

Kristīne LEVĀNE-PETROVA

*LU Matemātikas un informātikas institūts*

## LĪDZSVAROTS MŪSDIENU LATVIEŠU VALODAS TEKSTU KORPUSS UN TĀ TEKSTU ATLASES KRITĒRIJI

### Ievads

LU MII Mākslīgā intelekta laboratorijā (LU MII MIL) darbs pie valodas korpusa izstrādes tika aizsākts jau vairāk nekā pirms desmit gadiem, apkopojot un uzkrājot dažādus elektroniskos resursus. Mūsdienās šis darbs ir īpaši aktuāls, ņemot vērā tik straujo moderno tehnoloģiju attīstību un latviešu valodas konkurētspēju starp pasaulē plašāk lietotajām valodām.

Ar valodas korpusu tiek saprasts datorizētai analīzei pieejams apjomīgs tekstu kopums (VPSV, 196), kas izveidots atbilstoši noteiktiem kritērijiem.

Ar Latviešu valodas aģentūras (LVA) atbalstu LU MII ir izstrādāts vispārīgs *Līdzsvarots mūsdienu latviešu valodas tekstu korpus*, kas pieejams interneta vietnē <http://www.korpuss.lv>, izmantojot korpusa pārvaldības platformu *Manatee* (Manatee). Korpus ir automātiski morfoloģiski anotēts.

Rakstā, pamatojoties uz korpusa izveides kritērijiem, tiks aplūkots *Līdzsvarots mūsdienu valodas tekstu korpus* un tā uzbūves pamatprincipi, problēmjaucājumi, kas saistīti ar korpusa uzbūvi un tekstu atlases kritērijiem, rasti risinājumi un korpusa nākotnes izstrādes ieceres.

### 1. Vispārīgs mūsdienu latviešu valodas tekstu korpus

Darbs pie vispārīga mūsdienu latviešu valodas tekstu korpusa tika uzsākts 2005. gadā, kad pēc Valsts valodas aģentūras (VVA) pasūtījuma LU MII tika izstrādāta *Latviešu valodas korpusa koncepcija* (Koncepcija 2005). No 2007. līdz 2009. gadam projekta divos posmos tika izstrādāts ~ 3,5 miljonus vārdlietojumu lielais *Līdzsvarots mūsdienu latviešu valodas tekstu korpus*, kas ir:

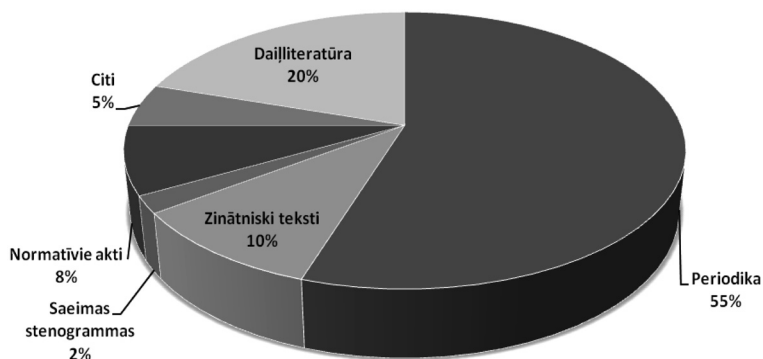
- vispārīgs – tā avoti netiek ierobežoti atbilstoši konkrētai tēmai, noteiktam valodas paveidam;
- līdzsvarots – valodas avoti korpusā ir ievietoti noteiktās proporcijās;
- tekstu korpus – šobrīd korpusā ir iekļauti tikai mūsdienu rakstu valodas avoti, izņemot Saeimas stenogrammas, kas ir runātas valodas pieraksts (bet ne transkripcija). Ar mūsdienu rakstu latviešu valodas tekstiem tiek saprasti teksti no 1991. gada līdz 2009. gadam;

- papildināms – korpuss laika gaitā tiek/tiks papildināts ar jauniem valodas avotiem;
- korpusā ir iekļauti gan pilni teksti, gan arī tekstu fragmenti;
- anotēts – korpuss ir automātiski morfoloģiski marķēts, tam ir pievienoti metadati.

### 1.1. Korpusa uzbūve

Projektējot valodas korpusu, ir jārisina dažādi ar tā uzbūvi saistīti jautājumi – korpusa apjoms, iekļaujamo valodas avotu proporcijas, to tematiskais daudzveidīgums u. c. –, tāpēc ir ļoti nozīmīgi, lai šāda veida korpuss tiktu veidots pēc skaidriem uzbūves principiem.

Izveidotais korpuss uzskatāms par līdzsvarotu – tajā no elektronisko tekstu arhīva, kas pieejams korpusa veidotājiem un kurā laika gaitā tiek uzkrāti dažādi elektroniskie resursi, teksti tika pārcelti noteiktās proporcijās. Korpusa proporcijas ir noteiktas atbilstoši *Latviešu valodas korpusa koncepcijai* (Koncepcija 2005) un ir redzamas 1. attēlā. Korpusu veido sešas sadaļas, katra no tām tiek dalīta sīkākās daļās.



1. attēls. Līdzsvarota mūsdienu latviešu valodas tekstu korpusa proporcijas

Periodikas sadaļā proporcionāli ir iekļauta nacionālo laikrakstu, reģionālo laikrakstu, interneta ziņu, speciālās periodikas un populārās periodikas daļa. Pēdējā daļa tiek sadalīta vēl sīkāk, iekļaujot tajā internetā publicētos izklaides resursus, blogus jeb emuārus un drukātos izklaides izdevumus (sk. attiecīgi 1. un 2. tabulu).

### 1. tabula. Korpusa periodikas sadaļā iekļauto daļu proporcijas

Nacionālie laikraksti	27 %
Reģionālie laikraksti	22 %
Interneta ziņas	14 %
Speciālā periodika	13 %
Populārā periodika	24 %

### 2. tabula. Korpusa periodikas sadaļas populārās periodikas daļas proporcijas

Interneta izklaides resursi	31 %
Blogi jeb emuāri	15 %
Žurnāli	54 %

Daiļliteratūras sadaļa tiek sadalīta trijās daļās, piešķirot aptuveni 80 % prozai, 15 % drāmai, bet 5 % lirikai (sk. 3. tabulu). Arī pārējās korpusa sadaļas ir sadalītas sīkāk, piemēram, zinātnisko tekstu sadaļā līdzīgās proporcijās iekļaujot dažādu zinātnes nozaru tekstus.

### 3. tabula. Korpusa daiļliteratūras sadaļā iekļauto daļu proporcijas

Proza	~ 80 %
Drāma	~ 15 %
Lirika	~ 5 %

Papildinot korpusu, tā proporcijas var tikt mainītas. Ja korpusa izstrādes pirmajā posmā, periodikas sadaļā tika iekļauti tikai drukāti nacionālie un reģionālie laikraksti, tad korpusa izstrādes otrajā posmā šīs sadaļas proporcijas tika mainītas, iekļaujot arī t. s. elektroniski dzimušo tekstu – dažādu ziņu portālu avotu un emuāru – daļu. Tāpat korpusa izstrādes pirmajā posmā tika atsevišķi izšķirta populārzinātnisko tekstu sadaļa, tomēr vēlāk tā tika iekļauta periodikas sadaļā, izveidojot speciālās periodikas daļu.

## 1.2. Korpusa tekstu atlases kritēriji

Lai korpusss būtu ne tikai līdzsvarots, bet tajā pēc iespējas daudzveidīgāk būtu atspoguļota mūsdienu latviešu rakstu valoda, pārklājot gan visus valodas paveidus, gan arī visplašāko tematiku, ir nepieciešami tekstu atlases kritēriji. Korpusa atlases kritērijus var iedalīt divās grupās – tādi, kas ir attiecināmi uz visu korpusu, un tādi, kas ir attiecināmi tikai uz noteiktu valodas paveidu.

### 1.2.1. Uz visu korpusu attiecināmi tekstu atlases kritēriji

Viens no nozīmīgākajiem tekstu atlases kritērijiem ir **valodas paveids**. Korpusa atbilstošajās sadaļās ir iekļauts vairums latviešu valodas paveidu, tostarp Saeimas stenogrammas. Sadaļā *Citi* tiek iekļauti dažādi citās korpusa sadaļās grūti klasificējami teksti, piemēram, mājaslapu, lietošanas instrukciju teksti u. tml.

Otrs nozīmīgākais tekstu atlases kritērijs ir **laiks**. Korpusā tiek iekļauti tikai tie teksti, kas publicēti kopš 1991. gada. Ja korpusā iekļaujamajam avotam ir vairāki izdevumi, tiek ņemts tā pirmizdevums. Daiļliteratūras sakarā svarīgs ir ne tikai darba izdošanas, bet arī sarakstīšanas gads – ja darbs ir uzrakstīts pirms 1991. gada, bet izdots pēc tā, tas korpusā iekļauts netiks.

Nav noteikts, cik valodas avotu no katra gada korpusā būtu ievietojami, tomēr netiek iekļauti kāda periodiskā izdevuma avoti tikai no vienas dienas / mēneša / gada.

Trešais nozīmīgākais tekstu atlases kritērijs ir **publikācija**. Tekstiem ir jābūt publicētiem vai nu elektroniskā formātā, piemēram, interneta portālu ziņas, vai drukātā formātā; nepublicēts romāna manuskripts korpusā netiks iekļauts.

Svarīgs aspekts ir iekļaujamā **valodas avota izmērs**. Tā kā izstrādātais korpus ir mazs, tajā tika izvēlēts likt nelielus, bet pilnus tekstus. To noteica arī dažu valodas paveidu tekstu apjoms, piemēram, ziņu portālu, reģionālo laikrakstu rakstu apjoms bieži ir apmēram 300–400 vārdlietojumu. Dalīt šāda apjoma rakstus nebūtu lietderīgi. Tāpat arī citās korpusa sadaļās, piemēram, normatīvo aktu sadaļā, priekšroka tiek dota nelieliem, bet pilniem tekstiem. Ja nav iespējams atrast korpusa sadaļai īsu reprezentatīvu valodas avotu, tiek ņemts valodas avota fragments. Parasti raksturīgais fragments tiek kopēts no valodas avota vidusdaļas, ievērojot rindkopu / nodaļu sadalījumu. Tekstu fragmenti parasti ir ievietoti daiļdarbiem (parasti romāniem), arī apjomīgiem zinātniskiem tekstiem.

Korpusā iekļaujamā avota maksimālais vārdlietojumu skaits nav noteikts, bet tam nevajadzētu būt apjomīgam – pārsniegt 50 000 vārdlietojumu –, lai viena autora / tēmas teksti korpusā nedominētu. Pagaidām, ņemot vērā izstrādātā korpusa izmēru, tik apjomīgi (vairāk par 50 000 vārdlietojumiem) tekstu fragmenti korpusā nav iekļauti.

Nav noteikts arī vārdlietojumu skaita ierobežojums viena autora darbiem. Lai viena autora teksti nedominētu korpusā, to apjomam korpusā arī nevajadzētu pārsniegt 50 000 vārdlietojumu. Tādējādi vairāki viena autora daiļdarbi,

raksti u. c. valodas paveidu izdevumi, kas kopsummā pārsniegtu 50 000 vārdlietojumu, korpusā iekļauti netiks. Piemēram, *British National Corpus* (BNC) (Burnard 2007) ierobežojums viena autora tekstu avotiem ir 120 000 vārdlietojumu, tomēr jāņem vērā, ka BNC apjoms ir 100 miljoni vārdlietojumu, kas ievērojami pārsniedz *Līdzsvarota mūsdienu latviešu valodas tekstu korpusa* apjomu.

Lai ilustrētu, cik lieli ir iekļautie pilnie valodas avoti korpusā, var minēt, ka šobrīd korpusā ir iekļauti 3938 teksti, no kuriem viena ievietotā teksta vidējais lielums ir 885 vārdlietojumi.

Korpusā ir iekļauti tikai **oriģinālteksti**, nevis tulkojumi.

Nozīmīgs tekstu atlasē kritērijs ir **pieejamība** – lai korpusu varētu uzkrāt ātrāk un teksti nebūtu jādigitalizē, kas ir laikietilpīgs un dārgs process, priekšroka tiek dota tekstiem, kas atrodami arī elektroniskā formātā (piemēram, interneta portālos atrodami drukāto izdevumu raksti). Pieejamības kritērijs ir nozīmīgs arī korpusa projektēšanas procesā – korpusā iekļaujamās valodas avotus nereti nosaka ne tikai korpusa veidotāju pieņemtie tekstu atlasē kritēriji, bet arī valodas avotu pieejamība.

Tā kā korpusā nav paredzēts tekstu lasīšanai un lietotājam ir redzama tikai viena konkordances rindiņa, svarīgs ir **formālas atbilstības kritērijs** – priekšroka tiek dota tekstiem, kuros nav dažādu *neteksta* daļu, piemēram, tabulu u. tml. Ja tomēr attiecīgais tekstu avots ir nozīmīgs, *neteksta* daļas tiek dzēstas.

Visi korpusā iekļautie teksti ir **autentiski**. Tajos netiek veikti nekāda veida labojumi, lai varētu atspoguļot valodu tādu, kāda tā ir. Ja nepieciešamais tekstu avots tiek digitalizēts un ja drukātajā formātā ir bijušas kādas kļūdas, tās tiek atstātas.

### 1.2.2. Uz noteiktu valodas paveidu attiecināmi tekstu atlasē kritēriji

Viens no vispārīgā korpusa izveides nosacījumiem ir visu iespējamo valodas paveidu iekļaušana un maksimāls tematiskais daudzveidīgums, tāpēc ir noteikti dažādi papildu kritēriji, kas svarīgi katra valodas paveida tekstiem.

#### **Periodika, normatīvie akti un zinātniski teksti**

Vairums korpusā iekļauto valodas avotu ir atlasīti tā, lai pēc iespējas plašāk aptvertu **tematisko daudzveidīgumu**.

Periodikas sadaļas laikrakstu un interneta ziņu daļās tiek iekļauti raksti par visdažādākajām tēmām (sports, ekonomika, politika, kriminālziņas u. tml.). Katra tēma tiek sadalīta sīkāk – sporta daļā tiek iekļauti raksti par visdažādā-

kajiem sporta veidiem un sporta notikumiem, politikas daļā – par visdažādākajiem politiskajiem notikumiem u. tml. Tiek iekļautas arī intervijas, apskati u. c. veidu periodisko izdevumu raksti.

Normatīvo aktu sadaļā tiek iekļauti likumi par visdažādākajām tēmām, bet zinātnisko tekstu sadaļā tiek iekļauti valodas avoti gan no medicīnas, gan fizikas, gan ķīmijas u. c. zinātņu nozarēm. Savukārt konkrētai zinātņu nozarei paredzētajā daļā tiks iekļauti tās apakšnozarēm veltītie darbi, piemēram, valodniecības daļā tiks iekļauti morfoloģijas, sintakses u. c. valodniecības nozaru jautājumiem veltītie zinātniskie darbi. Zinātnisko darbu sadaļā darbi atlasīti arī pēc zinātniskā darba veida – tiek iekļauti gan referāti, gan promocijas darbu fragmenti, gan promocijas darbu kopsavilkumi, gan raksti u. tml.

### **Periodika**

Tā kā periodikas sadaļa veido vairāk nekā pusi no visa korpusa apjoma, ir īpaši rūpīgi jāatlasa šajā sadaļā iekļaujамie valodas avoti.

**Ģeogrāfijas** kritērijs ir attiecināms tikai uz reģionālajiem laikrakstiem, kas tiek atlasīti no visiem Latvijas apgabaliem. Korpusā tiek iekļauti Latgales reģionālie laikraksti, piemēram, *Rēzeknes Vēstis*, *Latgales Laiks*, Kurzemes reģionālie laikraksti, piemēram, *Kursas Laiks*, *Kurzemes Vārds*, u. c. reģionālie laikraksti.

**Unikalitāte.** Periodikas sadaļā netiek ievietotas vairākas ziņas par vienu un to pašu notikumu. Ja dažādos elektroniskajos un drukātajos izdevumos tiek rakstīts par vienu un to pašu notikumu, piemēram, ievēlēto Latvijas prezidentu, korpusā tiks atlasīts tikai viens raksts par šo notikumu.

Gan laikraksti, gan ziņu portāli Latvijā ir aptveramā daudzumā, un korpusā ir iekļauti raksti no visiem lielākajiem Latvijā iznākošajiem elektroniskajiem un drukātajiem izdevumiem.

### **Daiļliteratūra**

Daiļliteratūras sadaļā ir iekļauti visi žanri – proza, drāma un lirika. Korpusā tiek iekļauti Latvijas literatūrā atzītu autoru darbi, piemēram, Laimas Muktupāvelas, Noras Ikstenas, Paula Bankovska u. c. atzītu autoru darbi.

### **Citi kritēriji**

Ja korpusā ievietojamais avots atbilst visiem iepriekšminētajiem tekstu atlases kritērijiem, bet pieejamo avotu joprojām ir vairāk nekā tiem korpusā atvēlētās vietas, tekstu atlase notiek pēc nejaušības principa. Šai sakarā var arī runāt par korpusa veidotāju subjektīvo kritēriju.

Šobrīd tiek izmantoti visi iepriekšminētie tekstu atlases kritēriji, tomēr laika gaitā esošie var tikt precizēti un papildināti ar jauniem. Tā kā tekstu atlases

kritēriji ir lielā mērā saistīti ar tekstu pieejamību, viens no tekstu kritēriju precizēšanas iemesliem būtu minams arī darbs ar lielākiem tekstu apjomiem.

### 1.3. Korpusa marķējums

Korpusam ir izstrādāta *metadatu kopa* (metadati jeb „dati par datiem” var aprakstīt kādu citu atsevišķu datu vienumu vai visu datu kopumu. Metadati tiek izmantoti, lai vieglāk varētu saprast, izmantot un pārvaldīt datus [Vikipēdija]). Katram korpusa tekstu avotam ir fiksēti noteikti metadati, piemēram, teksta avota nosaukums, darba autors, ja zināms, vārdlietojumu skaits u. tml., bet pagaidām korpusa lietotājiem šie dati nav redzami. Pieejama ir tikai informācija par konkrēta teksta avota valodas paveidu – vai konkrētais darbs ir daiļliteratūras, zinātnisks teksts u. tml.

Korpusss ir automātiski *morfoloģiski* marķēts, katram vārdlietojumam automātiski pievienojot tam atbilstošas morfoloģiskās pazīmes<sup>1</sup> un pamatformu (lemmu). Piemēram, vārdformai *pavasara* būs pievienots šāds morfoloģiskais marķējums *ncmsg2*, kas ir atšifrējams kā n – lietvārds, c – sugasvārds, m – vīriešu dzimte, s – vienskaitlis, g – ģenitīvs, 2 – 2. deklinācija.

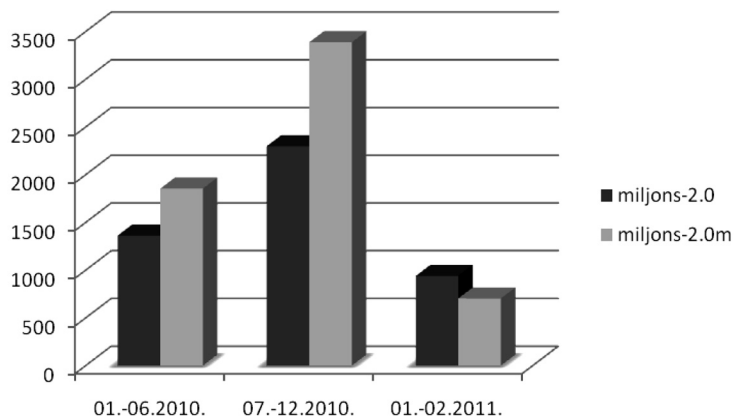
Jāņem vērā, ka latviešu valoda ir morfoloģiski daudznozīmīga un automātiskās marķēšanas procesā reizēm nav iespējams atšķirt konkrētas vārdformas vārdšķiru, tāpēc, lai saglabātu visu iespējamo informāciju, korpusā vārdformām, kurām marķētājs nespēj noteikt precīzu formu, ir atstāti visi sintaktiski iespējamie marķējuma varianti. Piemēram, ja marķētājs nav spējis noteikt, kāda ir vārda *sniegs* vārdšķira konkrētajā kontekstā, tam tiks atstāti vairāki morfoloģiskā marķējuma varianti, proti, attiecīgais vārds lietotājam tiks piedāvāts gan kā vīriešu dzimtes vienskaitļa 1. deklinācijas lietvārds nominatīvā, gan arī kā darbības vārda nākotnes forma.

### 1.4. Korpusa izmantošana

Korpusss ir indeksēts, izmantojot korpusa pārvaldības pārlūkprogrammu *Manatee*. Vaicājumiem korpusā ir izmantojama pārlūkprogramma *Bonito*. Korpusss ir pieejams gan ar morfoloģisko marķējumu, gan bez tā (attiecīgi miljons-2.0 un miljons-2.0m). Korpusss tiek plaši izmantots, par ko liecina unikālo vaicājumu skaits, turklāt laika gaitā korpusss tiek izmantots aizvien plašāk. Korpusa lietošanas dinamiku 2010. gada pirmajā un otrajā pusgadā, un 2011. gada sākumā sk. 2. attēlā.

---

<sup>1</sup> Latviešu valodas morfoloģisko pazīmju kopa, [http://www.semti-kamols.lv/doc\\_upl/TagSet.pdf](http://www.semti-kamols.lv/doc_upl/TagSet.pdf).



2. attēls. Korpusu lietošanas statistika laika gaitā, kur miljons-2.0 – 3,5 miljonu vārdlietojumu korpus bez morfoloģiskā marķējuma, miljons-2.0m – 3,5 miljonu vārdlietojumu korpus ar morfoloģisko marķējumu

### 1.5. Secinājumi un nākotnes ieceres

Pasaulē valodas korpusi jau sen ieņem stabilu vietu zinātnē. Arī Latvijā aizvien nozīmīgāka loma tiek piešķirta valodas korpusam, par ko liecina gan lielā interese par to, gan arī *Līdzsvarota mūsdienu latviešu valodas tekstu korpusa* plašā lietošana. Izstrādātais *Līdzsvarots mūsdienu latviešu valodas tekstu korpus* ir nozīmīgs valodas resurss, bet korpusa projektēšanas laikā ar tā uzbūvi saistīto jautājumu risināšana un iegūtā pieredze nodrošinās labu bāzi turpmākajam darbam pie korpusa attīstības.

Korpus sastāv ne tikai no tekstiem, bet arī no atbilstoša programnodrošinājuma, tāpēc liela uzmanība jāpievērš tā tehniskajiem risinājumiem. Šobrīd ar LVA atbalstu norisinās darbs pie korpusa vadības sistēmas izstrādes, kas ļaus sistematizēt korpusā iekļautos tekstus un atbilstīgos metadatus, kā arī pārvaldīt tos. Paralēli norisinās arī darbs pie korpusa standartizēšanas, proti, korpusa metadatu kopa tiek standartizēta atbilstoši TEI5 (Tei). Korpusa tehnisko risinājumu modernizēšana un standartizācija ļaus ne tikai ātrāk un mūsdienu standartiem atbilstīgāk attīstīt korpusu, bet tas būs arī iekļaujams pasaules zinātnes aprītē, piemēram, ar to varēs piedalīties dažādos starptautiskos projektos/tīklos.

Nākotnē *Līdzsvarots mūsdienu latviešu valodas tekstu korpus* ir noteikti papildināms ar jauniem valodas avotiem, sasniedzot vismaz 100 miljonus



vārdlietojumu, kas ir atbilstošs mūsdienās izstrādājamā korpusa apjoms, tomēr pasaules prakse rāda, ka šobrīd korpusu apjoms jau pārsniedz šo robežu. Korpusi arī būtu papildināmi ar runātas valodas transkripcijām. Lai ātrāk tāda apjoma korpusu varētu iegūt, nākotnē būtu jāpanāk vienošanās ar dažādiem resursu turētājiem par tekstu avotu elektronisku iegūvi mūsdienu latviešu valodas tekstu korpusa papildināšanai.

## THE BALANCED CORPUS OF MODERN LATVIAN AND THE TEXT SELECTION CRITERIA

### *S u m m a r y*

Recently *The Balanced Corpus of Modern Latvian* (~3.5 million running words) has been created in the Institute of Mathematics and Computer Science (IMCS) (see *www.korpuss.lv*). The Corpus has been compiled from printed and electronic materials created after 1990. The Corpus is automatically morphologically tagged: for each token all the syntactically valid interpretations are stored.

Texts for the Corpus were chosen according to different text selection criteria: for instance, time, media, domain, etc. This article discusses the text selection criteria chosen for this Corpus, problems related to Corpus design and text selection criteria, solutions found for these problems and future plans regarding the Corpus.

### LITERATŪRA

Burnard Lou 2007, Reference Guide for the British National Corpus (XML Edition), <http://www.natcorp.ox.ac.uk/docs/URG/>.

Čermák František 2002, Today's corpus linguistics. Some open questions, *International Journal of Corpus Linguistics* 7(2), 265–282.

Čermák František 2008, The case of the Czech national corpus: Its design and development, [www.clarin.lv/materiali/Cermak\\_Riga-en.doc](http://www.clarin.lv/materiali/Cermak_Riga-en.doc).

Koncepcija 2005 – Latviešu valodas korpusa koncepcija, 2005, <http://www.korpuss.lv/uzzinas/koncepcija.pdf>.

Kovalevskaitė Jolanta 2006, Dabartinės lietuvių kalbos tekstynas – 10 metų kaupimo ir naudojimo patirtis, *Prace Baltystyczne* 3, 231–241.

Kopa – Latviešu valodas morfoloģisko pazīmju kopa, 2009, [http://www.semtikamols.lv/doc\\_upl/TagSet.pdf](http://www.semtikamols.lv/doc_upl/TagSet.pdf).

Králík Jan, Michal Šulc 2005, The representativeness of Czech corpora, *International Journal of Corpus Linguistics* 10(3), 357–368.

Manatee – <http://www.textforge.cz/products>.  
Vikipēdija – Vikipēdija, Brīvā enciklopēdija, <http://en.wikipedia.org/wiki/Metadata>.  
TEI – Text Encoding Initiative, <http://www.tei-c.org/index.xml>.  
VPSV – Valentina Skujiņa (red.), Valodniecības pamatterminu skaidrojošā vārdnīca,  
Rīga: Latviešu valodas institūts, 2007.

*Kristīne LEVĀNE-PETROVA*  
*Mākslīgā intelekta laboratorija*  
*LU Matemātikas un informātikas institūts*  
*Raiņa bulvāris 29*  
*LV-1459, Rīga*  
*Latvia*  
*[[kristine.levane-petrova@lumii.lv](mailto:kristine.levane-petrova@lumii.lv)]*