://www.sketchengine.co.uk/>>.

glede na besede levo ali desno od iskane besede, filtriramo, izdelamo vzorec (npr. določenega števila naključno izbranih konkordanc, ki jih nato izvozimo za nadaljnjo analizo – gl. v nadaljevanju), analiziramo glede na pogostnost leme, besedne oblike, kombinacije morfosintaktičnih oznak itd. Poleg možnosti, ki nam jih ponuja konkordančnik, je za pridobivanje leksikalno-gramatičnih podatkov iz korpusa pomembna funkcija WSE, s pomočjo katere za izbrano lemo dobimo seznam kolokatorjev, ki ustrezajo določeni skladijski relaciji (npr. *a modifier*: (Prid-Sam): [lep, zgodovinski, trajen ...] spomin; *is obj3* (Glag-Sam3): [pokloniti, posvetiti, oddolžiti se...] spominu), možnost izbire konkretnih kolokatorjev iz seznama (TBL – *Tick Box Lexicography*) ter izvoz korpusnih zgledov, ki vsebujejo izbrane kolokatorje (GDEX – *Good Dictionary Examples*) (Kilgarriff et al. 2008).

Leksikalno-gramatični profil, kot ga omogoča WSE, predstavlja skupaj s prenosom kolokatorjev in zgledov rabe v programski vmesnik osnovo za pridobivanje skladijskih in kolokacijskih informacij v LBS. Pred tem izdelamo za vsako LE pomensko analizo, na podlagi katere določimo pomenske indikatorje in pomenske sheme.

spomin Fida PLUS 620m freq = 97524

<i>a modifier</i>	26132 1.3	post na-d	19227 13.1	prec v-d	15072 5.9	is obj3	1665 3.8	prec kot-d	409 3.3
<input checked="" type="checkbox"/> lep	4559 74.23	<input checked="" type="checkbox"/> otroštvo	340 54.13	<input checked="" type="checkbox"/> vstara	1783 88.53	<input type="checkbox"/> pokloniti	587 84.16	<input type="checkbox"/> ostati	27 19.5
<input type="checkbox"/> zgodovinski	1735 67.87	<input type="checkbox"/> žrtev	566 47.85	<input type="checkbox"/> prklicati	558 62.08	<input type="checkbox"/> posvetiti	496 59.6	<input type="checkbox"/> ohraniti	16 18.33
<input type="checkbox"/> trajen	632 58.4	<input type="checkbox"/> mladost	249 46.96	<input type="checkbox"/> ostati	2366 55.17	<input type="checkbox"/> oddolžiti	49 40.12	<input type="checkbox"/> biti	35 11.82
<input type="checkbox"/> nostalgčen	190 57.05	<input type="checkbox"/> preteklost	402 44.34	<input type="checkbox"/> slovesnost	261 44.81	<input type="checkbox"/> ubežati	14 25.51	>>	
<input type="checkbox"/> prijeten	609 50.56	<input type="checkbox"/> dogodek	735 43.05	<input type="checkbox"/> ohraniti	642 42.64	<input type="checkbox"/> posvečati	33 25.14		
<input type="checkbox"/> nepozaben	219 49.75	<input type="checkbox"/> holokavst	63 38.98	<input type="checkbox"/> komemoracija	35 35.9	<input type="checkbox"/> prepustiti	39 24.46		
<input type="checkbox"/> boleč	288 48.86	<input type="checkbox"/> rojak	129 38.54	<input type="checkbox"/> obeležje	58 35.44	<input type="checkbox"/> nameniti	77 24.32		
<input type="checkbox"/> grenek	218 48.75	<input type="checkbox"/> maturanec	15 37.9	<input type="checkbox"/> molk	64 32.98	<input type="checkbox"/> predajati	14 21.81		
<input type="checkbox"/> kolektiven	394 47.05	<input type="checkbox"/> čas	1183 35.82	<input type="checkbox"/> turnir	216 31.57	<input type="checkbox"/> priti	25 7.46		
<input type="checkbox"/> neizbrisen	69 45.72	<input type="checkbox"/> grozota	41 32.54	<input type="checkbox"/> zasidrati	64 30.13	>>			
<input type="checkbox"/> živ	397 43.52	<input type="checkbox"/> fobja	32 32.04	<input type="checkbox"/> svečanost	34 29.74				
<input type="checkbox"/> mladosten	140 42.31	<input type="checkbox"/> borec	106 31.28	<input type="checkbox"/> venec	51 28.91				
>>		>>		>>					

skladijske relacije in povezave na kodiran zapis

statistične vrednosti

frekvenca/povezave na konkordance

povezava na izbrane kolokatorje z zgledi

<i>is obj4</i>	9833 3.2	prec ob-d	415 3.0	prec prep	28986 2.8	prec brez-d	184 2.8	<i>is subj</i>	6523 2.7
<input type="checkbox"/> obupati	2520 102.63	<input type="checkbox"/> nasmejniti	29 32.9	<input type="checkbox"/> v	21710 47.57	<input type="checkbox"/> trg	58 35.1	<input type="checkbox"/> varati	373 68.81
<input type="checkbox"/> obuditi	1072 77.93	<input type="checkbox"/> biti	70 18.23	<input type="checkbox"/> ob	699 24.58	>>		<input type="checkbox"/> vezati	500 63.28
<input type="checkbox"/> počastiti	426 60.66	>>		<input type="checkbox"/> iz	997 22.93			<input type="checkbox"/> pežati	80 45.57
<input type="checkbox"/> osvoboditi	209 57.46			<input type="checkbox"/> kot	592 22.16			<input type="checkbox"/> seči	179 40.33
<input type="checkbox"/> buditi	44 33.82	post kot-d	358 2.9	<input type="checkbox"/> po	1089 20.71			<input type="checkbox"/> obupati	87 39.58
<input type="checkbox"/> ohranjati	135 31.34	<input type="checkbox"/> zgodba	29 26.61	<input type="checkbox"/> za	1578 15.93			<input type="checkbox"/> zble-deti	63 39.03
<input type="checkbox"/> prklicati	68 31.3	<input type="checkbox"/> človek	15 14.05	<input type="checkbox"/> brez	165 13.75			<input type="checkbox"/> ble-deti	38 35.77
<input type="checkbox"/> izbrnsati	65 29.88	>>		<input type="checkbox"/> z	1070 10.91			<input type="checkbox"/> ostajati	234 33.73
<input type="checkbox"/> oživljati	46 29.57			<input type="checkbox"/> razen	15 10.31			<input type="checkbox"/> ohranjati	123 32.95
<input type="checkbox"/> zbujsati	64 26.6			<input type="checkbox"/> zaradi	126 9.49			<input type="checkbox"/> obuditi	87 32.74
<input type="checkbox"/> ohraniti	148 25.49			<input type="checkbox"/> skozi	29 7.24			<input type="checkbox"/> lepiti	82 32.26
<input type="checkbox"/> oživiti	44 24.76			<input type="checkbox"/> med	146 7.07			<input type="checkbox"/> segati	100 30.48
>>		>>		>>				>>	

seznam kolokatorjev

možnost izbire kolokatorja (TBL)

Slika 3: Leksikalno-gramatični profil samostalnika spomin v WSE.

3.2 Postopek izdelave gesla v LBS

Potem ko pripišemo analizirani besedi v iztočnici gesla osnovne informacije, ki so na tej stopnji omejene zgolj na podatek o besedni vrsti, jo analiziramo z vidika pomenske razčlenjenosti in skladijskega obnašanja. Pri tem uporabljamo naslednji postopek: v SkE ali korpusu *FidaPLUS* izdelamo naključni vzorec konkordanc (oziroma se po potrebi odločimo samo za določene tipe besedil) s približno 150 do 300 konkordancami. Število konkordanc, ki jih je potrebno analizirati, da si ustvarimo sliko o njeni pomenski razčlenjenosti in tipičnih skladijskih realizacijah, ni fiksno in je odvisno od besedne vrste in pomenske kompleksnosti obravnavane besede. Konkordančni vzorec s pomočjo programa za konvertiranje²⁸ prenesemo v excelovo datoteko, kjer je pripravljen za nadaljnjo obdelavo.

še rasti . Zamenu je strah , nekaj ga je	stisnu	ven , lko ko gnoj iz mozola . Seveda ,	iztisniti snovitekočino iz česa - preneseno
vzamemo iz lonca in ju s pomočjo vilice	stisnemo	iz ovoja . Ker sta tako česen kot čebula	iztisniti snovitekočino iz česa v drugo stvar
na cedilu . Če pa sem ter tja v črpalko	stisnete	kapljico silikonske masti , bo delovanje	iztisniti snovitekočino iz česa v drugo stvar
segrevajte marelično marmelado , v katero ste	stisnili	malo limoninega soka . Nato zmes prellačite	iztisniti snovitekočino iz česa v drugo stvar
varuje pred prehladom . V ne pretopel čaj	stisnimo	sok limone . Prav tako prija topla limonada	iztisniti snovitekočino iz česa v drugo stvar
opozori , da ti je nekdo ^{da ravnaj} pjunil ali	stisnil	šarnpon . Zadnji trik sem doživel	iztisniti snovitekočino iz česa v drugo stvar
prebavni cikel , tako da mu moram občasno	stisniti	v danko (z nastavkom za modolex) nekaj	iztisniti snovitekočino iz česa v drugo stvar
pusitjo v posebnih sušilnicah , potem pa se	stisnejo	, in ko kvasovke še ne predelajo vseh	izželi kaj
jabolka je treba najprej umiti , razrezati in	stisniti	.Jabolčni sok nato pustimo stati v nepredušno	izželi kaj
slami sušili grozdje in ga po nekaj mesecih	stisnili	. Letos so s sušenjem grozdja poskusili	izželi kaj
je začeli dan s svežim sadnim sokom .	Stisnite	si limono , pomarančo , grenivko , melono	izželi kaj

Slika 4: Pomenska analiza konkordančnega niza z vidika udeleženske zgradbe za glagol stisniti.

Pomenski analizi sledi strukturalna in kolokacijska analiza, ki jo izdelamo na podlagi gramatično-kolokacijskega profila, kot ga omogoča WSE (**Slika 3**). Tudi tu imamo možnost kolokacijske informacije iz WSE konvertirati v excelovo datoteko, kjer beležimo komentarje in opombe ob posamezni skladijski relaciji in kolokatorjih. Pomembno pri tem je, da konvertirana datoteka ohranja povezavo s konkordancami, tako da je mogoče obnašanje in relevantnost kolokatorja in skladijske relacije, v katero tipično vstopa, preveriti v dejanski besedilni rabi.

²⁸ Program za konvertiranje sta za potrebe projekta izdelala sodelavca podjetja Amebis, Rok Rejc in Polonca Kocjančič.

konj 56415

a_modifier	18620	1.9	OK
paraden	1516	98.86	× FE
trojanski	845	88.15	× FE
islandski	410	74.11	× zloženka
nilski	228	73.85	× zloženka
povoden	352	71.81	× zloženka
hladnokrven	377	71.48	× zloženka
toplok rven	236	69.97	× zloženka
dirkalen	328	66.9	×
polnok rven	189	63.98	× zloženka
jahalen	261	63.93	×
vprežen	150	60.41	×
stepski	108	59.37	korpusni žum
vlečen	261	58.92	× FE
lipicanski	88	57.08	× zloženka
Marvin	105	56.1	korpusni žum
kočijaški	48	52.84	
bel	876	49.18	× FE
podarjen	110	47.14	× FE

Slika 5: Kolokacijska analiza samostalnika konj.

Ugotovljene jedrne pomene in podpomene za obravnavano LE skupaj s skladenjskimi strukturami, skladenjskimi vzorci pri glagolih in tipičnimi kolokacijami, ki potrjujejo zabeležene pomenske in skladenjske informacije, vnesemo v programski vmesnik DPS, ki ga na kratko predstavimo v nadaljevanju.

4 Programski vmesnik za izdelavo leksikalne baze

Za izdelavo LBS uporabljamo programsko opremo podjetja IDM,²⁹ *Dictionary Publishing System* (DPS). Gre za program, ki je namenjen sestavljanju slovarjev in podobnih jezikovnih virov (Kocjančič et al. 2006). Zasnovan je tako, da uredniki dostopajo do podatkovne baze prek spleta in z njo delajo bodisi neposredno na strežniku ali pa vnaprej pripravljeno gradivo prenesejo na svoj računalnik, uredijo in vrnejo v bazo, pri čemer za obdelavo uporabljajo programski vmesnik za delo v brezpovezavnem načinu (*Entry Editor*). Program DPS uredništvu omogoča nadzor nad posameznimi vsebinskimi segmenti, določitev faz obdelave, uporabniških pravic, razdeljevanje gradiva po različnih kriterijih, primerjavo stare in nove različice gradiva, preverjanje in preusmerjanje gradiva, kompleksna iskanja ter obdelave na celotnem slovarskem gradivu v formatu XML. Program omogoča tudi pretvorbo v pdf ter hiter izvoz in uvoz podatkov. Celotni programski paket sestavljata dve osnovni komponenti: DPS strežnik (*DPS Website*) in DPS vmesnik (*Entry Editor*). **DPS strežnik** je spletna funkcija, do katere uporabniki dostopajo s standardnimi orodji za navigacijo po spletu. Na strežniku so orodja za vodenje projektov, administriranje in konfiguriranje sistema ter uporabniški dostop do podatkov, **DPS vmesnik** pa je funkcija za uporabo v okolju Windows,

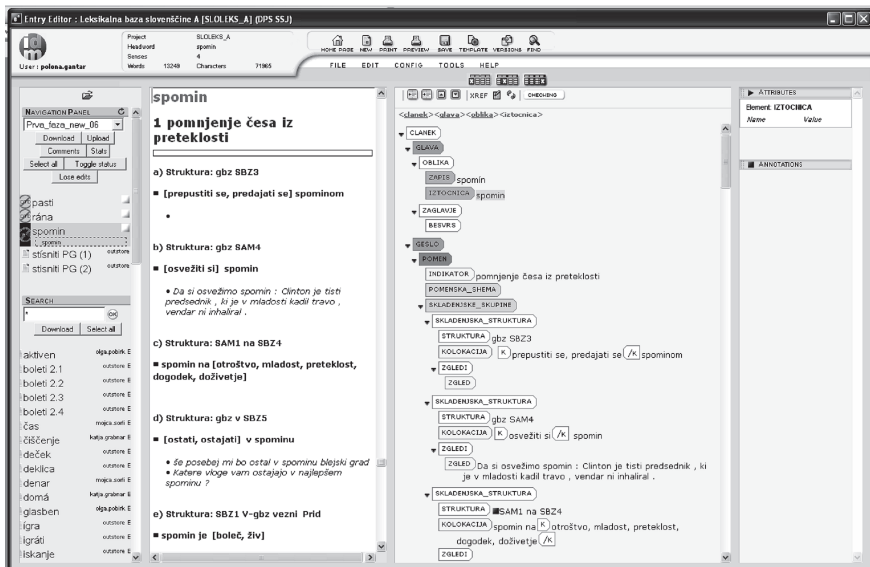
²⁹ <<http://www.idm.fr>>.

ki jo uporabnik naloži na svoj računalnik. Osnovni postopek dela je, da uporabnik z DPS strežnika pretoči podatke, jih obdela in vrne v skladišče (*Repository*). Ključne prednosti programa so nadzor nad opravljenim delom, varnost podatkov, večuporabniško delo na daljavo, upravljanje s centralno podatkovno bazo in uporabniškimi dovoljenji.

Za delo z LBS je pomemben predvsem DPS vmesnik (*Entry Editor*), kjer je celotno programsko okno razdeljeno na štiri dele:

- 1. okno** (*Navigation panel*) je namenjeno dostopu do gesel v dodeljenem paketu ter komunikaciji z DPS strežnikom.
- 2. okno** (*Wysiwyg view*) je namenjeno ogledu gesla brez XML elementov v pogledu za branje, pri čemer je mogoče okno tudi urejati.
- 3. okno** (*Tree view*) je namenjeno prikazu drevesne strukture gesla z XML elementi. To je glavno okno, ki je namenjeno urejanju gesla.
- 4. okno** (*Attributes and Annotations*) je namenjeno urejanju atributov pri XML elementih ter vstavljanju redaktorskih opomb k vsebini.

Zgornji del 1. okna prikazuje vsebino, ki je naložena na uporabnikov računalnik – je torej v DPS vmesniku – spodnji del 2. okna pa prikazuje vsebino na DPS strežniku. Hkrati uporabnik tudi vidi, ali so gesla zaklenjena s strani drugega uporabnika. 2. in 3. okno se sproti sinhronizirata, 1., 2. in 4. okno pa je mogoče izklopiti in tako prilagoditi površino DPS vmesnika urejanju gesla. Opisane lastnosti povzema spodnja slika.



Slika 6: DPS vmesnik (Entry Editor).

Zaključek

Leksikalna podatkovna baza za slovenščino, ki nastaja pri projektu *Sporazumevanje v slovenskem jeziku*, je namenjena izdelavi najpomembnejših in hkrati najpotrebnejših slovarskih in jezikovnih priročnikov za slovenščino, kamor sodijo zlasti eno- in dvojezični slovarji, specializirani (frazeološki, kolokacijski, sinonimni ...) in pedagoški slovarji, slovarji/priročniki za učenje slovenščine kot tujega jezika, slovnice, slogovni priročniki itd., ter računalniški obdelavi naravnega jezika. Strukturiranost podatkov v elektronski obliki, upoštevanje leksikalno-gramatičnega pristopa pri opisu LE, izhajanje iz teoretskih in praktičnih spoznanj pri oblikovanju tako slovarjev kot leksikalnih podatkovnih baz evropskih jezikov ter nenazadnje premisleki o vrsti informacij in njihovi notranji organizaciji, ki bi sledili specifičnim lastnostim slovenščine, omogoča po našem mnenju izgradnjo večkrat uporabnega in vedno znova nadgradljivega vira leksikalnih in slovničnih podatkov, ki bo lahko dolgoročno sledil in zadoščal zgoraj izpostavljenim ciljem.

Seznam uporabljenih kratic

CPA (*Corpus Pattern Analysis*) – projekt *Analiza korpusnih vzorcev*

DPS (*Dictionary Production System*) – računalniški sistem za izdelavo slovarjev

DTD (*Document Type Definition*) – definicija tipa dokumenta

EFL-slovarji (*English as Foreign/Second Language*) – slovarji za učenje angleščine kot tujega/drugega jezika

FE – frazeološka enota

FN – projekt: *FrameNet*

gbz ali **GBZ** – glagolska besedna zveza

GDEX (*Good Dictionary Examples*) – računalniška funkcija SkE za avtomatično izbiro slovarskih zgledov

LBS – leksikalna podatkovna baza za slovenščino

LE – leksikalna enota

LPB – leksikalna podatkovna baza

RONJ – računalniška obdelava naravnega jezika

sbz – samostalniška besedna zveza

SkE (*Sketch Engine*) – računalniški program za izdelavo leksikalno-gramatičnih profilov besed

SSJ – projekt *Sporazumevanje v slovenskem jeziku*

TBL (*Tick Box Lexicography*) – klikosikografija³⁰

WSE (*Word Sketch*) – računalniška funkcija SkE za izdelavo besednih skic

Literatura

Arhar, Špela, in Gorjanc, Vojko, 2007: Korpus FidaPLUS: nova generacija slovenskega referenčnega korpusa. *Jezik in slovstvo* 52/2. 95–110.

³⁰ Gre za delovni izraz, ki ga uporabljamo v specifikacijah za izdelavo LBS.

Arhar, Špela, 2007: *Kaj početi z referenčnim korpusom Fidaplus*. Ljubljana: Filozofska fakulteta. <http://www.fidaplus.net/Info/Info_index.html>.

Atkins, Sue B. T., in Rundell, Michael, 2008: *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

BNC – *British National Corpus*: <<http://www.natcorp.ox.ac.uk/>>.

Čermák, František, 2009: Leksikografski zapiski o korpusnem slovarju. *Jezik in slovstvo* 25–42.

FidaPLUS: <<http://www.fidaplus.net/>>.

Fillmore, Charles J., Johnson, Christopher R., in Petruck, Miriam R. L., 2003: Background to Framenet. *International Journal of Lexicography* 16/3. 235–250.

Hanks, Patrick, 1994: Linguistic norms and pragmatic exploitations, or Why lexicographers need prototype theory and vice versa. Kiefer, Ferenc, Kiss, Gabor, in Pajzs, Julia (ur.): *Papers in Computational Lexicography: Complex '94*. Research Institute for Linguistics, Hungarian Academy of Sciences.

Hanks, Patrick, 2009: Sestavljanje enojezičnega slovarja za domače govorce. *Jezik in slovstvo* 7–24.

Kilgarriff, Adam, Richlý, Pavel, Smrz, Pavel, in Tugwell, David, 2004: The Sketch Engine. Williams, G. in Vessier, S. (ur.): *EURALEX 2004 Proceedings*. Lorient: Université de Bretagne-Sud.

Kilgarriff, Adam, 2006a: Word senses. Agirre, E., in Edmonds, P. (ur.): *Word Sense Disambiguation: Algorithms and Applications*. New York: Springer. 29–45.

Kilgarriff, Adam, 2006b: Collocationality (and how to measure it). Corino et. al (ur): *Proceedings of 12th EURALEX International Congress*, EURALEX 2006. Alessandria: Edeizioni Dell'Orso.

Kilgarriff, Adam, Husák Miloš, McAdam, Katy, Rundell, Michael, in Rychlý, Pavel, 2008: GDEX: Automatically Finding Good Dictionary Examples in a Corpus. Bernal, Elisenda, in DeCesaris, Janet (ur.): *Proceedings of the XIII. Euralex international Congress (Barcelona, 15–19 July 2008)*. Barcelona: Institut universitari de lingüística aplicada Universitat Pompeu Fabra.

Kocjančič, Polonca, Krek, Simon, in Climent, Philippe, 2006: Bilingual dictionaries and IDM DPS: the development of a corpus-driven Slovenian-English pocket dictionary and its implementation in the IDM dictionary publishing system. Corino et. al (ur): *Proceedings of 12th EURALEX International Congress*, EURALEX 2006. Alessandria: Edeizioni Dell'Orso. 431–438.

Krek, Simon, in Kilgarriff, Adam, 2006: Slovene Word Sketches. Erjavec, Tomaž, in Žganec Gros, Jerneja (ur.): *Jezikovne tehnologije 5*. Ljubljana: IJS. 62–65.

Krek, Simon, 2008: *Framenet in slovenščina*. *Jezik in slovstvo* 53/5. 37–54.

LDOCE – *Longman Dictionary of Contemporary English*, 2003⁴. Harlow: Pearson Education Ltd.

MEDAL – *Macmillan English Dictionary for Advanced Learners*, 2007². Oxford: Macmillan Publishers Ltd.

Meer, Geart van der, 2006: It's about Time: On Coherence and Simplicity in Dictionary Entries. *English Studies*. Routledge. 602–616.

NODE – *The New Oxford Dictionary of English*, 1998. Oxford: Oxford University Press.

Pustejovsky, James, 1995: *The Generative Lexicon*. Cambridge, Massachusetts: MIT Press.

Sinclair, John McHardy, in Coulthard, Malcolm, 1975. *Towards an Analysis of Discourse: the English Used by Teachers and Pupils*. Oxford: Oxford University Press.

Vidovič Muha, Ada, 2000: *Slovensko leksikalno pomenoslovje – Govorica slovarja*. Ljubljana: ZIFF.

Vodušek, Božo, 1961: O leksikografskem ugotavljanju in urejanju besednih pomenov (Referat na III. jugoslovanskem slavističnem kongresu). *Jezik in slovstvo* 7/1. 5–10.

Spletne strani

Spletni slovarji:

<<http://www.sil.org/linguistics/GlossaryOfLinguisticTerms/Index.htm>>. (Dostop: 6. 7. 2009.)

<<http://www.thefreedictionary.com/lexical+database>>. (Dostop: 6. 7. 2009.)

Projekt SSJ:

<<http://www.slovenscina.eu/Vsebine/En/Dogodki/Slovarji/Program.aspx>>. (Dostop: 6. 7. 2009.)

<<http://www.slovenscina.eu/Vsebine/SI/Domov/Domov.aspx>>. (Dostop: 6. 7. 2009.)

FrameNet:

<http://videolectures.net/solomon_krek_fn>. (Dostop: 6. 7. 2009.)

<<http://www.icsi.berkeley.edu/~loenneke/>>. (Dostop: 6. 7. 2009.)

<<http://framenet.icsi.berkeley.edu/>>. (Dostop: 4. 8. 2009.)

British National Corpus: <<http://www.natcorp.ox.ac.uk/>>. (Dostop: 4. 8. 2009.)

Računalniški sistem za izdelavo slovarjev IDM DPS: <<http://www.idm.fr>>. (Dostop: 6. 7. 2009.)