

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ФИЛОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ

**МАТЕРИАЛЫ
XXXVI
МЕЖДУНАРОДНОЙ
ФИЛОЛОГИЧЕСКОЙ
КОНФЕРЕНЦИИ**

Выпуск 1

СЕКЦИЯ БАЛТИСТИКИ

**БАЛТИЙСКИЕ ЯЗЫКИ:
ПСИХОЛИНГВИСТИКА,
СОЦИОЛИНГВИСТИКА,
КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА**

Тезисы докладов

*5–6 марта 2007 г.
Санкт-Петербург*

**Филологический факультет
Санкт-Петербургского государственного университета
2007**

Ответственный редактор:

А. В. Андронов

© Коллектив авторов, 2007

© Филологический факультет

Санкт-Петербургского государственного университета, 2007

МАТЕРИАЛЫ
XXXVI Международной филологической конференции

Выпуск 1

СЕКЦИЯ БАЛТИСТИКИ

БАЛТИЙСКИЕ ЯЗЫКИ:
ПСИХОЛИНГВИСТИКА, СОЦИОЛИНГВИСТИКА,
КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

Тезисы докладов

5–6 марта 2007 г.
Санкт-Петербург

Лицензия ЛП № 000156 от 27.04.99

Авт. л. 2,49. Печ. л. 2,25. Тираж 100 экз. Заказ № 728.

Филологический факультет СПбГУ.
199034, С.-Петербург, Университетская наб., 11.

ОНУТ филологического факультета СПбГУ.
199034, С.-Петербург, Университетская наб., 11.



Санкт-Петербургский
государственный университет

5–6 марта 2007 г.

От редактора

Последняя, десятая конференция тематического цикла «Балтистика в Петербурге» посвящена рассмотрению балтийских языков в аспекте психолингвистики, социолингвистики, компьютерной лингвистики. Несмотря на известную нетрадиционность этих сравнительно новых разделов языкознания для классической балтистики в трудах участников конференции их проблематика оказалась представлена весьма равномерно и разнообразно. В первую очередь это касается докладов коллег, работающих в различных направлениях компьютерной лингвистики: излагаются проблемы как автоматического анализа языковых структур на уровне морфологии, синтаксиса и семантики, так и компьютерного синтеза звучащей речи, широко освещаются вопросы корпусной лингвистики (рассматриваются исторические, синхронные, параллельные корпуса текстов, демонстрируются возможности применения подходов корпусной лингвистики к решению вопросов теоретической грамматики). То же можно сказать о социолингвистических докладах, в которых наряду с рассмотрением теоретических вопросов показан и анализ конкретного материала — как отражающего современную языковую ситуацию, так и представляющего социолингвистический срез некоторых предшествующих периодов развития языка. Несколько в меньшей степени представлена психолингвистика, хотя её элементы, кроме собственно психолингвистических докладов, присутствуют и во многих работах, оказывающихся на пересечении разных наук (например, анализирующих фольклорные тексты). Можно надеяться, что материалы конференции окажутся полезными для определения ориентиров, выбора наиболее оптимальных моделей и уточнения в целом приоритетов дальнейшей работы в рассматриваемых областях балтийского языкознания (подчеркнём важность в этой связи докладов, посвящённых истории развития этих лингвистических дисциплин).

Организаторы и участники конференции благодарят Российский фонд фундаментальных исследований за финансовую поддержку (грант № 07-06-06011-г).

Конференции по балтистике в СПбГУ

(время проведения — первая неделя марта)

- 1998 г. Круглый стол «Балтийская филология»
- 1999 г. Фонетика и акцентология
- 2000 г. Морфология и словообразование
- 2001 г. Синтаксис
- 2002 г. Лексикология и лексикография
- 2003 г. Этимология и ономастика
- 2004 г. Историческая грамматика, история литературного языка
- 2005 г. Диалектология
- 2006 г. Стилистика и культура речи
- 2007 г. *Психолингвистика, социолингвистика, компьютерная лингвистика*
- 2008 г. Круглый стол «Балтийская филология»: 10 лет спустя

Председатель организационного комитета

Алексей Андронов

e-mail: aleksey.andronov@lycos.com

Санкт-Петербургский государственный университет
Филологический факультет
Кафедра общего языкознания
Университетская наб., 11
199034 Санкт-Петербург, Россия
тел. (7 812) 328 97 62, 328 95 10
факс (7 812) 328 97 65

St. Petersburg State University
Philological faculty
Department of General Linguistics
Universitetskaya nab., 11
199034 St. Petersburg, Russia
phone (7 812) 328 97 62, 328 95 10
fax (7 812) 328 97 65

<http://www.genling.nw.ru/baltist/baltist.htm>

Алексей Викторович Андронов (Санкт-Петербург)

На пути к корпусу латгальского языка России (1920–30-е гг.)

В числе крестьян, переселившихся из западных губерний на свободные земли в европейской части России и в Сибири в соответствии с государственной политикой к. XIX — н. XX в., были тысячи латгальцев. В первые десятилетия советской власти правительством уделялось особое внимание развитию малых народов: ими занимались специальные научные институты, разрабатывавшие в частности системы письменности, была создана сеть национальных начальных и (неполных) средних школ, выходили книги и периодика, работу с национальными меньшинствами курировали специальные отделы в советских и партийных органах. Эта политика в полной мере относилась и к латгальцам. Одним из аспектов рассмотрения их богатой истории и культуры в этот период является исследование языка латгальских изданий, которых с 1918 по 1937 г. в СССР было выпущено более сотни. Образуя весьма обширный и разнообразный, но закрытый и чётко очерченный круг текстов, эти источники представляют чрезвычайно благодарный материал для организации в виде современного лингвистического корпуса (http://www.genling.nw.ru/baltist/lg/lg_corp.htm). На пути к нему необходимо решить несколько задач.

1. **Составление библиографии** латгальских изданий рассматриваемого периода на основе указателей [LGĀ] и [LP2]¹. Список источников корпуса на настоящий момент включает 97 названий книг и 15 периодических изданий и может считаться вполне завершённым. Возможность в дальнейшем незначительных изменений, однако, не исключена, в частности, из-за того, что для некоторых источников (особенно для периодики) не ясно место издания (Латвия или Россия), некоторые книги (в основном школьные учебники) описаны по второму или третьему изданию при отсутствии библиографической информации о первом издании, возможность некоторых пропусков допускают и сами составители указателей [LGĀ: 15], не все издания им удалось описать de visu.

2. **Составление сводного каталога** латгальских изданий по материалам фондов крупнейших библиотек Латвии (Национальной и Академической), Москвы (РГБ) и Петербурга (РНБ и БАН), в которых сохранилось большинство таких изданий. 11 книг, однако, остаются известными лишь по библиографическим описаниям, поэтому актуальным является их выявление в других хранилищах, в частности в библиотеках, архивах, краеведческих музеях городов, в окрестностях которых активно селились переселенцы из Латгалии. К сожалению, наши поиски в Новосибирске (втором после Москвы центре издания латгальской литературы в рассматриваемый период), Красноярске и Омске не принесли пока ощутимых результатов, очередной задачей является обследование библиотек Томска и Иркутска. Ни в одной библиотеке нет полного комплекта главной газеты российских латгальцев — «Taisneiba», а 22 выпуска её (за 12 лет издания — с 1926 по 1937 гг.) пока не обнаружены вообще.

3. **Работа по сканированию текстов** начата в РНБ, с руководством которой согласовано подписание договора о совместном проекте «Латгальская книга в России», предполагающем изготовление цифровых копий всех имеющихся в библиотеке российских изданий на латгальском языке за 1918–1937 гг. (59 наименований — см.

¹ В библиографии В. Сейле (1935; 1936) латгальские издания Советской России представлены крайне скудно.

[ЛКР]) и публикацию их в сети интернет с историко-филологическими комментариями. Этот пример может послужить моделью для аналогичного сотрудничества с другими библиотеками, ведь для создания полного корпуса необходимо объединение ресурсов разных фондов.

4. По мере сканирования материалов будет осуществляться **автоматическое распознавание текстов, ручная проверка, макетирование и структурная разметка.**

5. Собственно филологической задачей является **разработка системы «версий» представления текста** (оригинал, транспонация в современную орфографию с исправлением опечаток, морфологическая аннотация, синтаксическое описание и т. д.). При этом необходимо учитывать специфику текстов, характеризующихся, в частности, большой вариативностью, по началу также не устоявшейся графической системой.

В числе перспективных направлений научной работы с корпусом латгальского языка России могут быть названы:

- создание грамматического описания и словаря латгальского языка на полностью документированном материале закрытого корпуса текстов;
- выявление принципов редакторской правки, унификации диалектных различий, критериев определения (формирования) нормы латгальского литературного языка в России (в сопоставлении с аналогичной работой в Латвии), истории этой работы;
- описание способов обозначения новых реалий, связанных с новыми природными и общественными условиями жизни, вопросов терминовтвора, особенностей латгальского языка России вообще и в том числе связанных с влиянием русского языка;
- изучение языковых особенностей разных стилей, оригинальных и переводных текстов.

Литература

- ЛКР — Латгальская книга в России (1918–1937). Каталог выставки / сост. А. В. Андронов, Л. Лейкума; Российская национальная библиотека, Санкт-Петербургский государственный университет. Санкт-Петербург, 2007 (в печати)
- LGĀ — Latviešu grāmata ārzemēs 1920–1940 : Bibliogrāfiskais rādītājs / sast. Ē. Flīgere, L. Lāce; atb. red. D. Ivbule; Latvijas Akadēmiskā bibliotēka. Rīga, 1998.
- LP2 — Latviešu periodika. 2. sējums. 1920–1940. Revolucionārā un padomju periodika / Ā. Brempele, Ē. Flīgere, V. Lūkina; Latvijas PSR ZA, Fundamentālā bibliotēka. Rīga, 1976.

Everita Andronova (Rīga — Sanktpēterburga)

No seno tekstu korpusa līdz vēsturiskajai vārdnīcai: iestrādes un problēmas

Pirmais publiski pieejamais latviešu valodas korpus ar atbilstošu programnodrošinājumu tika atklāts 2003. g. janvārī, kad, apvienojoties LU FF, LU LVI un LU MII pētniekiem, tika pārskatīti agrāk ievadītie teksti un pievienoti jauni 17. gs. avoti, kā arī izstrādāta korpusa sistēma. Teksta apstrāde un satura sagatavošana sevī ietver: strukturālo marķēšanu, marķējuma sintakses pārbaudi, vārdformu indeksu izveidi. Tika izstrādāti lietojumrīki: konkordances programma, meklēšanas iespējas, vārdformu konteksta aplūkošanas iespējas, vārd-

formu biežuma saraksti, inversā vārdnīca; nodrošināta navigācija korpusa saturā (pēc autora, avota un teksta kategorijas) (skat. Milčonoka 2003).

Latviešu valodas seno tekstu korpusa (www.ailab.lv/senie) mērķis ir veicināt valodas diahronisko izpēti, popularizēt agrākos latviešu rakstu pieminekļus, kā arī kalpot par pamatu topošajai latviešu valodas vēsturiskajai vārdnīcai. Ar LZP atbalstu 2004. g. uzsākta vārdnīcas teorētisko jautājumu izpēte un paraugšķirkļu izstrāde (projekta vadītājs P. Vanags). Vienlaikus tiek veidota arī šķirkļu elektroniskā versija un leksikogrāfa darba vietas sagatavošana, izmantojot mūsdienu tehnoloģiju iespējas.

Korpusa reprezentativitātes un līdzsvara jautājums ir svarīgs gan paša korpusa izveidē, gan arī vārdnīcas sastādīšanā. 2005. g. LU Pētniecības projekta laikā latviešu valoda seno tekstu korpusu tika papildināts ar 18. gs. populārzinātniskajiem, lietišķajiem tekstiem un pavisam nelielu daiļliteratūras avotu skaitu.

Vārdnīcas izstrādātāji ir saskārušies ar daudzām problēmām, šajā referātā netiks aplūkotas šķirkļa izstrādes teorētiskās nostādnes, galvenā uzmanība veltīta jautājumiem, kā no korpusa top vārdnīca.

Vārdnīcas sastādīšanas posmā līdz 2007. g. var izšķirt vairākas pakāpes:

- 1) vārdnīcas tipa un šķirkļa uzbūves nostādnes izstrāde (sīkāk par to Andronova, Trumpa, Vanags 2006);
- 2) leksikogrāfa darbs ar korpusu (meklēšana vārdformu indeksā, vārdformu atlase, biežuma un „adrešu“ uzkrāšana);
- 3) atlasīto piemēru analīze (izmantojot konkordances programmu un izvērstā konteksta iespējas);
- 4) šķirkļu pirmās versijas izstrāde atbilstoši mūsu izvēlētai šķirkļa struktūrai;
- 5) šķirkļu apspriešana seminārā vienreiz mēnesī, komentāru uzkrāšana elektroniskā un „papīra“ veidā;
- 6) struktūras precizēšana.

2006. g. nogalē, sadarbojoties leksikogrāfiem un sistēmanalītiķim, tika formalizēta šķirkļa struktūras gramatika, izveidota atbilstoša XML (extensible mark-up language) shēma.

Tādējādi papildu esošajiem vārdnīcas rakstīšanas uzdevumiem pievienojas:

- 1) esošo šķirkļu „pārrakstīšana“ izstrādātajā anotēšanas shēmā;
- 2) šķirkļa struktūras precizēšana un uzlabošana;
- 3) jaunu šķirkļu marķēšana, izmantojot struktūrai pielāgotu XML redaktoru;
- 4) šķirkļa rakstīšanas un marķēšanas instrukcijas sagatavošana.

LU MII speciālistiem jāatrisina vairākas ar vārdnīcas sistēmas izstrādi saistītas problēmas: metodika un rīku atbalsts datu plūsmas organizēšanai (katrs autors strādā ar katru šķirkli), kā arī atgriezeniskās saites ar korpusu nodrošināšana; vārdnīcas vizualizācija; interaktīvas meklēšanas iespēju izstrāde.

Tālākā nākotnē var domāt par modernas leksikogrāfa darba vietas izveidi, kurā tiktu apvienots gan korpusu un no tā izvilktu datu analīze, gan dokumentu melnraksti, gan šķirkļi un leksikogrāfu komentāri, gan arī dati no ārējiem resursiem.

Literatūra

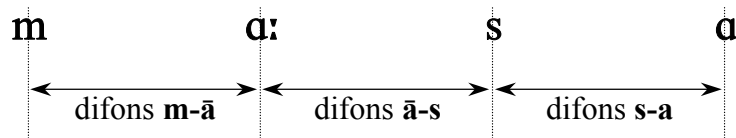
- Andronova, Trumpa, Vanags 2006 — E. Andronova, A. Trumpa, P. Vanags. Latviešu vēsturiskās vārdnīcas (16.–18. gs.) projekts: iestrādes un problēmas. // Letonikas pirmais kongress. Valodniecības raksti. Rīga: LZA, 2006— 9.–17. lpp.
- Milčonoka 2003 — E. Milčonoka. Latviešu valodas 17. gadsimta teksti internetā. // Baltu filoloģija XII₍₁₎. Rīga: LU, 2003. — 139.–150. lpp.

Segmentu datu bāzes izveide latviešu valodas konkatenatīvajam runas sintezatoram

Teksta—runas sistēma ietver kompleksu darbību virkni: teksta ievadi un priekšapstrādi, t. i., sagatavošanu automatizētai fonētiskai transkribēšanai, teksta analīzi, grafēmas—fonēmas atbilstmju likumu izstrādi, automatizētu transkribēšanu, runas ģenerēšanu utt. Lai teksta—runas sintezators darbotos, ir nepieciešami trīs moduļi: teksta apstrādes (teksta priekšapstrāde, teksta transkribēšanas sistēma), prosodijas un skaņas sintēzes modulis.

LU Matemātikas un informātikas institūta Mākslīgā intelekta laboratorijā ir izstrādāts teksta—runas sintezators, kas balstās uz segmentu savirknēšanas jeb konkatenācijas metodi. Runas segmentu savirknēšana jeb konkatenācija ir viena no populārākajām runas sintēzes metodēm. Tajā vajadzīgā secībā tiek savirknēti runas segmenti (vārdi, zilbes, puszilbes, fonēmas, difoni, pusfonēmas, zilbes un/vai trifoni), kas iegūti no konkrētas balss ierakstiem.

Visbiežāk par segmentiem konkatenatīvajā runas sintēzē izvēlas difonus. Difons tiek segmentēts no fonēmas centrālā, stabilā punkta līdz nākamās fonēmas centrālajam, stabilajam punktam, tādējādi var teikt, ka difonu veido divas blakus esošas pusfonēmas un tas ietver pāreju no vienas fonēmas uz nākamo fonēmu:



Arī latviešu valodas teksta-runas sintezatorā galvenokārt tiek izmantoti difoni. Atsevišķos gadījumos talkā tiek ņemtas pusfonēmas un trifoni.

Teorētiski iespējamo difonu skaits valodā atbilst fonēmu skaita kāpinājumam kvadrātā. Tā kā latviešu literārajā valodā ir 48 fonēmas — 26 līdzskaņi, 10 divskaņi un 12 patskaņi, teorētiski iespējami 2304 difoni. Tomēr dabiskajās valodās ir fonotaktiski ierobežojumi — daži fonēmu pāri vai pat veselu fonēmu tipu savienojumi nav iespējami. Arī latviešu literārajā valodā, analizējot VC, CV, VV un CC savienojumus, konstatētas fonēmas, kas savā starpā nevar savienoties, piemēram, platais patskanis /æ/ vai /æ:/ un palatāls līdzskanis ([æj], [æ:ɟ], [æɫ], [æ:ɫ] u. tml.). Tādējādi latviešu valodas runas segmentu datu bāzē iekļauto difonu skaits ir apmēram 1500.

Segmentējot difonus, lielākajā daļā fonēmu savienojumu par segmentēšanas vietu izraudzīts fonēmu centrs, tomēr ir fonēmas, kurās griezumus tiek veikts citur, piemēram, dalījuma vieta nebalsīgos troksneņos, eksplozīvos slēdžeņos /k/, /p/, /t/, /c/, /ts/, /tʃ/ ir slēguma beigu daļā, tieši pirms eksplozijas (skat. tab.).

Fonēmas	Segmentēšanas vieta fonēmās
Īsie patskaņi /i/, /e/, /æ/, /ɑ/, /u/, /o/	Fonēmas centrs
Nazālie slēdžeņi /m/, /n/, /ɲ/	
Spraudzeņi /f/, /v/, /s/, /z/, /ʃ/, /ʒ/, /j/, /x/, /l/, /ʎ/, /r/.	
Eksplozīvie slēdžeņi /p/, /b/, /t/, /d/, /c/, /ɟ/, /k/, /g/	Slēguma beigu daļa
Afrikatīvie slēdžeņi /ts/, /tʃ/, /dz/, /dʒ/	
Garie patskaņi /i:/, /e:/, /æ:/, /ɑ:/, /u:/, /o:/	1/3 : 2/3

Segmentēšanai ir izmantoti gan mākslīgi radīti beznozīmes vārdi, gan latviešu valodas vārdi. Tomēr jāatzīst, ka, mākslīgi radot beznozīmes vārdus un frāzes, ir iespējams izveidot tādas valodas vienības, no kurām var segmentēt pēc iespējas vairāk difonu, turklāt vēlāmā pozīcijā.

Māris Baltiņš (Rīga)

Latviešu terminoloģijas attīstības posmi un to raksturojums

Terminoloģijas attīstībā būtiskākā loma ir valodas sociolingvistiskajām funkcijām un attiecīgā laika posma dominējošajai kontaktvalodai, no kuras visaktīvāk tiek tulkoti teksti un pārņemti terminoloģiski risinājumi. Latviešu terminoloģijas vēsturē atbilstoši tautas sociālajai un politiskajai vēsturei varam izšķirt piecus periodus: (1) no pirmajiem rakstiem līdz XIX gadsimta vidum; (2) no XIX gadsimta vidus līdz 1918. gadam; (3) no 1918. līdz 1944. gadam; (4) no 1944. līdz 1990. gadam un (5) kopš 1990. gada. Visstraujāk terminoloģijas attīstība norisinās katra posma pirmajos gados, kad, mainoties valodas sociolingvistiskajām funkcijām, krasi pieaug sabiedrības vajadzības pēc terminiem.

Pirmā posma laikā, kad latviešu tekstus lielākoties sacerēja vācu tautības mācītāji, reāli termini veidojās tikai baznīcas dzīvē un tieslietās, bet citās jomās terminu vietā tika lietota perifrastiska to atveide. Tā plaši tika lietota kalendāros, padomgrāmatās un arī presē, tādēļ to var uzlūkot par apzinātu terminu atveides paņēmieni. Daudzos gadījumos tā spēja palielināt lietišķā teksta saprotamību.

Otrajā posmā terminoloģijas attīstību galvenokārt noteica populārzinātniski raksti, jo tajos atrodamā informācija bez pragmatiskiem mērķiem kalpoja lasītāja intelektuālo vajadzību apmierināšanai. Šo publikāciju sagatavošanā plaši tika izmantoti vācu un vēlāk arī krievu valodā publicēti materiāli, bieži iekavās norādot attiecīgā termina ekvivalentu šajās valodās vai vēl citas paskaidrojošas ziņas. Šim posmam raksturīga terminu atveides nekonsekvenca un autoru tieksme veidot savu terminu sistēmu, nevis pieturēties pie citu ieteikumiem. Tieši terminoloģijai veltīto publikāciju nav daudz, tās veltītas kādai šaurākai nozarei (botānikai, acs uzbūvei, jūrniecībai) vai ietilpst plašākos valodas kopšanai veltītos darbos. Terminoloģijas attīstības šai laikā raksturo pirmā pabeigtā enciklopēdija — Rīgas Latviešu biedrības Zinību komisijas „Konversācijas vārdnīca“ (1903–1921).

Trešā posma sākumā valodai pavērās būtiski jaunas sociolingvistikas funkcijas valsts pārvaldē, tiesību aktu izstrādē un izglītībā. Apzinoties vajadzības, 1919. gadā tika izveidota Izglītības ministrijas terminoloģijas komisija (tā darbojās līdz 1921. gadam). Šīs komisijas nozīmība saistīta ar pāreju no individuālas (vienas personas veiktas), privātas (bez noteikta pilnvarojuma un pienākumiem) un spontānas (atbilstoši konkrētā lietpratēja vajadzībām) terminrades uz kolektīvu (daudzu noteiktas nozares lietpratēju), plānotu (cenšoties aptvert nozari kopumā, nevis tikai dažus terminus konkrētai publikācijai) un institucionalizētu (strādājot kādas valsts pārvaldes iestādes pārraudzībā) darbu. Kaut gan turpmāk process atkal turpinājās mazāk koordinēti, tomēr šīs komisijas ietekme neaprobežojas ar 1922. gadā iznākušo „Zinātniskās terminoloģijas vārdnīcu“, jo daudzās nozarēs tās ieteikumi atstājuši būtisku iespaidu līdz pat mūsu dienām. Lai nodrošinātu pēctecību ar iepriekšējiem attīstības posmiem, šī laika materiālos parasti dots sastatījums ar vācu un krievu terminiem.

Laika posmā no 1944. līdz 1990. gadam vienlaikus vērojamas vairākas pretrunīgas tendences. Jau kopš 1945. gada nepārtraukti darbojas terminoloģijas komisija (kopš 1946. ga-

da Zinātņu akadēmijas Terminoloģijas komisija), kas veica nozaru terminu apkopošanu un daudzu terminoloģisku biļetenu izdošanu. Tās darbs bija orientēts uz klasisko terminoloģiskā darba modeli (laika ziņā precīzi neierobežots sistemātisks darbs kādas nozares vai apakšnozares ietvaros, veidojot plašas mācību grāmatas vai terminu vārdnīcas, kas paredzētas nozares lietpratējiem). Minētais darbs visai maz saistīts ar tiesību aktu izstrādi vai normatīvo aktu tulkojumiem, turklāt notika pakāpeniska terminoloģiskā darba reducēšana uz divvalodu (krievu un latviešu) terminu ekvivalentu sastatīšanu.

Kopš 1990. gada latviešu terminoloģijas attīstība bijusi visai strauja, turklāt būtiski mainījies terminu izstrādes raksturs. Par dominējošo kontaktvalodu kļuva angļu valoda, radot problēmas ar agrāko terminoloģisko tradīciju turpināšanu, jo dažās nozarēs anglosakšu terminu sistēma būtiski atšķīrās no līdz tam parastās kontinentālās Eiropas tradīcijas (kurā ietilpa kā vācu, tā krievu termini). Par galveno jaunu terminu avotu kļuvuši dažādu dokumentu tulkojumi, bet par galveno terminoloģisko resursu lietotāju — ierēdnis vai tulkotājs, kuram turklāt parasti bija nepieciešams straujš risinājums. Klasisko terminu izstrādes modeli, kas vairs pilnībā neatbilda reālajām prasībām, šais apstākļos aizstāja paātrinātais modelis, kurā bieži nav iespējams veikt sistēmisku izpēti, bet par galveno kvalitāti kļuvis darba ātrums. Turklāt informācijas tehnoloģiju attīstība prasa esošo resursu digitalizēšanu un elektronisku datubāzu veidošanu.

Оярс Бушс (Рига)

Займствование, иноязычная цитата, смена кода (к разграничению понятий на материале латышского языка)

На теоретическом уровне содержание упомянутых в заглавии общелингвистических и социолингвистических понятий и терминов представляется достаточно чётко разграниченным и определённым: **з а и м с т в о в а н и е** — это (если вести речь только о лексических заимствованиях) «слово, заимствованное из другого языка и в большинстве случаев получившее грамматическое оформление, соответствующее нормам заимствовавшего языка», **и н о я з ы ч н а я ц и т а т а** — «включённое в текст иноязычное слово или словосочетание, а также небольшое, относительно устойчивое высказывание на другом языке» (в данном случае речь идёт преимущественно об устном тексте), **с м е н а к о д а** — «переход от одного языка речевого текста к другому языку в пределах одного речевого акта».

Однако, как всегда бывает, на практике всё далеко не так просто, и распределение языковых или речевых фактов между смежными категориями (заимствование или цитата? цитата или смена кода?) нередко представляется архитрудной — если вообще выполнимой — задачей. Необходимость такого распределения — особенно в первом случае, т. е. между заимствованиями и цитатами — является не только теоретической, но и сугубо практической, связанной с лексикографической работой; эта необходимость особенно актуализируется в процессе работы над словарями сленга и диалектными словарями. О некоторых конкретных решениях по поводу классификации малоадаптированных русизмов латышского языка в связи с отбором материала для «Словаря сленга латышского языка» [Bušs, Ernstsone 2006] уже шла речь на конференции баллистики в Санкт-Петербурге, посвящённой культуре речи и стилистике [см. Бушс 2006: 7]. При этом в качестве примера приводились такие грамматически неадаптиро-

ванные слова (словоформы), как *dašlo, prikiņ, užas! začem? značit*. Несмотря на эту неадаптированность, упомянутые слова включены в словарь сленга, из чего явствует, что авторы словаря тем не менее считают эти слова сленгизмами латышского языка, а не иноязычными цитатами. Вопрос, правомерно ли это решение, остается, конечно, по-прежнему спорным. Какие аргументы «за» и «против» в данном случае учитывались? Основным аргументом «против» является отсутствие «грамматического оформления, соответствующего нормам заимствовавшего языка»; «грамматический» аргумент в ряде случаев усиливается тем, что классифицируемое слово сохранило признаки косвенных словоформ языка-источника. Однако соответствующее грамматическое оформление, как известно, является лишь типичным, но не обязательным признаком заимствования, а лексема (чаще всего глагол), заимствованная лишь в одной словоформе, может функционировать в заимствовавшем языке либо как слово с ущербной парадигмой, имеющей только одну форму, либо, например, как междометие (ср. русизм *davai!*, уже давно и бесспорно укоренившийся в латышском сленге, см. LŽV). С другой стороны, и окказионально использованное в латышской речи иноязычное слово может в данном речевом акте получить латышскоязычное грамматическое оформление, не переставая тем не менее быть иноязычной цитатой. Специфическая акцентуация — словесное ударение на первом слоге — в настоящее время также перестало быть надёжным признаком чужеродности для латышского языка — как из-за достаточно частого ударения на первом слоге в заимствованных именах собственных (об этом см. также Daugaviete 2007), так и в результате влияния акцентологических навыков билингвальных носителей латышского языка. Приходится констатировать, что более или менее надёжных формальных критериев проведения грани, отделяющей заимствования от иноязычных цитат, не существуют; на практике используется критерий частотности употребления [Бушс 2006: 7], который, однако, следует признать малонадёжным. Представляется, что максимально точное проведение упомянутой грани возможно лишь в идеале, при условии учёта психолингвистических процессов.

Ещё более расплывчатой является грань между иноязычной цитатой и сменой кода. Некоторые авторитеты социолингвистики (К. Истмен) полагают, что разграничить смешение кода, смену кода и заимствование иноязычных слов вообще не представляется возможным [Beikers 2002: 74]. Не разделяя столь подчёркнутого «нигилизма», в ряде случаев можно, среди прочего, для идентификации цитаты использовать возможное наличие металингвистического комментария (напр., *kā teiktu .., kā saka ..*, ср. Kvašīte 2007). Трудно согласится с утверждением, что смена кода может реализоваться и в виде всего лишь одного включённого в речь иноязычного слова [Beikers 2002: 74]; такое слово вряд ли можно назвать иначе, чем цитатой. О смене кода правомерно говорить лишь тогда, когда высказывание создается «на месте», а не при использовании «готовых» иноязычных сегментов (слов, штампов и т. п., ср. высказывание с цитатой *Viņš ir, kā krievi saka, большой шишка на ровном месте*).

Литература

- Бушс 2006 — О. Бушс. Из опыта работы над словарём сленга латышского языка. // Материалы XXXV международной филологической конференции. Выпуск 1. Секция баллистики. Балтийские языки: культура речи и стилистика. Тезисы докладов. 3–4 марта 2006 г. Санкт-Петербург. Санкт-Петербург: Филологический факультет Санкт-Петербургского государственного университета, 2006. — с. 5–8.
- Beikers 2002 — K. Beikers. Bilingvisma un bilingvālās izglītības pamati. No angļu valodas tulkojusi I. Druvieta. Rīga: Nordik, 2002.

- Bušs, Ernstsone 2006 — O. Bušs, V. Ernstsone. Latviešu valodas slenga vārdnīca. Rīga: Norden AB, 2006.
- Daugaviete 2007 — A. Daugaviete. Nepirmās zilbes uzsvars latviešu valodā. // Mūsdienu valoda mūsdienu sabiedrībā. Sociolingvistika un sinhronā valodniecība. Akadēmiķa Jāņa Endzelīna 134. dzimšanas dienas atceres starptautiskās zinātniskās konferences materiāli. Rīga: LU Latviešu valodas institūts, 2007. — 23–24. lpp.
- Kvašīte 2007 — R. Kvašīte. Metalingvistiskie komentāri *kā teiktu...* sociolingvistiskajā skatījumā. // Mūsdienu valoda mūsdienu sabiedrībā. Sociolingvistika un sinhronā valodniecība. Akadēmiķa Jāņa Endzelīna 134. dzimšanas dienas atceres starptautiskās zinātniskās konferences materiāli. Rīga: LU Latviešu valodas institūts, 2007. — 40–45. lpp.
- LŽV — Latviešu žargonā vārdnīca. / Sast. I. Mirovics, A. Dubaus. Mordialloc (Australia): AIVA, 1990.

Лина Бронислововна Воробьева (Псков)

Пословицы и поговорки как социолингвистический источник

Вопросы соотношения национальной культуры и информации, заложенной в языковых единицах, издавна привлекали лингвистов. Все особенности жизни страны (природные условия, географическое положение, социальное устройство, исторические события и т. д.), этапы её развития, характер народа, безусловно, находят отражение в языке. Поэтому можно утверждать, что язык представляет собой отражение культуры нации, он несёт в себе социокультурный код того или иного народа.

Пословицы, представляющие значительный пласт лексики любого языка и отражающие вековые обычаи, обряды, поверья и другие реалии, являются ценным источником сведений о том или ином народе, о его мировидении. Им присущ не только национальный, но и временной колорит. Особенно мощным социокультурным потенциалом обладают пословицы советского периода. При этом фиксируются они не только в русском языке, но в других языках мира, подтверждая этим всемирное значение происходящих в стране исторических событий. Социолингвистический феномен, который в течение 70 лет советской власти именовался «новоязом» или языком советской эпохи, привлекал и продолжает привлекать внимание лингвистов. Многие понятия того времени сейчас допускают неоднозначную интерпретацию, однако они должны быть переосмыслены и адекватно откомментированы с позиций постсоветского времени.

Сложно говорить однозначно о природе пословиц советского периода: они возникли исходя из системы ценностей того времени, создавались и функционировали по уже существующим языковым законам. Тематически они покрывают исключительно советскую действительность, в дореволюционных пословицах не обнаруживаются такие социокультурно детерминированные компоненты, как бригадир, агроном, колхоз, партия, революция, секретарь, трактор, ударник, новатор, фашист и т. д.

В центре многих идеологизированных пословиц — отражение труда колхозника и жизни колхозной деревни, при этом преимущества коллективного ведения хозяйства явно идеализируются: *Не хочешь тужить — в колхоз иди жить; Свой колхоз или чужой, а один устав — трудовой; В хорошем колхозе и плохая земля родит; В колхозе что ни год, то больше доход; Береги колхоз — получишь хлеба воз; В колхозе поживать — добро наживать; У хорошего колхоза нет поломанного воя; В колхозе язык не в зачёт, кто работает, тому и почёт.* Пафос коллективного труда также отражается в следующих выражениях: *Работать вместе радостно, жить в колхозе сладостно; В колхозную пору пошла жизнь в гору; Хороша нива только у коллектива. Без коллектива и жизнь несчастлива; В колхоз вступил — сапоги купил; В колхоз пришёл —*

кафтан нашёл. Иногда в подобных пословицах подчеркивается «то, что было» и «то, что есть», при этом первый член этой оппозиции отмечен негативно, а второй — позитивно: *Раньше жили — слезу лили, а теперь живем — счастье куем; Не жили — мотались, в колхоз не догадались; Не шли в колхоз — маялись, пошли — не покаялись.*

Данная тема отразилась и в литовских пословицах. Как правило, они записаны от информантов в 1960–1963 годы [Lietuvių tautosaka 1968]: *Kolūkis — mano ūkis, dėl to aš jį ginu; Kai kolūkyje būsi, niekuomet nežūsi; Vyk vagi iš kolūkio kaip kiaules iš purų; Tinginiui kolūkyje ne gyvenimas; Blogoje brigadoje vis bėda: sodina kukurūzus — balanda išauga; Arkliu tiek neapkulniuosi, kiek traktorium apvažiuosi.* При этом обнаруживаются русско-литовские параллели: *Был бы дружный колхоз, а урожай будет — Vieningas kolūkis — gausus ir derlius.* Изучение такого материала в сопоставительном лингвокультурологическом аспекте позволит сделать ценные выводы о межъязыковых и межкультурных контактах народов в эпоху тоталитаризма, выявить механизмы использования паремиологических единиц в качестве средства агитации и пропаганды.

Несмотря на то, что лингвисты постсоветского времени видят в «новоязе» средство подавления духовности, индивидуальности, свободы мысли, примеры показывают, что такие пословицы нельзя рассматривать однозначно. По ним можно судить о патриотических настроениях и взглядах людей того времени на труд, общество, мораль. С другой стороны, в связи с политическими, экономическими и идеологическими преобразованиями, происходящими в обществе, такие пословицы подвергаются трансформациям, которые выявляют естественное соотношение положительного и отрицательного в сфере паремиологии.

Литература

Lietuvių tautosaka 1968 — Lietuvių tautosaka. T. 5. Smulkioji tautosaka, žaidimai ir šokiai. Vilnius, 1968.

Anna Vulāne (Rīga)

Ļaužu valodas — būtisks latviešu pasaulinās komponents

Valoda pieder pie tautai konceptuāli nozīmīgām parādībām. Tā ir kultūras radīta un galvenais kultūras attīstības nosacījums. Valoda kā privātas informācijas iegūšanas un nodošanas līdzeklis latviešu valodā bieži vien tiek apzīmēta ar vārdu *runas, valodas* (vārda *runa, valoda* daudzskaitļa forma). Informācijas plūsmai sabiedrībā kā nozīmīgam zināšanu ieguves un uzvedības regulētājam faktoram latviešu tautas pasaulīnā atvēlēta ievērojama vieta, par ko liecina leksēmas *ļaužu valodas* (resp. *tenkas, baumas, aprunāšana*) plašais izmantojums folklorā, īpaši tautasdziesmās.

Kā norāda etnolingvists J. Bartminskis, ikdienišķais, naivais domāšanas un spriešanas modelis, uzvedības scenāriji, kuros atrodams gan ideālais paraugs, gan konservatīvi aizliegumi, arhaiskie un līdz ar to vienkāršie priekšstati par pasauli, kas atspoguļoti klišejeida verbālajās formulās, ir nozīmīgs mūsdienu valodas un nacionālās kultūras slānis. Tie funkcionē kā kolektīvo priekšstatu arhetipi un ir mūsdienu racionālo, kultūras elitārajā slānī ietilpstošo zinātnisko koncepciju spilgts fons.

Latviešu kultūrā līdzīgi kā daudzās citās kultūrās zināšanas pieder pie konceptuāli nozīmīgiem jēdzieniem. Folkloras materiālos atklāta ne tikai sakrālo, bet arī informatīvo zināšanu loma, kuras savukārt saistītas ar ētisko vērtību sistēmu konkrētā sabiedrībā. Informācija,

nonākusi svešinieku ziņā (*ļaužu valodās*), tiek apspriesta un vērtēta no sabiedrībā pieņemto uzvedības normu viedokļa un rezultātā veidojas attieksme pret konkrētu personu. Līdz ar to *ļaužu valodas* latviešu kultūrā izsenis ir cilvēka uzvedības modelēšanas noteicējfaktors, tāpēc jautājums — ko citi par to teiks? — ir bijis aktuāls mūsu senčiem un ir svarīgs arī mūsdienu cilvēkam.

Visplašāk jēdziena *ļaužu valodas*, arī *valodiņas*, saturs atklāts tautasdziesmās.

1. Neizbēgamā iespēja nokļūt *ļaužu valodās* (īpaši jaunai meitai un puisim, bāreņiem) ir gan biedējoša, nomācoša, gan arī uzvedību regulējoša — labāk dzīvot tā, lai nebūtu iemesla par tevi runāt, tāpēc jāsarģā savs gods, jābūt čaklam, izdarīgam, krietnam, nedrīkst būt lepnam, jāprot satikt ar cilvēkiem, bet, ja esi nokļuvis *ļaužu valodās*, tad jānovērš to cēlonis, respektīvi, jārikojas saskaņā ar sabiedrībā pieņemtajām normām. Piemēram:

<i>Runaj' ļaudis, ko runaja, Mani ņēma valodās. Vaj tad mani māmuliņa Samazgās mazgajuse? Vaj ūdeņa tai nebija Tekošā upitē, Tekošā upitē, Tīrajā avotā? 8846</i>	<i>Apkārt kalnu ar līkumu Gāj' pie savas līgaviņas, Lai ļautiņi neredzeja, Lai necēla valodiņas. 8324</i>	<i>Tu puisiti, es meitiņa, Abi ļaužu valodā; Mani sauca puišu bābu, Tevi meitu brāķeriti. 8984</i>
<i>Klusu, klusu audzināju Ciemiņos arajiņu, Lai ļautiņi nezināja, Lai neņēma valodās. 8532</i>	<i>Lepna, lepna ta meitiņa, Ne ta dzied, ne ta runā. Ta ar savu lepnumu Taps ļautiņu valodās. 830</i>	<i>Tu meitiņa, es puisitis, Abi ļaužu valodā; Iesim abi baznicā, Šķirsim ļaužu valodiņu. 14833</i>
<i>Labak sunim maizes devu, Ne kà tautu dēliņam: Sunits manas mājas sargā, Tautiets lika valodā. 31159-4</i>	<i>Labak sunim maizes devu, Ne kà tautu dēliņam: Sunits manas mājas sargā, Tautiets lika valodā. 31159-4</i>	<i>Ko, tautieti, nu daram, Nu valoda klajumā? Seglo savu kumeliņu, Jāj uz manu māmuliņu. 8543</i>

2. Aprunāšanas tieksme ir negatīva iezīme, ļaunums pret otru cilvēku, kas pelnījis nosodījumu, sodu. Īpaši negatīvi tiek vērtēta apzināta apmelošana, savu tuvinieku, ģimenes locekļu aprunāšana. Piemēram:

<i>Labak gāju dziedadama, Ne par otru runādama: Ne man kauna priekšā bija, Ne valodu pakaļā. 801-3</i>	<i>Valodnese, valodnese, Bāliņ, tava līgaviņa: Cik iedama ciemiņā, Mani nesa kukuļam. 9026</i>
<i>Valod', mana valodiņa, Kāda mana valodiņa! Ar tautām(i) runadama, Aprunāju bāleliņus. 9020</i>	<i>Valodiņu cēlajiņa, Nāc uz mani ciemoties; Es tev celšu dadžu krēslu, Div' asaru biķerites. 8701-4</i>

3. Māka pasargāt sevi un citus, norobežoties no tukšām runām, ļaunām mēlēm — laist pār galvu, izpurināt no vainaga vai seģenes, apzināties savu vērtību un taisnību, arī — samierināties vai stāties tām pretī — ir vēlama un augstu vērtējama.

<i>Pūš vējiņi, runā ļaudis Par to manu augumiņu; Smuidris mans augumiņš, Garam bira valodiņas. 8628-2</i>	<i>Ko darīt, japanes Izrunata valodiņa; Es būt' sirdi rādijuse, Kaut varetu parādīt. 8535</i>	<i>Izpurini, māmulite, Man sedzamu villainiti: Izpurini sirdēstiņus, Visu ļaužu valodiņas. 17033</i>
<i>Ai, bāliņ, ai bāliņ, Nevar vairs valodām. — Ej, māsiņa, droši pāri, Min valodas kājiņām. 8286</i>	<i>Dievs sodija tautu dēlu Devitāja ugunī, Kam preceja, neieņēma, Atstāj' ļaužu valodās. 8368</i>	<i>Vējiņš pūta apsitē, Apse lapu drebinaja; Dieviņ, devi tā drebēt Valodiņu cēlejam. 8379-1</i>

Varam secināt, ka latviešu pasaulainā *ļaužu valodas* atklājas kā ambivalenta parādība: no vienas puses, tā ir negatīva, ļauna rīcība, kas rada nepatiku pret runātāju, it īpaši, ja stāstītais neatbilst patiesībai (notiek aprunāšana), no otras puses — tām ir nozīmīga loma sabiedrības locekļu uzvedības regulēšanā — tās sekmē pieņemto normu ievērošanu. Vienlaicīgi izteikts arī brīdinājums, ka ikvienam, kas aprunā otru, jāpatur prātā, ka arī pats var nokļūt ļaužu valodās. Folkloras materiālos netiek izmantots apzīmējuma *ļaužu valodas* mūsdienu analogs *tenkas, baumas, aprunāšana*.

Татьяна Александровна Горностай (Рига)

Русско-латышско-русский словарь с элементами машинного перевода

В Латвии, где общество смело можно назвать билингвальным, возникает необходимость в программном продукте (ПП), который предоставит пользователю новые языковые технологии.

Компания «Tilde», зарекомендовавшая себя лидером в области лингвистического обеспечения языковых систем для языков стран Балтии, продолжает работу в направлении русско-латышско-русского машинного перевода, в частности полнотекстового перевода. В недалеком будущем ожидается выход нового ПП, что позволит расширить круг пользователей, повысит конкурентоспособность и укрепить позиции компании на рынке.

Первые шаги были сделаны давно, когда в свет вышел электронный словарь «Tildes Datorvārdnīca». Далее в течение почти двух лет коллектив отдела языковых систем трудился над проектом «Daudzvalodu Vārdnīca» (DVV, многоязычный словарь). В его состав вошли русско-латышский и латышско-русский словари, в создании которых использовалась технология перевода «Tildes Viedtulks» (интеллектуальный переводчик), внедренная в коробочный ПП «Tildes Birojs 2005».

В распоряжении пользователя два режима перевода: пословный перевод и перевод фраз. В первом случае выдается словарная информация о единице текста, на которую указывает курсор мышки (переводные эквиваленты с грамматическими комментариями). Во втором случае словарь предлагает перевод фрагмента текста:



между представителями этих языковых групп

Работа словаря DVV основывается на принципе модульной архитектуры, где каждая составляющая программы отвечает за определенную процедуру:

- текст переводного языка (source language) фрагментируется на предложения,
- предложение анализируется на уровне функциональных сегментов и слов, за что отвечают модули синтаксического и морфологического анализа,
- далее трансфер получает на входе синтаксическую структуру в виде дерева, в котором каждой единице приписана ее морфологическая и синтаксическая информация,
- входная структура на основе правил трансформируется в структуру языка перевода (target language),
- из словарной базы выбираются переводные эквиваленты,
- на последующих этапах структура обрабатывается модулями согласования (agreement) и снятия омонимии.

Для реализации модели русско-латышско-русского трансфера детально исследовались системы грамматических категорий (ГК) и функциональные свойства обоих языков. В области морфологии в системе частей речи РЯ и ЛЯ наблюдается существенное сходство, чего нельзя сказать о системе ГК, в которой не вполне совпадают сам набор ГК каждой части речи и способы их выражения:

- ГК рода существительного: в ЛЯ нет среднего рода;
- ГК одушевленности / неодушевленности существительного, вопросительно-относительных, отрицательных и неопределенных местоимений в РЯ и её отсутствие в ЛЯ;
- ГК рода латышских местоимений: указательное местоимение *это* в русском языке и латышские *tas, tā, tie, tās* в функции подлежащего;
- категория определенности в ЛЯ и ее отсутствие в РЯ;
- ГК наклонения глагола: изъявительное, повелительное и сослагательное РЯ VS *īstenības, vēlējuma, pavēles, vajadzības, atstāstījuma izteiksmes* ЛЯ, при этом последние два не имеют соответствий в РЯ, который эти значения выражает лексически *viņam jālasa grāmata — ему надо читать книгу, viņš lasīšot grāmatu — он, мол (говорят), будет читать книгу;*
- система времён: прошедшее, настоящее, будущее, сложное будущее РЯ VS *tagadne, saliktā tagadne, pagātne, saliktā pagātne, nākotne, saliktā nākotne* ЛЯ;
- форма глагола 3 лица единственного и множественного числа в РЯ VS одна форма в ЛЯ *viņš, viņa, viņi, viņas lasa, strādāja, darīs — он / она читает, они читают, он / она работал / работала, они работали, он / она будет работать, они будут работать;*
- ГК вида русского глагола VS её отсутствие в ЛЯ;
- ГК падежа: локатив в ЛЯ без предлога VS предложный падеж с предлогом в РЯ *viņš ir skolā / viņš ir darbā — он в школе / он на работе;* инструменталис с предлогом в ЛЯ VS творительный с предлогом и без в РЯ *griezt ar nazi / runāt ar draugu — резать ножом, говорить с другом.*

В области синтаксиса следует упомянуть, например, гипертрофию генитива в ЛЯ, что является особенностью синтаксиса ЛЯ, а также несвойственные РЯ многочисленные конструкции с дательным падежом.

Extraction of Basic Semantic Relations from the Dictionary of Latvian

The research problem discussed is the way how to re-organize an existing Latvian machine-readable dictionary from a traditional form into a semantic representation, where the lexical knowledge is stated out explicitly, in a formal way.

Semantic networks of lexical concepts are organized in terms of word senses and in a way how they are related to other concepts in our mental hierarchy. Typical semantic relations used to organize the taxonomy of such lexical databases are: hyponymy, synonymy, different types of meronymy, as well as basic case roles like agent, patient, theme etc. In spite of implicit presentation conventional dictionaries are rich sources of such knowledge.

The following assumption will be discussed: the genus words (hypernyms) in the definitions of nouns and verbs typically are heads of noun phrase / verb phrase. Syntactic analysis of definitions is necessary in order to find these heads; but again — as syntactic analysis is not a trivial problem, for sake of simplicity heuristics and patterns can be used also at a syntactical level.

A lot of useful information can be extracted by quite simple methods. To continue the same example: (in English) hypernym is a verb that follows *to* (i. e. infinitive); if *to* is followed by several verbs, all of them are hypernyms to keyword.

Parts of speech analyzed in the Dictionary of Latvian: nouns, verbs, adjectives and adverbs as most important parts of speech. In this paper the emphasis is put on verbs. Prevailing relations among verbs are hyponymy and synonymy.

For example (in Latvian): if meaning is defined by infinitive or several infinitives only (without modifiers), it is a definition by synonyms and these verbs (infinitives) are synonyms of headword. In this case definition may contain also: 1) *arī* ‘too’ before an infinitive; 2) string „*Par ...* —“ ‘About...’ at the beginning of the definition; 3) text in brackets, if it does not consist of adverbs.

Pattern for the definition by synonyms: $[\wedge;] \textit{infinitive} (, \textit{infinitive})^* [;.]$

If text in brackets consists of *piem.*, ‘for example’ and one or several infinitives, in most cases these verbs are hyponyms of headword. Pattern: $(\textit{piem.} ([.]^* \textit{infinitive})+ [.]^*)$

Not only taxonomic relations, but also other information can be extracted from the definitions. As it is not a purpose of this paper, only some examples will be given. Text that follows string *parasti par* ‘usually about’ in brackets gives information about selectional restrictions of the word.

Automatically acquired hierarchies of words are necessarily tangled due to polysemy of a large subset of words present in a lexicon. Another problem is caused by general senses of words, which are difficult to define. This leads to confusing multiple inheritances and cross-references. As the goal is taxonomy of senses (concepts) instead of words, the next major step after the extraction of relations is disambiguation of the fillers (words) at the ends of relations.

Morphological and Syntactic Analysis of Latvian: a Different Approach

Reported research is based on an interdisciplinary project at the Institute of Mathematics and Computer Science, University of Latvia, which aims at developing resources and methodologies for efficient integration of the Latvian language and the latest semantic web technologies. If successful, this approach promises to open vast possibilities, enabled by automatic meaning extraction from texts. Although being the aim, text meaning representation is based on several lower levels of language analysis, namely, morphology, syntax and lexical semantics, which have to be deeply understood in advance, unveiling the impact on semantics. Morphological analysis via formal methods is more or less solved problem for virtually any language group. However, comprehensive syntactic analysis and representation, which would facilitate semantic analysis, is still a challenge, especially for less studied languages (mostly those, which have significantly different grammar than English).

As known, Latvian is highly inflective synthetic language with rather free word order. We emphasize the word “rather” due to the fact that there is virtually no language with completely free word order and vice versa [Saussure 1966]. In terms of the grammar structure it is closely related to Lithuanian and also to Slavonic languages. Therefore the model we are proposing in this paper seems to be reasonable and convenient for other inflective languages as well.

Models for grammars with rather strict word order typically follow the top-down approach: sentences are split into constituents or phrases, which are then split into more fine-grained constituents. Conventionally, formalization of constituents is done by means of a context free grammar [Abney 1989; Chomsky 1957]. Languages with a rather free word order are typically described following bottom-up approach: from the surface to a model by drawing links connecting individual words. Conventionally, these links are formalized via a dependency grammar [Covington 2001; 1990].

Although constituency and dependency grammars are basically equivalent and mutually transformable [Gaifman 1965], they propose significantly different views on the parsing problem. Structure of any sentence is essentially based on a simple sentence structure. Considering analysis of a free word order, subordination relations are described between parts of a simple sentence, assuming that each part basically is mapped to a single word. In a result, dependency grammar is defined by a set of head-dependent pairs, where only the agreement of morphological forms between both parts is significant, but not the order in which they appear in a sentence.

Main benefits of the dependency model are the following: (i) considerably fewer rules have to be managed as it is not necessary to describe possible combinations of the parts of a sentence; (ii) rules are more clear-cut; (iii) heads and dependants of each dependency pair can be easily turned into restricted patterns of parts, stating conditions on morphological features and inflectional agreement [Covington 1994]; (iv) the structure of a dependency tree is close to the semantic structure of a sentence with a verb/event at the root, making the mapping more natural.

However, rather free word order means that there exist rather strict constructions as well, like analytical forms of verbs, prepositional phrases et. al. Apart from consequences that make an impact on parsing, this irregularity also helps to avoid several aspects of syntactical ambiguity and helps to synthesize smoother constituents. Therefore, our proposal is to extend

the dependency approach allowing usage of constituent patterns — analytical forms may be seen as complex words with their own (or inherited) morphological and semantic features, though still having simple sentence structure with word-to-word dependencies. A shallow illustration of a hybrid parse tree is presented in Figure 2.

In general, dependency parsing can be based on two simple tables [Covington 1990]: (1) a list of word forms and their morphological descriptions, (2) a list of possible head-dependent pairs, declaring which parts of a sentence may be linked by which syntactical roles. On the base of this infrastructure we are introducing one more table of “word forms” — a list of complex word form patterns what we call x-words (see Figure 3). The backbone of an x-word is formed via production rules analogous to those of the constituency grammar, however, explicitly declared and therefore in fixed order are only mandatory elements of an x-word, where an element is a pattern of a word, either simple or complex one. From the point of view of the dependency table, simple and complex heads/dependants are treated equally. Moreover, in our model subordinate clauses are seen as x-words as well — they link to the principal clause as a single part of a sentence (both syntactically and semantically), and typically they are dependants of a single word (simple or complex). Complex word model is also used to describe equal parts of a sentence and clauses being in coordination relationship. Although x-words are treated as single units, their internal structure is parsed exploiting the same assumption of a simple sentence structure.

Last but not least: we have introduced a convenient, semantic-enabled graphical notation — nested boxes — in addition to the classical tree representation. Each box corresponds to a single word (simple or complex) and has both syntactic and semantic annotations. A list of morphological features and syntactic role is given on the top of a word/box; a label of an ontological concept and semantic role is given at the bottom (see Figure 4).

References

- Abney 1989 — S. P. Abney. *A Computational Model of Human Parsing* // Journal of Psycholinguistic Research. 1989, Nr. 18. — pp. 129–144.
- Chomsky 1957 — N. Chomsky. *Syntactic Structures*. The Hague: Mouton, 1957.
- Covington 1990 — M. A. Covington. *A Dependency Parser for Variable-Word-Order Languages*. Research Report AI-1990-01, Artificial Intelligence Center, The University of Georgia, 1990.
- Covington 1994 — M. A. Covington. *GULP 3.1: An Extension of Prolog for Unification-Based Grammar*. Research Report AI-1994-06, Artificial Intelligence Center, The University of Georgia, 1994.
- Covington 2001 — M. A. Covington. *A fundamental Algorithm for Dependency Parsing* // Proceedings of the 39th Annual ACM Southeast Conference / Eds. John A. Miller and Jeffrey W. Smith. Athens, Georgia, 2001 — pp. 95–102.
- Gaifman 1965 — H. Gaifman. *Dependency Systems and Phrase-Structure Systems* // Information and Control. 1965, Nr. 8. — pp. 304–307.
- Saussure 1966 — F. Saussure. *Course in General Linguistics*. New York: McGraw-Hill Book Company, 1966.

Vasarā	lieli un mazi bērni	dodas	uz Baltijas jūru	, kur	viņi	var peldēties.
In the summer	big and small kids	are going	to the Baltic sea	where	they	can swim.

Figure 1. An example sentence (with direct English translation), which unveils several types of basic syntactic constructions that can be handled by our hybrid model. See further figures below.

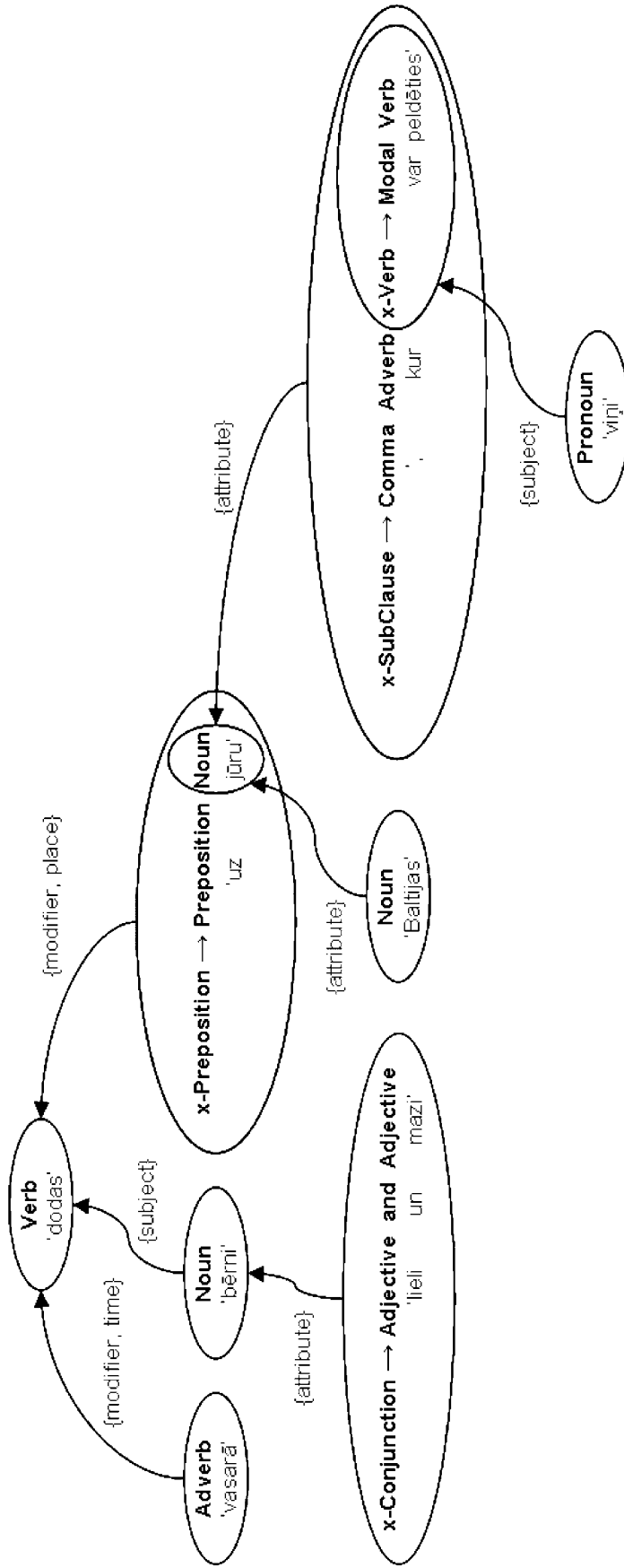


Figure 2. Tree illustration of the sentence in the Figure 1 according to the hybrid model: basically a dependency tree, where particular nodes consist of production rules of context-free grammar.

Word	Morphol. Features	Sememes
vasarā	{N, F, Sg, Loc}	{time}
Baltijas	{N, F, Sg, Gen}	{place}
jūru	{N, F, Sg, Acc}	{water, place}
var	{V, Modal, Trans}	
peldēties	{V, Main, Inf, Refl}	{activity}

Word	Morphol. Features	Sememes
lieli	{Adj, M, Pl, Nom}	{size}
kur	{Adv, Place}	{place}
viņi	{Pron, M, Pl, Nom}	
uz	{Adp, Pre}	
,	{Punct, Comma}	

x-Word	Morphol. Features	Sememes	Substituted by x-Word
x-Equality	{Adj, Gender, Case}		□{Adj, Gender, Case} un{Conj}
x-Preposition	{N, Sg, Acc}	{place}	□{N, Sg, Acc, place} uz{Adp}
x-Verb	{V, Main, Inf, Number}	{Sememes}	□{V, Main, Inf, Number, Sememes}
x-SubClause		{quality}	□{Punct, Comma} kur{Adv, Place} □{V, Main}

Syntactic Role	Head	Dependant	Semantic Role	Ordering
Modifier-of-Time	□{V, Main}	□{N, Loc, time}	Time	
Subject	□{V, Main, Number}	□{Pron, Number, Nom}	Agent	
Attribute	□{N}	x-SubClause{quality}	Quality	head-first
Attribute	□{N, Gen}	□{N}	Owner	head-last

Figure 3. Simplified grammar, which partially defines a sentence structure, such as presented in the Figure 2.

Notations: word $\{f_1, \dots, f_n\}$ — specific word form, whose morphological features conform to (is superset of) the specified list of features;
□ — any word form;

Feature — (i) agreement on the feature between head and dependant or among elements of an x-word,
(ii) value of the variable *Feature* is copied form elements of an x-word to the feature list of the x-word as whole.

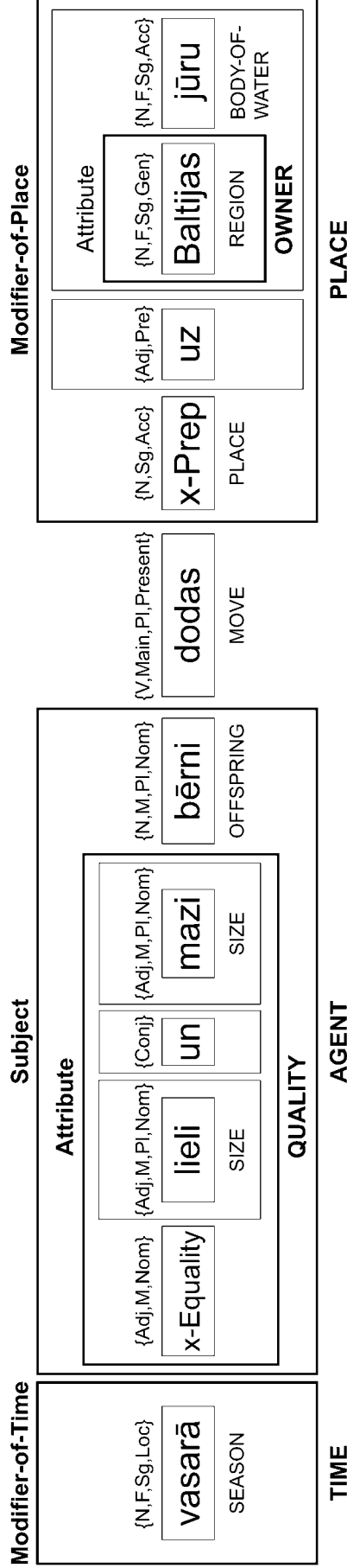


Figure 4. Innovative, graphical representation of syntactic and semi-semantic structures of the sentence presented in the Figure 2. The simplified and partial representation is acquired via grammar and mappings declared in the Figure 3.

Анна Дмитриевна Даугавет (Санкт-Петербург)

Состояние традиционных говоров в Латвии в 20–40 гг. XX в. (по материалам журнала «*Filologu biedrības raksti*»)

Вытеснение диалекта литературным языком или другим диалектом в принципе не отличается от взаимодействия «разных» языков: в том и другом случае имеются два (или более) конкурирующих идиома, один из которых расширяет сферу употребления за счёт другого (других). Сведения о сохранности говора в отдельных волостях, представленные в описаниях латышских говоров, которые публиковались в журнале «*Filologu biedrības raksti*» (FBR 1921–1940), дают некоторое представление о том, как этот процесс протекал в разных частях Латвии в период между двумя мировыми войнами. (Разумеется, основное внимание авторов этих статей сосредоточено на собственно языковых фактах — информация о типичных носителях говора, их возрасте, образовании и занятиях приводится лишь мимоходом и примерно в половине статей отсутствует.) В поле зрения исследователей латышских диалектов в 20–40 гг. попадают следующие моменты.

1) Степень владения говором в зависимости от поколения. Чаще всего встречаются указания на то, что уже среди старшего поколения невозможно найти хороших информантов, чей язык представлял бы собой «чистый» диалект. Заслуживает упоминания факт, что большинство таких утверждений относится к западной Видземе (волости Алоя, Пале, Вилькене, Лиепупе, Туя, Дунте, Стиене, Ладе, Набе, Мазстраупе, Лиелстрапе, Райскумс, Сталбе, Розула). В нескольких случаях старшее поколение противопоставляется младшему как относительно хорошо владеющее диалектом, и реже всего отмечается сохранение диалекта среди младшего поколения, хотя такие случаи тоже присутствуют.

2) В качестве причин отмирания традиционных говоров упоминается как влияние литературного языка, так и взаимодействие с соседними диалектами. Довольно характерной является ситуация, о которой говорит К. Анцитис применительно к волости Акнисте [Ancītis 1935: 161]: «Теперь во всей волости говорят примерно одинаково. По рассказам, прежде существовали различия между разными концами, но не знаю, какие именно».

3) Особый интерес представляет язык переселенцев из других волостей, который может по-разному взаимодействовать с местным говором. В частности, по данным Л. Латковскиса в г. Ливаны, куда переселились люди из окрестностей Крустпилса и верхней Земгале, сформировался поддиалект, вобравший в себя многие привнесённые элементы [Latkovskis 1940: 135].

4) В этом пункте можно объединить наблюдения, касающиеся взаимодействия традиционных говоров с другими языками, а также положения эстонских и литовских диалектов на территории Латвии. Для некоторых местностей конкуренцию традиционным говором составляет не столько литературный латышский, сколько русский и польский языки. Так, по свидетельству М. Алксниса [Alksnis 1932: 26], жители волости Краслава в обществе стремятся говорить по-польски или по-русски. Любопытно, что литовское население волости Гарсене владеет латышским языком в его литературной форме (освоенной в школе), а не местным латышским диалектом [Āboliņa 1927: 129–130].

Литература

- Āboliņa 1927 — O. Āboliņa. Susējas pagasta izloksne // FBR VII, 1927. — 129.–153. lpp.
Alksnis 1932 — M. Alksnis. Krāslavas izloksne // FBR XII, 1932. — 26.–48. lpp.
Ancītis 1935 — K. Ancītis. Par Aknīstes izloksni // FBR XV, 1935. — 161.–196. lpp.
Latkovskis 1940 — L. Latkovskis. Līvānu izloksne // FBR XX, 1940. — 134.–158. lpp.

Sociolingvistikas attīstība Latvijā

1. Sociolingvistika ir zinātne, kas pētī attiecības starp valodu un sabiedrību, starp valodas lietojumu un sociālajām struktūrām, kurās valodas lietotāji dzīvo. Sociolingvistika ir atjaunināta pieeja tūkstošiem gadu ilgušajai valodas izpētei, tajā plaši iesaistot socioloģijas, etnoloģijas, psiholoģijas u. c. dotumus. Mūsdienu sociolingvisti izmanto dažādas metodes visatšķirīgāko parādību izpētei, bet kopīgs ir viens: valodas lietojums tiek analizēts sociālā kontekstā.

2. Mikrosociolingvistika pētī korelāciju starp variantiem valodā un starp atbilstošajiem variantiem sociālajā grupā, kas tos lieto. Makrosociolingvistika valodu pētī kā kultūras un politisku fenomenu. Uzmanība tiek pievērsta ne valodas lietojumam indivīda un sociālo grupu līmenī, bet pašu valodu un to variantu attiecībām.

3. Klasiskā sociolingvistika kā zinātne attīstījusi ASV 60. gados (Labovs, Braits, Ervina-Tripa, Fergusons, Fišmens, Gampercs, Haimzs, Grimšo u. c.). Pašlaik sociolingvistu skolas pastāv gandrīz visās pasaules valstīs, iznāk ap 50 specializētu žurnālu, kā priekšmets tā tiek docēta visās lielākajās pasaules universitātēs.

4. Ilgi pirms klasiskās sociolingvistikas izveides jau iespējams runāt par sociolingvistisku pieeju daudzu tautu valodniecības vēsturē (piem., Humbolts, Boduēns de Kurtenē). Arī latviešu valodniecībā lingvopolitisku apsvērumu dēļ skaidri iezīmējas valodas un sabiedrības attīstības cieša saistība (Alunāns, Kronvalds, Velme, īpaši jau Mīlenbahs; 20.–30. gados Blesse, Šmits, Ķiķauka).

5. 20. gs. 70. gados sākas sociālās lingvistikas bums PSRS. Tas pamatojas uz galvenokārt 20.–30. gadu krievu valodniecības tradīcijām, īpaši sociālo variantu un literārās valodas izpētē, bet aizgūst arī Rietumu idejas. Gandrīz visu ievērojamāko sociolingvistu darbu tulkojumi, kaut ar šīs „buržuāziskās pseidozinātnes“ obligāto kritiku priekšvārdā un piezīmēs, ir pieejami krievu valodā. Rodas vērtīgi pētījumi mikrosociolingvistikā un normas teorijā (Šveicers, Krisins, Bondaļetovs, Leontjevs, Skvorcovs u. c.), kas lielā mērā pozitīvi ietekmē arī latviešu valodniecību. Turpretim makrosociolingvistika, īpaši valodas politika, kļūst par ideoloģizētu zinātnes nozari, kuras virsuzdevums ir teorētiski pamatot „nacionālo“ krievu divvalodību (Dešerijevs, Isajevs, Guboglo, Mihaļčenko u. c.). Arī Latvijā tiek izdots bibliogrāfiskais rādītājs „Latviešu-krievu divvalodība“ (Drīzule, Gerentoviča 1987) ar 459 vienībām. 1987. g. V. Drīzules vadībā Vissavienības programmas „Nacionālo valodu attīstība sociālisma periodā“ ietvaros tiek veikts pirmais apjomīgais sociolingvistiskais pētījums par divvalodību dažādās sociolingvistiskās funkcijās. Mainoties politiskajai iekārtai, pētījums publicēts netiek, bet tā galvenais secinājums: „Vairumā jomu latviešu valoda funkcionē kā latviešu-krievu divvalodības komponents uz plašas krievu valodas izplatības fona“ noder par argumentu Atmodas perioda cīņā par latviešu valodas tiesībām.

6. Atmodas gados (1988–1992) daudzi latviešu valodnieki pievēršas valodas politikai, tā veicinot arī šīs zinātnes nozares attīstību (Balode, Bušs, Druviete, Hirša, Ikere, Kuzina, Kvašīte, Skujiņa, Sīlis, Valdmanis, Veisbergs, īpaši Blinkena). Savu ieguldījumu sniedz trimdas zinātnieki (Dreifelds, Kārklīņa, Kangere, Ozoliņš). Latvijā tiek izstrādāta lingvistiskā likumdošana un izveidotas nepieciešamās institūcijas latviešu valodas tiesību aizsardzībai un valodu koeksistences nodrošināšanai.

7. 1992. gadā Latviešu valodas institūtā tiek izveidota Sociolingvistikas daļa (vad. I. Druviete), latviešu sociolingvisti sāk piedalīties starptautiskās konferencēs, stažēties ārval-

stu pētniecības centros, publicēties nozīmīgos sociolingvistikas žurnālos. Kopš 1993. gada latviešu pētnieku devums tiek atspoguļots bibliogrāfijā „Sociolinguistica. International Yearbook of European Sociolinguistics (Niemeyer)“. Kopš 1997. gada sociolingvistikas kurss tiek docēts Daugavpils universitātē, gadu vēlāk — vairākās Latvijas Universitātes fakultātēs, pašlaik arī Liepājas Akadēmijā. Ir tikuši īstenoti vairāki LZP projekti, kā „Valodas funkcionālie un sociolingvistiskie aspekti“, „Latvijas sociolingvistiskās situācijas monitorings“, „Lingvistiskā likumdošana Eiropas Savienībā un Latvijā“, „Latvijas sabiedrības lingvistiskā integrācija“. Nopietns darbs veikts sarunvalodas izpētē. Promocijas darbus sociolingvistikā aizstāvējušas V. Ernstsone, V. Poriņa, L. Lauze, D. Joma, L. Tidriķe, pašlaik doktorantūrā mācās 4 pētnieces.

8. Latviešu sociolingvisti plaši iesaistījušies Eiropas Savienības valodas situācijas un valodas politikas izpētē (Euromosaic III. Presence of Regional and Minority Language Groups in the European Union's new acceding countries); izpilda divus 6. ietvarprogrammas projektus, darbojas žurnāla „Language Policy (Springer)“ redkolēģijā un vairākās starptautiskās sociolingvistu organizācijās. Turpmāk nepieciešams padziļināt pētījumus mikrosociolingvistikā — valodas sociālo un teritoriālo variantu izpētē, valodas standartizācijā u. c.

Анжелика Дубасова (Минск — Санкт-Петербург)

О web-версии словаря балтийской терминологии

Разрабатываемый «Словарь балтийской терминологии (web-версия)» рассчитан на широкий круг пользователей: лингвистов, историков, географов и всех, создающих тексты, так или иначе связанные с балтийской тематикой.

В числе наиболее существенных отличий web-версии от печатной укажем:

- объём web-версии теоретически не имеет границ, что позволяет помещать в словарь все балтийские термины, независимо от области их применения, а также все имена собственные; таким образом, возможно полное представление терминов из области лингвистики, географии, истории и др., в то время как для печатной версии необходим отбор наиболее важных и частотных терминов;
- в web-версии возможно давать полную грамматическую информацию, указывать достаточно полно сочетаемость слов, а также информацию о системных связях слова;
- в web-версию возможно без ограничений включать графический и иллюстративный материал: карты (современные и исторические), таблицы, схемы и т. п.;
- web-версию можно постоянно обновлять и корректировать; со временем планируется включение большего числа языков (английского, польского, немецкого, французского, итальянского и др.).

Предварительная структура словаря содержит такие части, как: «О словаре», «Все термины в алфавитном порядке», «Лингвистика», «История и культура», «География», «Карты», «Нормативные документы», «Библиография», «Список сокращений». В дальнейшем возможно создание форума, посвящённого проблемам терминологии в целом и терминологии данной области.

Структура словаря предполагает не только толкование терминов и попытку их нормирования, но и более явное, по сравнению с печатной версией, совмещение сло-

варя с энциклопедическим справочником. Цель — толкование терминов в качестве справок, а также их систематизация и унификация. Начальная версия содержит термины из разных областей знания, объединённых балтийской тематикой.

Термины — как в общей части, так и в тематических частях — расположены в алфавитном порядке. Если термин представляет собой словосочетание, то он располагается по алфавиту первого слова, а не стержневого. Заголовочное слово даётся (для облегчения машинного поиска) прописными буквами; в случае, если термин пишется с прописной буквы, в скобках приводится его правильное написание. В статье приводятся все встреченные варианты термина (за исключением явных ошибок и опечаток); эти же варианты приводятся также на своём месте в алфавитном порядке с отсылкой к основному термину. О принципах выбора основного термина говорится в разделе «О словаре».

Поскольку определение термина не всегда может дать полное представление о его месте в системе данного раздела науки, о сфере его применения и т. д., по возможности даётся более развёрнутая статья, раскрывающая объём и содержание соответствующего явления, понятия и т. п. с иллюстрациями и ссылками (в том числе за пределы данного словаря); упоминаются наиболее важные теории и положения, в которых существует данный термин; приводятся географические и диалектологические карты.

На основании словаря предполагается значительное расширение и детализация статей по балтийской тематике в русской части пользующейся большой популярностью Свободной энциклопедии Wikipedia, что, в конечном счёте, послужит унификации балтийской терминологии.

Мария Вячеславовна Завьялова (Москва)

Словесные формулы в вербальных ассоциациях литовцев, поляков и русских

Ассоциативные реакции в большинстве своем предсказуемы и стереотипны. Можно сказать, все вербальные ассоциации представляют собой одни и те же словесные формулы, многократно воспроизводимые разными испытуемыми во время тестирования. Однако статус этих воспроизводимых формул различен. В докладе представлена попытка классификации подобных формул по разным параметрам на материале ассоциативного эксперимента с носителями разных языков (русского, литовского и польского) и билингвов.

Условно можно выделить две категории формульных ассоциативных реакций: это «формулы левого полушария» (основанные на категоризации, классификации и стереотипных представлениях), и «формулы правого полушария» — клише и штампы, основанные на часто воспроизводимых прецедентных текстах.

В реакциях разных групп испытуемых существуют различия в частоте воспроизведения стереотипных формул. В целом ответы русских более стереотипны, чем литовцев и поляков. В ответах билингвов по-русски их также больше, чем в ответах по-литовски. Интересно, что литовцы чаще воспроизводят прецедентные тексты на втором языке (т. е. на русском), чем на своем. Всё это может свидетельствовать о том, что

русский язык более ориентирован на воспроизводство и повторение таких текстов, если можно так сказать, более насыщен прецедентными текстами. Характерно и то, что билингвы, овладевая русским языком как вторым, с легкостью усваивают подобные штампы и воспроизводят их спонтанно. Это говорит о том, что немалую долю языковой компетенции составляет набор прецедентных текстов изучаемого языка.

Ieva Zuicena (Rīga)

Mūsdienu latviešu valodas vārdnīca (elektroniskā versija) — realitāte un iespējas

1. Viens no projektiem, pie kā patlaban strādā Latvijas Universitātes Latviešu valodas institūta darbinieki (I. Migla, L. Oldere, A. Roze, I. Šmidebergs, I. Zuicena) ir jaunas latviešu valodas skaidrojošās vārdnīcas veidošana (Mūsdienu latviešu valodas vārdnīca, apmēram 70 000 šķirkļu).
2. Ar Valsts valodas aģentūras finansiālu atbalstu internetā ir pieejama Mūsdienu latviešu valodas vārdnīcas sākumdaļa no *A* līdz *I* burtam (www.ailab.lv/LVV).
3. Vārdnīcai elektroniskā versijā ir vairākas priekšrocības:
 - Publicēšanai internetā nav nepieciešami lieli materiālie resursi.
 - Vārdnīcas tekstu pēc vajadzības var labot, papildināt, rediģēt. Pastāvīgi atjaunota, šāda vārdnīca nenoveco.
 - Ja elektroniskajā vārdnīcā katrs šķirkļa elements ir attiecīgi marķēts, vārdnīcas lietotājs var saņemt nepieciešamo informāciju jau sasistematizētā veidā (piem., sargrupētu pēc gramatiskajām pazīmēm, leksikosemantiskajām grupām, leksikostilistiskajām norādēm, cilmes).
4. Pagaidām Mūsdienu latviešu valodas vārdnīcas šķirkļi ir ievadīti internetā kā veselas vienības. Nākotnē būtu vēlams izstrādāt attiecīgu programmu šķirkļa elementu marķēšanai, lai arī šīs vārdnīcas lietotāji varētu baudīt XXI gadsimta informācijas tehnoloģiju radītās iespējas.

Referāta gaitā paredzēts demonstrēt fragmentus no Oksfordas angļu valodas vārdnīcas elektroniskajā versijā.

Regīna Kvašīte (Šauļi)

Lietuvieši Latvijā: sociolingvistiskais aspekts

Lietuvieši Latvijā apmetās uz dzīvi dažādos laikos politisku un ekonomisku apstākļu dēļ, un to procents visu laiku ir svārstījies starp 1,2–1,5 %. 2006. gada 1. janvārī pēc Pilsonības un migrācijas lietu pārvaldes Iedzīvotāju reģistra datiem Latvijā dzīvoja 31 307 lietuvieši resp. 1,4 % visu Latvijas iedzīvotāju. Tātad lietuvieši skaita ziņā pašlaik ir sestā mazākumtautība Latvijā.

Lietuviešu dzīves Latvijā, to kultūras, arī valodas īpatnību pētīšanai pievēršas gan Lietuvas, gan Latvijas dažādu nozaru zinātnieki. Līdz šim veiktie un izdotie pētījumi veltīti atsevišķu novadu lietuviešiem, kā arī skar dažādus to dzīves aspektus: pēdējā desmitgadē iznāku-

šas vairākas lielāka apjoma zinātniskas un populāras publikācijas, kurās aplūkota lietuviešu ieceļošana, iesakņošanās un dzīve Latvijā [Jēkabsons 2003; Merkienē 2005; Treija, Dribins 1998].

Valoda — viens no svarīgākajiem Latvijas lietuviešu etniskās pašapziņas rādītājiem. Apzinot Latvijas lietuviešu valodu kā vienu no nacionālās identitātes izpausmēm, valodnieki uzmanību galvenokārt pievērsuši aiz Lietuvas robežām, tostarp arī Latvijā, dzīvojošo lietuviešu izloksnēm. Iznācis pierobežas izlokšņu fonoloģijas apraksts, kur daudz uzmanības veltīts Latvijas lietuviešu valodas īpatnībām [Garšva 2005], par šiem jautājumiem, īpaši par Daugavpils apkārtnē dzīvojošo lietuviešu fonētiku un gramatiku, ir rakstīts arī iepriekš [Grinaveckienē 1980; 1983] u. c. Ir arī mēģinājumi skatīt valodu kontaktu izpausmes leksikā [Kvašytē 2003] u. c. Tie balstīti uz divu ekspedīciju vērojumiem Saldus un Dobeles rajonu apdzīvotajās vietās, attiecīgi Rubā (1993) un Bukaišos (1995), kur samērā kompakti dzīvo lietuvieši. Uz minēto ekspedīciju gaitā konstatētajiem faktiem un citiem personīgajiem vērojumiem balstīts arī šis referāts.

Sociolingvistiskajā skatījumā lietuviešu dzimtās valodas kvalitāti ir ietekmējuši vairāki apstākļi. Vispirms to nolēma laikposms, kad un kādā vecumā persona apmetās uz dzīvi Latvijā. Piem., tiem, kas apmetās uz dzīvi Latvijā padomju gados, nebija iespēju mācīties lietuviešu valodu skolā, jo tajā laikā tās (tāpat kā citu padomju republiku) teritorijā nebija mazākumtautību skolu. Tādējādi, ja lietuvietis izeļoja no Lietuvas brieduma gados, viņš bija apmeklējis skolu un viņam bija lietuviešu valodas pamati vai pat iegūta dažāda līmeņa izglītība. Visa tā nebija un nevar būt lietuviešiem, kas dzimuši Latvijā. Ne mazākā mērā atšķirības nolemj ģimenes nacionālais sastāvs un personīgais skatījums uz dzimto valodu kā vērtību. Tāpēc īpaši svarīgi ir arī tuvinieku un saziņas partneru skatījums uz citvalodiski runājošu personu un viņa dzimto valodu. Personām, kas nav mācījušās skolā, ir grūti perfekti mācēt literāro valodu — viņu valodā vienmēr būs jūtama sarunvalodas nokrāsa vai izloksnes ietekme, jo mazākumtautības valoda neveic visas valodai raksturīgās sabiedriskās funkcijas, kā to dara savā zemē lietojamā valoda.

Lietuviešu valoda Latvijā ir izstumta no publiskās dzīves sfērām un arvien noteiktāk ir kļuvusi par mājas valodu. Pašlaik notiek zināmas pārmaiņas šajā jomā — pēc 1990. gada dažādās Latvijas pilsētās (Rīgā, Jelgavā, Bauskā, Daugavpilī) ir izveidotas lietuviešu skolas, svētdienas skolas, nodibinātas lietuviešu biedrības, darbojas mākslas kolektīvi u. tml. Tomēr iespēja lietot dzimto valodu citvalodu vidē pārsvarā saistīta ar individuālām un ģimenes, neliela draugu loka vajadzībām (mazliet plašāka šī iespēja ir teritorijās, kur lietuvieši dzīvo kompakti). Tāpēc Latvijas lietuviešu starpā visvairāk izplatīta lietuviešu sarunvaloda resp. literārās valodas paveids, kas tiek lietots ikdienas saziņā un kam raksturīgas saziņas sfēras nolemtas iezīmes.

Nav šaubu par to, ka citvalodu vidē pastāvīgi dzīvojošiem cilvēkiem arvien nākas apzināti izvēlēties saziņas valodu (ja viņi prot vairāk nekā vienu — dzimto — valodu) un konkrētās valodas izteiksmes formu. Ekspedīciju laikā izdarītajos pierakstos atrodami fakti, ka respondenti ļoti cenšas runāt pareizā (tā, kā viņi to iedomājas) lietuviešu valodā un bieži vien atkārtoti, ka nezina lietuvisko vārdu vajadzīgajai reālijai nosaukt, vai paši uzsver, ka vairs neprot runāt lietuviski. Tātad var secināt, ka iztaujātāji psiholoģiski ietekmē respondentu runu, kaut arī to nelabo un nekā nevērtē tās kvalitāti. Domājams, ka visu valodas līmeņu (izņemot fonētikas) īpatnībām fiksēt vislabāk der tāds valodas pieraksts, kad cilvēks nezina, ka viņa runa tiek vērota.

Kvantitatīvās un datorlingvistikas aizsākumi Latvijā 20. gs. 60.–70. gados

Matemātiskās lingvistikas (tā laika terminoloģijā) aizsākumos būtisks faktors ir tas, ka 1959. g. tika nodibināts LVU Skaitļošanas centrs (tagad — LU Matemātikas un informātikas institūts) un 1960. g. LPSR ZA Elektronikas un skaitļošanas tehnikas institūts (ESTI).

Pārsteidzoši, ka jau 1959. g. vasarā LVU Vēstures un filoloģijas fakultātes latviešu valodas un literatūras specialitātes bukletā reflektantiem bija minēta iespēja iegūt zināšanas matemātiskajā lingvistikā, tiesa — bez reāla pamata un nodrošinājuma. Tomēr trīs ieinteresētas studentes pēc pāris gadiem ar toreizējā Latviešu valodas katedras vadītāja, arī LVU Zinātņu prorektora doc. A. Ozola atbalstu savus kursadarus izstrādāja par latviešu valodas formalizētas analīzes jautājumiem. Tos vadīja Krievu valodas katedras docents V. Juriks, kas kopā ar A. Gobzemi, V. Gorobecu un T. Jakubaiti 1960. g. septembrī ESTI bija piedalījies tulkošanas mēģinājumā ar ESM LM-3 no krievu valodas latviešu valodā. Turpmākos kursadarus studentēm S. Kļaviņai un R. Prūsei vadīja Ļeņingradas universitātes aspirants N. Mecs no Latvijas. Tā studentes 1963. g. nonāca ĻVU Eksperimentālajā mašintulkošanas laboratorijā (ELAMP), kas N. Andrejeva vadībā veidoja starpniekvalodu automātiskai tulkošanai. 1964. g. tika izstrādāti diplomdarbi latviešu valodas morfoloģiskajā analīzē atbilstoši šī projekta stratēģijai. S. Kļaviņa tika atstāta universitātē aspirantūrā, bet R. Prūse uzsāka darbu LPSR ZA Valodas un literatūras institūtā, kur 1965. g. T. Jakubaites vadībā nodibinājās Matemātiskās lingvistikas laboratorija, lai izstrādātu „Latviešu valodas biežuma vārdnīcu“. No 1966. līdz 1976. g. tika publicēti šīs vārdnīcas 4 sējumi un apvienotais sējums. Vārdnīca rāda vārdu un to gramatisko formu lietošanas biežumu mūsdienu latviešu literārās valodas trīs funkcionālajos stilos: zinātnes, publicistikas un daiļliteratūras tekstos. Kā papildproduktu 1976. g. laboratorija publicēja arī „Latviešu valodas inverso biežuma vārdnīcu“.

S. Kļaviņa, kļuvusi par Ļeņingradas profesora R. Piotrovskas aspiranti, iekļāvās Vissavienības grupā „Статистика речи“ un sagatavoja latviešu publicistikas biežuma vārdnīcas manuskriptu (Kļaviņa 1968), arī mācību līdzekli „Statistiskās metodes valodniecībā“ (Rīga: LVU, 1968; 2., pārstrādāts un papildināts izdevums ar nosaukumu „Statistika valodniecībā“ — Rīga, 1980). LU Vēstures un filoloģijas fakultātē aktīvi darbojās Studentu zinātniskās biedrības Matemātiskās lingvistikas pulciņš kā saistošs papildinājums studijām — ar Maskavas un Ļeņingradas vieslektoriem (J. Apresjans, B. Suhotins, R. Piotrovskis u. c.), ar Vissavienības matemātiskās lingvistikas studentu zinātniskajām konferencēm Kijevā, Tbilisi u. c.

60. gadu 2. pusē un 70. gados latviešu valodas pētnieki galvenokārt nodarbojās ar skaitlisko datu uzkrāšanu un statistisko parametru iegūšanu. Empīrisku datu korektas matemātiskas apstrādes paraugus sniedza prof. I. Liepas grāmata „Biometrija“ (Rīga 1974), arī Tartu universitātes doc. J. Tuldavas u. c. publikācijas.

Pārskats par šā perioda kvantitatīvās lingvistikas veikumu rodams S. Kļaviņas rakstā [Kļaviņa 1999], bet t. s. matemātiskās lingvistikas darbu bibliogrāfija — V. Drīzules sastādītajos rādītājos: līdz 1975. g. [Дризул 1976], no 1976. g. [Дризуле 1985], kā arī prof. R. Kēlera (Köhler) sastādītajā [Bibliography 1995].

Kvantitatīvā analīze tika lietota ne tikai mūsdienu valodas leksikas un gramatikas elementu lietojuma izziņai (Ceplītis 1964; Jakubaite 1977, 1978; Kļaviņa 1968, 1976), bet arī literārās valodas vēstures (Freidenfelds 1970, Miķelsone 1967) un skolēnu rakstu valodas attīstības (Freidenfelds 1976) pētījumos. Tika noteikts vārdšķiru biežuma statistiskais

sadalījums (Kļaviņa 1969), ar korelācijas koeficientiem tika novērtēta korelācija starp dažādu funkcionālo stilu tekstiem vārdu krājuma un vārdšķiru lietojumā (Kļaviņa 1976).

Uz topošās Latviešu valodas inversās vārdnīcas (Soida, Kļaviņa 1970) un uz publicistikas biežuma vārdnīcas (Kļaviņa 1968) inversijas pamata radās iespēja veikt pilnīgu vārdrašanās elementu, modeļu un tipu kvantitatīvo analīzi, šķirot vārdnīcas un teksta aspektu, resp. nosakot produktivitāti un lietošanas biežumu, kā arī to savstarpējās attiecības (Kļaviņa 1969).

60.–70. gadu darbu vainagoja aizstāvētās zinātņu kandidāta (pašreizējā terminoloģijā — doktora) disertācijas. Automātiskai latviešu zinātniski tehnisko tekstu morfoloģiskai analīzei tika veltīta V. Drīzules disertācija (Minskā 1975), bet latviešu valodas funkcionālo stilu lingvostatistikai izpētei — S. Kļaviņas disertācija (Viļņā 1977).

Literatūra

Bibliography 1995 — Bibliography of Quantitative Linguistics. Amsterdam / Philadelphia: John Benjamins Publishing Company, 1995.

Kļaviņa 1999 — Kļaviņa S. Latviešu valoda kvantitatīvās lingvistikas spoguļi. // Baltu filoloģija, VIII. Latvijas Universitātes Zinātniskie raksti, 619. sēj. Rīga, 1999. — 19.–26. lpp.

Дризул 1976 — Дризул В. А. Использование методов математической лингвистики для изучения латышского языка (библиографический обзор). Рига: Зинатне, 1976.

Дризуле 1985 — Дризуле В. А. Лингвостатистическое и прикладное изучение балтийских языков. Библиографический обзор. 1975–1983 гг. Рига: Зинатне, 1985.

Kristīne Levāne-Petrova (Rīga)

Paralēlā korpusa nozīme Eiropas Savienības tiesību aktu tulkošanā

Paralēlo valodas tekstu korpusu aizsākumi meklējami 20. gadsimta 90. gados, kad arī radās jaunas tehniskas tekstu apstrādes iespējas un arī vajadzība pēc jauna veida tekstu krājumiem.

Paralēlo tekstu korpusam nav tikai paralēli uzkrāti parasti divvalodu teksti, bet tie ir arī īpaši apkopotī, apstrādāti. Tāpat korpusam parasti tiek izstrādāta īpaša programmatūra dažādu nepieciešamo datu ieguvei.

Domājams, ka pagaidām Latvijā nav īsti noformēta paralēlo tekstu korpusa, kas būtu plaši lietojams dažādiem mērķiem, piemēram, pētījumiem dažādās nozarēs, izmantošanai tulkošanas procesā u. tml.

Eiropas Savienības dokumenti pieder pie lietišķo rakstu valodas stila juridisko tekstu paveida un, tulkojot Eiropas Savienības dokumentus, ir jāņem vērā „Tulkošanas un Terminoloģijas centra“ izstrādātajā „Eiropas Savienības tiesību aktu tulkošanas rokasgrāmatā“ izvirzītie principi.

Eiropas Savienības tekstu un citiem tulkotājiem internetā ir publiski pieejams paralēlais tekstu korpus „EUR-lex“ (http://eur-lex.europa.eu/RECH_menu.do), kas gan nav sastādīts un kam nav pieejami dažādi programmrīki, tomēr tajā ir iespējamas dažādas meklēšanas iespējas. Šīs iespējas tulkotājam ir pietiekamas, lai varētu korpusu izmantot gan kā vienkāršu tulkojošo vārdnīcu, gan terminu datu bāzi, gan palīgu dažādu grūti saprotamu konstrukciju atveidē. Korpusam ir noderīgs arī tāpēc, ka latviešu valodas tulkojošās vārdnīcas bieži vien neatbilst mūsdienu valodas prasībām un ir novecojušas vai arī ļoti nepilnīgas attiecībā uz populārākajiem valodu pāriem, piemēram, angļu-latviešu, latviešu-angļu.

Nereti tulkojošajās vārdnīcās atrodamie ekvivalenti nav tieši izmantojami, tulkojot tekstus, piemēram, angļu valodas darbības vārda *provide* atbilstošais „Tildes“ tulkojošajā vārdnīcā atrodamais ekvivalents ir *apgādāt, nodrošināt, paredzēt (notikumus)*. Tomēr vārdkopā *to provide information* atbilstošais ekvivalents ir *sniegt informāciju*, bet vārds *sniegt* vārdnīcā nav norādīts kā vārda *provide* ekvivalents.

Strādājot ar ES dokumentiem, paralēlajā korpusā „EUR-lex“ atrastas reālas jaunas vārdu nozīmes, kas nav atspoguļotas nevienā no pašreizējām angļu-latviešu, latviešu-angļu tulkojošajām vārdnīcām, piemēram, *overburden* vārdnīcā minētā atbilde ir *pārslodze*, bet reāli, tulkojot tekstu par kalnrūpniecību, atbilstošais ekvivalents ir *noderīgie iezī*. Jāmin, ka daudzi no atrastajiem vārdiem ir pietiekami specifiski un varētu arī nebūt iekļauti parastajā tulkojošajā vārdnīcā, bet, piemēram, specifiskā terminoloģijas vārdnīcā.

Tāpat arī atrasti daudzi vārdi un vārdu savienojumi, kuru nozīme ir atrodama tikai paralēlajā korpusā, piemēram, *companion animals — mājdzīvnieki*.

Ekstragējot un citādi apstrādājot šāda veida korpusu, varētu papildināt jau esošās tulkojošās vārdnīcas un izveidot arī jaunas terminu vārdnīcas. Šeit gan jāmin, ka „EUR-lex“ ir tikai lietišķā stila teksti, tāpēc tie nebūtu attiecināmi uz visu valodu kopumā.

Vineta Poriņa (Rīga)

Valsts valodas lietojuma jautājums imigrantu integrācijā

Imigrācija ir viena no jaunākajām ES politikām. Pēdējo gadu laikā imigrācija ir kļuvusi par satraucošu un strīdīgu jautājumu, jo daļa Eiropas iedzīvotāju, un ne maza daļa, saskata tajā draudus nacionālajai identitātei. Tomēr tā arī tiek uzskatīta par darbaspēka trūkuma problēmu risinātāju, jo, kā prognozēts demogrāfiskajos pētījumos, iedzīvotāju skaits ES valstīs turpinās augt tikai līdz 2025. gadam un pēc tam tas samazināsies. Latvijā 2030. gadā jau būs tikai 2,0 milj., un 2040. gadā un vēlāk — 1,9 milj. iedzīvotāju. Tomēr neviens nav apgalvojis un garantējis, ka imigrācija pilnībā atrisinās darbaspēka trūkumu ES dalībvalstīs.

Pēdējo gadu migrācijas procesi liecina, ka no Latvijas emigrējušo skaits ir lielāks par imigrējušo skaitu, bez tam — darbaspējas vecumā emigrējušo vidū sieviešu skaits pārsniedz vīriešu skaitu, kas arī neveicina jaunākās paaudzes palikšanu Latvijā, jo bērni lielā daļā gadījumu dzīvo kopā ar mātēm. Tas viss vedina secināt, ka esošo darba roku trūkumu būs spiesti aizpildīt iebrucēji no citām valstīm.

Latvija nevar izcelties uz t. s. veco Eiropas dalībvalstu fona ar lielu reģistrēto bēgļu skaitu — to 2004. gadā nebija neviena, 2003. gadā — 5, savukārt patvēruma meklētāji ir bijuši 6 cilvēki 2004. gadā. Šie skaitļi Latviju ierindo Eiropā unikālu valstu vidū, jo ekonomiski attīstītās valstīs gan bēgļu, gan patvēruma meklētāju skaits mērāms simtos un tūkstošos. Latvijā tie ir nelieli skaitļi, tāpēc ar etnisko un lingvistisko integrāciju saistītajā izglītībā joprojām aktuālākais ir nevis nedaudzo bēgļu un patvēruma meklētāju izglītības jautājumi, bet gan Latvijā dzīvojošo mazākumtautību cilvēku iespējas un vēlme lietot latviešu valodu.

Ņemot vērā gan esošās migrācijas tendences, gan Latvijā dzīvojošo mazākumtautību un latviešu vajadzību pēc valsts valodas lietojuma, ES finansētā projekta ietvaros, kurā Latvijas Universitātes pētnieki un mācītspēki sadarbojās ar Rietumeiropas lielāko valstu mācītspēkiem, ir izveidotas izglītības programmas imigrantu integrācijai valstī. Latvijai veidotā programma veltīta integrācijai uz valsts valodas pamata. Viens no programmas pamatuzde-

vumiem ir trenēt psiholoģisko gatavību lietot valsts valodu valodas izvēles situācijās. Programma ir domāta imigrantu integrācijai Latvijā, bet tā ir labi izmantojama arī citām mērķa grupām: programmas mērķa grupas ir ne tikai imigranti un daļa mazākumtautību pārstāvju, kas ir funkcionāli monolingvi krievu valodā, — viena no mērķa grupām ir latvieši. Latvijā šī ir pirmā imigrantu integrācijas izglītības programma, kas līdztekus citiem uzdevumiem, pirmām kārtām trenē psiholoģisko gatavību lietot valsts valodu.

Erika Rimkutė (Kaunas)

Automatinē morfologinē lietuvių kalbos analizė

Pranešime bus pristatyta automatinė morfologinė lietuvių kalbos analizė, Vytauto Didžiojo universiteto Kompiuterinės lingvistikos centre sukurtas morfologinis anotatorius, iš daugiau kaip 111 mln. žodžių sudarytas morfologiškai anotuotas tekstynas, jame išryškėjusios lietuvių kalbos ypatybės.

Pirmoji automatinės morfologinės lietuvių kalbos analizės programos versija atpažintiems žodžiams pateikdavo antraštinius žodžius (lemas) ir morfologines pažymas, todėl renkant morfologiškai anotuotą tekstyną buvo susidurta su morfologiniu daugiareikšmiškumu. Beveik pusė lietuvių kalbos žodžių ir žodžių formų yra morfologiškai daugiareikšmiai. Tai reiškia, kad automatinės morfologinės analizės programos jiems pateikia dvi ar daugiau lemuų arba dvi ar daugiau morfologinių pažymų (pvz., formai *laisvės* morfologinės analizės programa pateikia dvi lemas: *laisvė* ir *laisvėti*. Ir daiktavardžio, ir veiksmažodžio formos dar yra morfologiškai daugiareikšmės: daiktavardžio forma *laisvės* gali būti vienaskaitos kilmininkas arba daugiskaitos vardininkas, o veiksmažodžio forma — tai vienaskaitos arba daugiskaitos trečiasis asmuo).

Neseniai baigtas kurti lietuvių kalbos morfologinis anotatorius iš kelių galimų variantų pasirenka labiausiai tikėtiną tame kontekste. Statistiniu pagrindu veikiantis morfologinis anotatorius padėjo išspręsti didelę lietuvių kalbos morfologinio daugiareikšmiškumo problemą. Dėl to automatinė morfologinė analizė labai paspartėjo: per kelias minutes galima išanalizuoti tūkstančius žodžių. Morfologinis anotatorius veikia 94 % tikslumu. Vis dar lieka problema neatpažinti žodžiai: kai kurie tikriniai daiktavardžiai, sutrumpėję žodžiai, užsienio kalbų intarpai.

Išanalizavus anotuoto tekstyno duomenis, paaiškėjo įdomių faktų apie lietuvių kalbą, kurie gali praversti besimokantiems lietuvių kalbos. Lietuvių kalba dažniausiai yra pristatoma kaip archajiška, sudėtinga fleksinė kalba, tačiau morfologiškai anotuoto tekstyno duomenys rodo, kad realiai vartojama labai nedaug kaitybinių formų — maždaug tik 2,34 formos (šis skaičius gautas visų lemuų skaičių padalinus iš skirtingų formų skaičiaus). Nors dauguma lietuvių kalbos daiktavardžių teoriškai gali turėti 14 formų (7 vienaskaitos, 7 daugiskaitos linksnius), morfologiškai anotuotame tekстыne pavartota maždaug 2,54 formos. Tai reiškia, kad dažniausiai vartojame tik kilmininko, galininko ir vardininko — dažniausių linksnių — formas, o įnagininkas, naudininkas, vietininkas ir šauksmininkas sudaro labai nedidelę visų linksnių dalį. Veiksmažodis — dar sudėtingesnė kalbos dalis, turinti dar daugiau skirtingų formų, bet morfologiškai anotuotame tekстыne tų formų vidutiniškai pavartota tik 2,63.

Kadangi morfologiškai anotuotas tekstynas buvo sudarytas iš gana įvairių dabartinės lietuvių kalbos rašytinių tekstų, galima daryti prielaidą, kad tokios žodžių formų vartojimo

tendencijos yra ir visoje lietuvių kalboje. Šie duomenys leidžia daryti prielaidą, kad vis labiau išryškėja pagrindinės ir periferinės lietuvių kalbos dalių formos, kad lietuvių kalboje galima rasti vis daugiau analitinių formų, atsirandančių dėl anglų kalbos įtakos.

Pranešime bus pristatomi ir kiti duomenys, gauti ištyrus morfologiškai anotuotus tekstus: kalbos dalių, gramatinių kategorijų pasiskirstymas, dažniausių žodžių vartoseną. Dažniausia kalbos dalis morfologiškai anotuotame tekстыne yra daiktavardis (sudaro net 40 %), veiksmažodžiai sudaro apie 20 %, kitos kalbos dalys vartojamos gerokai rečiau. Dažniausiai vartojamos vyriškosios giminės formos (sudaro apie 57 % visų giminų formų). Vienaskaitos formos dvigubai dažnesnės negu daugiskaitos (atitinkamai 65 % ir 32 %). Veiksmažodžių asmenuojamosios ir neasmenuojamosios formos yra pasiskirsčiusios gana tolygiai (atitinkamai 50,7 % ir 49,3 %). Iš asmenuojamųjų formų dažniausiai vartojama tiesioginė nuosaka (apie 91,5 %), iš neasmenuojamųjų — dalyviai (56,2 %) ir bendratys (33,6 %).

Dažniausios žodžių formos, t. y. kaitybinės formos, yra šios: *ir, tik, kad, į, su, buvo, iš, o, tai, yra, kaip*. Dažniausios lemos, t. y. antraštiniai (žodyniniai žodžiai) yra šie: *ir, jis, būti, kad, į, aš, kuris, su, šis, Lietuva*. Dažniausi bendriniai daiktavardžiai yra šie: *metai, žmogus, darbas, diena, respublika, valstybė, vaikas, šalis, laikas, įstatymas*. Dažniausi būdvardžiai: *didelis, naujas, svarbus, geras, aukštas, valstybinis, įvairus, bendras, mažas, pagrindinis*. Dažniausi veiksmažodžiai: *būti, turėti, galėti, nebūti, reikėti, sakyti, norėti, manyti, negalėti, tapti*.

Erika Rimkutė, Jolanta Kovalevskaitė (Kaunas)

Sustabarėjusių kalbos vienetų problema: morfolginės samplaikos

Pranešime nagrinėsime dažnai lietuvių kalboje vartojamus junginius, sudarytus iš kelių nekaitomų (kartais ir kaitomų) kalbos dalių, atliekančių vienos nekaitomos kalbos dalies funkciją. Tokius sintaksiškai ir semantiškai susijusius junginius vadinsime *m o r f o l o g i n ė m i s s a m p l a i k o m i s*, nors jų terminija dar nėra iki galo nusistovėjusi. Palyginsime lietuvių kalbos morfolgines samplaikas su kitų kalbų panašiais stabiliais junginiais, aptarsime klasifikacijos ypatumus.

Su morfolginėmis samplaikomis kaip tam tikra problema susidurta tvarkant morfolgiškai anotuotą tekstyną. Problemiška analizuoti ir morfolgiškai anotuoti šiuos junginius dėl dviejų priežasčių: pirma, sunku nustatyti, kokiomis kalbos dalimis laikytini morfolginę samplaiką sudarantys nekaitomi žodžiai, pvz., junginį *kai kur* sudaro jungtukas arba dalelytė *kai* irrieveiksmis arba dalelytė *kur*. Pagal „Dabartinės lietuvių kalbos žodyne“ ir „Dabartinės lietuvių kalbos gramatikoje“ pateiktus apibrėžimus nelengva skirti šias nekaitomas kalbos dalis. Antroji priežastis, dėl kurios išsamiau imtos nagrinėti morfolginės samplaikos, yra ta, kad išskaidžius kai kuriuos junginius į atskirus dėmenis, neaišku, kokią nurodyti antraštinę formą (lema) ir morfolginę pažymą, pvz., nežinia, kas yra *tolo* junginyje *iš tolo*, *mažens* junginyje *nuo mažens* ir pan. Morfolginės analizės metu pasirodė nelogiška junginį *be galo* išskaidyti į prielinksni *be* ir daiktavardį *galas*, nes šis junginys turi bendrą reikšmę.

Iš Dabartinės lietuvių kalbos tekstyno išrinkta apie 650 morfolginių samplaikų (į ši skaičių įeina visos kaitomų žodžių, pvz., įvardžių, veiksmažodžių formos (*kas nors, ko nors, kam nors, ką nors* ir pan.; *turėti omeny(je), turiu omeny(je), turės omeny(je)* ir pan.).

Ne visada lengva nustatyti, ar junginys vartojamas gana stabiliai ir turi bendrą reikšmę, kad būtų galima laikyti morfologine samplaika. Buvo nustatyti penki kriterijai, pagal kuriuos kelių žodžių junginiai laikytini morfologinėmis samplaikomis: morfologinis neskaidomumas — pagrindinis samplaikiškumo kriterijus. Išskaidžius samplaiką į atskirus dėmenis, neįmanoma jų analizuoti morfologiškai, pvz., *iš anksto*, *iš mažens*, *iš paskos* — neaišku, kas yra *anksto*, *mažens* ir *paskos*, negalima nustatyti šių formų lemų. Taigi negalima atskirai analizuoti reikšmės požiūriu neskaidomo junginio dėmenų. Kiti kriterijai: leksinės reikšmės bendrumas (vientisumas), morfologinių samplaikų sustabarėjimo laipsnis, neapibrėžta leksinė aplinka ir gramatinės sudėties pastovumas.

Kai kurie junginiai vienais atvejais vartojami pastoviai, kitais — kaip atskiri žodžiai, pvz.: *merginą priglaudė ir po to nuvedė į policiją viena danė* (abu dėmenys vartojami kaip samplaika, reiškianti „tada“) vs. *o kokias išvadas po to darbo padarė mūsų darbininkai?* (čia pavartoti du atskiri žodžiai: prielinksnis *po* ir įvardžio forma *to* — šis junginys nelaikomas samplaika). Tik nedidelė analizuotų junginių dalis visada vartojama kaip pastovieji junginiai. Dauguma jų, atsižvelgiant į kontekstą, sakinyje gali būti pavartoti kaip keli savarankiški žodžiai.

Lietuvių kalbos morfologines samplaikas bandyta skirstyti pagal tai, kokių kalbos dalių funkciją jos atlieka. Šie junginiai dažniausiai vartojami kaiprieveksmiai, pvz., *iš anksto*, *iš arti*, *kada nors* (82 % visų morfologinių samplaikų); dalelytės, pvz., *kažin ar*, *ko gero*, *tarp kitko* (6 %); jungtukai, pvz., *kaip tik*, *vis dėlto*, *ne tik ... bet ir* (4,8 %); prielinksniai, pvz., *ligi pat*, *prie pat*, *nuo pat* (3,2 %). Kai kurių samplaikų negalima priskirti jokiai kalbos daliai, pvz., *ir taip toliau*, *ir kita*, *ir panašiai*. Tokie junginiai sudaro 3,6 % visų morfologinių samplaikų.

Problemiška morfologines samplaikas versti į kitas kalbas, ypač galvojant apie vis spartėjančius mašininio vertimo darbus. Reikia turėti kalbos, iš kurios verčiama, ir kalbos, į kurią verčiama, morfologinių samplaikų sąrašus. Taip pat reikia aiškių kriterijų, kada tas pats žodžių junginys vartojamas stabiliai, taigi laikytinas morfologine samplaika, o kada turėtų būti analizuojamas kaip du atskiri žodžiai.

Giedrė Čepaitienė (Šiauliai)

Kalbos etiketo bendrumai ir skirtumai baltų ir slavų patarlėse

Šiame pranešime aiškinamasi, į kokias kalbos etiketo situacijas kreipiamas dėmesys baltų (lietuvių bei latvių) ir slavų (rusų, baltarusių, lenkų) patarlėse, kaip šios situacijos vertinamos bei kokius adresato ir adresanto santykius jos atskleidžia. Minėtų kalbų atranka neatsitiktinė: šiomis kalbomis kalbančios tautos gyvena gana kompaktiškoje teritorijoje, jas jungia daugelis istorinių faktų nuo Viduramžių iki pat šių dienų. Straipsnyje kreipiamas dėmesys ir į patarlių struktūrą, ir į jų kalbinę raišką, nes esama ir bendrų, ir visiškai (ar iš dalies) skirtingų meninių vaizdų. Neretai patarlės turi ritmą, jose tarsi žaidžiama žodžių sąskambiais, jų formomis ar reikšmėmis. Tai neretai įvardijama kaip šnekamosios kalbos bruožas apskritai [Ernstson 1997: 77], o šnekos akte tai traktuotina kaip poetinės funkcijos apraiška. Bronys Savukynas teigia, kad poetinė funkcija esti tada, kai tekstas visų pirma orientuotas į komunikatą [Savukynas 1977: 17]. Su tuo galima sutikti tik iš dalies: kiekvienas tekstas paprastai pirmiau-

siai būna orientuots ī adresatā, bet tada, kai siekiama didesnio jo paveikuma (kartais — dēl kitu priēzasēiū), esama papildomas orientācijas ī komunikatā, t. y. patī tekstā, jo sāranga.

Patarliū lyginimas padeda nustatyti, kokie panašūs ir kokie skirtingi lietuviū ir latviū, baltarusiū, rusū ir lenkū kalbēsenos bei māstysenos bruožai atsiskleidžia kalbant apie panašias kalbos etiketo situācijas. Latviū kalbos medžiaga darbuī rinkta iš Elzēs Kuokarēs knygos „Latviešu un lietuviešu sakāmvardu paralēles“ (Ryga, 1980). Rusū kalbos patarlēs rinktos iš internetinio rinkinio (<http://www.aphorism.ru/dal/>), baltarusiū ir lenkū — iš Kazio Grigo knygos „Patarliū paralelēs“ (Vilnius, 1987). Lietuviū kalbos medžiaga iliustruojama pavyzdžiais iš tautosakos rinkinių bei autorēs gimtosios zanavykū tarmēs.

Patarlēs atskleidžia nemaža mūsū tautū požiūrio ī kalbėjimą bei kalbos etiketā bendrumū. Jos parodo, kad nuo seno lietuviū ir latviū bendruomenēs turējo tvirtas bendravimo nuostatas. Jos vertē laikytis tam tikros kalbinēs elgsenos. Patarlēs parodo, kokia kalbinē elgsena yra derama, o kokia — smerktina, kokiū dalykū būtina paisyti, kad parodytum pagarbā pašnekovui ir pats būtum gerbtinas. Lygintū kalbū patarlēse išryškēja 6 bendros kalbos etiketo normas apibrėžiančios temos: 1) gebėjimas parinkti tinkamā žodį; 2) žodžio įtaka žmoniū santykiams; 3) požiūris ī apkalbas; 4) požiūris ī gyrimąsi; 5) požiūris ī dalyko išmanymā; 6) keiksmo vartojimas. Nors visos jos teigia panašias moralines nuostatas, tačiau jų raiška atskleidžia savitā kiekvienos tautos požiūrī ī kalbinę elgsenā.

Literatūra

Ernstsonē 1997 — V. Ernstsonē. Valodas spēle latviešu sarunvalodā funkcionālā skatījūmā // *Linguistica Lettica*, 1. 1997. — 76.–99. lpp.

Savukynas 1977 — B. Savukynas. Leidinio paskirtis ir kalba // *Žmonēs ir kalba*. Vilnius, 1977. — p. 15–22

Lilija Jurgīte (Rīga)

Krievijas 1897. g. tautskaites materiāli par lingvistisko situāciju 19. gs. beigās dažos Dienvidlatgales pagastos

1895. g. 5. jūnijā Krievijas imperators Nikolajs II parakstīja „Nolikumu par Pirmo vispārējo tautas skaitīšanu“, kas notika 1897. g. 28. janvārī. Tautskaites materiāli N. Trojņicka redakcijā tika publicēti laika posmā no 1899. g. līdz 1905. g. Šie materiāli izmantoti dažādos pētījumos, tai skaitā arī lingvistiskos. Tomēr jānorāda, ka reti tiek analizētas primārās uzskaites kartītes, kas daudzviet ir iznīcinātas.

Latvijas Valsts Vēstures arhīva 2706. fondā tiek glabātas Krievijas 1897. g. tautas skaitīšanas Kurzemes, Vitebskas un Vidzemes guberņu primārās uzskaites kartes, kas sniedz bagātu informāciju par iedzīvotājiem. Tajās ir fiksēts uzskaitītās personas: vārds, uzvārds, tēv-vārds; ģimenes stāvoklis; statuss attiekmē pret ģimenes galvu; dzimums; vecums; kārta; profesionālā piederība; dzimšanas vieta; pieraksta vieta; dzīvesvieta; dzimtā valoda; valodprasmes; nodarbošanās un fiziskie trūkumi. Izmantojot minētos materiālus, valodnieks var pētīt ne tikai vietvārdus un personvārdus, bet arī lingvistisko situāciju attiecīgajā teritorijā. Šī ziņojuma autore pievērsusies Dienvidlatgales lingvistiskajai situācijai 19. gs beigās, jo bez zināšanām par šī laika īpatnībām nav iespējams izprast daudzus 20. gs. procesus un arī valodu lietošanas īpatnības mūsdienās.

Dienvidlatgale, proti, bijušais Vitebskas guberņas Daugavpils apriņķis izsenis bijis novads ar ārkārtīgi raibu iedzīvotāju etnisko sastāvu. Tas atspoguļojas arī uzvārdos, kuru cil-

me sakņojas dažādās baltu un slāvu valodās. Novadā blakus latviešiem (latgaliešiem) izsenis dzīvojuši krievi, baltkrievi, poļi, ebreji un pat lietuvieši. Tomēr vēsturnieki maz pievērsušies empīriskā materiāla izpētei, jo Latgales vēstures avotu ir maz, tie ir fragmentāri, kādēļ arī grūti pētāmi. Visu šo iemeslu dēļ 1897. g. tautskaites materiāli būtu pētāmi un analizējami dažādos aspektos, tai skaitā sociolingvistiskā.

LVVA 2706. fondā ir Daugavpils apriņķa Asūnes, Dagdas, Izvaltas, Skaistas un Višķu pagastu, kā arī grāfa Plātera, S. Mola u. c. zemes īpašnieku teritorijā dzīvojošo iedzīvotāju uzskaites kartes. Tā ir informācija gandrīz par 900 ģimenēm, kuras runā dažādās valodās: baltu — latviešu (augšzemnieku dialekta izloksnē) un lietuviešu, slāvu — krievu, baltkrievu, poļu, kā arī ebreju un vācu valodā.

Aptaujātie ir norādījuši ne tikai dzimto valodu, bet arī to, vai prot lasīt un rakstīt. Materiāli rāda, ka novadā ir diezgan daudz analfabētu, bet tos, kas šīs valodprasmes ir apguvuši, var dalīt divās grupās: latvieši (latgalieši) norāda, ka lasa latviski, krievi — krieviski, poļi — poliski (tātad dzimtajā valodā), bet baltkrievi, kuri par dzimto valodu uzskata baltkrievu, lasa tikai poliski. Atrodama arī informācija par to, kur attiecīgā persona ieguvusi izglītību, tai skaitā arī valodprasmes. Tas viss tiks detalizētāk analizēts referātā.

Содержание

От редактора	3
Конференции по балтистике в СПбГУ	4
<i>A. B. Андронов (Санкт-Петербург)</i> . На пути к корпусу латгальского языка России (1920–30-е гг.)	5
<i>E. Andronova (Rīga — Sanktpēterburga)</i> . No seno tekstu korpusa līdz vēsturiskajai vārdnīcai: iestrādes un problēmas	6
<i>I. Auziņa (Rīga)</i> . Segmentu datu bāzes izveide latviešu valodas konkatēnātajam runas sintezatoram	8
<i>M. Baltiņš (Rīga)</i> . Latviešu terminoloģijas attīstības posmi un to raksturojums	9
<i>O. Буис (Puza)</i> . Заимствование, иноязычная цитата, смена кода (к разграничению понятий на материале латышского языка)	10
<i>Л. Б. Воробьева (Псков)</i> . Пословицы и поговорки как социолингвистический источник	12
<i>A. Vulāne (Rīga)</i> . Ļaužu valodas — būtisks latviešu pasaulainas komponents	13
<i>T. A. Горностај (Puza)</i> . Русско-латышско-русский словарь с элементами машинного перевода	15
<i>N. Grūzītis, G. Nešpore, B. Saulīte (Rīga)</i> . Extraction of Basic Semantic Relations from the Dictionary of Latvian	17
<i>N. Grūzītis, G. Nešpore, B. Saulīte (Rīga)</i> . Morphological and Syntactic Analysis of Latvian: a Different Approach	18
<i>A. Д. Даугавет (Санкт-Петербург)</i> . Состояние традиционных говоров в Латвии в 20–40 гг. XX в. (по материалам журнала «Filologu biedrības raksti»)	22
<i>I. Druviete (Rīga)</i> . Sociolingvistikas attīstība Latvijā	23
<i>A. Дубасова (Минск — Санкт-Петербург)</i> . О web-версии словаря балтийской терминологии	24
<i>M. В. Завьялова (Москва)</i> . Словесные формулы в вербальных ассоциациях литовцев, поляков и русских	25
<i>I. Zuicena (Rīga)</i> . Mūsdienu latviešu valodas vārdnīca (elektroniskā versija) — realitāte un iespējas	26
<i>R. Kvašīte (Šauļi)</i> . Lietuvieši Latvijā: sociolingvistiskais aspekts	26
<i>S. Kļaviņa (Rīga)</i> . Kvantitatīvās un datorlingvistikas aizsākumi Latvijā 20. gs. 60.–70. gados	28
<i>K. Levāne-Petrova (Rīga)</i> . Paralēlā korpusa nozīme Eiropas Savienības tiesību aktu tulkošanā	29
<i>V. Poriņa (Rīga)</i> . Valsts valodas lietojuma jautājums imigrantu integrācijā	30
<i>E. Rimkutė (Kaunas)</i> . Automatinė morfologinė lietuvių kalbos analizė	31
<i>E. Rimkutė, J. Kovalevskaitė (Kaunas)</i> . Sustabarėjusių kalbos vienetų problema: morfologinės samplaikos	32
<i>G. Čepaitienė (Šiauliai)</i> . Kalbos etiketo bendrumai ir skirtumai baltų ir slavų patarlėse	33
<i>L. Jurģīte (Rīga)</i> . Krievijas 1897. g. tautskaites materiāli par lingvistisko situāciju 19. gs. beigās dažos Dienvidlatgales pagastos	33