

Latviešu valodas verbu valences marķēšana *Annotation of verbal valency in Latvian*

Gunta Nešpore, Baiba Saulīte

Latvijas Universitāte

Matemātikas un informātikas institūts

Raiņa bulvāris 29, Rīga, LV-1459

E-pasts: {gunta.nespore, baiba.saulite}@lumii.lv

Rakstā apskatītas latviešu valodas verbu valences marķēšanas galvenās problēmas: aplūkoti izvēlētie marķēšanas līmeņi – sintaktiskā valence, semantiskā valence un leksiskā valence, tāpat analizēts izmantoto semantisko lomu kopums, kā arī ieskicēti izsecinātie verba valences modeļi un tajos iekļaujamās verba apkaimes galvenās semantiskās lomas. Īsi raksturots arī programmrīks, kas izveidots marķēšanai un tās rezultātu apstrādei. Darba rezultātā lietotājiem plānots piedāvāt vairāku veidu datus, piem., verba valences modeļus – verba apkaimei raksturīgās semantisko lomu kombinācijas –, kā arī apstrādātu tekstu, kur marķēti verba paplašinātāji ar norādītām semantiskajām lomām, tām atbilstošajām morfoloģiskajām formām un ontoloģiskajām kategorijām.

Atslēgvārdi: semantiskā valence, sintaktiskā valence, semantiskās lomas, valences modeļi, apkaimes marķēšana.

LU Matemātikas un informātikas institūtā, manuāli marķējot latviešu valodas materiālu pamatā no Līdzsvarotā mūsdienu latviešu valodas tekstu korpusa, tiek veidota verbu valences datubāze¹. Marķējamais valodas materiāls sastāv no divām daļām. Tajā ir iekļauti apmēram 300 minētajā tekstu korpusā visbiežāk sastopamie verbi, kā arī viena verbu tematiskā grupa – pārvietošanās verbi. Raksta centrā ir marķēšanas metodikas izveide, tajā ņemta vērā citu valencei veltītu valodas resursu pieredze, kas atbilstoši nepieciešamībai kombinēta un papildināta. Tie ir dažāda apjoma un izmantojamības resursi, kas verbu apkaimei apraksta dažādos aspektos, piem.: *Lietuviešu valodas verbu saistāmības vārdnīca* (Sližienė 1994), čehu valodas saistāmības vārdnīca *Vallex* (Lopatková et al. 2006), angļu valodas situāciju semantikas leksikons *FrameNet* (Ruppenhofer et al. 2006), kā arī Patrika Henksa (*Patrick Hanks*) veidotā *Angļu valodas verbu modeļu vārdnīca* (*Pattern Dictionary of English Verbs*; CPA 2004), kas iecerēta kā papildinājums *FrameNet* sistēmai.

Latviešu valodas verbu marķēšanā izvēlēti trīs apraksta līmeņi (sk. nodaļu *Verbu valences apraksta līmeņi*), bet rezultātā piedāvāti verbu valences modeļi ar konkrētam verbam tipiskajām semantiskajām lomām.

¹ ERAF projektā Nr. 2011/0009/2DP/2.1.1.1.0/10/APIA/VIAA/112

Verbu apkaimes marķēšanai ir īpaši izveidots programmrīks (autore Ginta Garkāje), ar kura palīdzību tekstā sastopamās semantiskās lomas var iekrāsot dažādās krāsās, kā arī pievienot informāciju par morfoloģiskajām formām, ar ko izteiktas semantiskās lomas (sk. 1. att.), un apkopot marķēšanas rezultātus (sk. turpmāk).

Marķējot verba apkaimei, tiek nošķirtas arī verba nozīmes (sk. 1. att. – taisnstūri kreisajā pusē, kas atšifrēti labās puses lejasdaļā). Pašlaik tas tiek darīts pamatā empīriski – pēc marķētāja ieskatiem, izmantojot marķēto valodas materiālu, verbu apkaimei, arī vārdnīcu datus – pēc iespējas izvairoties no pārāk sāka dalījuma nozīmēs. Piem., ja vārdnīcā ir minētas verba nozīmes, ko tekstā ir grūti nošķirt, vai ir grūti izprast/definēt atšķirības starp šīm nozīmēm (valodas materiālā kopumā, nevis atsevišķos gadījumos), ir pamats apšaubīt šāda dalījuma jēgu. Kad marķēšana visos trīs marķēšanas līmeņos būs pabeigta, verbu lietojumu tekstā plānots grupēt pēc saistāmības vai vismaz eksperimentāli izmēģināt šāda grupējuma lietderību. Marķēšanu sākot, tiek izvirzīta arī hipotēze – ja dažādos teikumos, kur kāds verbs ir lietots, neatšķiras verba apkaime – vismaz verba paplašinātāju semantiskās lomas un leksiskās nozīmes kategorijas –, nevajadzētu atšķirties arī nozīmei, kādā šis verbs ir lietots.

1. Verbu valences apraksta līmeņi

Sākotnēji verba apkaime tiek aprakstīta divos līmeņos: pirmkārt, tiek marķētas morfoloģiskās formas, netieši marķējot arī teikuma locekļus jeb sintaktisko valenci, otrkārt, marķētajām formām tiek pievienotas semantiskās lomas. Nākamajā marķēšanas kārtā tiks marķēta verba paplašinātāju vispārinātā leksiskā nozīme.

1.1. Sintaktiskā valence

Valodas materiālā tiek marķēti verba paplašinātāji, kuri atrodas verba apkaimē un ar kuriem izteiktas semantiskās lomas.

Šī raksta mērķis nav definēt un nošķirt verba obligātos un neobligātos paplašinātājus (par sintaktisko obligātumu sk., piem., Lokmane 2004), tomēr, veidojot verba valences modeli, ir svarīgi saprast, kādus verba paplašinātājus tajā iekļaut. Verba paplašinātājus ir pieņemts iedalīt valentajos un nevalentajos paplašinātājos.

Analizējot sintaktisko valenci, visbiežāk tiek runāts par verba argumentu struktūru, jo argumenti tiek uzskatīti par teikuma locekļiem, ko teikumā piesaista izteicējs, tātad teikuma priekšmets, papildinātājs, apstāklis (sk. VPSV 2007, 44). Argumenti var būt obligāti un neobligāti, piem., teikumā *Policija^{Nom} masu medijiem^{Dat} informāciju^{Acc} <do> nelabprāt^{Adv}* obligātie argumenti ir teikuma priekšmets nominatīvā un papildinātājs akuzatīvā, savukārt neobligātie – netiešais papildinātājs datīvā un veida apstāklis.

Marķējot sintaktisko valenci latviešu valodā, izmantota nedaudz atšķirīgā čehu valodnieku (Sgall 1986, 127; Hajičová 1993, 16–17) pieeja, kas tuvāka verba saistāmības pamatlicēja – Lisjēna Tenjēra (*Lucien Tesnière*) – koncepcijai (Tesnière 1988, 117–118), kur visi valentie locekļi dalīti aktantos, t. i., dzīvās būtnēs un priekšmetos, kas piedalās procesā, un sironstantos, kas norāda uz

<Lūgšim>	<Dievu>	lūgšim, – mudināja sievieti ar vistas knābim līdzīgo	
V1	Acc		
Esi stipra, <lūgšu>	<Dievu>	par tevi, bērniņ... tu man vienmēr esi bijusi kā pašas meita	
V1	Acc	par-Acc	
Vīna	<lūdza>	domet materiālo palīdzību, lai varētu segt izdevumus par dzīvesvietu	
Nom	V3	Dat	Acc
Pils rajona iedzīvotājiem, komisijas priekšsēdētājs Ainars Vasilis	<lūdza>	domes juristu informēt	
Inf	Nom	V3	Acc
policiju, ka situācijās, kad pārkāpumi nav			
Inf			
Vīnam	<lūdza>	dziedāt vēl un vēl. Noslēgumā gan sekoja vilšanās,	
Dat	V3	Inf	
Zālnī, ka viens mans dzejolis tiešām bijis publicēts, taču man	<lūdza>	izdomāt pseidonīmu.	
	Dat	V3	Inf

1. attēls. Verbu apkaimes marķēšanas rīks.

apstākļiem, kādos process norisinās (Tesnière 1988, 117–118). Čehu valodnieku pieejā aktanti ir iekšējie dalībnieki (*inner participants*), tātad gan obligātie, gan neobligātie argumenti, izņemot apstākļus, bet sirkonstanti ir brīvie paplašinātāji (*free modifications*), tātad apstākļi – gan obligāti, gan fakultatīvi. Čehu valodnieki šādu dalījumu balsta galvenokārt uz diviem kritērijiem – pirmkārt, vai verba apkaimē paplašinātājs var parādīties vienu reizi vai vairākas un, otrkārt, vai paplašinātājs raksturo konkrētu verbu vai var piesaistīties jebkurai situācijai.

T. s. iekšējiem dalībniekiem raksturīga piesaistīšanās tikai konkrētam verbam, turklāt verba apkaimē katrs no tiem var parādīties tikai vienu reizi² (*Latvijas prezidente Vaira Vīķe-Freiberga*_{Nom} pirmdien intervijā *Lielbritānijas raidkorporācijai BBC*_{Dat} <skaidroja> *iemeslus*_{Acc} savas kandidatūras izvirzīšanai *ANO ģenerālsekretāra amatam*). Turpretim brīvie paplašinātāji³ galvenokārt ir fakultatīvi, turklāt vienā teikumā var parādīties vairāk nekā vienu reizi. Tipiski brīvie paplašinātāji ir ar vietas un laika nozīmi, jo tie var raksturot gandrīz jebkuru situāciju (piem., divi vietas apstākļi teikumā *Protams, beigās saimes istabā*_{Loc} *uz galda*_{uzAcc} <gaida> *silta un veselīga zāļu tēja*).

Katrā valodas piemērā tiek marķēti visi „iekšējie dalībnieki” (gan sintaktiski obligātie, gan neobligātie, piem., verbam *tirgot* tiek marķēts gan tiešais papildinātājs, gan netiešais papildinātājs – *Vajadzētu sodīt pārdevējus, kas*_{Nom} <*tirgo*> *tabakas izstrādājumus*_{Acc} *nepilngadīgajiem*_{Dat}). Atsevišķu verbu struktūrā obligāti var būt arī brīvie paplašinātāji, piem., *Pēc stundas Aleksis* <*nokļuva*> *pie sava dzīvokļa durvīm*_{pieDat} – arī tie tiek marķēti (visi minētie parasti atbilst centrālajām semantiskajām lomām, sk. turpmāk).

Fakultatīvi brīvie paplašinātāji tiek marķēti pēc marķētāja ieskatiem, bet tie netiek iekļauti verba valences modelī, piem., *Pavisam nesen*_{Adv} *mana draudzene* <*apmeklēja*> *manikīra meistari* .. (perifērās lomas, sk. turpmāk).

Lai iegūtu verba valences tipisko modeli, marķēšanai parasti tiek atlasīti verbi finītās formās. Netiek atlasīti teikumi, kuros predikāta finītā funkcija ir izteikta sintaktiski, t. i., nevis verba formā, bet analītiskā formā (piem., saliktās laika formas, ciešamās kārtas un vajadzības izteiksmes formas). Piem., ciešamās kārtas konstrukcijā gramatiskais objekts jeb tiešais papildinātājs (TĒMA⁴) ir nominatīvā, bet teikuma priekšmets (AGENSS⁵) nav izteikts (.. *finālā tiek* <*atrasts*> *Adas mirušais ķermenis*_{Nom}). Šādas konstrukcijas nav lietderīgi aprakstīt, lai rezultātā neveidotos verba valences modelis, kur verba apkaimei raksturīga TĒMA nominatīvā, bet neparādās AGENSS, kas šī verba dziļajai struktūrai ir raksturīgs.

Marķējot verba valento locekļu morfoloģiskās formas, netieši tiek marķētas arī valento locekļu sintaktiskās funkcijas – lietvārdam akuzatīvā tipiska ir papildinātāja funkcija, lietvārdam lokatīvā – apstākļa (*Vai bērns*_{Acc} <*atrod*> *kāpostu laukā*_{Loc}). Protams, tiek marķētas arī atšķirīgās formas, piem., ja gramatiskais subjekts (šajā gadījumā AGENSS) formāli izteikts nevis ar nominatīvu, bet, piem., ar ģenitīvu,

² Vienlīdzīgi teikuma locekļi un pielikums tiek uzskatīti par vienu valento vienību.

³ Latviešu valodas verba valences aprakstā līdzīgi kā čehu valodas saistāmības vārdnīcā (Žabokrtský 2005, 9) par brīvo paplašinātāju teikumā tiek uzskatīts jebkurš apstāklis.

⁴ TĒMA – objekts, kurš tiek pārvietots vai kura atrašanās vieta tiek raksturota.

⁵ AGENSS – situācijas dalībnieks, kas veic darbību apzināti, pēc savas gribas, ir darbības iniciators.

ko nosaka atkarīgais komponents – visbiežāk apstākļa vārds vai skaitļa vārds –, tas marķējumā tiek norādīts (*Arvien vairāk māsaienieču^{Gen} gatavo un <cep> maizi^{Acc} savā virtuvē ..*), kaut, protams, tipiskā sintaktiskā valence šim verbam būs teikuma priekšmets nominatīvā.

1.2. Semantiskās lomas

Tiek marķētas verba apkaimē sastopamās semantiskās lomas, piem., verbs *dot* piesaista darītāju (AGENSU), tiešo objektu (TĒMU) un adresātu (BENEFICIENSU⁶). Semantisko lomu izpratne un definīcijas ir veidotas, ņemot vērā Džona Saīda (*John I. Saeed*), Frenka R. Palmera (*Frank R. Palmer*), Čārlza Dž. Filmora (*Charles J. Fillmore*), Jurijs Apresjana (*Юрий Апресян*) teorētiskos uzskatus un semantisko lomu definīcijas (Fillmore 1968; Palmer 1994; Апресян 1995; Saeed 2007).

Semantisko lomu marķēšanā svarīgs jautājums ir izmantojamais semantisko lomu kopums, semantisko lomu vispārinājuma pakāpe. Šī jautājuma risināšanā ir vairākas pieejas. Piem., *FrameNet* projektā katras situācijas (verbu grupas) semantiskās lomas (*FrameNet* terminoloģijā – *frame elements* ‘situācijas elementi’) tiek uzskatītas par unikālām, tiek sauktas salīdzinoši konkrētas nozīmes vārdos (piem., *strīdnieks1*, *strīdnieks2*, *strīdnieki*, *iesesls* strīdēšanās situācijā) un sakritība to nosaukumos tiek uzskatīta tikai par sakritību (Ruppenhofer et al. 2006). Populārāka pieeja (no šeit minētajiem – *Vallex*, *Lietuviešu valodas verbu saistāmības vārdnīca*) ir izmantot mazāku skaitu (30–35) vispārīgākas nozīmes semantisko lomu (piem., AGENS, PACIENS, ADRESĀTS) un uzskatīt, ka dažādu verbu apkaimē sastopamo situācijas dalībnieku pildītās funkcijas (semantiskās lomas) var sakrist.

Latviešu valodas verbu valences marķēšanā tiek izmantotas vispārīgākas nozīmes semantiskās lomas. Šī pieeja ir teorētiski sarežģītāka un darbietilpīgāka, jo jāvispārina verbu valento locekļu nozīme, jāizprot, kuras no vispārinātajām semantiskajām lomām ir raksturīgas konkrētā verba apkaimei. Piem., kādu lomu pilda verba *mīlēt* objekts – vai tas ir PERCEPTS vai SATURS? Vai verba *aicināt* subjekts uzskatāms par AGENSU, ja tā ir, piem., valsts, tātad nav tiešā veidā dzīva būtne, kas spējīga uz apzinātu darbību (tomēr to pārstāv cilvēki, kam šādas spējas piemīt), piem., .. *ASV <aicinātu> Latviju uzņemt kādu no Gvantanamo atbrīvotajām personām*. Vispārinātas semantiskās lomas nodrošina lielākas verbu apkaimes (valences modeļu) salīdzināšanas, grupēšanas, analīzes iespējas, kā arī katras semantiskās lomas apraksta iespējas. Darba gaitā reizēm tiek izmantotas arī konkrētākas nozīmes lomas, lai nezaudētu marķējuma korektumu, ja nav īsti skaidrs, kura no vispārinātajām semantiskajām lomām konkrētajā gadījumā būtu atbilstoša. Piem., verba *sāpēt* apkaimē pagaidām nav pilnībā skaidrs, kura no vispārinātajām semantiskajām lomām būtu piemērotākā sāpošās ķermeņa daļas nosaukumiem (*Jau kuro mēnesi man <sāp> auss; Man tik ļoti <sāp> kakls*), tāpēc pašlaik šī informācija tiek saukta par ĶERMEŅA DAĻU. Tas tiek uzskatīts par pagaidu marķējumu, pieņemot, ka, marķējot līdzīgus verbos un savācot vairāk valodas materiāla, attiecīgo informāciju vispārināt (izvēlēties piemērotāko semantisko lomu) varētu būt vieglāk.

⁶ BENEFICIENS – situācijas dalībnieks (dzīva būtne), kura labā darbība tiek veikta.

1.3. Leksiskā nozīme

Kad semantiskās lomas un morfoloģiskās formas jau būs nomarkētas, plānots pievienot informāciju par to, kādas leksiskās nozīmes vārdus verbs parasti piesaista, t. i., norādīt verbu valento locekļu vispārinātu leksisko nozīmi jeb ontoloģiskās kategorijas (*FrameNet* terminoloģijā – semantiskais tips). Piem., verbs *skaldīt* piesaista dzīvu darītāju (AGENSU) un fizisku objektu (PACIENSU⁷), verba *jautāt* AGENSU un ADRESĀTS⁸ parasti ir cilvēks, savukārt verba *ēst* PACIENS ir fizisks objekts.

Šeit galvenā problēma ir līdzīga kā semantisko lomu marķēšanā, tikai ar neskaidrākiem risinājumiem – cik sīki iedalīt leksisko nozīmi (līdz ar to arī pasaulē sastopamās reālijas), cik precīzi norādīt verba apkaimei raksturīgo nozīmi, lai šī informācija palīdzētu atklāt verba nozīmi, bet vienlaikus būtu pietiekami vispārīga, lai būtu precīza un „neierobežotu” verba lietojumu. Piem., kāda informācija būtu norādāma pie verba *domāt* AGENSA? Vai tas būtu tikai cilvēks, jebkura dzīva būtne vai varbūt arī mehānismi, piem., dators, uz ko domāšanu mēdz attiecināt pārnēstā nozīmē? Dažādos projektos leksiskās nozīmes jeb ontoloģisko kategoriju skaits svārstās no mazāk nekā desmit līdz vairākiem simtiem (apskatu par dažādos projektos izmantotajām ontoloģiskajām kategorijām sk. Žabokrtský 2005, 15–29).

Latviešu valodas saistāmības aprakstā plānots rīkoties līdzīgi kā *Lietuviešu valodas verbu saistāmības vārdnīcā* (sk. **Sližienē 1994**), kur izmantotas četras pamatkategorijas – cilvēks (Hum), dzīvs objekts (+ Anim), nedzīvs objekts (-Anim) un abstrakts objekts (Abstr), kā arī papildu informācija par specifiskākas leksiskās nozīmes argumentiem, ja verba nozīme tāds prasa, piem., *uzgērbt* objekts ir fizisks objekts (apģērbs). Tādā veidā plānots papildu ontoloģisko kategoriju kopumu izveidot empīriski, marķējot reālus valodas datus, nevis iepriekš teorētiski definēt iespējamās kategorijas. Lai arī grūti prognozēt šādas papildu kategoriju marķēšanas rezultātus (tos var ietekmēt pieejamo datu apjoms, marķētāja subjektivitāte u. tml.), arī tikai četru pamatkategoriju marķējums palīdz verbu nozīmju nošķiršanā, automātiskā valodas analizē u. tml.

2. Verbu valences modeļi

Izmantojot verbu apkaimes marķējumu, tiek veidoti verbu valences modeļi, centrā liekot semantiskās lomas. Arī šajā ziņā dažādos projektos ir liela daudzveidība – *Lietuviešu valodas verbu saistāmības vārdnīcā* centrā ir morfoloģisko formu kombinācijas, *Vallex* – semantiskās lomas, savukārt *Angļu valodas verbu modeļu vārdnīcā* – ontoloģiskās jeb leksiski semantiskās kategorijas.

Lai izveidotu verbu valences modeļus, jāizvērtē, kāda verbu apkaimē esošā informācija (kuras semantiskās lomas) tajos iekļaujama, līdz ar to – kāda informācija noteikti ir jāmarķē. Semantisko lomu iedalīšanai tiek izmantota *FrameNet* sistēmā izveidotā semantisko lomu tipoloģija, kur tiek šķirtas centrālās un perifērās semantiskās lomas (Ruppenhofer et al. 2006, 26).

⁷ PACIENS – objekts, ko fiziski ietekmē darbība.

⁸ ADRESĀTS – situācijas dalībnieks, kam tiek sniegta informācija.

Semantiskās lomas tiek šķirtas pēc tā, vai attiecīgā informācija ir konceptuāli svarīga situācijā, vai bez tās situācija var pastāvēt, arī tad, ja šī informācija tieši neparādās teikumā.

Centrālās semantiskās lomas raksturo situācijā konceptuāli nepieciešamus elementus, kas atšķir situāciju no citām (Ruppenhofer et al. 2006, 26). Piem., tirdzniecības situācija nevar pastāvēt bez pircēja (AGENSS), pārdevēja (KONTRAGENSS⁹), preces (TĒMA).

Perifēro semantisko lomu nosauktā informācija neatšķir situāciju no cita tipa situācijām, bet sniedz papildinformāciju par situāciju, tās apstākļiem, šīs semantiskās lomas līdzīgi var raksturot jebkuru semantiski atbilstošu situāciju (Ruppenhofer et al. 2006, 27). LAIKS, VEIDS, CĒLONIS u. tml. līdzīgi var raksturot gandrīz jebkuru darbību (vairāk par semantisko lomu iedalījumu latviešu valodā un pārvietošanās vārdiem sk. Nešpore 2010).

Verbu valences modeļi pamatā tiek veidoti no centrālajām, tātad situācijā konceptuāli svarīgajām, semantiskajām lomām. Ja marķētājs ir izvēlēties marķēt arī perifērās lomas (sk. *Vieta?* 1. att.), tās tiek īpaši apzīmētas un parādās tikai konkrētajos piemēros, nevis verba valences modelī.

1. izteikt vēlēšanos; lūgt kaut ko darīt			
329			
AGENSS	ADRESĀTS	verbs	5
AGENSS	verbs	ADRESĀTS	SATURS 76
verbs	ADRESĀTS	SATURS	35
verbs	SATURS		56
verbs			1
AGENSS	verbs	SATURS	147
AGENSS	verbs		9
2. (dievu) 21			
AGENSS	verbs	ADRESĀTS	SATURS 8
verbs	ADRESĀTS	SATURS	3
AGENSS	verbs	ADRESĀTS	4
verbs	ADRESĀTS		2
verbs	SATURS	AGENSS	3
verbs			1

2. attēls. Verba *lūgt* valences modeļi.

⁹ KONTRAGENSS – viens no diviem vienādi aktīviem situācijas dalībniekiem (otrs ir AGENSS), kas veic apzinātu darbību.

Ar rīku automātiski tiek ģenerēti visi marķētajā tekstā sastopamie verbu valences modeļi, norādot, cik bieži katrs no tiem parādās (sk. 2. att.). Tāpat tiek piedāvāts arī katrai semantiskajai lomai atbilstošo morfoloģisko formu saraksts (sk. 3. att.). Kad būs marķētas leksiskās nozīmes kategorijas, arī tās tiks iekļautas šajā sistēmā.

Kategorija	Skaitis	Grupas	
AGENSŠ	252	Grupa	Skaitis
		Nom	252
SATURS	328	Grupa	Skaitis
		Adv	1
		Inf	189
		S	52
		Acc	75
		parAcc	8
		Loc	3
ADRESĀTS	133	Grupa	Skaitis
		Dat	54
		Acc	79
Vieta?	2	Grupa	Skaitis
		Loc	2
Verbs	350	Grupa	Skaitis
		V3	230
		V1	106
		V2	14

3. attēls. Semantiskajām lomām atbilstošās morfoloģiskās formas: verbs *lūgt*.

3. Plānotais rezultāts

Lietotājam tiks piedāvāti vairāku veidu dati:

- 1) verbu valences modeļi – verba apkaimei raksturīgās semantisko lomu kombinācijas;
- 2) semantisko lomu, morfoloģisko formu un leksiskās nozīmes kategoriju atbilstes;
- 3) marķēts teksts, no kā verbu valences modeļi izsecināti, ar meklēšanas iespējām pēc semantiskajām lomām vai ontoloģiskās informācijas. Līdz ar to rodas plašas iespējas pētīt verbu apkaimes, atbilstes starp semantiskajām un sintaktiskajām lomām, analizēt teikumu struktūras shēmu realizāciju, sintaktiskajā struktūrā izteiktos paplašinātājus u. tml.

Vietnē <http://www.tezaurs.lv/valence/preview/> var aplūkot marķēšanas rezultāta paraugus – marķēto valodas materiālu un verbu valences modeļus. Šobrīd darba variantā ir marķēts 180 verbu (apmēram 25 tūkstoši lietojuma piemēru), bet datubāze arvien tiek papildināta ar jauniem vārdiem.

Saīsinājumi

Acc	akuzatīvs
Adv	apstākļa vārds
CPA	Corpus Pattern Analysis
Dat	datīvs
Gen	ģenitīvs
Loc	lokātīvs
Nom	nominatīvs
VPSV	Valodniecības pamatterminu skaidrojošā vārdnīca

Avots

Līdzsvarots mūsdienu latviešu valodas tekstu korpuss; miljons-2.0m. Izstrādātājs: LU Matemātikas un informātikas institūts. Pieejams: www.korpuss.lv.

Literatūra

1. Апресян, Юрий. 1995. Д. *Лексическая семантика (синонимические средства языка)*. 2-е дополненное издание. Москва: Языки русской культуры.
2. *Corpus Pattern Analysis*. 2004. Pieejams: http://nlp.fi.muni.cz/projects/cpa/#hanks_2004.
3. Fillmore, Charles J. 1986. The Case for Case. *Universals in Linguistic Theory*. London: Holt, Rinehart and Winston, 1–88.
4. Hajičová, Eva. 1993. *Issues of Sentence Structure and Discourse Patterns*. Praha.
5. Lokmane, Ilze. 2004. Valence un sintaktiskais obligātums. *Vārds un tā pētīšanas aspekti*. 8. Liepāja: LiePA, 268–275.
6. Lopatková, Markéta, Žabokrtský, Zdeněk, Benešová, Václava. 2006. *Valency Lexicon of Czech Verbs VALLEX 2.0. Technical Report 34*, UFAL MFF UK.
7. Nešpore, Gunta. 2010. Pārvietošanās verbu apkaimi raksturojošo semantisko lomu tipoloģija un to nošķiršanas problēmas. *Vārds un tā pētīšanas aspekti: rakstu krājums*. 14 (1). Liepāja: LiePA, 189–196.
8. Palmer, Frank R. 1994. *Grammatical roles and relations*. Cambridge: Cambridge University Press.
9. Ruppenhofer, Josef, Ellsworth, Michael, Petruck, Miriam, R. L., Johnson, Christopher, R., Scheffczyk, Jan. 2006. *FrameNet II: Extended Theory and Practice*. Pieejams: <http://framenet.icsi.berkeley.edu/book/book.pdf>.
10. Saeed, John I. 2007. *Semantics*. Blackwell Publishing, 2nd ed.
11. Sgall, Petr, Hajičová, Eva, Panevová, Jarmila. 1986. *The Meaning of the Sentence and Its Semantic and Pragmatic Aspects*. Dordrecht: Reidel Publishing Company.
12. Sližienė, Nijolė. 1994. *Lietuvių kalbos veiksmazodžių junglumo žodynas* T. 1: A–M. Vilnius: Mokslo ir enciklopedijų leidykla.
13. Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. Paris: Klincksieck. (Tulkojums: Теньер Л. *Основы структурного синтаксиса*. Ред. Гак, В. Г. Москва: Прогресс, 1988.)
14. *Valodniecības pamatterminu skaidrojošā vārdnīca*. 2007. Sast. Bušs, O., Joma, D., Kalnača, A., Lokmane, I., Markus, D., Pūtele, I., Skujiņa, V. (red.). Rīga: LU Latviešu valodas institūts.

15. Žabokrtský, Zdeněk. 2005. *Valency Lexicon of Czech Verbs*. PhD thesis. Prague: Charles University.

Summary

Development of valency resource for Latvian was started recently. The chosen approach takes into account the experience of other projects dealing with verbal valency (such as *FrameNet* for English, the valency dictionary of Lithuanian by N. Sližienė and *Vallex* for Czech) and is supplemented with our own ideas.

The article describes our approach to the verbal valency annotation – the valency types (syntactic, semantic valency, selectional restrictions), the set of the semantic roles, and presents the tool developed for the valency annotation.

Initially two annotation layers are chosen – semantic valency (semantic roles) and syntactic valency – the morphological description of the dependents of the verbs.

The valency models (combinations of semantic roles) are generated from the annotated core semantic roles only. Non-core semantic roles are not used in this process because they are not specific enough to particular verbs.

In future the annotation of one more valency layer is planned – very general ontological categories (animate, inanimate, abstract object, physical object, human) will be added.

An online tool is developed for manual annotation of valences within a pre-selected corpus. The valency models, as well as morphological forms corresponding to each semantic role are suggested automatically.

Summarized information of the valency of each verb, as well as a corpus of annotated sentences will be available as a result of the project. Currently valency of approx. 180 verbs (more than 25 000 utterances) is annotated using data from the Balanced Corpus of Modern Latvian.