Latviešu valodas leksiskā tīkla izveides vadlīnijas

Sagatavoja: Laura Rituma, Ilze Lokmane, Agute Klints

1.Vārda nozīmju nošķiršana pētījuma ietvaros 2
1.1. Projektā izvēlētie vārda nozīmju nošķiršanas kritēriji 3
1.1.1.Semantisko elementu analīze 3
1.1.2.Semantiskās saistāmības analīze 4
1.1.3.Sintaktiskās saistāmības analīze 4
1.1.4.Iespēja aizstāt ar sinonīmu vai ekvivalentu citā valodā 5
1.1.5.Nozīmes nianšu izpratne 6
1.2. Vārda skaidrojuma izveide 9
1.3. Vārda gramatisko pazīmju noteikšana 11
1.4. Šķirkļa vārdu nošķiršana atsevišķā šķirklī 11
1.4.1. Substantivējušies vārdi 12
1.4.2.Divdabja formas šķirkļa vārdi 13
1.4.3. Skaitļa vārdu kategorija un robežgadījumi 14
1.4.4. Vārds lietots citas vārdšķiras nozīmē 14
1.5. Vārda nozīmju rediģēšana un piemēru atlase online rīkā wordnet.tezaurs.lv 14
1.5.1. Nozīmes piemēru atlase un pievienošana 17
1.5.2. Vārda lietojuma ierobežojuma pazīmes 18

2. Leksiskā tīkla saišu veidošana 18
2.1. Marķējamo vārdu saraksta veidošana 20
2.2. Sinonīmu kopas 20
2.3. Semantisko attieksmju/saišu veidi 21
2.3.1. Sinonīmija 21
2.3.2. Hiponīmija - virsjēdziens un apakšjēdzieni 21
2.3.3. Meronīmija un tās izpratne 22
2.3.4. Aptuvena sinonīmija 22
2.3.5. “Saistīts ar” 22
2.4. Sastatīšana ar Princeton WordNet 22

3.Rīks leksiskā tīkla veidošanai 24

Literatūras saraksts: 30

1.Vārda nozīmju nošķiršana pētījuma ietvaros

Latviešu valodas leksiskā tīkla izveidē galvenais resurss ir Tēzaurs - skaidrojošā vārdnīca, kurā apkopots līdzšinējo skaidrojošo vārdnīcu un citu resursu materiāls. Četri primārie avoti Tēzaurā ir "Latviešu valodas skaidrojošā vārdnīca" (1972-1996, turpmāk tekstā - LLVV), K. Mīlenbaha "Latviešu valodas vārdnīca" (1923-1932) un tās papildinājumi (1934-1946), kā arī "Latviešu konversācijas vārdnīca" (1927-1940). Biežāk lietoto vārdu skaidrojums pamatā mantots no LLVV, tātad veidots pagājušā gadsimta otrajā pusē. LLVV veidojis autoru kolektīvs, šķirkļi sadalīti starp autoriem pēc alfabētiskā, nevis semantisko grupu principa, turklāt pirmā sējuma izdošanu no pēdējā šķir krietns laika sprīdis, tāpēc nozīmju nošķiršanas kritēriji un skaidrojumu izveide nav īstenota konsekventi. Kā pētījumā par nozīmīgākajām vienvalodas vārdnīcām konstatējis Juris Baldunčiks, jau pirmā LLVV sējuma recenzenti norādījuši, ka nozīmes reizēm skaidrotas pārāk sīki, reizēm pietrūkst vispārīgāka skaidrojuma. Dažkārt konstatējami lieki nozīmes komponenti, bet citkārt kāda būtiska komponenta trūkst (Baldunčiks 2012, 118).

Kopš LLVV izdošanas vārdu nozīmes var būt arī mainījušās - kāda nozīme nākusi klāt, kāda kļuvusi vispārīgāka, izveidojies jauns nozīmes pārnesums. Daļu no šīm pārmaiņām ir fiksējuši "Mūsdienu latviešu valodas vārdnīcas" (2003-2019, turpmāk tekstā - MLVV) veidotāji, tomēr, pētot MLVV iekļauto materiālu, nākas secināt, ka arī tur par pamatu daudzreiz izmantots LLVV nozīmju dalījums. Daudznozīmīgiem vārdiem nozīmju skaitu mēģināts mazināt, piemēram, verbam *iet* LLVV ir dotas 15 nozīmes, bet MLVV - 10, pakārtojuma saiklim *ka* LLVV ir 10, bet MLVV - 5 nozīmes. Kā norāda vārdnīcas autori, nozīmju skaita samazinājums nenozīmē, ka līdz ar to būtu nepilnīgi atklāta attiecīgā vārda semantiskā struktūra. Atbilstoši vispārinot skaidrojumu, var panākt, ka tas aptver lielāku attiecīgās leksēmas semantiskās struktūras daļu (Zuicena 2010, 370).

Minēto iemeslu dēļ projekta sākumā tika izvirzīts mērķis pārskatīt un revidēt 5000 vārdu nozīmes. Lai noprastu, cik liela būtu atšķirība, ja nozīmju sadalījumu veidotu pilnībā no jauna, tika veikts eksperiments, kurā desmit daudznozīmīgiem vārdiem tika izveidots nozīmju dalījums, balstoties tikai uz lietojumu Korpusā un *LR 5.-12. Saeimas sēžu stenogrammu korpusā* (turpmāk tekstā - Saeimas korpuss). Divi valodnieki (tabulā V1 un V2) neatkarīgi viens no otra noteica savu nozīmju šķirumu, pēc tam salīdzinot vienojās par galarezultātu, tomēr nozīmju skaits vairākos gadījumos krietni atšķīrās no LLVV un MLVV dotā (skat.1. tabulu).

Vārds	V1	V2	Kopīgi nolemtais nozīmju skaits	Nozīmju skaits Tēzaurā	Nozīmju skaits MLVV	Nozīmju skaita starpība attieksmē pret Tēzauru	Nozīmju skaita starpība attieksmē pret MLVV
domāt	7	6	**6**	5	4	+1	+2
krist	8	8	**8**	24	18	-16	-10
gāzties	4	3	**4**	4	4	0	0
spēlēt	4	3	**4**	4	3	0	+1
dot	5	4	**4**	10	9	-6	-5
skriet	6	4	**5**	8	7	-3	-2

kristies	3	2	**2**	2	2	0	0
aizvērt	4	3	**4**	2	1	+2	+3
ass	7	5	**9**	9	7	0	+2
asums	7	5	**10**	11	3	-1	-7

1. tabula. Lingvistiskā eksperimenta rezultāti

Tika secināts, ka projektā iesaistītie valodnieki pēc korpusa piemēriem četriem vārdiem izšķīruši mazāk nozīmju nekā dotajās vārdnīcās, diviem - vairāk nozīmju, diviem - tikpat nozīmju, bet pārējiem diviem - tikpat cik Tēzaurā, bet vairāk nekā MLVV. Paredzamais darbs veidot nozīmju šķīrumu pilnībā no jauna, balstoties tikai uz lietojumu korpusā, būtu ļoti laikietilpīgs, un tas nav projekta pamatmērķis, turklāt tas daļēji pārklātos ar Latviešu valodas institūta jau veikto darbu, tāpēc kopumā tika nolemts izmantot Tēzaurā esošo nozīmju dalījumu kā pamatu, kas tiktu pārstrādāts pēc vienotiem nozīmju šķiršanas kritērijiem. Vienlaicīgi MLVV tiek izmantots kā būtisks papildu avots, lai salīdzinātu un ņemtu vērā leksikogrāfu paveikto mūsdienās.
Ir saprotams, ka korpusos parādīsies biežāk lietotās nozīmes, bet reti lietotām nozīmēm piemēri korpusā var nebūt. Tādā gadījumā, valodniekam jāvar uzrakstīt piemēru kaut vai pašiem.
Turpmākās apakšnodaļās tiks aprakstīts, kādi kritēriji izvēlēti nozīmju dalījuma pārskatīšanai projekta gaitā (1.1. nodaļa), vārda skaidrojuma izveides būtiskākie aspekti (1.2. nodaļa), dažādi vārda lietojuma aspekti, kas parādās vārdnīcā - vārdu gramatiskās pazīmes, šķiršana atsevišķos šķirkļos, (1.3. un 1.4. nodaļas). Tāpat sniegtas praktiskās vadlīnijas nozīmju rediģēšanai un piemēru atlasei un pievienošanai tiešsaistes rīkā (1.5. nodaļa).

1.1. Projektā izvēlētie vārda nozīmju nošķiršanas kritēriji

“Latviešu valodas leksiskā tīkla” izveidē par lietderīgiem verbu nozīmju nošķiršanā izvirzīti šādi kritēriji:

- semantisko elementu analīze (sk., piem., Cruse 2004, 239–254),
- saistāmības analīze – tiek ņemta vērā gan semantiskā, gan sintaktiskā saistāmība (sk., piem., Freimane 1983, 34–57),
- iespēja aizstāt ar sinonīmu (retāk ekvivalentu citā valodā) (sk., piem., Jackson 2002, 94).

Turpmākajās apakšnodaļās tikai pieminēsim katras metodes priekšrocības un trūkumus.

1.1.1.Semantisko elementu analīze

Ar šīs metodes izmantojumu saistītas vairākas problēmas. Pirmkārt, problemātiska ir semantisko elementu noteikšana, ņemot vērā, ka tiem var būt dažādas vispārinājuma pakāpes (Cruse 2004, 254–257). Vislabāk semantiskos elementus var noteikt, salīdzinot, piemēram, divu vārdu nozīmes vai viena vārda lietojumu dažādos kontekstos.
Otrkārt, problēmas rada semantisko elementu nosaukšana, jo jāizmanto dabiskās valodas vārdi un no vārdu izvēles var būt atkarīga arī semantisko elementu identificēšana. Šo problēmu cenšas risināt, izvēloties ierobežotu skaitu vārdu, kurus izmanto citu vārdu nozīmju skaidrošanai (vairāk par to sk., piem., Wierzbicka 1996, Goddard 1998). Tomēr tāda semantisko elementu inventāra, kas būtu izmantojams visu kādas valodas vārdu skaidrošanai, nav, un jādomā, ka tāds nav iespējams vai arī tas būtu pārāk apjomīgs, lai būtu ērti izmantojams.

Treškārt, semantisko elementu skaits nav galīgs; praksē katrs pētnieks izvirza sava pētījuma mērķim atbilstošu semantisko elementu komplektu. Tomēr leksikogrāfa darbā un arī elektronisko resursu veidošanā šāda pieeja nav atrisinājums, jo jāaptver visa valodas leksika.
Ceturtkārt, pat ja detalizēta semantisko elementu analīze ir iespējama, tomēr nav konkrēti nosakāms, cik un kādiem semantiskajiem elementiem jāatšķiras, lai vārdnīcā fiksētu dažādas nozīmes. Šeit konsekvents risinājums nav iespējams, un leksikogrāfa darbā parasti tiek izmantota intuīcija, lai noteiktu, kuri semantiskie elementi ir tik būtiski, ka to maiņa rada jaunu nozīmi. Ja katru gadījumu, kad kāds semantiskais elements atšķiras, uzskatītu par jaunu nozīmi, iegūtu pārāk sīku nozīmju dalījumu.

1.1.2.Semantiskās saistāmības analīze

Verbu semantiskā saistāmība ietver aprakstītās situācijas dalībnieku jeb argumentu semantiskās lomas (piemēram, darītājs, cietējs, izjutējs, ieguvējs, adresāts, līdzeklis) un vispārīgās semantiskās pazīmes (piemēram, dzīvs / nedzīvs, persona / ne-persona, abstrakts / konkrēts, skaitāms / neskaitāms). Līdzīgi kā semantisko elementu skaits, arī semantisko lomu skaits nav galīgs un dažādās teorijās un dažādos pētījumos lieto atšķirīgus semantisko lomu kopumus (Saeed 2000, 149–178).
Tomēr galvenā problēma šīs metodes izmantošanā ir tā, ka nav precīzi nosakāms, kuras semantiskās lomas un semantiskās pazīmes ir pietiekami būtiskas, lai tās ņemtu vērā nozīmju nošķiršanā, piemēram, vai semantiskā opozīcija cilvēks / citas dzīvas būtnes ļauj atšķirt nozīmes vai ne. Šādu semantisko nianšu ir daudz, un, ja nozīmes analīzē mēģinātu aprakstīt katru vārda saistījumu ar noteiktām semantiskām lomām un vispārinātām apkaimes elementu semantiskajām pazīmēm, tad nozīmju un to nianšu būtu tik daudz, ka vārda šķirklis būtu ļoti plašs, nepārredzams un grūti uztverams. Turklāt tas būtu pretrunā ar kognitīvās lingvistikas pieņēmumiem, ka cilvēkam pietiek zināt kategorijas tipiskos locekļus, lai spētu atpazīt pārējos. Skaidrojošās vārdnīcas mērķis būtu sniegt galvenās nozīmes un nozīmju nianses, kas ļauj saprast vārda nozīmi konkrētā lietojumā.

1.1.3.Sintaktiskās saistāmības analīze

Sintaktiskā saistāmība rāda, kādās sintaktiskās konstrukcijās vārds iesaistās, piemēram, vai tam iespējams saistījums ar tiešo vai netiešo objektu, noteiktas semantikas apstākļiem u.c. Te jāņem vērā ne tikai formālais aspekts (piemēram, kāds ir objekta locījums vai kāds saiklis ievada sekojošu palīgteikumu), bet arī gramatiskā semantika – piemēram, vai teikuma priekšmeta pozīcijā ir darītāja nosaukums vai kāda cita semantiskā loma. Tātad sintaktiskā saistāmība tiek analizēta kompleksā ar iepriekš aplūkoto semantisko saistāmību. Darbības vārdu nozīmju nošķiršanā nolemts par būtiskām atzīt gramatiskā subjekta un gramatiskā objekta semantiskās lomas un semantiskās pazīmes (dzīvs / nedzīvs, cilvēks / citas dzīves būtnes, konkrēts / abstrakts), kas var būt par pamatu nozīmju vai to nianšu šķiršanai. Dažkārt ņemtas vērā arī konkrētākas semantiskās pazīmes.
Sintaktiskā saistāmība var būt īpaši svarīga atsevišķu verbu nozīmju nošķiršanā. Piemēram, verba *būt* nozīmes vārdnīcā tiek nošķirtas, balstoties galvenokārt uz konstrukcijām, kurās tas ietilpst. Tas ir saprotams, jo *būt* ir visvairāk desemantizējies verbs un kalpo galvenokārt

gramatisko nozīmju (laika, personas, modalitātes) izteikšanai, leksiskā semantika ir ļoti vispārināta un bez sintakses atbalsta te nav iespējams iztikt.
Viens no sintaktiskās saistāmības aspektiem, kas tradicionāli tiek rādīts vārdnīcas šķirklī, ir pārejamība (transitivitāte) vai nepārejamība (intransivitāte). Tā parāda, vai darbības vārdam ir iespējams saistījums ar tiešo objektu akuzatīvā (transitīvi verbi), vai tāda nav (intransitīvi verbi). Daudznozīmības sakarā svarīgi, ka vienam un tam pašam verbam var būt transitīvas un intransitīvas nozīmes. Šis aspekts kā viens no nozīmju šķīruma kritērijiem aplūkots arī leksikogrāfijas teorijā (sk., piem., Ozola 1970). Tomēr secināts, ka ne vienmēr iespējams vilkt skaidras robežas starp transitīvu un intransitīvu lietojumu, tāpēc šī pazīme izmantojama tikai kompleksā ar citām (par to tuvāk skat. apakšnodaļā par nozīmes nianšu izpratni).
Kaut arī sintaktiskās saistāmības formālais aspekts varētu būt objektīvs (jo ir tieši novērojams) un samērā drošs nozīmju šķiršanas kritērijs, tomēr jāatzīst, ka dažkārt divas dažādas nozīmes var realizēties vienā un tajā pašā sintaktiskajā konstrukcijā, un otrādi – viena nozīme dažādās konstrukcijās.

1.1.4.Iespēja aizstāt ar sinonīmu vai ekvivalentu citā valodā

Tā kā aizstāšana ar sinonīmu ir tradicionāls un plaši lietots paņēmiens nozīmes skaidrošanā, to var izmantot arī nozīmju vai to nianšu nošķiršanā. Piemēram, vārds *spēlēt* ir aizstājams ar *atskaņot* saistījumā ar skaņdarba vai mūzikas nosaukumu (spēlēt / atskaņot skaņdarbu, mūziku), bet ne saistījumā ar mūzikas instrumenta nosaukumu (spēlēt vijoli, bet ne atskaņot vijoli). Tas ir pietiekams pamats nozīmes nianses šķiršanai. Arī hiponīmiskas attieksmes ar vārdiem *trinkšķināt*, *čīgāt* un citiem instrumentu spēlēšanas veidu nosaukumiem veidosies tikai šai vienai nozīmes niansei.
Ņemot vērā "Latviešu valodas leksiskā tīkla" struktūru un nepieciešamību fiksēt jēdzieniskās attieksmes starp vārdiem un to nozīmēm, sinonīmijas jēdziens iegūst īpašu nozīmību, un nozīmju skaidrojumos izmantotie sinonīmi var tieši norādīt uz sinonīmu kopām. Tomēr jāpatur prātā, ka valodniecībā nav vienotas sinonīmu izpratnes. Tradicionāli ar sinonīmiem saprot vārdus, kam ir "vienādas vai tuvas nozīmes" (Skujiņa 2007, 354), tomēr pēc šāda skaidrojuma sinonīmi ir grūti identificējami. Mēģinājumi definīciju precizēt, piemēram, sinonīmi ir "vārdi (resp., vārdu leksiski semantiskie varianti) ar vienādiem nozīmes galvenajiem komponentiem" (Veidemane 1970, 36) vai "vārdi, kuru semantiskā līdzība ir ievērojamāka par semantiskajām atšķirībām" (Cruse 2004, 154), tāpat nedod skaidrus kritērijus sinonīmu noteikšanā. Sinonīmu identificēšanā iespējams izmantot arī aizstāšanas metodi, bet šādi iespējams identificēt tikai absolūtos sinonīmus, kam ir vienādas nozīmes un kas savstarpēji nomaināmi jebkurā kontekstā. Pārējos gadījumos aizstāšana stilsitisku vai jēdzienisku nianšu dēļ var būt iespējama tikai dažos kontekstos. Tāpēc jāatzīst, ka sinonīmija būtībā ir relatīvs jēdziens – vārdu nozīmes var būt vairāk vai mazāk sinonīmiskas, tajās var būt vairāk vai mazāk kopīgā.
Ekvivalentu salīdzināšanu citās valodās var izmantot tikai kā palīgmetodi, piemēram, kad jāizvēlas starp vairākiem nozīmju dalījuma variantiem.

1.1.5.Nozīmes nianšu izpratne

Vārda nozīmes nianse ir diezgan neskaidrs jēdziens, kaut gan leksikogrāfijas praksē tiek plaši izmantots. "Valodniecības pamatterminu skaidrojošajā vārdnīcā" dota šāda definīcija: "No vārda

pamatnozīmes vai sekundārās nozīmes nedaudz atšķirīgs jēdzieniskais saturs, kas neveido patstāvīgu leksiski semantisko variantu, piemēram, vārda *ceļš* pamatnozīmes ‘speciāli izveidota zemes strēle sauszemes transportlīdzekļu, arī gājēju kustībai’ nianse ‘vieta (upē, ezerā, jūrā) ūdens transportlīdzekļu kustībai’.” (Skujiņa 2007, 263) Definīcijā minētais “nedaudz” ir grūti definējams.

Praksē latviešu leksikogrāfi ir rīkojušies dažādi. Ne 1987. gadā izdotajā “Latviešu valodas vārdnīcā”, ne labotajā un pārstrādātajā šīs vārdnīcas versijā, kas iznāk 2006. gadā, nozīmes nianses vārdu nozīmju skaidrojumos atsevišķi nav izdalītas, ievadā norādīts vienīgi, ka “nozīmes nianses skaidrojumā parādītas aiz semikola” (Bāliņa 2006, 10). Pats nozīmes nianses jēdziens nav skaidrots.

Savukārt “Latviešu literārās valodas vārdnīcas” priekšvārdā teikts: “Par nozīmes niansi uzskatīta relatīvi patstāvīga nozīmes sastāvdaļa. Pamats nozīmes nianses izdalīšanai parasti ir sīkākas semantiskas atšķirības, kas attiecināmas uz vienu un to pašu nozīmi, kā arī vārda gramatiski funkcionālās īpatnības. Piemēram, verba aizplūst 1. nozīmei Plūstot attālināties (piemēram, par šķidrumu, straumi) viena no niansēm ir // Izplatīties (par skaņām). Rādot nozīmes nianses, vārdnīcā iespējams veidot detalizētāku pārskatu par nozīmi, sīkāk raksturot vārda lietošanas nosacījumus.” (Priekšvārds 1972, 11)

Elektroniskās “Mūsdienu latviešu valodas vārdnīcas” šķirkļos nozīmes nianses, sekojot “Latviešu literārās valodas vārdnīcas” paraugam, tiek rādītas, bet vārdnīcas instrukcijā nianses jēdziens nav skaidrots: “Nozīmju skaidrošanai MLVV izmantots trīspakāpju semantiskais dalījums, kas pirmo reizi latviešu leksikogrāfijā lietots „Latviešu literārās valodas vārdnīcā”. Vārda nozīmes daudznozīmju vārdiem apzīmētas ar arābu cipariem, bet sīkākas semantiskās atšķirības (nozīmes nianses) – ar paralēļu zīmi (//) vai semikolu. Paralēļu zīme likta, ja nozīmes nianse no nozīmes atšķiras diezgan ievērojami, bet semikols lietots gadījumos, kad atšķirības ir nelielas.” (https://tezaurs.lv/mlvv/#/instrukcija) Tātad nianšu nošķiršanas kritēriji nekur nav skaidri definēti, un vienīgi no šķirkļu struktūras var censties secināt, kādas semantiskas vai gramatiskas īpatnības bijušas tās pamatā.

Līdz šim Tēzaurā iekļautās nozīmes nianses pārņemtas no LLVV, tās var būt ļoti dažādas un pēc attieksmes pret virsnozīmi dalāmas tādos pašos tipos kā atvasinātās nozīmes attieksmē pret pamatnozīmi. Pirmkārt, nianse no virsnozīmes var atšķirties ar kādu semantisko elementu. Piemēram, verbam *uztvert* ir 3. nozīme ‘satvert (ko krītošu, mestu u. tml.)’ ar nozīmes niansi ‘satvert un atsist (volejbola, tenisa u. tml. spēlē bumbas servi, gremdi, piespēli)’ (LLVV), tātad pievienots semantiskais elements ‘atsist’.

Otrkārt, nianse no virsnozīmes var atšķirties ar semantisko saistāmību, proti, situācijas dalībnieku semantiskajām lomām vai semantiskajām grupām, pie kurām tie pieder. Piemēram, jau minētajai verba *uztvert* 3. nozīmei ‘satvert (ko krītošu, mestu u. tml.)’ ir arī otra nozīmes nianse ‘būt tādam, kurā (kas) tiek satverts, aizturēts u. tml., (tam) krītot, plūstot u. tml.’ (LLVV), tātad virsnozīme paredz dzīvu darītāju, savukārt nozīmes nianse - nedzīvu. Līdzīgi verbam *rakt* 1.nozīme ‘durt, parasti lāpstu, zemē un celt, pārvietot ar to zemi (par cilvēkiem)’ attiecas uz cilvēku kā darbības veicēju, savukārt nozīmes nianses atklāj iespējamus citus darbības veicējus - iekārtas niansē ‘par iekārtām, ierīcēm, mašīnām ar lāpstveida darbīgajām daļām’ vai dzīvniekus otrā nozīmes niansē ‘ar kājām vai purnu raust zemi (piemēram, ko meklējot) – par dzīvniekiem’ (LLVV). Vienlaikus var atšķirties arī līdzekļa semantika - cilvēki rok ar lāpstu, bet dzīvnieki - ar purnu vai kājām. Gadījumos, kad dzīvs darītājs aizstāts ar nedzīvu, varam runāt par metaforisku nozīmes pārnesumu, tomēr minētais pārnesums - iekārtas vai priekšmeti

darītāja lomā - ir visai ierasts, tāpēc tēlainību runātāji vairs neuztver un nozīmes nianses ir tiešas, nevis pārnestas. Dažos gadījumos līdzīgs nozīmes pārnesums kvalificējams kā tēlains, tādējādi nozīmes nianse atšķirībā no virsnozīmes ir pārnesta, piemēram, verbam *staigāt* viena no 1. nozīmes 'vairākkārt vai ilgstoši ejot, pārvietoties (parasti dažādos virzienos) – par cilvēku' niansēm ir pārnesta – 'virzīties (par parādībām dabā, parasti vēju, mākoņiem)' (LLVV). Tēlainības saglabāšanās saistāma ar retāku dabas parādību nosaukumu lietojumu darītāja pozīcijā.

Subjekta pārnesumā balstītu metonīmiju vajadzētu nošķirt kā atsevišķu nozīmes niansi. Piemēram, vārda *skaidrot* pamatnozīmes subjekts var būt cilvēks, nozīmes nianses subjekts ir cilvēka radīts informatīvs materiāls, kas satur informāciju par skaidrojamo reāliju.

Līdzīgā veidā var atšķirties arī cietēju semantiskās grupas. Piemēram, verbam *durt* ir pamatnozīme 'virzīt (kur iekšā), parasti ar piepūli (ko asu, smailu)' (LLVV), piemēram, durt mietu zemē. Šīs nozīmes niansei 'virzīt (ko) virsū (kam tievam, smailam)' cietējs ir priekšmets, kuru uzdur, piemēram, durt desu uz iesma, savukārt niansei 'veidot (ar adatu, īlenu u. tml.) dūrienus, caurumus' cietējs ir priekšmets, kas duršanas rezultātā rodas.

Projekta gaitā nolemts kā atsevišķu nozīmes niansi no kādas institūcijas skaidrojuma nodalīt ēku, kurā darbojas attiecīgā institūcija, piemēram, šķirkļos *pašvaldība, banka*. Nianse nodalīta, ja lietojuma piemēros atrodami vismaz 5 piemēri ar vārda lietojumu ēkas nozīmē.

Treškārt, nianse no virsnozīmes var atšķirties ar sintaktisko saistāmību, piemēram, verbam rakt pamatnozīme ir uzskatīta par intransitīvu, jo iespējams lietojums bez tiešā objekta, piemēram, Sapratu, ka jārok dziļāk! (LVK2018). Divas šīs pašas nozīmes nianses LLVV kvalificētas kā transitīvas – viena bez jebkāda papildu semantiska skaidrojuma, acīmredzot, tāpēc, ka verbu var lietot arī ar tiešo objektu, piemēram, Kārlis raka siltumnīcā zemi (LVK2018). Otrai transitīvajai nozīmei dots arī savs skaidrojums 'durt lāpstu zemē un vērst augsnes virskārtu (piemēram, ko veidojot augsnes virskārtā)' (LLVV), piemēram, rakt puķu dobi.

Par nozīmes niansi nebūtu jāuzskata transitīva verba lietojums bez tiešā objekta, ja tiešais objekts ir noprotams no konteksta vai situācijas, vai arī ir tik vispārīgs, ka netiek nosaukts. Piemēram, vārdam dzert sniegta transitīva pamatnozīme 'ieņemt mutē un norīt (šķidrumu)' (LLVV) un intransitīva nozīmes nianse, kuru ilustrē tikai piemēri: Dzert gribi? (LVK2018). Vienīgi tad, ja intransitīvā lietojumā verba nozīmei pievienojas kāds jauns semantiskais elements, ir pamats nozīmes niansi šķirt, piemēram, verbam *lasīt* ir transitīva pamatnozīme 'uztvert, saprast rakstu valodā (tekstu); uztverot, saprotot rakstu valodā (tekstu), runāt (to klausītājiem)' un intransitīva nianse 'būt tādam, kam piemīt spēja, prasme uztvert, saprast rakstu valodā (tekstu)' (Tēzaurs). Šajā gadījumā papildus saistāmības atšķirībai nianse ieguvusi semantisko elementu 'prast'. Arī citos gadījumos formālā sintaktiskā saistāmība pati par sevi nav pietiekams pamats šķirt nozīmes nianses, ja tas nav saistīts ar semantiskām atšķirībām.

Gan uz sintaktiskās, gan semantiskās saistāmības atšķirībām (precīzāk – uz semantiskās un sintaktiskās saistāmības savstarpējām attieksmēm) būtu attiecināmi gadījumi, kad ar vienu un to pašu verbu dažādos lietojumos tā pati situācija tiek atspoguļota no dažādiem skatu punktiem. Šādu gadījumu raksturošanai valodniecībā lieto terminu diatēze, kas nozīmē, ka situācijas dalībnieki paliek tie paši, bet mainās to sintaktiskais statuss. Piemēram, rakšanas darbībā ir iesaistīts gan darītājs (cilvēks rok), gan līdzeklis (rakt ar lāpstu), gan divu dažādu veidu cietēji - pārvietojamais (rakt zemi) un radāmais (rakt bedri). Sintaktiski var realizēt tikai vienu no tiem, bet situācija kopumā nemainās. Sintaktiskā subjekta pozīcijā var būt arī līdzeklis (lāpsta labi rok, ekskavators rok). Dažādi diatēzes gadījumi semantikas pētījumos aplūkoti visai plaši

(Paducheva 2004, 51–79) un dalīti tipos, kas katrā valodā ir mazliet atšķirīgi. Citās semantikas teorijās šādi viena verba nozīmes pārnesumi ir raksturoti kā metonīmiski (Pustejovsky 1998, 31–33), savukārt kognitīvajā semantikā šis process raksturots kā perspektīvas maiņa vai profilēšana (Saeed 2000, 328–330).

Lai arī kā šī parādība tiktu interpretēta semantikā, leksikogrāfijā joprojām aktuāls paliek jautājums, vai šādas lietojuma atšķirības vārdnīcā vispār būtu jāparāda (jo atšķiras tikai saistāmība, bet ne semantiskie elementi, kas ietilpst vārda nozīmē) un, ja jā, tad vai tās ir atsevišķas nozīmes vai vienas nozīmes nianses. "Latviešu valodas leksiskā tīkla" veidošanā šī problēma pilnībā nav atrisināma, jo tam vajadzīgs plašāks diatēzes tipu apkopojums visā valodas sistēmā, nevis tikai biežāk lietoto verbu grupā. Tomēr gadījumos, kad semantiskie elementi nemainās, autori sliecas šīs atšķirības kvalificēt kā nozīmes nianses, nevis dažādas nozīmes. Kaut arī šādi metonīmiski pārnesumi nereti ir ļoti regulāri un valodas lietotājiem vispārzināmi, tomēr precīzu informāciju, vai konkrētam vārdam tādu pārnesumu lieto vai nelieto, iespējams uzzināt tieši vārdnīcā.

No aplūkotajiem piemēriem redzams arī, ka LLVV nozīmes nianses tiek veidotas tā, ka tās var tieši neatbilst galvenajai nozīmei – nozīme var iekļauties pamatnozīmes aprakstā, bet var arī atšķirties ar kādu semantisku elementu, nonākot pretrunā ar pamatnozīmē ietverto informāciju, kas vārdnīcas lietotāju var mulsināt. "Latviešu valodas leksiskā tīkla" veidotāji, pārskatot nozīmes, cenšas pārveidot un vispārināt pamatnozīmes raksturojumu tā, lai tas būtu vispārīgāks un atbilstu nozīmes niansēm, nenonākot pretrunā ar tām.

Jāpiebilst, ka nozīmes nianšu skaits var būt atkarīgs no vārda nozīmju skaita vispār. Protams, nepieciešamība veidot pārskatāmu un ērti uztveramu šķirkli var radīt vēlmi īpaši daudznozīmīgu vārdu sīkās nozīmes nianses vārdnīcā nešķirt vai samazināt. Tomēr, ja vēlamies īstenot iespējami konsekventu pieeju, tad kritēriji vienādā mērā jāatiecina uz visiem vārdiem. Arī projekta mērķis veidot sinonīmu kopas liek izmantot vienādus principus līdzīgas nozīmes vārdu nianšu izšķiršanā. Piemēram, vārdam *domāt* viena no daudzajām nozīmēm ir nozīme 'rūpēties, gādāt (par kādu, par vēlamu rezultātu)' (LLVV) bez kādām nozīmes niansēm. Savukārt vārdam *rūpēties* ir šķirtas divas nozīmes - atkarībā no tā, vai rūpju objekts ir cilvēks vai vēlamais rezultāts, bet verbam *gādāt* šī nozīme tiek attiecināta tikai uz cilvēku kā objektu. Lai izveidotu starp šīm nozīmēm atbilstošas sinonīmu kopas, būtu nepieciešams arī vārdam domāt šķirt vismaz nianses atkarībā no rūpju objekta.

Tātad var secināt, ka kritēriji, kurus izmanto gan atsevišķu nozīmju, gan nozīmes nianšu nošķiršanā, ir vieni un tie paši – nozīmi veidojošie semantiskie elementi, semantiskā saistāmība ar dažādu semantisko lomu un dažādu semantisko grupu argumentiem, sintaktiskā saistāmība, kā arī semantiskās un sintaktiskās saistāmības savstarpējās attieksmes. Tādēļ nav iespējams droši šķirt atsevišķu nozīmi no nozīmes nianses, un tas ir atkarīgs no konkrētā leksikogrāfiskā avota autoru koncepcijas.

1.2. Vārda skaidrojuma izveide

Skaidrojumu projekta laikā nāksies veidot no jauna tad, ja a) ir izveidojusies jauna, vārdnīcās neminēta nozīme, kas parādās korpusos; b) nepieciešams apvienot vairākas vārdnīcās minētās nozīmes, jo tās ir pārāk sīki sadalītas.

Skaidrojumā būtu jāizmanto vienkārši, saprotami, viennozīmīgi vārdi, kurus viegli tulkot. Būtu jāizvairās no zinātniskiem skaidrojumiem, piemēram, 'kaza - pārnadžu kārtas dzīvnieks [..]'

(Tēzaurs). Specifiskus konkrētas jomas terminus pēc iespējas skaidrot ar vienkāršiem vārdiem, bet tiem pieļaujami arī sarežģītāki skaidrojumi.

Lai veidotu WordNet sastatījumu, jaunizveidotā skaidrojuma beigās jāpieraksta arī tās nozīmes, kas LLVV bijušas pārāk acīmredzamas, nav skaidrotas vai ir skaidrotas ļoti vispārīgi ar bultiņu uz citu šķirkļavārdu vai nozīmi. Iespējams, esošās shematiskās definīcijas būs jāizmanto, lai automātiski iegūtu potenciālos atvasinājuma vai cita veida saišu kandidātus, un pēc tam jālabo par pilnvērtīgiem skaidrojumiem, lai lietotājam nav jāpārslēdzas uz citu šķirkli, lai uzzinātu esošā šķirkļa nozīmes pamatu.

Praktiskajā leksikogrāfijā pastāv vairāki nosacījumi, kā ir veidojami skaidrojumi jeb definīcijas skaidrojošajās vārdnīcās, kas paredzētas plašam lietotāju lokam: 1) skaidrot vārda (šajā gadījumā - termina) nozīmi, nevis jēdzienu; 2) skaidrojumam jābūt pēc iespējas īsam, vienā teikumā; 3) skaidrojumam jābūt vienkāršam un saprotamam (leksikogrāfam jāspēj vienkārši un saprotami izskaidrot arī ļoti sarežģīta termina nozīmi); 4) skaidrojumam jābūt pareizam; 5) skaidrojumam jābūt objektīvam, t.i., leksikogrāfam skaidrojumā nevajadzētu paust savu attieksmi. (Zuicena, 2008).

Pētījumā par LLVV tiek norādīts, ka jau pirmā sējuma recenzenti norādīja, ka nozīmes reizēm skaidrotas pārāk sīki, reizēm pārāk vispārīgi. Jau toreiz recenzenti norādījuši, ka shematiskā nozīmju attēlošana vārdu atvasinājumiem (piemēram, “smalkums” nozīmes pēc numura saistītas ar “smalks” nozīmēm pēc numura u.tml.) ir pārāk mehāniska un neatvieglo materiāla uztveri. Dažos skaidrojumos arī minēts kāds lieks nozīmes komponents vai trūkst kāda būtiska nozīmes komponenta, piemēram, vārda *šampūns* skaidrojumā trūkst norādes, ka tas ir matu mazgāšanai. (Baldunčiks, 2012, 118). Tiešsaistes versija Tezaurs.lv šīs problēmas risina, shematiskā skaidrojuma gadījumos piedāvājot saiti uz norādīto vārdu, tomēr arī šajā gadījumā jāspēj orientēties nozīmes numuros, nesajaukt; ja kādas vārdnīcas skaidrojums licies pilnīgāks par LLVV piedāvāto, tad vārdnīcā Tēzaurs.lv lietots tas.

Pētījumā par LVV (Latviešu valodas vārdnīca, izdota divreiz, otrais izdevums Avotam 2006.gadā) norādītas galvenās nepilnības vārdu skaidrojumos:

1) Nesaskaņoti skaidrojumi radniecīgu (savstarpēji saistītu) vārdu grupās,
2) Būtiskas pazīmes trūkums,
3) Liekvārdība,
4) Skaidrojumu papildinoša piemēra trūkums,
5) Neizskaidrota vārda vai vārda nozīmes lietošana cita vārda skaidrojumā,
6) Novecojusi informācija,
7) Pilnīgi vai daļēji aplams skaidrojums.

Pirmie četri skaidrojami ar vāju definēšanas prasmi un paviršību, pēdējie divi ar nekompetenci kādā jomā. Visi šie trūkumi atrodami arī citās skaidrojošajās vārdnīcās. Gan LVV, gan LLVV visbiežāk nepilnības vērojamas sporta, jūrniecības un militāras sfēras leksikas vienību traktējumā. (Baldunčiks, 2012, 135)

Rakstā par MLVV nozīmju skaidrojumiem tiek norādīts, ka autori ir centušies izvairīties no pārāk sīka nozīmju dalījuma (Zuicena, 2010). Piemēram, verbam *iet* llvv ir dotas 15 nozīmes, bet mlvv - 10, pakārtojuma saiklim *ka* llvv 10, bet MLVV - 5. Nozīmju skaita samazinājums nenozīmē, ka līdz ar to būtu nepilnīgi atklāta attiecīgā vārda semantiskā struktūra. Atbilstoši vispārinot skaidrojumu, var panākt, ka tas aptver lielāku attiecīgās leksēmas semantiskās struktūras daļu. (Zuicena, 2010, 370). Rakstā atsauce, ka savulaik LLVV tapšanas laikā tika formulēti galvenie skaidrojumu veidi, kas izmantojami skaidrojošo vārdnīcu izstrādāšānā: t.s. Reālais skaidrojums

(definīcija un apraksts) un sinonīmiskais skaidrojums. Šāds iedalījums leksikogrāfijas teorijā pamatā ir saglabājies līdz pat mūsu dienām. (Zuicena, 2010, 370). Bo Svensons vārdnīcu skaidrojumu iedala trīs tipos: parafrāze (skaidrojums ar citu vārdu vai vārda savienojumu, kurā ietilpt arī skaidrojums ar sinonīmu un sinonīmiem), reālais skaidrojums (definīcija) un kombinētais skaidrojums (reālo skaidrojumu vai parafrāzi papildina viens vai vairāki sinonīmi. (atsaucas uz Svenson 1993, 121-122 - Svensen, Bo. Practical Lexicography. Oxford University Press, 1993).

Latviešu leksikogrāfijā visbiežāk tiek izmantots reālais skaidrojums definīcijas, apraksta vai salīdzinājuma veidā. Veidojot skaidrojumu, vispirms jāatrod kategorija, kurā ietilpst skaidrojamā lieta vai parādība, resp., virsjēdziens. Tad jānosauc atšķirīgais, kas šo konkrēto lietu vai parādību atšķir no citām, šajā pašā kategorijā ietilpstošajām lietām. Svarīgi skaidrojumā atrast vistipiskākās, pašas raksturīgākās atšķirīgās pazīmes, lai nebūtu iespējams skaidrojamo lietu vai parādību sajaukt ar citām. Specifisko pazīmju skaitam jābūt pietiekamam. (Zuicena, 2010, 370)

Adjektīviem piemīt augstāka abstrakcijas pakāpe nekā lietvārdiem, tie neveido hierarhisku sistēmu. Sevišķi abstraktu adjektīvu skaidrojumā izmanto salīdzinājumu, piemēram, krāsām. Atvasinātu adjektīvu skaidrojumos izmanto attiecīgo pamatvārdu: dzeltenīgs - ar dzeltenu nokrāsu. (Zuicena, 2010, 371)

Verbu skaidrojumi ir īpaši sarežģīti, jo verbs teikumā ieņem centrālo vietu un nosaka tā struktūru. Jāņem vērā ne tikai vārda semantiskā struktūra, bet arī saistījums ar citiem vārdiem teikumā, resp., sintakse. Skaidrojumu nosaka arī leksiski gramatiskās kategroijas - transitivitāte, refleksivitāte, veida kategorija (pabeigtība), kā arī tas, kas ir darbības veicējs. Visvairāk verba skaidrojumu ietekmē tas, vai verbs ir transitīvs vai nē. Neņemot vērā šo kategoriju, nav iespējams izveidot adkevātus skaidrojumus. Ir verbi, kuriem ir tikai transitīvas vai tikai intransitīvas nozīmes, bet ir tādi, kuriem ir dažādas nozīmes. Reizēm nozīmes nianses nav saistītas ar transitivitāti - *klausīt* var būt ar objektu (klausīt māti), vai bez (klausīt bez ierunām), bet nozīme uzskatāma par vienu. (Zuicena, 2010, 371-372) [arguments tas, ka abus var apvienot vienā konstrukcijā - klausīt māti bez ierunām]

Saikļu, prievārdu, partikulu un izsaukmses vārdu nozīmi skaidrojošās vādnīcās nevar aprakstīt citādi, kā tikai aprakstot attiecīgā šķirkļa vārda funkcijas, izmantojot skaidrojumā darbības vārdus “lieto”, “norāda”, “saista”, u.tml.

Izmantojot skaidrojumā polisēmisku sinonīmu, sinonīms jāpapildina ar kādu citu sinonīmu, lai skaidrojums būtu precīzs un neradītu pārpratumus, iemēram, “klapēt - sist, dauzīt”.

Nereti leksikogrāfi izmanto kombinēto skaidrojumu.

1.3. Vārda gramatisko pazīmju noteikšana

Vārdu morfoloģiskās tabulas virsrakstus piešķir pēc paradigmas nosaukuma. Tam vajadzētu nebūt maldinošam. Šobrīd paradigmas ir tādas, kādas ir analizatorā definētas. Vārda *pēdējais* un vēl dažu citu īpašu vārdu gadījumā morfoloģijas tabulai tiek piešķirts nosaukums “substantivējies īpašības vārds”, bet tas neatbilst patiesībai, jo šo vārdu lieto kā īpašības vārdu, kas raksturo lietvārdu. “Substantivējies” nav paradigmas īpašība, vēl nepieciešams pārskatīt morfoloģijas paradigmu nosaukumus, lai tie ir atbilstoši visiem grupas locekļiem un nesniedz nepatiesu informāciju.

Divdabjus ar galotnēm -is, -usi, kuriem nav verba pamatformas, piemēram, “jaundzimis”, nedrīkst morfoloģiski jaukt ar parastiem īpašības vārdiem, jo tad tiktu ģenerētas nepareizas galotnes.
Tēzaurs.lv šobrīd nešķir, kuriem adverbiem veido un neviedo pārākās pakāpes, tāpēc šobrīd visi adverbi ir tādā paradigmā, kurā neveido pārākās pakāpes, un tās netiek atspoguļotas morfoloģijas tabulā. Jānolemj, vai mēs vēlamies, lai apstākļa vārdiem tiek rādītas pārākās un vispārākās pakāpes formas morfoloģijas lauciņā, ja jā, tad jāatrod veids, kā nošķirt tos adverbus, kuriem netiek veidotas pakāpes. Problēma būs ar adverbiem, kas ir veidoti no divdabjiem, piemēram *uzvaroši*, *saraustīti*. Ne visiem divdabjiem veidojas pārākās pakāpes, tāpāt arī ne visiem adverbiem, kas no tiem veidoti, var veidot pārākās pakāpes. Tika realizēta otra ideja - atlasīti vārdi no Tēzaura, kuriem ir klāt gramatiskā norāde, ka tie ir apstākļa vārdi. Tādu bija 7652 vārdi. Daļa no tiem ir kategorijā *word*, daļa - *reference*, daļa *derived* - varbūt kaut kur pārklājas. Visiem, kas ir sadaļā *word*, manuāli tika norādīts, vai var veidot pakāpes. Par pārējiem notiek diskusija, vai automatizēt no īpašības vārdiem iegūtos vārdus (06.05.2020).

1.4. Šķirkļa vārdu nošķiršana atsevišķā šķirklī

Viens no pamatprincipiem Tēzaurā - ja vienādām vārda formām ir atšķirīga vārdšķira, tad tiek veidoti divi šķirkļi jeb **homonīmi**. Piemēram, ja vārds tiek lietots gan kā lietvārds, gan kā īpašības vārds, jābūt diviem šķirkļiem homonīmiem, piemēram, vārdam *aerobs*.
Daudznozīmība jeb polisēmija robežojas ar homonīmiju. Polisēmiska vārda atsevišķa nozīme var atrauties no pārējām vārda nozīmēm un kļūt par homonīmu. Teorētiski izvirzīts kritērijs - daudznozīmīga vārda vienība saglabājas tik ilgi, kamēr tā leksiski semantiskos variantus apvieno kopīgs semantisks moments, kamēr netiek sarauti atsevišķo variantu iekšējie sakari. Šāds kritērijs neko daudz nepaskaidro. Iespējams, ka polisēmijas priekšnoteikums ir piederība vienai vārdšķirai. Tas nepalīdz atrisināt homonīmiju vienas vārdšķiras robežās. (Laua, 1969, 31-32) Reizēm viena vārda daudzskaitļa vai vienskaitļa forma jau ir ieguvusi tik attālinātu nozīmi no sākotnējā šķirkļa, turklāt to lieto tikai šajā konkrētajā skaitlī, ka to jau var uzskatīt par atsevišķu vārdu un likt atsevišķā šķirklī. Piemēram, vārds *brilles* vārdnīcā parasti dots daudzskaitlī, reizēm ir norādīts, ka arī vienskailti sarunvalodā mēdz lietot ar šo nozīmi, savukārt vienskaitļa forma *brille* ar nozīmi “tualetes pods” tiek uzsaktīts par atsevišķu šķirkli. Abas vienskaitļa formas šajā gadījumā ir homoformas. Bet, tā kā šķirkļa vārda forma atšķiras, šos vārdus nesauc par homonīmiem un neuzrāda homonīmu logā ar attiecīgu homonīma numuru. Līdzīgi ir ar vārdu *atteicības*, kur nozīme par sakariem starp cilvēkiem piemīt tikai daudzksaitļa formai un šķiet jau krietni attālinājusies no vienskaitļa nozīmes par dažāda veida matemātiskām attieksmēm starp parādībām. Šādos gadījumos var pieņemt lēmumu daudzksaitļa nozīmi pārcelt uz atsevišķu šķirkli ar šķirkļavārdu daudzskaitlī un gramatiskajām norādēm, kas nosaka, ka šo vārdu lieto tikai daudzskaitlī.
Problemātiski ir gadījumi, kad valodniecības teorijā atšķiras viedokļi par vārdu piederību noteiktai vārdšķirai. Ne vienmēr vārdnīcās norādītā vārdšķira sakrīt ar Tēzaura veidotāju komandas valodnieku viedokli, ņemot vērā, ka valodniecības teorija attīstās un arī šie jautājumi ik pa laikam tiek pārskatīti. Tādos gadījumos ir iespējama šķirkļa vārda vārdšķiras maiņa, papildinot šķirkļa avotus ar norādi, ka Tēzaura veidotāji arī ir to mainījuši. Turpmāk ir norādīti

dažādi gadījumi, kad tieši vārdšķiras noteikšanas problemātika arī ietekmē šķirkļu vārdu nošķiršanu.

1.4.1. Substantivējušies vārdi

Problemātiska ir substatntivējušos vārdu vārdšķiras noteikšana un jautājums par to, vai tiem jābūt atsevišķā šķirklī. Vai vārdam *cietušais* nepieciešams atsevišķs šķirklis, kaut arī tas morfoloģiski ir divdabis no verba *ciest*? Daudzi divdabji Tēzaurā ir kā atsevišķi šķirkļi, kuros viena nozīme ir uzrādīta divdabja forma no verba, otra nozīme - īpašības vārda nozīme. *Cietušais* kā atsevišķs šķirklis ir tikai substantivējusies nozīme, bet divdabja nozīme pat netiek uzrādīta.

Ilzei Lokmanei ir raksts par substantivēšanās pazīmēm un iespējamo iedalījumu (Lokmane, 2015). Tajā norādīts, ka valodniecībā ir problemātiski noteikt, kad konversija ir notikusi pilnībā un kad daļēji (kontekstuālā konversija). Autore atsaucas uz Emīlijas Soidas teoriju, ka substantivējumi var būt raudšies sintaktiskās redukcijas rezultātā, un šai sakarā Lokmane piedāvā substantivējumu dalījumu atkarībā no tā, vai nozīme ir kontekstam, vai reducētās vārdkopas elementu ir iespējams noteikt vai nē. Raksts uzrāda problemātiku, kā arī satur apkopojumu par līdzšinējiem kritērijiem pilnīgas vai daļējas konversijas noteikšanā. Nozīmes maiņa ir būtisks aspekts - iespējams, ka kaut kas no minētā noder kā kritēriji atsevišķa šķirkļa izveides nosacījumiem Tēzaurā. Divdabjiem gan rakstā nav pievērsta īpaša uzmanība, tāpēc šo problēmu tas nerisina.

1.4.2.Divdabja formas šķirkļa vārdi

Imants Šmidebergs savā pētījumā par divdabjiem ar priedēkli “ne-” skaidrojošajās vārdnīcās (Šmidebergs, 2009) apskata, kādos gadījumos LLVV, LVV un MLVV veido atsevišķus šķirkļus noliegtu verbu divdabjiem (netiek veidoti atsevišķi šķirkļi noliegtiem verbiem, jo tiek uzskatīts, ka verba noliegums nerada izmaiņas verba pamatsemantikā), jo uzskata, ka laviešu leksikogrāfijā kritēriji šo šķirkļu izveidei vēl nav definēti. Analizējot LLVV materiālu, secināts, ka atsevišķā šķirklī parasti skaidroti tie noliegtie divdabji, kuru semantiskajā struktūrā konstatējams atributīvs lietojums ar tādu apzīmējamo priekšmetu, parādību loku, kas atšķiras no pamatverbam raksturīgā darbības objektu kopuma. Divdabis ar ne- tādā gadījumā apzīmē un konkretizē tkai vienu šaurāku segmentu no tā, ar ko var būt saistīta ar verbu izteiktā darbība. Kopumā šajā vārdnīcā divdabjiem ar priedēkli ne- tiek veidoti atsevišķi šķirkļi šādos gadījumos (Šmidebergs, 2009, 164):

- Noliegtajam divdabim ir kāda terminoloģiska, ar noteiktu lietošanas sfēru saistīta nozīme, piemēram, “nepiesātināts”;
- Noliegtajam divdabim ir kāda nozīmes nianse, kas nekādā veidā nevar būt pretstatāma ar verbu izteiktajai darbībai, piem., “nesamaksājams” par cilvēku;
- Eksistē vārdnīcā rādāms frazeoloģisms, kurā ietilpst noliegtais divdabis, piemēram, “nevēlams”;

Divdabim parasti norādīta viena nozīme kā noliegta verba divdabja forma, otra nozīme - skaidrojums ar “tāds”, kas skaidro divdabja adjektīvisku lietojumu. Gramatiskās norādes šiem divdabjiem ir kā īpašības vārdiem. LVV šādus divdabjus, ja tiem ir izveidots šķirklis, jau dēvē par īpašības vārdiem, izņemot divus gadījumus. Savukārt MLVV izvēlas formulējumu “divdabis

īpašības vārda nozīmē", uzskatot, ka tas vislabāk atbilst valodā noteikošajam konversijas procesam, kurā nav iespējams skaidri noteikt robežas, kad divdabis jau būtu jāuzsakta par adjektivizējušos. Turklāt MLVV tiek atrastas arī jaunas noliegto divdabju adjektīviskas nozīmes, kuru dēļ tiek veidoti jauni šķirkļi, piemēram, *necaurskatāms* (lēmumu pieņemšanas process), *necenzēts* (literārais darbs) (Šmidebergs, 2009, 165).
Arī Ārijas Ozolas rakstā par divdabju ar izskaņu -ošs, -oša atspoguļojumu skaidrojošajās vārdnīcās (Ozola, 2010) minēts, ka divdabju vārdšķiras noteikšanā vērojama atšķirīga pieeja katrā vārdnīcā - LLVV nav norādes, ka būtu īpašības vārds, bet nozīmēs viena ir norāde uz divdabja formu no verba, bet otra - nozīme ar adjektīvam raktsurīgo skaidrojumu "tāds, kas…". LVV šķirkļa vārdi ar izskaņu -ošs, -oša tiek uzsaktīti par divadbjiem, kas adjektivējušies, tātad par īpašības vārdiem ar potenciāli pieļaujamu verbalitāti (norāde par divdabja formu no verba iekavās). MLVV esot divējāda pieeja - no burta A līdz burtam I šādi divdabji ar izskaņu -ošs, -oša tiek pielīdzināti īpašības vārdam, bet no burta J tiek lietota MLVV jaunāka pieeja - norāde, ka tas ir divdabis īpašības vārda nozīmē, bet par tādiem, kas adjektivējušies, uzskata tikai tos, kuriem vairs nav attiecīgo darbības vārdu semantiskās saistības (Ozola, 2010, 220).
Skaidrs kritērijs divdabju formu nošķiršanai par atsevišķu šķirkli ir tad, ja neeksistē verbs, no kura divdabja forma veidota - "jaundzimušas" - nav tāda verba "jaundzimt", tāpēc šo divdabja formu definē kā īpašības vārdu. Tēzaurā tam visam tiek rādīta morfoloģijas tabula "adjektivizējies divdabis", bet MLVV norāda, ka viena nozīme ir kā parastam īpašības vārdam (*jaundzimušais jēriņš*), otra nozīme substantivējusies, kad īpašības vārds lietots lietvārda nozīmē. Šobrīd (12.03.2020.) nav iespējams atrast šajās vārdnīcās pēc formas "jaundzimušais", jo ir šķirklis "jaundzimis". (09.02.2021.) - iespējams atrast divdabjus arī ar noteikto galotni, pat ja šķirkļa vārds ir ar nenoteikto galotni, jo meklētājs meklē arī pēc vārda formas.

1.4.3. Skaitļa vārdu kategorija un robežgadījumi

WordNet vārdu biežuma statistikas sarakstā iekļuvuši tikai pāris skaitļa vārdu, tādēļ par skaitļa vārdiem un to robežgadījumiem projekta ietvaros pētīts maz. Paredzams, ka nāksies saskarties ar sarežģītiem skaitļa vārdu gadījumiem, piemēram, izzudušo nenoteikto skaitļa vārdu kategoriju, nelokāmiem skaitļa vārdiem, kas, nosaucot pulksteņa laiku, kļūst lokāmi u.c.

1.4.4. Vārds lietots citas vārdšķiras nozīmē

Vēl viena liela problemātiska grupa - norādes šķirklī kā atsevišķas nozīmes, ka vienas vārdšķiras vārds lietots citas vārdšķiras vārda nozīmē, bieži attiecas uz adverbiem, kas lietoti saikļa nozīmē, piemēram, adverbs *tādēļ*. MLVV divdabja šķirkļu raksturojumā lieto jēdzienu "divdabis lietots īpašības vārda nozīmē". Ir divdabji, kuriem ir savs šķirklis, bet ir divdabji, kurus rāda pie verba šķirkļa ar atzīmi "īpašības vārda nozīmē". Tāpat lietvārdu šķirkļos mēdz būt nozīme, kura realizējas tikai noteiktā lietvārda locījumā un klāt ir atzīme, ka lietots "īpašības vārda nozīmē" vai "apstākļa vārda nozīmē". Visi šie gadījumi attiecas uz teorijā lietoto terminu "korversija". Tā apzīmē kāda vārda vai vārda formas pāreju citā vārdšķirā, un tas ir pakāpenisks process, kuram nav ļoti precīzu kritēriju, kad to var uzskatīt par pabeigtu (un tādā gadījumā vārdam pienāktos savs atsevišķs šķirklis vārdnīcā), vai kad tas ir vēl tikai pusceļā, jo izpaužas arī vēl kādas sākotnējās vārdšķiras pazīmes, funkcijas vai semantika. Leksiskā tīkla sakarā

piederība vienai vai otrai vārdšķirai var nebūt tik svarīga, jo saites tiek vilktas sarp nozīmēm, neatkarīgi no to vārdšķiras, pieļaujot gadījumus, kad semantiskā saite pastāv starp dažādu vārdšķiru vārdiem, piemēram, starp īpašības vārdu un kādu divdabi.

1.5. Vārda nozīmju rediģēšana un piemēru atlase online rīkā wordnet.tezaurs.lv

Projekta ietvaros tika izveidots rīks, kas apvieno Tēzaura redaktora piedāvātās iespējas rediģēt šķirkļus un tajā esošās vārdu nozīmes un iespēju pievienot piemērus no dažādiem latviešu valodas korpusiem. Rīks nepieciešams, lai rediģētu Tēzaurā esošās vārdu nozīmes, kuras tālāk tiek izmantotas WordNet saišu veidošanai.
Tēzaurs pamatā ir balstīts uz šādiem avotiem:

1) *Latviešu literārās valodas vārdnīca*. 1.–8. Rīga, Zinātne, 1972.–1996. (LLVV)
2) Mīlenbahs K. *Latviešu valodas vārdnīca*. Rediģējis, papildinājis, turpinājis J. Endzelīns. Rīga, Kultūras fonds, 1923.–1932. (ME)
3) Endzelīns J., Hauzenberga E. *Papildinājumi un labojumi K. Mīlenbaha Latviešu valodas vārdnīcai*. Rīga, Kultūras fonds; VAPP, 1934.–1946.(EH)
4) *Latviešu Konversācijas vārdnīca*. 1.–21. Rīga, A. Gulbja apgāds, 1927.–1940.; papildināts faksimilizdevums 1.–22. Rīga, 2000.–2004. (KV)
5) Vikipēdija un citi interneta resursi (Wi)

Izmantotās skaidrojošās vārdnīcas ir pietiekami senas, lai pieņemtu, ka valodā tikmēr ir notikušas vārdu nozīmju pārmaiņas - dažas nozīmes saplūst, citas izveidojas no jauna. Jaunākā šāda veida informācija par aktuālajām nozīmēm ir Mūsdienu latviešu valodas vārdnīcā (MLVV). Kuras veidotāji norāda, ka centušies veidot mazāk saskaldītas nozīmes nekā LLVV. Tai pašā laikā MLVV satur daudz mazāk šķirkļus nekā Tēzaurs, tāpēc viņu dati var noderēt tikai daļai Tēzaura vārdu. Mēs gan ceram, ka biežākie 5000 vārdi, kurus plānojam projekta ietvaros rediģēt un kuriem plānojam veidot Wordnet saites, būs MLVV pārstāvēti.
Darba plūsma plānota šādi:

1) Salīdzināt Tēzaurā un MLVV dotās vārdu nozīmes, koriģēt tuvāk MLVV nozīmēm, ja vien marķētāja valodas izjūta nav pretrunā ar to.
    - Tēzaurā neņemt nost terminu un apvidvārdu nozīmes. Lai to noteiktu, pievērst uzmanību avotu sarakstam šķirkļa apakšā.
    - Sākotnēji, lai nezaudētu datus, ir nolemts no LLVV mantotās nozīmes, kuras vēlas likvidēt, pielikt pie kādas no aktuālajām nozīmēm kā apakšnozīmi, tomēr darba gaitā šo pieeju ne vienmēr izdodas realizēt.
    - Nozīmes, kuru skaidrojums ir nesaprotams un pēc sākotnējā avotā minētajiem piemēriem joprojām nešķiet būtiski šķiramas, var tikt likvidētas.
    - Tāpat tiek likvidētas tukšās apakšnozīmes, kurās trūkst skaidrojuma teksta, bet ir tikai norādes par citāda veida lietojumu transitivitātes aspektā.
    - Nozīmes, kurām nav iespējams atrast piemērus aktuālajos korpusos, bet ir saprotams, ka tās bijušas spēkā senākos tekstos, tiek paturēts, bet noteikti tiek pievienots karodziņš “novecojis”.

- Pārskatīt stabilu vārdu savienojumus šķirklī un gadījumā, ja kāds no tiem sastāv tikai no viena vārda, pārvērst to par nozīmi šķirklī, likvidēt stabila vārdu savienojuma šķirkli.

2) Piemēru apskate korpusos un pievienošana nozīmēm:
   - vispirms Līdzsvarotajā korpusā jāapskata visi piemēri, ja tās ir dažas lapas, vai izlases kārtā kādas 5-10 lapas no visām piemēru lapām, cenšoties pievienot piemērus visām nozīmēm;
   - Saeimas korpusā jāapskata visi piemēri, ja tās ir dažas lapas, vai izlases kārtā kādas 5 lapas no visām piemēru lapām;
   - Papildus Emuāru vai Interneta korpusā jāapskata izlases kārtā vairākas lapas, vai neparādās vēl kāda sarunvalodas nozīme;
   - Lai atrastu kādu retāk lietotu nozīmi, izmanto visus korpusus un var izmantot piemēru meklētājā vārdu savienojumus;
   - Ja kādai nozīmei neatrod piemēru, var meklēt citur internetā, iekopēt piemēru piemēra laukā un iekopēt saiti avota laukā;
   - Piemērus pievieno arī šķirklī esošajiem stabiliem vārdu savienojumiem - pagaidām tikai tos, kurus pamana pārskatītajās piemēru lapās, speciāli izteicieniem lietojuma piemēri netiek meklēti.
3) Nozīmju koriģēšanas vadlīnijas:
   - Ja vairākiem piemēriem ar līdzīgu nozīmi nav skaidrs, pie kuras nozīmes pievienot, jo der divām vai vairākām, jāapsver iespēja apvienot šīs nozīmes vienā - sapludināt, labojot skaidrojumu, vai, ja skaidrojumā ir kāda nianse, kuru bail pazaudēt, vienu atstāt kā apakšnozīmi pie otras.
   - Ja piemērs neder nevienai esošaja nozīmei, tad a) jāpārbauda, vai tas nav minēts kā frazeoloģisms; b) jāpārbauda, vai nav kādā apakšnozīmē; c) jāveido jauna nozīme, uzrakstot jaunu skaidrojumu. Skaidrojuma izveidei ņemt vērā līdzīgas nozīmes vārdu skaidrojumus, kā arī svarīgi skaidrojumā iekļaut būtisko, ar ko šī nozīme atšķiras no citām.
   - Rediģēt nozīmju skaidrojumus, kuri sastāv tikai no viena vārda - sinonīma, lai neveidotos apļveida skaidrojumi. Veidot aprakstošu nozīmes skaidrojumu, sinonīmu norādot aiz semikola.
   - Regulārās metonīmiskās nozīmju pārmaiņas netiek marķētas kā atsevišķas nozīmes, bet izplatītiem metonīmizkiem nozīmes pārnesumiem iespējams veidot apakšnozīmi - kurt krāsni un kurt māju, cilvēks spēlē mūziku/ radio spēlē mūziku.
   - Pēc iespējas neatstāt apakšnozīmes bez skaidrojuma - no LLVV mantojumā ir apakšnozīmes, kurām vienīgā piezīme klāt ir, ka tās ir “intrans.”. Šādos gadījumos jāapsver šādas apakšnozīmes nepieciešamība - ja nekādi semantiskie komponenti nemainās un potenciālais objekts ir vispārīgs, tāpēc teikumā netiek minēts, tad to var attēlot ar piemēru, nevis atsevišķu apakšnozīmi. Pretējs piemērs - ja ir nozīmes papildkomponents par spēju, prasmi darbību darīt - *Vai tavs sešgadīgais bērns jau lasa?*.
   - Ja nevar atrast kādai nozīmei piemēru dotajos korpusos, tad piemēru pievieno no citiem avotiem, ja tas ir iespējams, piemēram, atrodot Google vai Periodika.lv.
   - Ja ir grūtības ar skaidrojumu precīzi atklāt nozīmes nianses, var pievienot viegli uztveramu izdomātu piemēru, norādot tam avotu TWN.

4) Piemēru atlasē vajag piemērus ar dažādu saistāmību un nozīmes niansēm. Ja ir stabils vārdu savienojums, kas bieži parādās korpusos, tad arī tam pievieno piemērus, lai veidotu labākus datus iespējamajai datora apmācībai. Piemēru dažādībai un nozīmju nošķiršanai vai nianšu atklāšanai jāņem vērā:
    - vai darītājs jeb subjekts ir dzīvs/nedzīvs;
    - vai darītājs jeb subjekts ir konkrēts/abstrakts;
    - vai darītājs jeb subjekts ir skaitāms (diskrēts) / neskaitāms (viela vai abstrakts jēdziens);
    - vai ir objekts;
    - pārējoša/nepārejoša nozīme;
    - objekta semantiskais tips - konkrēts/abstrakts, semantiskā grupa;
    - apstākļi kā semantiski obligāti argumenti kādas nozīmes izteikšanai - “slēgt vaļā”, “krist nost”;
    - argumentu esamība/neesamība, vai nozīme izpaužas arī bez argumentiem;
    - piesaistītais arguments ir dažādos locījumos - parādīt arī piemēros a) nozīme nemainās, jānorāda dažādi saistījumi: “cerēt darīt”, “cerēt uz kaut ko”, “cerēt, ka”); b) nozīme mainās.

1.5.1. Nozīmes piemēru atlase un pievienošana

Projekta ietvaros šķirkļos atsevišķām nozīmēm tiek pievienoti piemēri no dažādiem korpusiem. Sarunvalodas nozīmēm piemēri visticamāk būs no Saeimas un interneta resursu korpusiem. Tā kā piemēri nepieciešami, lai apmācītu datoru automātiski šķit nozīmes, tad tos katrai nozīmei pievieno vairāk, nekā tiks rādīti publiskā versijā Tēzaura lietotājiem. Piemēri automātiski tiek atzīmēti kā “paslēpti”, un nozīmes reprezentēšanai publiski tiek izvēlēti tikai daži piemēri. **Izvēlēto piemēru daudzums** atkarīgs no tā, cik atšķirīgās apkaimēs vārds lietots. Atsevišķs piemērs noteikti nepieciešams, ja ir atšķirīga saistāmība, piemēram, verbam rādīt piemēru gan ar pieveinotu papildinātāju, gan ar pievienotu palīgteikumu, lai cilvēki redz dažādu sintaktisko lietojumu. Darbības vārdu nozīmēm ir būtiski parādīt dažādus iespējamos darbības veicējus - šķirt dzīvus subjektus no nedzīviem, ja tas ir vienas nozīmes ietvaros, lai gan iespējams, ka tam ir paredzēta atsevišķa nozīmes nianse. Atklāt metonīmiskos pārnesumus - kad var veikt cilvēks, bet kad cilvēku kopumu apzīmējošs vārds (mūziku spēlē kungs vai mūziku spēlē orķestris). Parādot objektu semantisko saistāmību, svarīgi parādīt, ja var būt piesaistīts dažādu semantisko grupu objekts. Sākotnēji plāns bija nepiešķirt piemērus konkrētām apakšnozīmēm, bet norādīt tām atsbilstošus piemērus pie galvenās nozīmes, tomēr ar laiku tika secināts, ka apakšnozīmes bieži funkcionē kā patstāvīgas nozīmes, tikai to nozīmes nianses statuss norāda uz nozīmju tuvumu. Tāpēc piemēri tiek likti gan pie nozīmes, gan pie katras nozīmes nianses. Pie nozīmes tiek likti vairāk piemēru **(10 - 20)**, bet pie nozīmes nianses, ja tā reti lietota, var būt maz piemēru. Projekta sākotnējā posmā izvēlējāmies pielikt piemērus arī stabiliem vārdu savienojumiem, kas iekļauti atbilstošā vārda šķirklī, tomēr laika gaitā tika secināts, ka apstrādāt arī visas stabilu vārdu savienojumu nozīmes šķirklī prasa pārāk daudz laika un neļauj pildīt pamatuzdevumu strādāt ar vārda nozīmēm, kas izpaužas tomēr dažādos kontekstos. Ja kādi piemēri korpusu materiāla apskatē ar stabilu vārda savienojumu parādās regulāri, tad tos

pievieno, bet speciāli visiem stabiliem vārdu savienojumiem piemēri netiek meklēti. Šo lēmumu nākotnē var mainīt automātiskās vārda nozīmju nošķiršanas eksperimentu rezultāti.
Piemēri netiek pievienoti ļoti retām nozīmēm, kuras nevar atrast visos projektā iesaistītajos korpusos. Ja marķētājam ir pārliecība, ka nozīmi lieto sarunvalodā, tāpēc neparādās korpusos, atbilstošus piemērus var pievienot no tīmekļa, atrodot ar google meklētāja palīdzību. Tāpat piemēri var nebūt specifiskām terminu nozīmēm, kuras lieto kādā šaurā jomā.
Reprezentatīvi piemēri - tādi, kas nav pārāk gari, ieteicams ne vairāk par 2 - 3 teikuma daļām. Nepieciešams, lai atklājas nozīmei būtiskās sēmas un lai nav norāde uz kontekstā nosauktām lietām, kas teikumā nav minētas. To parasti dara ar norādāmiem vientiekvārdiem tas, šis, tāds utt. Piemēram jābūt neatkarīgam no konteksta. Piemēram jāatklāj tipiskā saistāmība šīs nozīmes attēlošanai.
Piemēriem nevajadzētu saturēt daudzus svešvārdus vai vecvārdus, lai lietotājs varētu viegli uztvert saturu. Slikts piemērs vecvārdu dēļ: *Godīgs burlaka, ja tam kaut grasis ķešā, lai kāds laikmets būtu sētsvidū, uzreiz spersies uz krogu un dzers par grasi — un tikai pēc tam sāks domāt, kur ņemt dālderi, lai dzertu par dālderi.* [Guntis Berelis, Mīnotaura medības. Rīga, Atēna, 1999.]
Piemēriem nevajadzētu saturēt necenzētu leksiku vai vienkāršrunu, ja vien tas nav nepieciešams noteiktās nozīmes atklāšanai vai vienkāršrunas vārda šķirklim. Slikts piemērs: *Pirmais, ko ik rītu pamodies domāju − kā lai ar savu sarūsējušo zobenu konkurentiem nocērtu pautus?* [Norberts Klaucēns, Nauda, bauda un citas dēkas. Rīga, Madry, 2003.]

Piemērs nedrīkst saturēt atbilstošo šķirkļa vārdu divreiz, ja katrā reizē vārds lietots citā nozīmē. Slikts piemērs vārdam “domāt”: *Es domāju, ka mēs visi, loģiski domājoši cilvēki, saprotam, ka ārkārtas situācija ir tad, kad nepieciešams ātri, operatīvi rīkoties.* [No Saeimas stenogrammām 24.07.2014., Jānis Klaužs (ZZS)]

Publiskā versijā rādāmie piemēri - šķirkļa rediģētājs izvēlas reprezentatīvus piemērus no visiem pievienotajiem piemēriem. Pamatnozīmēm ieteicams vismaz 3 piemērus, nozīmes niansēm var arī tikai vienu. Piemēru skaits var būt lielāks, ja nepieciešams atklāt dažādu sintaktisko vai semantisko saistāmību vienas nozīmes ietvaros**.** Publiski rādāmajos piemēros nedrīkst iekļaut reālu personu vārdus un uzvārdus, lai nebūtu problēmu ar privāto datu aizsardzību.

1.5.2. Vārda lietojuma ierobežojuma pazīmes

Baldunčiks pētījumā par latviešu literārās valodas vādnīcu norāda, ka pazīme “novecojis” reizēm likta nepamatota, un atsaucas arī uz citiem pētījumiem, kas norāda to pašu (Baldunčiks, 2012, 117). Valodas nozīmju izmaiņas ir viens no visātrāk notiekošajiem valodas maiņas procesiem, tāpēc ir iespējams, ka kāda nozīme ir atguvusi savu aktualitāti, vai kāda novecojusi. Marķētāja ziņā ir šo norādi par nozīmes aktualitāti atstāt, noņemt vai likt klāt, ņemot vērā valodas materiālu un savu valodas izjūtu.
Daudziem vārdiem pazīme, ka tas ir apvidvārds, ir pievienota tāpēc, ka tie nāk no izlokšņu vārdnīcām. Taču tas nenozīmē, ka šāda nozīme nav sastopama arī literārajā latviešu valodā vai sarunvalodā. Piemēram, “aizvērt” nozīme “verot aizvirzīt aiz kā, kam cauri” ir norādīta kā apvidvārda nozīme, tomēr tādā nozīmē to lieto ne tikai izloksnēs. Ja nozīmju rediģētājs saprot,

ka nozīme ir plaši lietota vispārējā sarunvalodā vai rakstu valodā, tad šo atzīmi, ka tā ir apvidvārda nozīme, var ņemt nost.

2. Leksiskā tīkla saišu veidošana

WordNet ir leksiska datubāze, kurā vārdi tiek grupēti sinonīmu kopās, kur katra izsaka noteiktu konceptu. Starp sinonīmu kopām tiek norādītas dažāda veida attieksmes jeb vilktas saites. Būtiski, ka saites (gan sinonīmu kopas iekšienē, gan starp kopām) tiek veidotas starp atsevišķām vārdu nozīmēm, nevis starp vārdnīcas šķirkļiem. Atšķirībā no sinonīmu vārdnīcām, WordNet grupē vārdus arī pēc citiem semantiskajiem konceptiem, ne tikai pēc to līdzības pēc nozīmes. Sinonīmija ir pamatsaite starp vārdiem WordNet.
Latviešu valodas lesiskajā tīklā (Latvian WordNet) ir izvēlētas šādas semantiskās attieksmes starp sinonīmu kopām:

1) Hiponīmija - starp virsjēdzienu un apakšjēdzienu (atteicīgi - hiperonīms un hiponīms);
2) Meronīmija - starp veselo un daļu (attiecīgi - holonīms un meronīms);
3) Antonīmija - pretstatāmi jēdzieni. Sākotnēji plānots izmantot klasiskiem īpašības vārdu pāriem.
4) Aptuvenā sinonīmija - sinonīmi, kuri tomēr nav gluži aizstājami kontekstā, līdzība tikai aptuvena.
5) Saistīts ar - ja nav skaidrs, kāda semantiskā saite iederētos, bet ir vēlme saiti norādīt, lai varētu precizēt vēlākos pētījumos.
6) Gradācijas jēdzienu grupa - apvieno vienā kopā sinsetus, kas nosauc vienu pazīmi vai parādību dažādās pazīmes piemitības pakāpēs. Grupā ietilpst gan atsevišķas nozīmes, gan sinonīmu kopas. Iekļaušana grupā netraucē veidot atsevišķu saiti starp grupas locekļiem, lai norādītu antonīmiju vai hiponīmiju.

Antonīmiju nav plānots marķēt visās tās izpausmēs, šobrīd paredzēts lietot īpašības vārdiem. Nav pānots marķēt kauzāciju jeb attieksmi par to, vai darbība ir veikta ar nodomu, speciāli izraisīta, vai notiek pati no sevis.
WordNetā semantiskās attieksmes izpratne var atšķirties atkarībā no vārda vārdšķiras, piemēram, hiponīmija lietvārdiem atsķiras no hiponīmijas verbiem. Princeton WordNet leksiskajā tīklā verbu hiponīmiju sauc savādāk - troponīmija, un ar to izprot noteikta veida attieksmes, kas varbūt aptver šaurāku loku, neklā latviešu valodas leksiskā tīklā marķētās hiponīmiskās attieksmes. Tomēr mēs nolēmām, ka šķīrumu nosaukumiem pēc vārdšķiras var izveidot arī vēlāk automātiski, ja tas izrādīsies nepieciešams un ja būs precizēta šo jēdzienu atšķirība. Tāpat jāuzsver, ka Princeton WordNet adjektīviem neveido hiponīmiskas attieksmes, bet “līdzības” attieksmes, savukārt Poļu WordNet tomēr veido adjektīviem hiponīmiskas attieksmes. Saišu lietojums pa vārdšķirām un pa dažādu valodu leksiskajiem tīkliem var atšķirties, tāpēc ir būtiski katram atsevišķam leksiskam tīklam noskaidrot lietotās semantiskās attieksmes un to izpratni.
Vispārīgas semantiskas attieksmes izsaka regulārie atvasinājumi. Vārddarināšanā varētu būt sistēmiskas atbilsmes - plēst un plīst - viens ir ar nodomu (kauzatīvs), otrs pats plīst kaut kādu ārēju apstākļu ietekmē. Vācu wordnet ir šāda kauzācijas attieksme (causation). Latviešu valodā tādu varētu veidot vārda degt nozīmēm - degu un dedzu. Šim nepieciešams atsevišķs pētījums.

Vārddarināšanas attieksmes varētu paust kādu vispārīgu semantisku attieksmi starp vārdiem, kas varētu būt kā cits paralēls slānis, un kuru nav paredzēts attīstīt šajā projektā.
Projekta laikā pēc biežuma Latviešu valodas līdzsvarotajā morfoloģiski marķētajā tekstu korpusā tika atlasīti 5000 vārdi, kuriem paredzēts veidot semantiskās attieksmes leksisakajā tīklā. Šiem vārdiem pēc nepieciešamības tika pārskatītas nozīmes, lai veidotu konsekventāku nozīmju dalījumu daudznozīmīgiem vārdiem, pievienotu trūkstošas nozīmes vai nozīmju nianses, apvienojot pārāk līdzīgas nozīmes, kur nav iespējams atšķirt smalkās nianses. Tika ņemts vērā, ka nozīmēm jāveido sinonīmu kopas, tāpēc tika meklēts kompromiss starp nepieciešamību veidot vispārīgās nozīmes, kas būtu vieglāk uztveramas, un nepieciešamību veidot atsevišķas nozīmes, kuras derētu noteiktā sinonīmu kopā, bet nederētu, ja nozīme tiktu vispārināta.
Pēc nozīmju pārskatīšanas tām tiek norādītas semantiskās attieksmes. Daļai vārdu ir ieteikumi no sinonīmu vārdnīcas, tomēr ne visi ieteikumi der sinonīmu kopai, daļa no tiem atbilst hiponīmijas attieksmēm, leksiskā tīkla veidotājam tas rūpīgi jāizvērtē. Sinonīmu rindā vārdnīcā turklāt ir uzskaitīti dažādu nozīmju jeb sinonīmu kopu sinonīmi, arī to jāizvērtē darba gaitā.
Ņemot vērā, ka projektā plānots arī sastatīt latviešu valodas leksiskā tīkla sinonīmu kopas un to attieksmes ar Princeton Wordnet sinonīmu kopām, tad iespējams, ka šis solis vēl ietekmēs marķēšanas principus. Tomēr, šķirot nozīmes un veidojot sinonīmu kopas, sākotnēji angļu leksiskā tīkla piemēri netiek ņemti vērā, lai neietekmētos un izveidotu pēc iespējas autentiskāku latviešu valodas tīklu. Ļoti ticams, ka noteiktas nozīmes un semantiskās attieksmes jau tiek pārņemtas no angļu valodas un sāk parādīties sarunvalodā un interneta resursos, kas oriģināli latviešu valodas vārdiem nebija, bet tas pieder pie starpvalodu kontaktiem un valodas attīstības.
Latviešu leksiskajā tīklā sinonīmu kopas tiek veidotas, apvienojot vārdu nozīmes, bet neveidojot kādu vispārīgu skaidrojumu visai sinonīmu kopai. Princeton WordNet ir viens vispārīgais skaidrojums visiem sinonīmu kopas dalībniekiem. Šajā projektā nav paredzēts veidot vispārīgo skaidrojumu , tas varētu būt uzdevums nākamajam projektam.

2.1. Marķējamo vārdu saraksta veidošana

Lai noteiktu, kuri vārdi tiks apstrādāti projekta ietvaros (gan nozīmju analīze, gan WordNet saišu izveidošana), tika izmantoti vairāki avoti. Pamatavots ir vārdu biežuma saraksts, kas veidots pēc Mūsdienu latviešu valodas līzsvarotā korpusa datiem. No šī saraksta tiek atlasīti to vārdšķiru vārdi, kuriem tiks veidotas WordNet saites - **lietvārdi, verbi, adjektīvi, adverbi**. No lietvārdiem netiek marķēti īpašvārdi, jo tie nav paredzēti šajā projektā. No adverbiem netiek marķēti vietniekvārdu cilmes adverbi - *kad, tad, kur, tur, cikos, tikos, kāpēc, tāpēc, kādēļ, tādēļ, kā, tā, tādējādi, te, tātad, tikpat, kamēr, tikmēr* utt. Sarakstā netiek iekļauti arī saīsinājumi, piemēram, “nr.”, “plkst.”. “SIA”, “a.”, “g”, “j.”. Skaitļa vārdi no “viens” līdz “deviņi” izņemti no saraksta.

Pēc šīs atlases pirmie 5000 vārdu šai sarakstā tiek uzskatīti par vārdu sarakstu. Vārdi, kuri paliek pēc atlases paliek biežuma sarakstā 5001. - 6000. pozīcijā, tiek pārskatīti manuāli. Ja valodniekam pēc valodas izjūtas liekas, ka kādam no tiem varētu būt biežāks lietojums, nekā uzrāda korpusa saraksts, tiek pārbaudīts, vai vārds atrodams “**Pamatvārdnīcā**” un/vai “**3000 sarunvalodā nepieciešamie vārdi**”. Ja jā, tas tiek pievienots marķējamo vārdu sarakstam kā kandidāts.

Saraksta augšgalā ir iekļuvuši homonīmi adverbi un atbilstošās partikulas, saikļi, kuriem nevar nošķirt, lietojuma biežumu vienā vai otrā funkcijā, nozīmē. Iespējams, arī tos nepiciešams izņemt no saraksta, jo adverbu funkcijā tie noteikti būs retāk, piemēram, “tieši”, “vēl”.

2.2. Sinonīmu kopas

Angļu WordNet veidotāji norāda, ka iespēja aizstāt vārdus kontekstā ir vēlama prasība sinonīmu kopu izveidē, bet tā tomēr nav obligāta, runājot par jebkuru kontekstu. Šādu jebkurā kontekstā aizstājamu sinonīmu valodā ir maz, tos sauc par absolūtiem sinonīmiem. Atbilstošāks apgalvojums ir, ka sinonīmi vienā sinonīmu kopā ir aizstājami **dažos** kontekstos. Svarīgākais, lai sinonīmu jēdzieniskā satura galvenie komponenti sakrīt un visa sinonīmu kopa reprezentē vienu leksisku konceptu.
Absolūtie sinonīmi un salīdzinoši tuvi sinonīmi veido WordNet galveno vienību - “sinsetu” (no angļu ‘SynSet’, kas no ‘set of synonyms’) jeb sinonīmu kopu. Pārējās semantiskās saites tiek veidotas jau starp sinonīmu kopām, ne konkrētiem vārdiem.

2.3. Semantisko attieksmju/saišu veidi

WordNet redaktora rīkā starp vārdu nozīmēm un sinsetiem iespējams norādīt semantiskās attieksmes, novelkot starp nozīmēm saites. Viena nozīme var saistīties ar dažādām citām nozīmēm ar dažāda veida saitēm. Pagaidām pieejamas šādas semantiskās saites: sinonīmija, aptuvena sinonīmija, antonīmija, hiponīmija, metonīmija, gradācijas jēdzienu grupa, “saistīts ar”.

2.3.1. Sinonīmija

Sinonīmija latviešu valodniecībā līdz šim ir lietota ļoti plašā izpratnē, iekļaujot gan absolūtos sinonīmus, gan vārdus, kuri drīzāk ir hiponīmiskās attieksmēs (virsjēdziens un apakšjēdziens), izteikti darbības vārdiem, kas nosauc dažāda veida pārvietošanos. R. Veidemane min, ka sinonīmi var atšķirties pēc nozīmes elementiem, kas nav tās galvenie komponenti – pēc jēdzieniskām niansēm, stilistiskās, emocionālās nokrāsas, vides, vietas vai laika pazīmes.
Robežas noteikšanas kritēriji, kad vārds jau ir iekļaujams sinonīmu kopā un kad nav, visticamāk atklāsies projekta gaitā. Saistāmībai ir liela nozīme. No vienas puses tiek sagaidīts, lai vairumā piemēru saistāmība būtu tāda pati vai līdzīga. Tai pašā laikā ir gadījumi, kad tā nav, tomēr nozīme tiešām sinonīmiska, it sevišķi, ja vienu un to pašu nozīmi izsaka gan ar vienu vārdu, gar ar citu vairāku vārdu savienojumu. Kritērijs nav absolūts.
Saistāmība un nozīme sakrīt, piemēram, sinsetos ‘*izdarīt, padarīt, paveikt’* un ‘*nomirt, atstirināties, aiziet mūžībā, aiziet pie tēviem’*.
Vārdu “gribēt” un “gribēties” nozīmes ir sinonīmiskas, taču to sintaktiskā saistāmība nesakrīt, tāpat arī vārdiem “kārot” un “kāroties”.

2.3.2. Hiponīmija - virsjēdziens un apakšjēdzieni

Hiponīmija ir semantiska attieksme, kas parāda, kā vārdi ir sakārtoti leksiskajā hierarhijā. Tā ir galvenais hierarhiskās struktūras veidotājs leksiskajā tīklā. Svarīgi, ka šī attieksme veidojas

starp leksiskiem konceptiem jeb sinonīmu kopām, nevis starp noteiktiem vārdiem, jo vārdi var būt daudznozīmīgi. Tā kā lietvārdiem parasti ir viens virsjēdziens, leksikogrāfi to bieži iekļauj definīcijā. Tāpēc arī ir mēģināts šādus virsjēdzienus iegūtu automātiski no definīcijām. Latviešu valodas tīklam tas tiks darīts manuāli. Nozīmes skaidrojumā lietotais vārds bieži veido hiponīmiskas attieksmes ar skaidrojamo vārdu, tāpēc tas jāizvērtē kā hiperonīma kandidāts.
Tiek pieņemts, ka vienmēr ir iespējams noteikt starpību starp hiponīmiju un sinonīmiju. Praksē gan tas ne vienmēr ir tik skaidrs, bet WordNet veidotāji pasaulē redz risinājumu. Piemēram dots (Wordnet grāmata, 27.lpp.), ja board var būt gan dēlis, gan arī dažos kontekstos tieši sērfošanas dēlis, tad veidojas virsjēdziens dēlis 1.nozīme un apakšjēdziens sinonīmu kopa, kas sastāv no sērfošanas dēlis un dēlis 2.nozīme, kurā dēli izmanto tieši kāda konkrēta izstrādājuma apzīmēšanai. Tas ir piemērs grūtībām noteikt hiponīmiju.
Angļu valodas WordNet sadala lietvārdus vairākās hierarhijās, katrai no tām ir savs sākotnējais virsjēdziens. Viens no pamata kritērijiem virsjēdziena noteikšanai - tas aptver visus kategorijā ietilpstošos lietvārdus. Sākumā izveidotas 25 kategorijas, no kurām 11 ir pamata virsjēdzieni - lai attēlotu leksisko hierarhiju, kad no konkrētiem specifiskiem jēdzieniem saites aiziet uz plašākām semantiskām grupām, un vēl plašākām grupām. Parasti vairāk par 12 līmeņiem hierarhijā neveidojas. Latviešu valodas leksiskā tīkla izveidē nolemts iet pretējo ceļu, nenoteikt uzreiz gatavu semantisko hierarhiju, kurā mēģināt nozīmes piekārtot, bet gan ļaut hierarhijai veidoties pašai ar hiponīmisko attieksmju palīdzību. Pamatojums - lai notvertu valodas lietotāju dabisko valodas izjūtu par jēdzienu hierarhiju, kas tautām un valodām var atšķirties. Protams, iespējams, ka rezultāts vienalga veidosies ļoti līdzīgs semantikas literatūrā aprakstītājām hierarhijām.
Virsjēdziens savā jēdzienā ietver apakšjēdziena jēdzienisko saturu, piemēram, sinonīmu kopa *bērns, atvase* ir vispārīgāka un ietver sevī hiponīmu *dēls*; *pārvietoties* ir, piemēram, verbu *iet*, *peldēt* hiperonīms. Apakšjēdziens ir specifiskāks par virsjēdzienu; tie var atšķirties ar kvalitatīvām vai kvantitatīvām īpašībām. Hiponīmija var veidoties arī vārddarināšanas ceļā, piemēram, sinseta *darīt, veikt* apakšjēdziens ir sinsets *izdarīt, paveikt*.

2.3.3. Meronīmija un tās izpratne

Meronīmija latviešu valodniecībā maz pētīta. Pagaidām meronīmijas saites tiek vilktas dažiem lietvārdiem, piemēram, starp vārdiem *dzīvoklis - istaba* un *istaba - pusistabas,* arī *likums - pants*. Meronīmija starp verbiem nepastāv.

2.3.4. Aptuvena sinonīmija

Aptuvenā sinonīmija tiek izmantota, lai izveidotu saiti starp ļoti tuvām nozīmēm, kuras tomēr praktiski nekad nevar aizstāt vienā kontekstā. Projekta otrajā pusē nepieciešams izvērtēt, kādas nozīmes nonākušas šajā kategorijā un vai nav nepieciešams precizēt sinonīmu kopas izveides principus, lai kopas paplašinātu, vai gluži otrādi - precizētu, kāda veida tuvās nozīmes ietilpst šajā kategorijā.

2.3.5. “Saistīts ar”

Šobrīd šī saite tiek izmantota gadījumos, kad gribam norādīt semantisko saistību, bet tā neatbilst nevienai no iepriekš minētajām semantiskajām attieksmēm, tomēr ir skaidrs, ka pastāv semantiskās attieksmes starp nozīmēm. Kategorija noteikti izmantojama turpmākiem pētījumiem par citām iespējamām semantiskajām saitēm, kuras varētu ieviest latviešu valodas leksiskajā tīklā nākamos projektos.

2.4. Sastatīšana ar Princeton WordNet

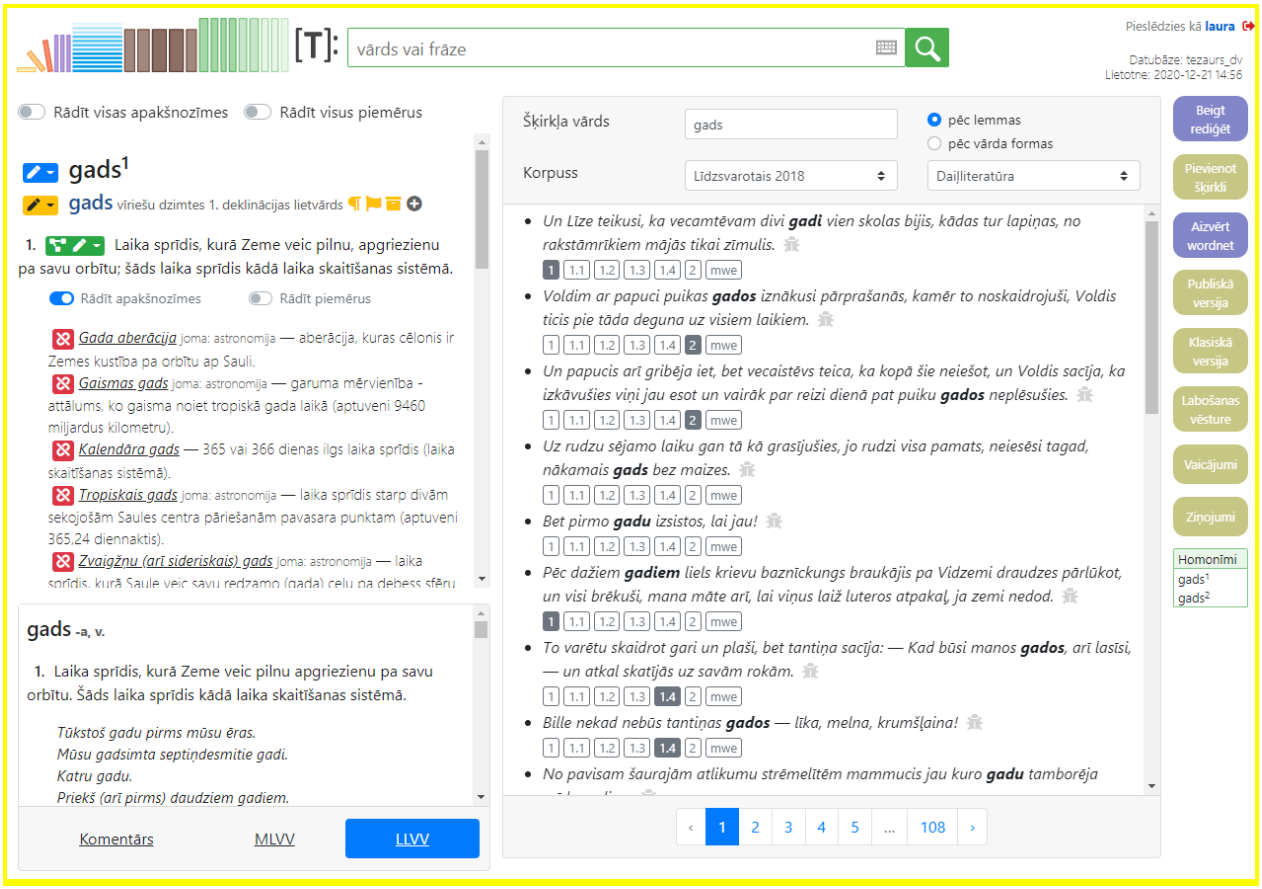
Sastatīšanai tika izveidotas papildu funkcijas tēzaura un WordNet redaktorā. Sinonīmu kopu veidošanas logā ir iespēja atzīmēt, ka jāveic meklēšana Angļu wordnetā (meklējumiem angļu valodas specifiskākos vārda gadījumos, jāatzīmē ar zvaigznīti), tad jāievada atbilstošais angļu valodas vārds, kam varētu būt atbilstoša nozīme, un no piedāvātā saraksta jāizvēlās atbilstoša nozīme, ja tāda ir atrasta. Angļu vārdu izvēlas cilvēks pēc savām zināšanām, meklē vārdnīcās vai arī seko ieteikumam.

Sastatīšanā iespējami vairāki gadījumi:

1) latviešu nozīme ar angļu valodas nozīmi atbilst precīzi, tad pievieno atteicīgo angļu sinonīmu kopu ar guļošo astotnieciņa jeb saites pogu;
2) latviešu sinonīmu kopas nozīme ir plašāka - prasās pievienot vairākus angļu sinsetus, jo tie visi atbilst plašajai lv nozīmei - tad izvēlas no bultiņas "LV plašāka nozīme";
3) angļu nozīme daļēji atbilst, bet tajā vēl arī citas specifiskākas lietas - pievienojam šo nozīmi, izvēloties "LV nozīme šaurāka" saites pogu;
4) netiek atrasta neviena atbilstoša angļu WordNet sinonīmu kopa. Tādā gadījumā jāpārbauda vēl citi iespējamie tulkojumi, vai tiem neparādās atbilstoša nozīme. Var būt, ka nav atbilstoša sinseta, tādā gadījumā nepievieno neko;
5) tiek pamanīts, ka angļu wordnet ir vēl citi sinseti tam pašam vārdam, kādi nav mūsu valodā - vai nu neesam iedomajušies, un tādā gadījumā jāpievieno jauna nozīme vai nozīmes nianse arī latviešu valodas vārda šķirklī, vai arī angļu valodā tam vārdam ir vēl cita nozīme, kāda nepiemīt atbilstošam mūsu valodas vārdam, un to nepievienojam. Šeit svarīgi rūpīgi pārdomāt, lai neieviestu latviešu valodā tādas nozīmes nianses, kādas mēs paši dabiski nebūtu šķīruši;
6) Latviešu valodas nozīmei atbilst divas vai vairākas angļu valodas nozīmes, kurām , mūsuprāt, būtu jābūt sinsetā, nesaprotam, kāpēc nošķirts - reti gadījumi, iespējams, ka Princeton datu kļūda, paši nav apvienojuši. Tādā gadījumā liekam abas nozīmes klāt ar ekvivalences saiti.

3.Rīks leksiskā tīkla veidošanai

Nozīmju pārskatīšanai tiek izmantots Tēzaura redaktora rīks, kas papildināts ar īpašām funkcijām. Tajā ir iespējams 1) rediģēt šķirkli, mainot vārda nozīmes, to secību un skaidrojumu, 2) aplūkot nozīmju skaidrojumu divās galvenajās latviešu valodas skaidrojošajās vārdnīcās, 3) aplūkot piemērus ar šķirkļa vārdu no dažādiem korpusiem un pievienot tos atbilstošām šķirkļa vārda nozīmēm vai ar šķirkļa vārdu saistītiem stabiliem vārdu savienojumiem, 4) veidot sinonīmu kopas un citas semantisko attieksmju saites starp dažādu vārdu atsevišķām nozīmēm (skat. 1. attēlu).



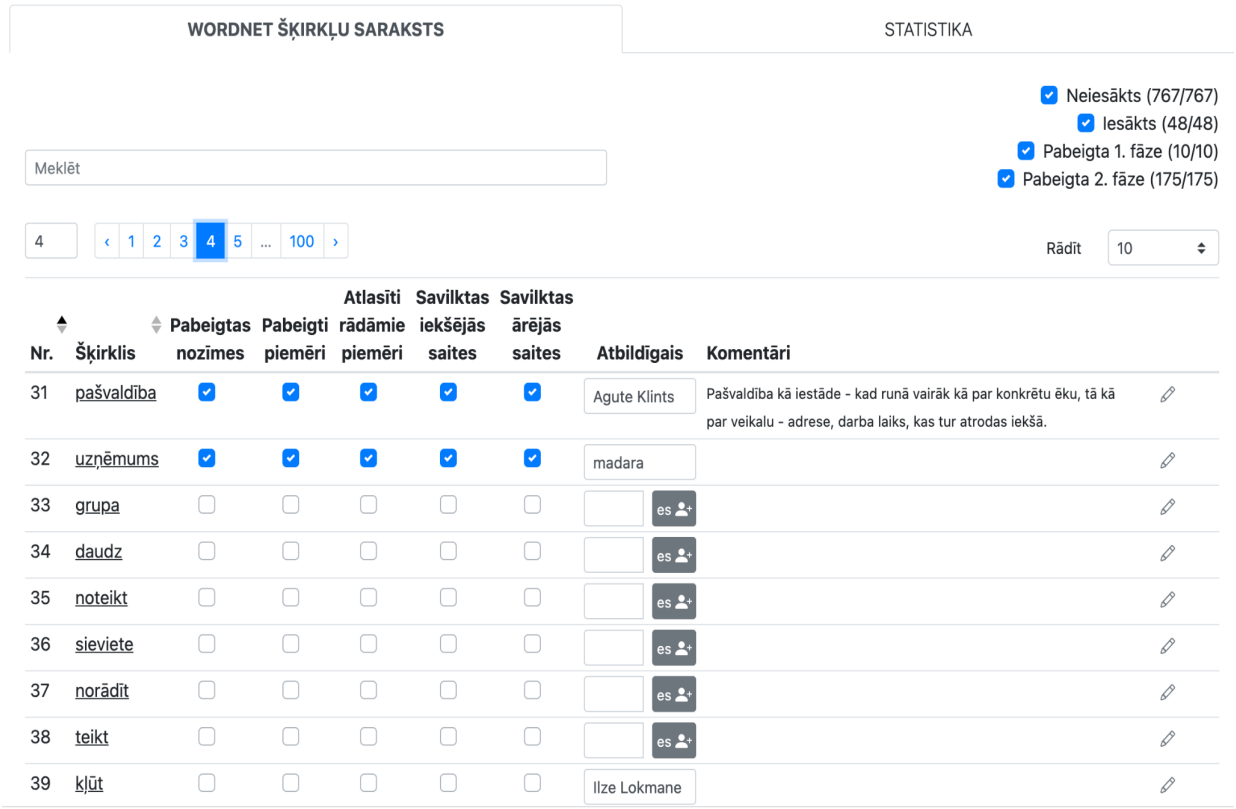
1. attēls. Tēzaura redaktora rīks

Sīkās nianses, kas rīkā ieviestas projekta laikā:

- Piemērus pievieno, uzspiežot uz nozīmes vai apakšnozīmes numuriņa zem piemēra. Ja pārdomā, var nospiest vēlreiz vai uz cita numuriņa, un piemērs tiks noņemts vai pārvietots.
- Šķirklī esošajiem stabiliem vārdu savienojumiem pievienot piemēru var, uzspiežot uz podziņas MWE zem piemēra - tad atveras atsevišķs logs, kurā var izvēlēties, kuram stabilam vārdu savienojumam no šķirkļa pievienot šo piemēru. Ņemot vērā, ka stabils vārdu savienojums ir atsevišķs šķirklis, tā bija būtiska funkcija, lai nebūtu jāiet katru reizi uz konkrēto šķirkli, lai pievienotu piemēru.

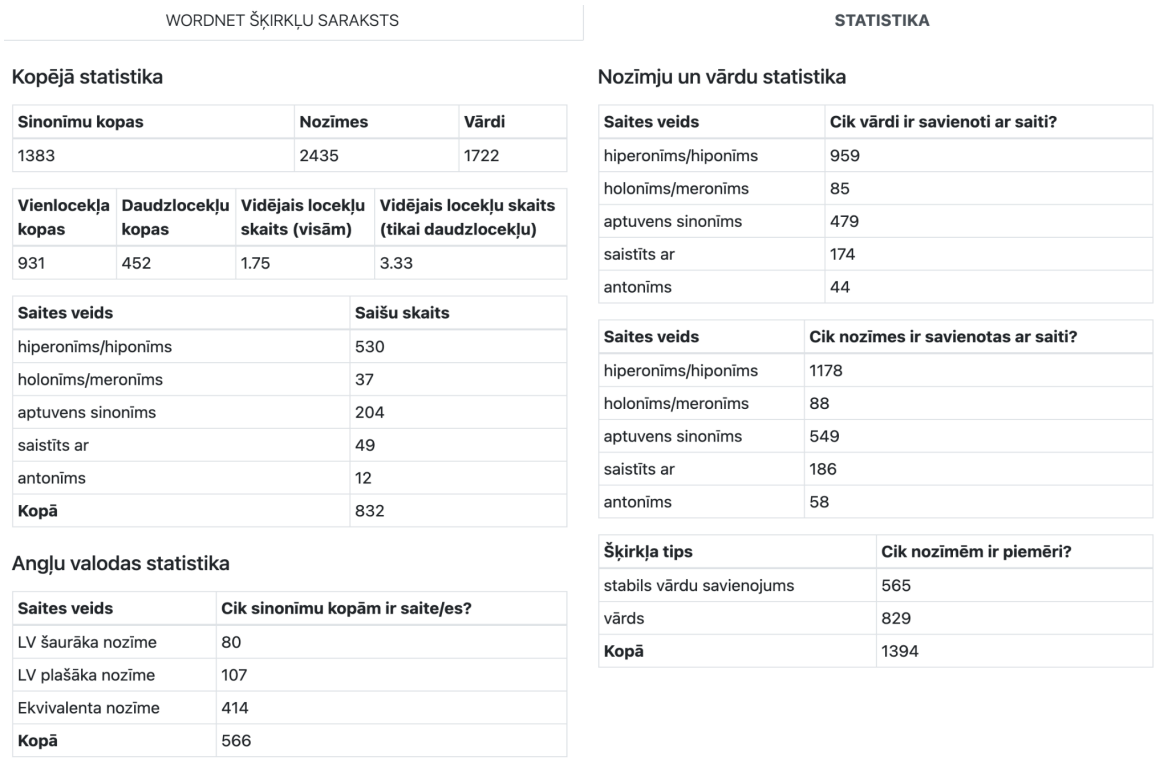
- Lejā kreisajā pusē ir komentāru lauks, kurā rediģētājs var pievienot jebkādas piezīmes par šķirkli un darbu ar to, lai pats nākamā reizē varētu atcerēties vai arī kāds cits var izlasīt, kādas problēmas un pārdomas bija attiecīgajā šķirklī.
- Piemērus pievieno vairāk, nekā plānots rādīt tēzaura lasītājam, tāpēc viens no rediģēšanas procesiem ir norādīt, kuri piemēri tiks publiksoti - to izdara, spiežot rediģēšanas režīmā uz piemēra galā esošās acs ikonas - ja tā ir nosvītrota, piemēru nerādīs, bet ja pelēka - piemērs būs redzams tēzaura nākamajā relīzē publiski.
- Gan nozīmju, gan nozīmes nianšu, gan piemēru secību var mainīt.
- Korpusos var meklēt gan pēc pamatformas jeb lemmas, gan pēc noteiktas vārdformas. Tāpat var meklēt pēc divām vai vairāk lemmām vai vārdformām blakus.
- Līdzsvarotajā korpusā ir iespējams meklēt atsevišķā žanra sadaļā vai visās kopā.
- Projekta laikā tika pievienots arī jaunākais interneta korpuss Tīmeklis 2020. Jāņem vērā, kad tas ir milzīgs un nav atmesti dublikāti, tāpēc meklējumi aizņem ilgu laiku un rezultāti mēdz atkārtoties.
- WordNet piemēru rediģēšanas skatā var redzēt arī homonīmus labajā pusē, tāpēc, ja piemērs atbilst homonīmam, nevis šķirkļa vārda nozīmēm, var pārslēgties uz homonīmu un pievienot piemēru tur.
- Lai rediģētāji nepievienotu pārāk daudz piemērus no viena avota, tika ieviesta funkcija, ka, ar peli uzbraucot uz piemēra, var redzēt tā avotu, kā arī iespēja pašam ievadīt, kuru piemēru lapaspusi skatīties, lai varētu ātrāk pāriet uz citu avotu.
- Lai šķirklis joprojām būtu pārskatāms, ir iespējams visam šķirklim un atsevišķi nozīmēm norādīt, vai šobrīd rādīt nozīmes nianses un vai rādīt piemērus. Visu var paslēpt, lai labi pārskatītu pamatnozīmes šķirklī.
- Šķirkli var skatīt Tēzaura redaktora skatā un WordNet skatā, rediģētājs var mainīt starp tiem. Tēzaura redaktora skatā sākotnēji nerāda paslēptos piemērus, bet ir iespējams ieslēgt pogu, lai rāda visus šķirklim pievienotos piemērus. Tiesa, pie katras šķirkļa izmaiņas, kad notiek saglabāšana un ielāde no jauna, visi nerādāmie piemēri atkal ir paslēpti.
- Netiek paslēpti piemēri, kad notiek šķirkļa saglabāšana un ielāde no jauna (t.i., atrisināta iepriekšējā punktā minētā problēma).
- Pievienota opcija paslēpt logu, kurā rādās komentāri / vārda nozīmes no MLVV / LLVV, lai šķirklis būtu vieglāk pārskatāms arī WordNet skatā.

Speciāli WordNet vajadzībām ir izveidota tēzaura atsevišķa sadaļa “Wordnet saraksts”. To atverot, ir pieejamas divas cilnes. Vispirms tiek parādīts to vārdu saraksts, kurus ir plānots projekta gaitā apstrādāt (šobrīd 25.05.2021. Tur ir iekļauti 1000 līdzsvarotajā korpusā biežāk lietotie vārdi), bet otrā cilnē ir pieejama statistika par projekta laikā apstrādiem vārdiem, to nozīmēm, semantiskajām saitēm.
Sarakstā esošajiem vārdiem ir paredzēti vairāki darba etapi, kuriem rediģētājam jāiet cauri un jāatzīmē, kad tie ir pabeigti. Šie etapi ir 5 - nozīmju sakārtošana, piemēru pievienošana, piemēru atlasīšana publiskajai versijai, iekšējo saišu savilkšana un ārējo saišu savilkšana. Ja vārdam ir pabeigti pirmie 4 etapi, tam ir pabeigta pirmā fāze. Ja vārdam ir pievienotas arī angļu valodas saites, tad ir pabeigta arī otrā fāze. Vārdus iespējams meklēt pēc pamatformas vai marķētāja vārda, ievadot to meklētāja laukā. Vārdus iespējams arī atlasīt pēc tā, vai tiem ir jau pabeigts kaut viens etaps (iesākti), vai ir pabeigta pirmā fāze vai arī otrā fāze. Katram vārdam ir iespējams norādīt marķētāju jeb valodnieku, kas ar to strādā, lai apstrādē esošu vārdu nepaņemtu cits marķētājs. Tāpat katram vārdam ir komentāra lauks, kas sakrīt ar piemēru pievienošanas logā esošo komentāra lauku - labojot vienā vietā, komentārs parādīsies abās



(skat. 2. attēlu).
2. attēls. Apstrādājamo vārdu saraksta logs rīkā

Statistikas cilnē pieejama plaša informācija par to, cik vārdiem un nozīmēm ir izveidotas sinonīmu kopas, cik saites to starpā, cik daudz katra veida saišu, utt. (skat. 3. attēlu)



WORDNET ŠĶIRKĻU SARAKSTS

STATISTIKA

Kopējā statistika

Sinonīmu kopas	Nozīmes	Vārdi
1383	2435	1722

Vienlocekļa kopas	Daudzlocekļu kopas	Vidējais locekļu skaits (visām)	Vidējais locekļu skaits (tikai daudzlocekļu)
931	452	1.75	3.33

Saites veids	Saišu skaits
hiperonīms/hiponīms	530
holonīms/meronīms	37
aptuvens sinonīms	204
saistīts ar	49
antonīms	12
Kopā	832

Angļu valodas statistika

Saites veids	Cik sinonīmu kopām ir saite/es?
LV šaurāka nozīme	80
LV plašāka nozīme	107
Ekvivalenta nozīme	414
Kopā	566

Nozīmju un vārdu statistika

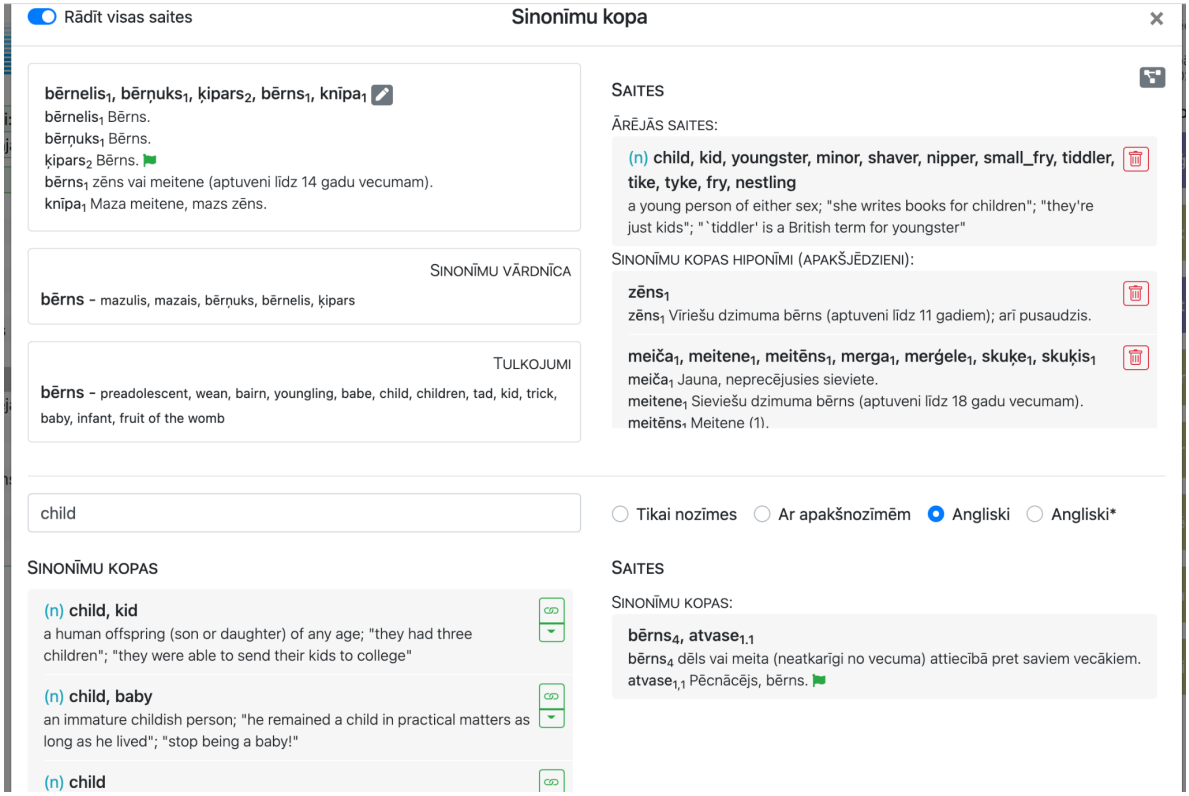
Saites veids	Cik vārdi ir savienoti ar saiti?
hiperonīms/hiponīms	959
holonīms/meronīms	85
aptuvens sinonīms	479
saistīts ar	174
antonīms	44

Saites veids	Cik nozīmes ir savienotas ar saiti?
hiperonīms/hiponīms	1178
holonīms/meronīms	88
aptuvens sinonīms	549
saistīts ar	186
antonīms	58

Šķirkļa tips	Cik nozīmēm ir piemēri?
stabils vārdu savienojums	565
vārds	829
Kopā	1394

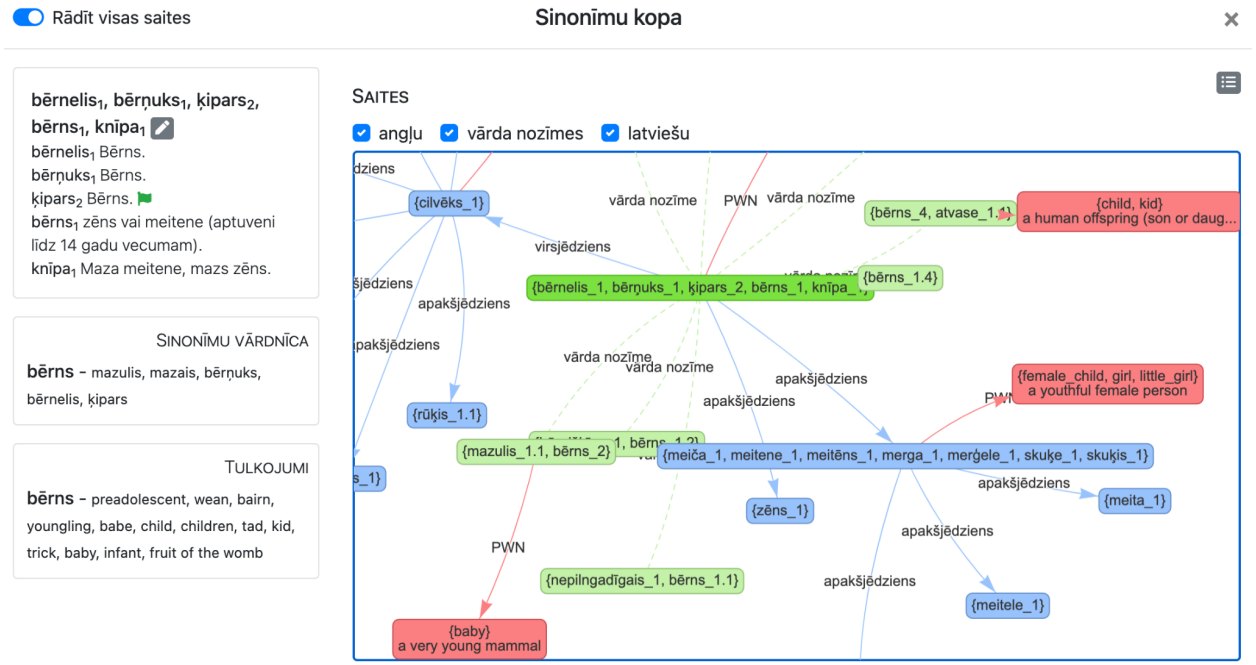
3. attēls. Leksiskā tīkla sinonīmu kopu un semantisko saišu statistikas logs rīkā

Nozīmīga funkciju daļa ir atsevišķa **logā "Sinonīmu kopas un saites"**, kas atveras, ja rediģēšanas režīmā uzspiež uz zaļas grafa zīmītes pie nozīmes. Šajā logā ir a) sinonīmu ieteikumi no vārdnīcas, b) iespēja meklēt citas vārdu nozīmes un apakšnozīmes Tēzaurā, c) atrastās nozīmes pievienot sinonīmu kopai ar aktīvo nozīmi vai arī norādīt kādu citu semantiskās saites veidu ar to, d) ieteiktie angļu valodas tulkojumi atvērtajam vārdam, e) iespēja meklēt Princeton WordNet atbilstošu sinonīmu kopu un izveidot ar to saiti (ekvivalences saiti, vai norādīt, ka LV nozīme ir plašāka vai šaurāka (sk. 2.4.)).
Uzspiežot uz atrastās angļu nozīmes ir iespējams redzēt tai pievienotās saites latviešu valodā, ja tādas jau ir.



4. attēls. Sinonīmu kopu un semantisko saišu izveides logs rīkā

Tāpat būtisks jaunums rīka attīstībā ir projekta laikā izveidotais leksiskā tīkla grafiskais attēls, kuru iespējams aplūkot sinonīmu kopas saišu logā. Tajā var redzēt visa šķirkļa nozīmes, to saites ar citām nozīmēm līdz otrajam līmenim, kā arī sastatīšanas saites ar angļu WordNet sinonīmu kopām. Tādā veidā valodniekam ir iespējams ērti pārlūkot, vai izveidotās saites nav pretrunā viena ar otru, ar kādiem virsjēdzieniem katra nozīme saistās.



5. attēls. Semantisko saišu grafs sinonīmu kopas rediģēšanas logā

Literatūras saraksts:

- Baldunčiks, Juris. 2012. Pārskats par nozīmīgākajām vienvalodas vārdnīcām: skaidrojošās vārdnīcas, svešvārdu vārdnīcas, etimoloģijas vārdnīca, slenga vārdnīca. // *Vārdnīcu izstrāde Latvijā*. Pētījums 1991-2010. Rīga: LVA, 248 lpp. [tiešsaiste]
- Bāliņa, Rita (red.). 2006. Ievads // *Latviešu valodas vārdnīca*. Rīga: Avots, 7–11 lpp.
- Ceplītis, Laimdots, Sangrevica, Melita (atb. red.). 1972-1996. *Latviešu literārās valodas vārdnīca*. 1. - 8. sēj. Rīga: Zinātne
- Cruse, Allan. 2004. *Meaning in Language. An Introduction to Semantics and Pragmatics.* Oxford: Oxford
- Goddard, Cliff. 1998. *Semantic Analysis. A Practical Introduction.* Oxford, New York: Oxford University Press
- Guļevska, Dainuvīte (atb. red.). 1987. *Latviešu valodas vārdnīca*. Rīga: Avots
- Henrich, Verena, Hinrichs, Erhard. 2010. "GernEdiT - The GermaNet Editing Tool" // *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010)*, Valletta, Malta, 2010. g. maijs, 2228-2235 lpp.
- Kuplā, Ieva. 2014. Kompleka pieeja daudznozīmīga verba semantiskās struktūras analīzē. // *Valoda: Nozīme un forma*. 4. Kategoriju robežas gramatikā. Rīga: LU Akadēmiskais apgāds, 77-92 lpp.
- Laua, Alise. 1969. *Latviešu leksikoloģija*. Rīga: Zvaigzne, 254 lpp.
- Lokmane, Ilze. 2015. Vārdkopu reducēšanās un konversija: Emīlijas Soidas interpretācija. // *Valoda: nozīme un forma*. 6. Valodas sistēma un lietojums. Rīga: LU Akadēmiskais apgāds, 26-34 lpp.
- *Mūsdienu latviešu valodas vārdnīca*. Pieejams: www.tezaurs.lv/mlvv [tiešsaiste]
- Ozola, Ārija. 1970. Par verba transitivitāti un intransitivitāti. *Latviešu valodas un literatūras problēmas*. Rīga: Zvaigzne, 95–102 lpp.

- Ozola, Ārija. 2010. Divdabju ar izskaņām -ošs, -oša lietojums mūsdienu latviešu valodā un atspoguļojums skaidrojošajās vārdnīcās. // *Vārds un tā pētīšanas aspekti*, 14(1). Liepāja: LiePA, 216-222 lpp.
- Paducheva, Elena. 2004. *Dinamicheskie modeli v semantike leksiki*. Moskva: Yazyki slavyanskoj kul'tury
- Pustejovsky, James. 1998. *The generative lexicon.* Cambridge, Massachusetts: The MIT Press
- Saeed, John I. 2000. *Semantics*. Oxford, Massachusetts: Blackwell Publishers.
- Skujiņa, Valentīna (red.). 2007. *Valodniecības pamatterminu skaidrojošā vārdnīca*. Rīga: LU Latviešu valodas institūts
- Šmidebergs, Imants. 2009. Verbi un to divdabji ar priedēkli ne- skaidrojošajās vārdnīcās. // *Vārds un tā pētīšanas aspekti*, 13(1). Liepāja, LiePa, 161-166 lpp.
- Veidemane, Ruta. 1970. *Latviešu valodas leksiskā sinonīmija*. Rīga: Zinātne
- Wierzbicka, Anna. 1996. *Semantics. Primes and universals*. Oxford, New York: Oxford University Press
- Zuicena, Ieva. 2008. Terminu atveides problēmas skaidrojošajās vārdnīcās. // *Vārds un tā pētīšanas aspekti*, 12(2). Liepāja: LiePA, 277 lpp.
- Zuicena, Ieva. 2010. Vārda nozīmes skaidrojums “Mūsdienu latviešu valodas vārdnīcā”. // *Vārds un tā pētīšanas aspekti*, 14(1). Liepāja: LiePA, 369-374 lpp.