Everita Andronova (LU MII MIL), Anna Frīdenberga (LU HZF LaVI), Lauma Pretkalniņa (LU MII MIL), Renāte Siliņa-Piņķe (LU HZF LaVI), Elga Skrūzmane (LU HZF LaVI), Anta Trumpa (LU HZF LaVI), Pēteris Vanags (LU HZF LaVI)

VADLĪNIJAS

latviešu seno tekstu (16.–18. gs.) konvertācijai mūsdienu rakstībā[1]

Saturs

Ievads

Divos Valsts pētījumu programmu projektos („Humanitāro zinātņu digitālie resursi", Nr. VPP-IZM-DH-2020/1-0001, un „Atvērtas un FAIR principiem atbilstīgas digitālo humanitāro zinātņu ekosistēmas attīstība Latvijā", Nr. VPP-IZM-DH-2022/1-0002) laikā no 2020. līdz 2025. gadam, veicot latviešu valodas seno tekstu korpusa *SENIE* (https://korpuss.lv/id/Senie) modernizāciju, tika konvertēti gandrīz visi (175)[2] korpusā esošie avoti. Korpuss katru gadu tiek papildināts ar jauniem avotiem, tādēļ darbs pie avotu konvertācijas nevar tikt uzskatīts par pabeigtu. Šīs vadlīnijas ir veidojuši visi projekta darba grupā iesaistītie pētnieki, balstoties uz savu pieredzi šajā darbā. Tās ir sarakstītas ar mērķi fiksēt līdz šim izstrādātos konvertācijas principus, lai turpmākais darbs būtu konsekvents un lai atvieglotu tā apguvi iespējamiem jauniem grupas dalībniekiem. Tāpat šīs vadlīnijas var būt par paraugu katram, kurš vēlas pārveidot latviešu senos (16.–18. gadsimta) tekstus mūsdienu rakstībā.

Uzreiz pēc ievada sniegts vadlīnijās lietoto specifisko, ar konvertāciju saistīto, terminu skaidrojums. Vadlīniju pirmajā nodaļā ir vispārīgo principu apraksts un tehniskās piezīmes. Tā kā triju gadsimtu laikā latviešu ortogrāfija ir ievērojami mainījusies, tad vadlīniju otrā nodaļa veltīta 16., 17. un 18. gadsimta tekstu konvertācijas specifikai, katram gadsimtam paredzot savu apakšnodaļu, bet 16. un 17.

[1] Izstrādātas VPP projekta „Atvērtas un FAIR principiem atbilstīgas digitālo humanitāro zinātņu ekosistēmas attīstība Latvijā" (Nr. VPP-IZM-DH-2022/1-0002, 2023.–2025. g.) ietvaros.

[2] Bībeles grāmatas skaitītas kā atsevišķi avoti.

gadsimta avotu konvertācijas apraksti sadalīti vēl sīkāk. 16. un 17. gadsimta avotu konvertācijas aprakstos apakšnodaļās ir aplūkoti spilgtākie piemēri, proti, ir raksturotas konkrētu avotu tabulas, kurās, protams, ir ievēroti konvertācijas vispārīgie principi, savukārt 18. gadsimta avotu konvertācijas apraksts ir vairāk vispārīgs, likumi ir izsecināti, analizējot visu korpusā ietverto 18. gadsimta avotu konvertācijas tabulas.

Lietotie termini

Avota kods – avota nosaukuma saīsinājums (skat. sarakstu https://senie.korpuss.lv/).

Individuālie / leksiskie likumi – likumi, kuros seno tekstu vārdi vai vārdu daļas aizstātas ar vārdiem vai saknēm mūsdienu rakstībā.

Konvertācija – seno tekstu pārveide mūsdienu rakstībā, izmantojot konvertācijas likumus.

Konvertācijas likumi – pārveides likumi (viennozīmīgie, pozicionālie, individuālie) saskaņā ar kuriem seno tekstu burti, burtkopas, vārdu saknes vai veseli vārdi tiek pārvērsti mūsdienu rakstībā.

Konvertācijas rīks / konvertors – programmatūras rīks, ar kura palīdzību tiek konvertēti teksti.

Pagaidu aizstājējzīme / substitūts – rakstzīme, ar kuru tiek aizstāta kāda burtkopa (piemēram, dubultie līdzskaņi), lai tā netiktu konvertēta, realizējot turpmāko likumu; pēc tam aizstājējzīmi pārvērš atpakaļ par sākotnējo burtkopu.

Pozicionālie likumi – likumi, kuros ievērota burta vai burtkopas atrašanās vieta vārdā (sākumā, vidū vai beigās).

Regulārā izteiksme – simbolu virkne, kas definē meklējamos teksta fragmentus.

Reģistrjutība – spēja atšķirt augšējā un apakšējā reģistra rakstzīmes (lielos un mazos burtus u. c.).

Unicode – universāls rakstzīmju kodēšanas standarts, kas nodrošina unikālu numuru katrai rakstzīmei neatkarīgi no platformas, ierīces, lietojumprogrammas vai valodas.

I. Vispārīgie principi

Konvertācijas mērķis ir seno tekstu korpusa (https://korpuss.lv/id/Senie) tekstus pārveidot nosacītā mūsdienu rakstībā, lai šos tekstus varētu izmantot par pamatu meklētājam Nacionālās korpusu kolekcijas vietnē korpuss.lv. Proti, lai, meklētājā ievadot vaicājumu, parasti vārda sakni, mūsdienu rakstībā, korpusā būtu iespējams atrast seno tekstu vārdus to vēsturiskajā pierakstā (kas oriģinālā ir rakstīti no mūsdienām ievērojami atšķirīgā ortogrāfijā).

Par mūsdienu rakstību to var saukt tikai nosacīti, jo konvertējot netiek mūsdienīgots viss teksts – ja vien iespējams, tiek atstātas neskartas seno tekstu morfoloģiskās un fonētiskās īpatnības, piemēram, vārds *dabujama* netiek konvertēts par *dabūjama*, pieņemot, ka tā ir nevis rakstības, bet agrāko laiku valodas īpatnība (kas mūsdienās var būt saglabājusies dialektos, proti, darbības vārds *dabuit* ‘dabūt’). Tāpat ne vienmēr ir mūsdienīgotas vārdu izskaņas vai galotnes, piemēram, vārds *mūžigas* parasti netiek konvertēts par *mūžīgas*, netiek ņemti vērā lielie burti, arī īpašvārdos.

Lai programmētāja varētu veikt automātisku tekstu konvertāciju, valodnieki katram avotam sagatavo konvertācijas likumu tabulas. Tajās ir vairākas kolonnas – 1) „Nr.” – likuma numurs pēc kārtas, 2) „*Unicode*” – simbols vai to virkne no senā teksta *Unicode* fontā, 3) „Konvertētais” – konvertētais simbols vai to virkne, 4) „Reģistrjutība” – tā galvenokārt jāatzīmē garajam nepārsvītrotajam ſ vai kombinācijām

ar šo simbolu. Dažkārt reģistrjutību var atzīmēt attiecībā uz lielajiem burtiem, proti, ja nomaiņa jāveic tikai vārdā, kurā ir lielais burts, 5) „*Unicode* kods" – tikai tad, ja ir kāds īpašs simbols, 6) „Pārcēlumu skaits", 7) „Piezīmes" (skat. 1. attēlu).

Nr.	*Unicode*	Konvertētais	Reģistrjutība	*Unicode* kods	Skaits	Piezīmes
1.	ah	ā			2190	
2.	Gabrie	Gabr%			2	
3.	Danie	Dan%			2	
4.	ie	ī			1618	
5.	%	ie			4	
6.	eeh	ie			2	
7.	eh	ē			1523	
8.	ee	ie			4902	
9.	uih	ui			1	
10.	ih	ī			59	
11.	jh	ī			21	
12.	yh	ī			27	
13.	uh	ū			1230	
14.	yoh	jo			15	
15.	oh	o			1230	
16.	äh	ē			1089	
17.	ädam	ēdam			29	
18.	Thäw	tēw			9	
19.	ſchäl	žēl	1	17F	114	

1. attēls. Georga Manceļa darba „Lettische geistliche Lieder vnd Psalmen" (Manc1643_LGL) konvertācijas tabulas fragments.

Likumus raksta un nomaina secīgi, līdz teksts atbilst vēlamajam rezultātam. Var būt dažādas pieejas tekstu konvertācijā: 1) var vispirms nomainīt viennozīmīgos burtus vai burtkopas un pēc tam veidot likumu un to izņēmumu kopas daudznozīmīgajiem burtiem; 2) var vispirms veidot likumus daudznozīmīgajiem burtiem vai burtkopām un daudzajiem izņēmumiem un pēc tam pārvērst viennozīmīgos burtus vai burtkopas; 3) var konvertācijas likumus veidot jauktā secībā, kā tas tika darīts sākotnēji, meklējot atbilstošāko konvertēšanas metodi. Prakse rāda, ka 3. variants īsti izmantojams tikai īsajiem tekstiem.

Tāpat var vispirms konvertēt patskaņus, pēc tam līdzskaņus vai otrādi, var konvertāciju veikt jauktā secībā. Likumu skaitu un konvertācijas kvalitāti tas īpaši neietekmē. Jebkurā gadījumā konvertācijas tabulas beigās tiek ietverti likumi, kas novērš veiktās konvertācijas rezultātā radušās kļūdainās atveides. Tomēr ir svarīgi ievērot likumu secību tajos gadījumos, kad no tās ir atkarīgs konvertācijas rezultāts, piemēram, *w* pārvērst par *v* tikai tad, kad tekstā esošais *v* būs pārvērsts par *u* vai *ū*, iespējams, ka pirms tam būs jānovērš vēl kādi izņēmuma gadījumi.

Paralēli konvertācijas likumiem tiek gatavots t. s. pārcēlums *word* dokumentā, kurā, ievērojot konvertācijas likumus, seno tekstu burti, burtkopas, vārdu daļas vai veseli vārdi pakāpeniski tiek nomainīti pret mūsdienu atbilsmēm. Vārdu daļu vai veselu vārdu nomaiņa parasti notiek jeb tā sauktie leksiskie / individuālie likumi tiek rakstīti tad, ja burtu un burtkopu likumiem ir izņēmumi, kā arī pašās konvertācijas beigās, kad tiek meklētas konvertācijas kļūdas. Retos gadījumos, lai konvertācijā nodalītu avotā vienādi rakstītu vārdu nozīmes (*auksts* no *augsts*, *mūsam* no *mūžam*, *sāl* no *zāl*, *kazas* no *kāzas*), var veidot likumus divu vai pat vairāku vārdu savienojumiem. Piemēram, pēc likumiem *aukſta[]top→auksta top* un *Milti[]aukſti→milti auksti* var rakstīt likumu *aukſt→augst.*

Katru likumu pārbauda visā tekstā, tādā veidā nosaka izņēmumus. Piemēram, pirms pārvērst *ſ* par *z*, jāatrod visi izņēmumi (gan atsevišķas simbolu kombinācijas, gan veseli vārdi), kad *ſ* mūsdienās atbilst cita zīme. Šie izņēmumi jāraksta vispirms.

Pārcēlumam tiek izmantots *Unicode* fails ar paplašinājumu „Unicode_unhyphened" (šeit vārdu daļas, kuras atdala pārnesuma zīmes, saplūdinātas kopā vienā vārdā), kuru iekopē *word* dokumentā. Pārbaudot tekstu, tiek izsecināts, kādi konvertācijas likumi nepieciešami.

Pārcēlumu veido, izmantojot „Find and Replace" funkciju un burtus, burtkopas vai veselus vārdus vai vārdu daļas nomainot pret mūsdienu atbilsmēm citā krāsā.

Izstrādātā tabula pusautomātiski (ar regulāro izteiksmju palīdzību, skat. https://github.com/LUMII-AILab/Senie/blob/main/Unicoding/Docs/add-translit-table-from-word_readme.md), tiek pārveidota par *Perl* programmkoda fragmentu un iekļauta konvertācijas rīkā. Šis rīks pieejams https://github.com/LUMII-AILab/Senie/tree/main/Unicoding, tabulas glabājamas failā https://github.com/LUMII-AILab/Senie/blob/main/Unicoding/LvSenie/Translit/SimpleTranslitTables.pm, kur tās tiek indeksētas pēc avotu kodiem. Tad šādi papildinātais konvertors tiek darbināts, par ieejas datiem izmantojot atbilstošo avotu. Rezultātu pārlasa tabulas autors un novērtē konvertācijas kvalitāti. Ja atrodas konvertācijas kļūdas, tabulu precizē, pievienojot jaunus likumus vai labojot esošos, un tad šajā rindkopā aprakstītā konvertācijas rīka papildināšana un darbināšana un rezultāta novērtēšana tiek veikta vēlreiz. Procesu atkārto, līdz konvertācijas rezultāts sasniedz vēlamo kvalitāti.

Gari teksti (200 un vairāk lappušu) var būt mazāk noslīpēti.

Kā jau minēts, gandrīz katram avotam līdz šim tika veidota sava konvertācijas tabula. Izņēmums ir atsevišķas Bībeles grāmatas, kurām var izmantot vienu tabulu, savukārt 18. gadsimta tekstus turpmāk varēs konvertēt, par pamatu ņemot projekta darba grupas veidoto 18. gadsimta pilotkonvertoru (http://hdl.handle.net/20.500.12574/140). Konvertācijas likumu tabulu izveides, likumu secības principi ir atšķirīgi dažādiem gadsimtiem, dažādiem rakstības periodiem.

Tehniskās piezīmes

Konvertācijas tabulas faila nosaukumā jābūt ietvertam precīzam avota kodam – nosaukums jāsāk ar avota kodu, to papildinot ar datumu dilstošā secībā, piemēram, Baum1699_LVV-20231201.docx. Arī paša dokumenta pirmajā rindā ir vēlams uzrakstīt precīzu avota kodu.

Pirms konvertācijas tabulas ir vēlams uzskaitīt izmantotos substitūtu (pagaidu aizstājējzīmju) simbolus, tas palīdz vēlāk identificēt un izlabot kļūdas (atkļūdot tabulu). Pagaidu aizstājējzīme var būt jebkura zīme, kas nav alfabētā un kas nav izmantota seno tekstu korpusā, piemēram*, %, &, #, @, $*.

Visiem konvertācijas likumiem jābūt uzskaitītiem vienā tabulā, tajā nedrīkst būt tukšu rindu vai apvienotu šūnu. Ja nepieciešams sev darba laikā uzsvērt noteiktas tabulas daļas, var izmantot dažāda treknuma līnijas, iekrāsojumus un piezīmju lauku – šie elementi neietekmēs tālāko tabulas pārveidošanu par konvertācijas kodu.

Konvertācijas tabulas kolonnām „*Unicode*", „Konvertētais" un „Reģistrjutība" ir jābūt blakus un tieši šādā secībā. Tajās nedrīkst lietot pārnesumus jaunā rindā (*Enter*). Kolonnās „*Unicode*" un „Konvertētais" nedrīkst būt tukšu šūnu. Ja ir nepieciešams konvertācijas likums, kas kaut ko aizvieto ar neko, programmētāja jāpabrīdina atsevišķi.

Kolonnā „Registrjutība" vērtība 1 nozīmē reģistrjutīgu meklēšanu, bet 0 vai tukšums – reģistrnejutīgu. Jārēķinās, ka šo norādi izmantos programmatūra (*Perl*

regulāro izteiksmju dzinējs), kas vadās pēc *Unicode* reģistrjutības specifikācijas[3] un specifiskiem burtiem piemēro mūsdienu interpretācijas, nevis vēsturiskās. Piemēram, „s" (U+0073) reģistrnejutīgi sakrīt ne tikai ar „S" (U+0053), bet arī ar garo *s* „ſ" (U+017F), savukārt „ẞ" (U+1E9E) un „ß" (U+00DF) sakrīt ne tikai savā starpā kā attiecīgi lielais un mazais burts, bet arī ar burtkopu „ss" (U+0073 U+0073).

Konvertācijas likumos var izmantot šādus mehānismus:

- var meklēt vārda sākumā vai beigās, izmantojot simbolus *[]*, piemēram, *[]vārds* vai *vārds[]* un *[]vārds[]* atradīs precīzi šo vārdu, šiem likumiem vārdu robežas iezīmē arī pieturzīmes un arī biedruzīme =.
- var meklēt rindas sākumā, izmantojot *[^]*, piemēram, *[^]diezgan* atradīsies rindā „diezgan laba rinda", bet neatradīsies rindā „šī rinda ir diezgan laba"
- var meklēt rindas beigās, izmantojot *[$]*.

Ja tiek lietotas šīs norādes (piemēram, *[]wahrd*), tās rakstāmas kolonnā „*Unicode*", jo tās ir norādes, kas palīdz atrast labojamo vietu. Kolonnā „Konvertētais" raksta tikai tiešām lietojamos aizvietošanas simbolus, bez šīm papildu norādēm (piemēram, *vārd*).

Konvertācijas likumi nedrīkst mainīt tekstvienību (vārdu un pieturzīmju) skaitu rindā. Šim nolūkam biedruzīme = netiek skaitīta par atsevišķu vārdu, to uzskata par vārda daļu.

Konvertācijas likumi tiek piemēroti secīgi katrai avota rindai (izņemot daļas svešvalodās un funkcionālos blokus kā @a{}) tādā secībā, kādā tie uzskaitīti tabulā.

II. Dažādu gadsimtu tekstu konvertācijas specifika

1. 16. gadsimts

16. gadsimta tekstus var iedalīt divās grupās – īsie teksti, proti, tēvreizes tulkojumi, kas ir paši pirmie rakstītie teksti latviešu valodā (1507–1593), un jau garāki teksti, pirmās drukātās grāmatas (1585–1587). Abas šīs grupas vieno ortogrāfijas nekonsekvence. Ir vispārēji ortogrāfijas principi un atsevišķi viennozīmīgi burtu vai burtkopu lietojumi, kas raksturīgi gandrīz visiem šī gadsimta tekstiem un atspoguļojas vienotos konvertācijas likumos, piemēram, *bb→b*, tomēr lielākā daļa burtu vai burtkopu 16. gadsimtā ir daudznozīmīga, turklāt daudzu vārdu rakstība ir kļūdaina.

1.1. Tēvreižu tulkojumi

16. gadsimta īso tekstu – tēvreižu – konvertācijai ir nepieciešams salīdzinoši neliels likumu skaits (27–41), kas saistīts ar to, ka pats tēvreizes teksts ir neliels, tajā ir aptuveni 50 vārdu, no kuriem daļa tekstā atkārtojas, Gisberta tēvreizē vēl mazāk – 27. Tēvreižu tekstu konvertācijai galvenokārt tiek izmantoti leksiskie jeb individuālie likumi, proti, vesels vārds vai vārda daļa vēsturiskajā rakstībā tiek aizstāta pret veselu vārdu vai vārda daļu mūsdienu rakstībā.

Tēvreižu rakstību izprast palīdz to unificētais saturs, kura dēļ ir iespējams atpazīt atsevišķus neskaidri pierakstītus vārdus. Nereti oriģinālā atsevišķi vārdi ir sarakstīti kopā. Šie gadījumi vēl pirms konvertācijas jāatpazīst un jau *Unicode* dokumentā jāapzīmē kā kļūda, koprakstījumu sadalot atsevišķos vārdos. Šādam sadalītam tekstam pēc tam ir iespējams veidot leksiskos konvertācijas likumus, piemēram, *Girkade→Gir kade→jir kāde*, *Cademes→Cade mes→kāde mēs* (Gr1520_PN). Savukārt viens vārds oriģinālajā tekstā var būt sadalīts vairākās atsevišķās daļās, kas pirms konvertēšanas

[3] https://unicode.org/Public/UNIDATA/CaseFolding.txt

jāapvieno: *Prettaune kans→Prettaunekans→pretiniekiems* (Gr1520_PN), *AEth peſti→AEthpeſti→atpestī* (Br1520_PN).

Izmantoto burtu kopums ir daudzveidīgs, proti, pierakstā vērojama ļoti liela dažādība. Piemēram, mūsdienu grafēmai *v* tēvreizēs atbilst gan *w* (Gis1507_PN, Gr1520_PN), gan *v* (Meg1593_PN), gan *vv* (Laz1557_PN, Meg1593_PN), gan *vvv* (Laz1557_PN), gan *u* un *uu* (Thev1575_PN), u. tml.

Tomēr jau tēvreizēs vērojami precīzāka pieraksta meklējumi viena avota robežās.

Atsevišķās tēvreizēs saskatāmi centieni nodalīt īsos un garos patskaņus, piemēram, burtkopa *aa* atbilst mūsdienu grafēmai *ā*, tomēr pārējo burtu un burtkopu daudznozīmības dēļ konvertācijai ir izmantoti leksiskie likumi, piemēram, *vvaarcz→vārds*, *praatz→prāts* (Laz1557_PN), *vuaartz→vārds*, *praatz→prāts* (Thev1575_PN), *vvaartz→vārds*, *praatz→prāts* (Meg1593_PN), *eÿ→ē*, piemēram, *Zweÿtz→svētīts* (Gr1520_PN), *ä→ē*, piemēram, *Täbes→Tēves*, *gräke→grēke* (Has1550_PN), *y→ī*, piemēram, *Swetytz→svētīts* (Br1520_PN), *debbeſyß→debesīs*, *ſwetyt→svētīt*, *walſcyby→valstībī* (Gis1507_PN), *ÿ→ī*, piemēram, *Sÿmmes→zīmes*, *Worsünÿ→virsūnī* (Gr1520_PN), *ũ→ū*, piemēram, *mũs→mūs* (Meg1593_PN).

Dažkārt novērojams līdzskaņu dubultojums, tādēļ var tabulās ietvert, piemēram, likumus *bb→b*, *mm→m*, *rr→r*, tomēr kopumā arī līdzskaņu pieraksts ir ļoti daudzveidīgs un tikai likumu *ß→s* var izmantot visu šobrīd korpusā *SENIE* ietverto tēvreižu konvertācijā.

1.2. Pirmās grāmatas

Arī lielajiem 16. gadsimta avotiem, pirmajām izdotajām grāmatām, katram tiek veidota sava konvertācijas likumu tabula.

Konvertācijas vadlīnijas ir veidotas, balstoties uz senākās latviešu valodā izdotās un līdz mūsdienām nonākušās grāmatas „Catechismvs Catholicorum..” (CC1585) konvertācijas likumu tabulas, tomēr tas ir tikai kā paraugs, jo citu avotu tabulās lielākoties ir citi likumi. Katra teksta īpatnības nosaka unikālu konvertācijas gaitu. Avotos kopumā izmantoti ļoti dažādi burti un burtkopas, turklāt katra rakstības nosacīti kļūdainā novirze no avotam tipiskā ienes arvien jaunas papildu izmaiņas, liekot ar tām rēķināties cauri visam konvertēšanas procesam līdz beigām, krietni vairojot likumu skaitu. Lai arī teorētiskie principi ir tie paši, tieši šo individuālo gadījumu neproporcionāli lielais daudzums ir tas, kas neļauj unificēt 16. gadsimta tekstu konvertāciju.

Šī avota konvertācijā ir ievērots 2. princips, tabulas iesākumā veidojot likumus daudznozīmīgajiem līdzskaņu burtiem, piemēram, *ſ*, *x* un *c*, kuriem ir neskaitāmi izņēmumi, kuru dēļ ir jāveido daudzi leksiskie likumi.

Svarīgi ievērot vēl kādu papildu principu – no lielākā uz mazāko. Proti, vispirms tiek pārvērstas burtkopas, kuru sastāvā ir vairāk zīmju, un tikai tad, kad veiktas šīs konvertācijas un to izņēmumi, var pārvērst vismazāko vienību, proti, vienu burtu.

Lai varētu veikt veiksmīgu teksta konvertāciju, katram burtam vai burtkopai tiek noteikta biežāk sastopamā atbilsme, piemēram, *ſch→š*, pirms tam veicot virkni izņēmuma likumu jeb leksisko likumu, kuros *ſch* = [s] vai [z], piemēram, *ſcheſt→sest*, *ſchwan→zvan*. Tiek pārvērsts *dſ→dz*, bet pirms tam realizēti leksiskie likumi, kuros *dſ* = [ž] vai [z], piemēram, *dſelix→žēlīgs*, *dſem→zem*.

Pirms likuma *ſ→s* realizēti aptuveni 100 burtkopu vai leksiskie likumi, kuros *ſ* konvertēts kā *z* vai *ž*. Daļa no šiem likumiem ir pozicionāli, piemēram, *[]iſ→iz*, *[]vſ→uz* (*[]* rāda, ka likums attiecas tikai uz burtkopām, kas atrodas vārda sākumā). Tikai tad, kad visi šie likumi realizēti, var pārvērst *ſ→s*. Visos likumos, kur *ſ* ir viens pats, divu grafēmu savienojumā vai vārda sākumā, noteikti jāatzīmē reģistrjutība 1.

Dažos burtkopu vai leksiskajos likumos *x→gs*, *ks* vai *kš*, piemēram *vnx[]→ungs*, *exs[]→ēks*, *exs→iekš*.

Pirms likuma *sch→š* realizēti vairāki likumi, kuros *sch→s*, piemēram, *[]schwe→svē*.

Vairākos burtkopu vai leksiskajos likumos *c→k*, *s*, *c* vai *č*, piemēram, *[]cat→kat*, *[]cag→sag*, *[]cen →cien*, *[]cettr→četr*.

Pēc daudznozīmīgo līdzskaņu likumiem seko daudznozīmīgo patskaņu (gan burtu, gan burtkopu, gan leksiskie) likumi.

Virkne likumu, kuros *eh→ie*, piemēram, *ehle→ielie*.

Pirms likuma *ah→a* realizēti vairāki burtkopu vai leksiskie likumi ar *ah→ā*, piemēram, *[]tah[]→tā*.

Dažādi likumi ar burtu *i*, piemēram, *aij→āj*, *eij→ēj*, *ijt→īt*. Vairākās burtkopās *i→j*, piemēram, *ia→ja*, *[]iu→ju*.

Vairākos leksiskajos likumos *e→ie*, piemēram, *[]pect→piekt*.

Patskaņu burtkopa pirms *w→*patskanis, piemēram, *ouw→av*, *euw→ev*.

Pirms likuma *ee→ē* realizēti vairāki burtkopu vai leksiskie likumi ar ar *ee→e* vai *ie*, piemēram, *[]bees[]→bez*, *[]ween[]→vien*.

Pirms likuma *ae→ā* realizēti vairāki burtkopu vai leksiskie likumi ar *ae→a*, *ē*, *ai*, piemēram, *aen→an*, *aedi→aidi*, *aeecz[]→ēc*.

Pēc tam atkal seko vairāki leksiskie un burtkopu likumi, kas attiecas uz daudznozīmīgajiem līdzskaņiem: *tcz→ds*, *ts* vai *c*, piemēram, *[]wartcz[]→vārds*, *ietz[]→īts*, *[]titcz[]→tic*, un *tz→ts* vai *c*, piemēram, *otz[]→ots*, *petz→pēc*.

Vairākos leksiskajos un burtkopu likumos *c→k*, piemēram, *doc→dok*, vai *c*, piemēram, *ceni→cienī*.

Vairākos leksiskajos un burtkopu likumos *v→u*, piemēram, *[]vs→uz*, *vnd→und*. Burts *v* jāpārvērš par *u*, pirms tiek realizēts likums *w→v*.

Seko vairāki viennozīmīgo burtu un burtkopu likumi (gan patskaņu, gan līdzskaņu).

Vairākas divu patskaņu burtkopas tiek konvertētas par patskani vai divskani: *ue→ū*, *uo→ū*, *oe→o*, *ih→ī* un *oh→o*.

Līdzskanis savienojumā ar *h→*līdzskanis, piemēram, *bh→b*, *dh→d*, *mh→m*, *nh→n*, *th→t*, savukārt *ch→h*. Burtkopa *ng→ņ*.

Dubultie līdzskaņi: *gg→g*, *bb→b*, *dd→d*, *kk→k*, *nn→n*, *mm→m*, *pp→p*, *rr→r*, *tt→t*, *šš→š* (šis likums attiecināts uz jau agrāk pārvērstiem burtiem), *w→v*. Pirms likuma *ll→l* jāveic likumi ar aizstājējzīmi saknēs, kurās arī konvertētajā variantā jābūt *ll*, piemēram, *allaž→a&až*, *celle→ce&e*. Kad veikts likums *ll→l*, visus *&* var pārvērst par *ll*.

Tā kā konvertējot tekstā daudzas saknes vēl pēc šo likumu veikšanas ir atšķirīgas no mūsdienu lietojuma (16. gadsimta tekstos tikpat kā nav šķirti garie un īsie patskaņi, divskaņa *ie* vietā *e* utt., ir daudz acīmredzami kļūdainu lietojumu), tad, pēc iepriekš veikto likumu realizācijas pārskatot tekstu, izveidoti vēl aptuveni 150 leksiskie likumi, lai ar meklētāja palīdzību attiecīgās saknes tomēr būtu iespējams atrast, piemēram, *[]len→lēn*, *[]vard→vārd*, *[]pep→piep*. Tabulā kopā vairāk nekā 500 konvertācijas likumu.

Kā jau minēts, citos lielāka apjoma 16. gadsimta latviešu avotos likumi ir atšķirīgi, ir vairāki citi daudznozīmīgi burti, piemēram, *ä*, *ö*, *ß*, tomēr pamatprincipi visiem 16. gadsimta avotiem ir līdzīgi: 1) maz viennozīmīgo burtu vai burtkopu likumu, 2) daudz tādu likumu, pirms kuru realizācijas jārealizē virkne leksisko likumu ar izņēmumiem, 3) daudz pozicionālo likumu (tikai vārda sākumā vai beigās, blakus

noteiktam burtam). Salīdzinājumā ar 17. un jo īpaši 18. gadsimta tabulām leksisko likumu ir neproporcionāli daudz.

Šie lielāka apjoma teksti ar nekonsekvento rakstību un daudzajām kļūdām pēc konvertācijas var būt mazāk standartizēti salīdzinājumā ar īsākiem tekstiem un vēlāko gadsimtu tekstiem, koncentrējoties galvenokārt uz sakņu normalizēšanu, lai tās meklētājā būtu iespējams atrast.

2. 17. gadsimts

Arī 17. gadsimta latviešu rakstu avoti nav viendabīgi. Rakstības kvalitātes ziņā tie ir atšķirīgi, līdz ar to atšķiras arī pieeja to konvertācijai. 17. gadsimta sākumā rakstībā joprojām ir lielas nekonsekvences, raksturīga burtu un burtkopu daudznozīmība, tādējādi, līdzīgi 16. gadsimta beigu avotiem, katra avota konvertācijas tabula ir jāveido atsevišķi ar tam raksturīgiem konvertācijas likumiem, daudz ir t. s. individuālo atbilsmju, kad sakne tiek nomainīta pret sakni vai vesels vārds pret veselu vārdu.

17. gadsimta vidū Georga Manceļa darbos (1631–1654) rakstība jau ir vairāk izstrādāta, tomēr arī ar vairākām problemātiskām rakstu zīmēm.

Savukārt 17. gadsimta beigu latviešu rakstu avotos, lielā mērā saistībā ar Ernsta Glika Bībeles tulkojumu (1685–1694), jau ir raksturīga salīdzinoši lielāka konsekvence ortogrāfijā, līdz ar to vairākām Bībeles grāmatām iespējams izmantot vienu konvertācijas tabulu. Arī citu 17. gadsimta beigu avotu konvertācijas secība ir līdzīga, vismaz pamatam iespējams izmantot vienotu pieeju, pēc tam tabulu pielāgojot konkrētajam avotam.

Tomēr jāņem vērā, ka 17. gadsimta sākumā un vidū vairāki latviešu rakstu autori ir darbojušies paralēli, viņu lietotā rakstība atšķiras un līdz ar to atšķiras arī katra darbu konvertācijas gaita. Piemēram, savdabīga rakstība lietota Georga Elgera darbos (1621, 1672 u. c.), 17. gadsimta vidū savu rakstības sistēmu izveidoja Kristofs Fīrekers, citu atšķirīgu rakstu valodas modeli piedāvājis Vidzemes mācītājs Jānis Reiters savā Bībeles tulkojuma paraugā (1675). Un, protams, Georgs Mancelis, kurš pirmais ķērās pie agrākās latviešu rakstu valodas reorganizācijas, radot jaunu grafētikas un ortogrāfijas sistēmu, kas bija sistemātiskāka par līdzšinējo.

Līdz ar to atsevišķi jāaplūko darbs pie šo autoru darbu konvertācijas tabulām.

2.1. Georgs Mancelis

Materiāla analīze ir ļāvusi izdalīt divus skaidri nošķiramus Manceļa tekstu periodus, kā arī pārejas periodu starp tiem. Pie pirmā posma pieder 1631. gadā izdotie darbi, pie otrā posma pieskaitāmi 1643.–1644. gada izdevumi, kā arī spredikļu grāmata trijās daļās, kas iznāk 1654. gadā. Savukārt 1637.–1638. gada tekstus var raksturot kā pārejas periodu. Tiem piemīt gan pirmā, gan otrā posma iezīmes. Tomēr pat aptuveni vienā laikā tapušiem Manceļa darbiem ir pietiekami daudz katram savu rakstības īpatnību un vienu tabulu vairākiem avotiem izmantot nav iespējams.

Veidojot 1643.–1644. gada atjaunotā un pārlabotā rokasgrāmatas „Lettisch Vade mecum” (Manc1643_44_LVM) izdevuma tabulas, par pamatu varēja ņemt 1631. gada izdevumu tabulas, tās attiecīgi izmainot, ņemot vērā Manceļa paša ortogrāfijas uzlabojumus.

Piemēram, 1631. gada „Lettisch Vade mecum” (Manc1631_LVM) konvertācijas tabulā ir 404 likumi, 1637. gada „Die Sprüche Salomonis” – 301 likums, 1643–44. gada „Lettisch Vade mecum” – 365 likumi.

Tālāk dota Manceļa darbu konvertācijas likumu aptuvenā secība.

1. Burtkopa *ſch* ir daudznozīmīga. Tāpēc, pirms nomaina *ſch→š*, veic vairākas citas burtu kombināciju nomaiņas. Vispirms jānomaina lielākas burtu kombinācijas, leksēmas vai leksēmu daļas, kurās burtkopai *ſch* mūsdienās atbilst līdzskanis *ž* (*[]daſſch[]→dažs*, *muhſch→mūž*, *ſchähl→žēl*, *[]eſch→ež*, *[]meſſcha→meža*, *allaſch→a%až* u. c.). Jākonvertē *ſſch→š*, *ſſ→s*, *ſſch→š*, *tſch→č*, *ſch→š* (pirms tam *ſchk→šķ*), pēc tam *ſch→š.*

2. Lai veiktu nomaiņu *tz→c* un *z→c*, iepriekš jāveic vairākas citas maiņas.

Vārdos, kur burtkopa *ds[]* atbilst *dz[]*, *z* aizstāj ar aizstājējzīmi, piemēram, *&* (*lieds→līd&*, *neds→ned&*, *dauds→daud&*).

Konvertē vārdus, kur *tz[]* vārda beigās apzīmē *ds[]* vai *ts[]* (*Ƨirrtz→sirds*, *wahrtz→wārds*, *kahtz→kāds*, *[]patz[]→pats*, *deßmittz→desmits* u. c.), kā arī vārdus, kas beidzas ar *-ots*, *-īts*, *-āts*, *-ēts* (*ohtz[]→ots*, *ietz[]→īts*, *ahtz[]→āts*, *ähtz[]→ēts*), arī citus izņēmumus.

Konvertē *cz→c* (tikai Manc1631_LVM), *tz→c*, *z→c*, *&→z*, *ſ→s.*

3. Nomaina atsevišķas leksēmas vai leksēmu daļas, kur *s* atbilst *z* (*semm→zem*, *sinn→zin*, *siem→zīm*, *sohb→zob*).

4. Šīs maiņas notiek ar domu, ka kaut kur tālāk *ſ* tiks pārvērsts par *z* (tā kā izņēmumu ir daudz, tad šis likums tiek atstāts uz beigām).

Jākonvertē *ſkieſt→šķīst*, *ſki→šķi* (varbūt vēl kādas kombinācijas vai leksēmu daļas, kur *ſk* atbilst *šķ*), *ſſ→s*, *[]ſa→sa*. Visos likumos, kur *ſ* ir viens pats, divu grafēmu savienojumā vai vārda sākumā, noteikti jāatzīmē reģistrjutība 1.

5. Likums *iex[]→īgs* (pirms tam *ißniex→iznīks*, varbūt vēl kāds izņēmums).

6. Patskaņu konvertācija:

yahß[]→jās (ja nomaina *ah→ā*, pēc tam šo *j* nevar atrast), *à→ā*, *ie→ī*, *ee→ie*, *ahy→āj*, *ah→ā*, *äh→ā*, *ä→e* (pirms tam nomaina leksēmu daļas, kur *ä→ē*, *ghräk→grēk*, *cillwä→cilwē* u. c.), *eh→ē* (pirms tam *ehy→ēj*), *ë→e*, *ih→ī*, *yh→ī*, *oh→o*, *ö→e*, *uh→ū*.

7. Līdzskaņu dubultojums u. c. līdzskaņu burtkopas:

bb→b, *dd→d*, *dſ→dz*, *gg→g*, *gh→g*, *gk→ķ*, *ck→k*, *ll→l* (pirms tam ar aizstājējzīmi nomaina *ll* tajos vārdos, kur tam ir jāpaliek, piemēram, *pills→pi%s* (reģistrjutība 1), *elles→e%e*s u. c.), *ļļ→l*, *mm→m*, *mb→m* (pirms tam sameklējot tādus vārdus kā *kambaris*, *septembris* u. c., kur *mb* jāpaliek), *m̃→m*, *nn→n*, *ņņ→ņ*, *nh→n*, *ñ→n*, *pp→p*, *rr→*r (pirms tam aiztājot *rr* vārdos, kur tam jāpaliek, piemēram, *myrrhes→my#hes* u.tml.), *ŗŗ→ŗ*, *rŗ→ŗ*, *tt→t*, *th→t* (pirms tam var *[]tha[]→tā*), *dt→t*.

8. Pirms *v* pārvērš par *u*, jākonvertē *v* priedēkļos (piemēram, *[]vs→uz*, *[]vhs→ūz*, *[]vß→uz*, *[]vhß→ūz*). Konvertē *vh→ū*, pēc tam *v→u*.

9. Konvertē *ings[]→iņš*, *ing→iņ*, *nj→ņ*, *ny[]→nī*, *ny→ņ*.

10. Konvertē *ſw→sv*, *[]jaw→jau*, *w→v*, *qu→kv*, *q→k.*

11. Konvertē *[]py→pi*, *[]by[]→bij*, *[]gir[]→jir.*

12. Pirms *ß* pārvērš par *s*, jākonvertē izņēmuma gadījumi priedēkļos u. c. (piemēram, *[]iß→iz*,*[]i hß[]→īz*, *[]beß→bez*, *[]aiß→aiz*). Pēc tam *ß→s.*

13. Konvertē *Chr→Kr*, *JEſ→Jēz*.

14. Pirms *x* pārvērš par *ks*, jākonvertē izņēmumi (piemēram, *[]aux→augs*, *=aux→augs*, *kunx→kungs*, *draux→draugs* u. c.). Pēc tam *x→ks.*

15. Jākonvertē *[]ney[]→nei*, *ya→ja*, *yi→ji*, *yu→ju*, *ay→ja*, *ey→ej*, varbūt vēl kādi citi savienojumi ar *y.* Pēc tam var pārvērst *y→ī.*

16. Pirms *ſ* pārvērš par *z*, jākonvertē burtkopas, kur *ſ* atbilst *s*:

ſp→sp, *ſt→st*, *ſk→sk*, *ſl→sl*, *ſm→sm*, *ſ→z*, *Ƨ→s* (vēlākiem avotiem pirms tam *Ƨch→š*).

17. Jāveic maiņa *ch→h*.

18. *J* atbilsmes jāmeklē pēc leksēmām, piemēram, *Jr→ir*, *Jk→ik*, *MJ→mī* u. c.

Šie ir tikai galvenie bloki, katrā no tiem iespējami vēl kādi izņēmumi vai vārdi, kas jākonvertē pilnībā. Visur, kur lietotas aizstājējzīmes, tās pēc tam jāpārvērš atpakaļ par burtu, kuru tās aizstājušas.

2.2. Georgs Elgers

Georga Elgera tekstu konvertācijas likumi kardināli atšķiras no visiem iepriekš (un turpmāk) aprakstītajiem, jo viņa rakstības pamatā ir nevis vācu, bet gan poļu ortogrāfijas principi. Daudz un daudznozīmīgi lietotas ir arī dažādas (pārsvarā patskaņu) diakritiskās zīmes, savukārt patskaņu garumi saknē atzīmēti daudz retāk nekā, piemēram, Georga Manceļa darbos, tāpēc jāveido daudz individuālo likumu. Īpaši problemātiska šajā ziņā ir Elgera vārdnīca (Elg1683_DictPLL), jo tajā daudz unikālu vārdu lietojumu. Vērojams arī, ka Elgera dziesmu grāmatā (Elg1621_GCG), kas iznākusi viņa dzīves laikā, rakstība ir konsekventāka nekā pēc viņa nāves izdotajos darbos.

Kā piemērs Elgera tekstu konvertācijas gaitai izmantota katehisma (Elg1672_Cat) konvertācijas tabula. Minēti nav pilnīgi visi likumi, bet tikai lielākās to grupas un tipiskākie, kā arī sarežģītākie gadījumi.

Poļu rakstības specifikas dēļ Elgera tekstos ir daudz grafēmas *i* lietojuma. Tā ir daudznozīmīga, jo apzīmē gan patskani *i*, gan līdzskani *j*, gan – lietota kopā ar līdzskaņiem *l*, *n*, *k* u. c. – kalpo par līdzskaņu mīkstinājuma zīmi pēc tiem. Tādēļ šīs daudznozīmības mazināšanai tika izlemts sākumā maksimāli atrast un pārveidot visus gadījumus, kur *i→j*, piemēram, vārdu sākumus *[]ia→ja*, *[]ie→je*, *[]iu→ju*, tāpat arī *i* starp patskaņiem, piemēram, *aia→aja*, *aiu→aju*, *aie→aje*, *uiu→uju*, *eie→eje* u. c., kā arī *oi→oj*.

Arī gafēma *y* ir daudznozīmīga – apzīmē gan mūsdienu *j*, gan *i*, gan *ī*. Tā kā pārsvarā gadījumu ir garais *ī*, sākumā tiek pārvērsti visi atšķirīgie gadījumi, piemēram, *ya→ija*, *yi→iji* u. tml., arī *ay→ai*, *áy→ai*, *[]yk→ik*, *[]yz→iz*, *[]yt→it*, *[]ayz→aiz* u. tml., arī darbības vārds *[]by[]→bij*. Tad seko vārdu saknes, kur *y=i*, bet pēc tam likums *y→ī*.

Lai atvieglotu sakņu konvertēšanu, sākuma posmā jāmēģina arī apzināt visi tekstā lietotie patskaņi ar diakritiskajām zīmēm, kas jākonvertē vispirms. Nereti arī tie ir daudznozīmīgi (apzīmē gan garos, gan īsos patskaņus), tāpēc ideāls risinājums nav iespējams, bet jāizvēlas izplatītākais veids, piemēram, *á*, *à* un *ȧ→a*, *â→ā*, bet *ê*, *é* un *è →ie*, *ė→ē* u. c. Kļūdainos garumus saknēs vēlāk var labot, konvertējot visu sakni. Šāda pieeja var izrādīties vienkāršāka nekā mēģinājums atlasīt visus izņēmumus, ja to ir daudz.

Viennozīmīgi gadījumi līdzskaņu konvertēšanā ir vairāku dubulto līdzskaņu vienkāršošana: *gg→g*, *kk→k*, *nn→n*, *rr→r*, *zz→z*, *żż→ž*, *cc→c*, *ww→v*, arī *ch→k*, *th→t*. Pirms *ww→v* gan jāveic konvertācija *v→u*. Savukārt pirms *ll→l* jāizmanto aizstājējzīme tādos vārdos kā *allaž*, *elle* u. c. Tad var konvertēt līdzskaņus *ß→š*, *w→v* un *x→ks*.

Jāpārbauda arī līdzskaņi ar diakritiskajām zīmēm, piemēram, poļu valodas ietekmē lietotais *ł→l*, arī *ź→ž* un *ż→ž*, kā arī poļu rakstībai raksturīgās burtkopas *cz→č, sz→š*.

Tāpat kā citos šā laika tekstos, problemātiska ir grafēmas *ſ* konvertēšana. Vispirms jāatlasa visi gadījumi, kur *ſ* nav *s*, piemēram, *[]iſſ→izs*, *neiſſ→neizs*, *[]beſſ→bezs*, *ſſ→s*, *[]iſs→izs*, *ſs→s*, *ſki→šķi*, *aiſto→aizto* u. c. Pēc tam var konvertēt *ſ→s*. Visos likumos, kur *ſ* ir viens pats, divu grafēmu savienojumā vai vārda sākumā, noteikti jāatzīmē reģistrjutība 1.

Nekonsekvents ir arī burtkopas *ae* lietojums. Tas apzīmē gan (lielākoties plato) īso patskani *e*, gan garo – *ē*. Tā kā *ae→ē* gadījumi dominē, vispirms ir jāpārveido burtkopas un saknes, kur *ae→e*, piemēram, *vaeln→veln*, *[]vaec→vec*, *raedz→redz*, *caetur→cetur*, *[]vaesel→vesel*, *daer→der*, *paeln→peln* u. c., bet pēc tam var konvertēt *ae→ē*.

Līdzīgi jādara ar burtu *e*, kas atbilst gan mūsdienu *e,* gan *ē*, gan *ie*. Tomēr pirms tam jāpārliecinās, vai tekstā nav palicis kāds *ie*, kur vajadzīga konvertācija *ie→je*. Šādi un līdzīgi gadījumi var uzrasties, jo daudzo diakritisko zīmju dēļ visus gadījumus uzreiz nevar apzināt. Šajā brīdī arī jāpārveido visi gadījumi, kur kombinācija „līdzskanis+i+e" nozīmē „mīkstināts līdzskanis+e", piemēram, *nie[]→ņe*, *nies→ņes*, *lie[]→ļe*, *liem→ļem*, *[]ēlie→ēļe*, *kie→ķe* u. tml. Tas jādara, jo vēlāk *ie* būs tik daudz, ka šos gadījumus atrast būs ļoti laikietilpīgi.

Kad tas izdarīts, vēlams pārveidot *[]e→ie*, pirms tam ar aizstājējzīmi saglabājot vārdu sākumus ar *e*, piemēram, *[]esmu*, *[]esot*, *[]eita*, *[]eliza*, *[]engel*, *[]ezer* u. c. Tad seko desmitiem gadījumu ar vārdu sakņu maiņu, kur *e→ē* vai *e→ie*, piemēram, *pec→pēc*, *pedig→pēdīg*, *devet→dēvēt*, *bern→bērn* un *[]let→liet*, *lecib→liecib*, *melest→mielast*, *tek→tiek*, *prec→priec* u. c. Tā kā tie ir individuāli gadījumi, reizēm vienlaicīgi ir iespējams novērst arī citas rakstības kļūdas vārda saknē. Reizēm vispiemērotākā ir kombinēta pieeja, piemēram, sakne *dev* lielākoties (~200 gadījumu) attiecas uz vārdu *dievs*, tāpēc lietderīgi ir atlasīt citus vārdus ar *dev*, piemēram, *deve*, *devit*, *devin* un konvertēt tos ar aizstājējzīmi, bet tad pārcelt visus atlikušos *dev→diev*, un pēc tam aizstājējzīmi nomainīt atpakaļ uz *e*.

Līdzīgi jāpārbauda visi īso patskaņu lietojumi un jāpārvērš saknes, kurās nav apzīmēts garais patskanis, piemēram, *[]mat→māt*, *nak→nāk*, *mil→mīl*, *[]mus→mūs*, *muž→mūž*. Šādu gadījumu ir ļoti daudz. Arī šeit, iespējams, jālieto kombinēta pieeja. Piemēram, lai, pārveidojot *nak→nāk*, saglabātos īsais patskanis vārdā *nakts*, vispirms jālieto aizstājējzīme *nakt→%*, bet pēc *nak→nāk*, *%→nakt*.

Kad lielākā daļa vārdu sakņu ir pārceltas, vieglāk atrast gadījumus „līdzskanis+i" = „mīkstināts līdzskanis". Vispirms vēlams pārbaudīt tādas kombinācijas kā *lia*, *liu*, *lio*, arī ar līdzskaņiem *r* un *n*. Ne vienmēr tas ir mīkstināts līdzkanis. Ja *lia→ļa*, *liu→ļu* un *lio→ļo*, tad *riu* ir gan *ŗu* (retāk), gan *riju* (biežāk). Tāpat arī ar *n*, tāpēc vispirms jāstrādā ar retākajiem gadījumiem, pārceļot saknes vai vārdu daļas, bet tad ar likumu *riu→riju* un *niu→niju*.

Šīs ir tikai galvenās norādes, katrā gadījumā ir iespējami vēl kādi izņēmumi vai vārdi, kas jākonvertē pilnībā.

2.3. Bībeles tulkojums

Līdz ar Bībeles tulkojuma izdevumu (1685–1694) un citiem 17. gadsimta beigu izdevumiem pakāpeniski nostiprinās samērā konsekventa rakstības sistēma, kurā ir daudz mazāk svārstīgu gadījumu. Tomēr pats Bībeles izdevums vēl nav pilnīgi viendabīgs. Īpaši izšķiras Mateja un Marka evaņģēliji, kuros visai plaši izmantota diferencējošā jeb loģiskā rakstība, mēģinot vairākas fonētiski vienādas, bet gramatiski atšķirīgas formas arī rakstīt atšķirīgi, piemēram, adjektīva vīriešu dzimtes daudzskaitļa nominatīva forma *labbi* 'labi', bet adverba forma *labbì* 'labi'. Tālākās Bībeles grāmatās šādu paņēmienu skaits samazinās. Turpmāk sniegts vispārējs Bībeles konvertācijas likumu uzskaitījums, kas atsevišķās grāmatās var būt mazāks, bet dažās pievienojami vēl citi likumi.

1. Tā kā vārda sākumā ar *J* apzīmē gan *J*, gan *I*, vispirms jākonvertē lielais *J→I,* tad kombinācijās ar sekojošu patskani par *J* (piemēram, *Ia→Ja, Io→Jo* u. c.).
2. Jākonvertē *w→v* visos gadījumos.

3. Tikai Marka un Mateja evaņģēlijos jāveic atbrīvošanās no mēmā *h*, kas izmantots diferencējošajā rakstībā, vispirms *tha→tā*, tad *th→t.*

4. Tad var konvertēt patskaņu savienojumus ar *h* (piemēram, *ah→ā, eh→ē)*, kā arī divskaņa apzīmējums *oh→o*.

5. Tālāk konvertējami patskaņu garumu apzīmējumi lokatīva galotnēs *â→ā, ê→ē, û→ū,* tāpat arī *ô→o*. Tā kā ar jumtiņu tiek apzīmēts arī vietniekvārdu lokatīvs, piemēram *taî*, tad vispirms veicama konvertācija *aî→ai*, kurai seko *î→ī*.

6. Tad konvertējami ar gravi apzīmētie adverbu formu gala patskaņi: *à→ā, ì→i, è→e.*

7. Nākamais solis ir pārbaudīt visu patskaņu, kā arī līdzskaņu *m*, *n* iespējamu apvienojumu ar diakritiku ˜, piemēram, *ã*, *ñ*, un veikt to konvertāciju, vadoties no katra atsevišķa zīmes lietojuma gadījuma, piemēram, *ã→an* vai *ã→am, ñ→nn.*

8. Tad var konvertēt divskani *ee→ie.*

9. Pēc tam jāatbrīvojas no citām patskaņu burtu diakritikām – trēmas (¨) un punkta (˙), piemēram, *ä→a, ȧ→a.*

10. Tālāk pārveidojamas kombinācijas *aij→aj* un *eij→ej.*

11. Jāveic dubulto līdzskaņu burtu aizstāšana ar vienu burtu šādos gadījumos *dd, ģģ, ķķ, ļļ, mm, ņņ, ŗŗ,* piemēram*, bb→b*.

12. Tā kā vairākos gadījumos morfēmu robežā vai saknē arī mūsdienu valodā iespējama divu līdzskaņa burtu rakstība, veicama pārbaude. Proti, vispirms pārbauda, vai tekstā nav tādu vārdu kā *allaž*, *elle*, *vells* ‘velns’, *mells* ‘melns’, *pills* ‘pilns’ u. tml. Ja tādi atrodami, šais gadījumos lietojama palīgmaiņa ar aizstājējzīmi, piemēram, *%.* Tad var konvertēt *ll→l*, pēc tam atpakaļ *%→ll.*

13. Tāpat jāpārbauda, vai tekstā ir vietniekvārdi *ikkatrs*, *ikkurš*. Šai gadījumā vispirms veicama palīgmaiņa *[]ikk→[]i%*, tad var konvertēt citus *kk→k*, beidzot *%→kk.*

14. Līdzīgā veidā jāpārbauda, vai tekstā ir vārdi ar priedēkļiem *ap-*, *at-*, *pār-*, kam seko sakne ar tādu pašu līdzskani. Arī šais gadījumos vispirms jāveic palīgmaiņa, aizstājot šos dubultojumus ar aizstājējzīmi, piemēram, *[]app→[]ap%*. Tad var vienkāršot dubultos *pp*, *tt*, *rr*, pēc tam aizstājējzīmes vietā atjaunot dubultos līdzskaņus morfēmu sadurā.

15. Dubults *nn* var būt ne tikai pozīcijā aiz patskaņa tā īsuma apzīmēšanai, bet arī aizgūtos vārdos, īpaši personvārdos, piemēram, *Anna*, *Annas*, *Suſanna*. Ja tekstā ir šādi vārdi, vispirms tajos veicama palīgmaiņa ar aizstājējzīmi. Tad var vienkāršot pārējos dubultos *nn→n,* un pēc tam atjaunot dubulto *nn* aizguvumos.

16. Pirms konvertēt *ʧch* par *č*, jāveic formu ar piedēkli *at-* pārveide – *[]atʃch→atš*. Tad var veikt konvertāciju *ʧch→č.*

17. Tālāk jāveic pārsvītroto *ſ* un *S* pārveide burtu kombinācijās *ſch→š, Sch→š.*

18. Jāpārbauda, vai vārda galā kombinācija *-ņš* nav rakstīta *-ņſch*. Ja parādās šāds rakstījums, jāveic konvertācija *ņſch→ņš.* Pēc tam var veikt pārveidi *ſch→ž* un *Sch→ž*.

19. Jāpārbauda vārda *vecs* nominatīva formas *wezz*. Ja šādas formas atrodamas, jāveic pārveidojums *vezz[]→vecs[]* un pēc tam *vezz→vecs*. Pēc tam jāpārveido *z→c*, kā arī jāvienkāršo dubultnieku rakstība *cc→c.*

20. Jāveic visu priedēkļu ar *s* galā – *ais-*, *bes-*, *is-*, *us-*, *ūs-* gala līdzskaņa konvertācija, piemēram, *[]ais→[]aiz*.

21. Veicama burtu kombināciju *ſk, ſl, ſp, ſt* konvertācija, piemēram, *ſk→sk*. Tāpat jāpārbauda, vai ar *ſ* nav apzīmēts nebalsīgais *s* arī pirms līdzskaņa *n*.

22. Pēc tam var visus pārējos *ſ* pārveidot par *z – ſ→z*, kā arī *S→z*. Taču tad jāpārveido vēl vārda sākumā radušās kombinācijas *zk, zl, zp, zt* par *sk, sl, sp, st*, proti, *[]zk→[]sk* utt.

23. Lai veiktu pēdējo lielo konvertāciju, pārveidojot pārsvītroto *ſ*, vispirms jāpārbauda, vai kombinācija *ſs* nav lietota pozīcijā, kur šodien rakstāms viens *s*. Šādā gadījumā veicama vienkāršošana, piemēram *puſs→pus*. Pēc tam var vienkāršot un konvertēt dubulto *ſſ→s*. Pašā nobeigumā konvertējami *ſ→s* un *ß→s*.

24. Pēc tam jāveic pārceltā teksta pārbaude, atsevišķu gadījumu pārveide. Papildu likumi katrā avotā ir atšķirīgi.

3. 18. gadsimts

18. gadsimtā atšķirībā no 16. un 17. gadsimta tekstiem rakstība ir konsekventāka, ir mazāk daudznozīmīgu burtu un burtkopu, mazāk izņēmumu. Tomēr, veidojot konvertācijas tabulas, tāpat kā iepriekšējo gadsimtu tabulās, ir jālieto aizstājējzīmes dubulto līdzskaņu (piemēram, *ll*, *pp*) u. c. gadījumos un jāveido izņēmumu likumi pirms dažu daudznozīmīgo burtu un burtkopu konvertācijas (piemēram, *ſ*, *ſch*, *ſ*). Turpmākajos punktos ir uzskaitīti galvenie konvertācijas likumu tipi. Protams, jāņem vērā, ka katrā avotā likumi var nedaudz atšķirties. Ir likumi, kuru secībai nav nozīmes, piemēram, ir vienalga, vai vispirms nomaina *ah→ā*, vai *eh→ē*, bet ir vairākas pārveides, kas ir jāveic stingri noteiktā secībā, piemēram, vispirms jāpārvērš *tſch→č* un tikai pēc tam *ſch→š*.

Šeit ir apkopoti 18. gadsimta pilotkonvertoram paredzētajā tabulā izmantotie likumi un saglabāta tajā ievērotā secība.

1. Vispirms visi garie patskaņi, kur *h* rāda pagarinājumu, jāpārvērš par mūsdienu garajiem patskaņiem (piemēram, *ah→ā*, *eh→ē*, *ëh→ē*, *ih→ī*, *Jh→ī*, *uh→ū*, *üh→ū*).

2. Jāpārvērš divskaņi (piemēram, *ee→ie*, *oh→o*).

3. Pēc tam jāpārvērš visi patskaņi ar diakritiskajām zīmēm (piemēram, *à→ā*, *â→ā*, *ã→ā*, *ä→ā*, *è→ē* vai *e*, *ê→ē*, *é→e*, *ë→e*, *î→ī*, *ï→ī*, *ô→o*, *ö→o*, *û→ū*, *ü→i*, *ī* vai *ū*).

4. Jāpārveido *j→i* dažādās kombinācijās (piemēram, *Jk→ik*, *Jr→ir*, *Js→iz*, *Jt→it*).

5. Dubultie līdzskaņi jānomaina pret vienu līdzskani visur, kur mūsdienās dubulto nav (par izņēmumiem skat. 6. punktā) (piemēram, *bb→b*, *dd→d*, *gg→g*, *ģģ→ģ*, *ķķ→ķ*, *ll→l*, *ļļ→ļ*, *mm→m*, *nn→n*, *ņn→ņ*, *ņņ→ņ*, *pp→p*, *rr→r*, *rŗ→ŗ*, *ŗŗ→ŗ*, *ſſ→ſ*, *tt→t*, *w→v*, *zz→c*).

6. Mūsdienu latviešu valodā ir vairākas leksēmas, kuru saknē vai arī morfēmu sadurā tomēr ir dubultais līdzskanis. Lai šajos izņēmumos dubultais līdzskanis nepareizi netiktu pārvērsts par vienu līdzskani, konvertējot ir jāizmanto aizstājējzīme. Piemēram, *ll* nevar automātiski nomainīt pret *l*, vispirms jāpārbauda, vai nav tādi vārdi, kā *elle*, *vells*, *mells*, *allaž*. Mainot tos, jāizmanto aizstājējzīme, piemērm, *%*. Tā kā, mainot *ell*, var pārmainīties arī citi vārdi, kuros ir šis burtu savienojums, tad jānorāda, ka *ell* ir vārda sākumā. To norāda ar *[]ell*: *[]ell→e%*. Kad visi gadījumi, kuros jāsaglabā *ll*, pārvērsti par *%*, var nomainīt *ll→l*. Pēc tam *%→ll*.

Dubultie līdzskaņi, kas jāsaglabā mūsdienu rakstībā, var būt, piemēram, šādās leksēmās vai to daļās: *ikkatr*, *ikkur*, *ikkuŗ*, *allaſch*, *allaſch*, *Bulli*, *della*, *[]ell*, *kelle*, *krell*, *lalla*, *mella*, *mellu*, *papill*, *pill*, *rull*, *ſtall*, *ſtell*, *Stella*, *Svamm*, *trall*, *well*, *willa*, *villane*, *willu*, *Ann*, *kann*, *winn*, *appufchko*, *nerr*.

7. Kad *zz→c*, var veikt maiņu *z→c* (*cc→c*).

8. Līdzskaņi ar tildi jānomaina pret parastiem līdzskaņiem (piemēram, *m̃→m*, *ñ→n*).

9. Līdzskanis savienojumā ar sekojošu *h* jānomaina pret līdzskani (*mh→m*, *gh→g*).

10. Divskaņa un līdzskaņa *j* savienojums jāpārvērš par patskaņa un līdzskaņa *j* savienojumu (piemēram, *aij→aj*, *eij→ej*). Pirms tam ar aizstājējzīmēm jānomaina, piemēram, *paij→p%j*, *maij→m%j*. Pēc tam, kad veiktas iekavās dotās pārveides, *%→ai*.

11. Jānomaina *iņg→iņ*. Pirms tam ar aizstājējzīmi jānomaina, piemēram, *vingr→v%r*. Kad nomainīts *iņg→iņ*, tad *%→iņg*.

12. Grafēma *ſ* ir daudznozīmīga, tādēļ vispirms jāpārmaina visas divu burtu kombinācijas, kurās grafēmai *ſ* mūsdienās atbilst līdzskanis *s* (piemēram, *[]ſa→sa*, *ſk→sk*, *ſl→sl*, *ſm→sm*, *ſp→sp*, *ſſ→s*, *ſs→s*, *ſf→s*, *ſt→st*). Jāatceras: ja *ſ* ir viens pats vai divu grafēmu savienojumā, noteikti jāatzīmē reģistrjutība 1.

13. Pēc tam jānomaina veselas leksēmas vai to daļas, kurās grafēmai *ſ* mūsdienās atbilst līdzskanis *s* (piemēram, *dvēſ→dvēs*, *eſſi→esi*, *Eſaj→esaj*, *klauſ→klaus*, *paſaul→pasaul*, *ſac→sac*, *ſakr→sakr*, *ſarg→sarg*, *ſav→sav*, *ſest→sest*, *ſlav→slav*, *ſlēp→slēp*, *taiſn→taisn*, *ſod→sod*, *ſirm→sirm*, *viſ→vis*). Ja grafēma *ſ* atrodas vārda sākumā, arī šajā gadījumā ieteicams atzīmēt reģistrjutību 1.

14. Jāveic maiņa: *tſch→č*.

15. Arī burtkopa *ſch* ir daudznozīmīga. Vispirms jānomaina lielākas burtu kombinācijas, leksēmas vai leksēmu daļas, kurās burtkopai *ſch* mūsdienās atbilst līdzskanis *š* (piemēram, *Iekſch→iekš*, *keniſchi→ķenīši*, *muſchi→muši*, *pestiſch→pestīš*, *Pīſch→pīš*, *ſchan→šan*, *ſchk→šķ*, *vīluſch→vīluš*, *Viņſch→viņš*). Ja grafēma *ſ* atrodas vārda sākumā, ieteicams atzīmēt reģistrjutību 1.

16. Jāveic maiņa: *ſch→ž*. Ja grafēma *ſ* atrodas vārda sākumā, ieteicams atzīmēt reģistrjutību 1.

17. Jāveic maiņa *ſ→z*. Jāatzīmē reģistrjutība 1.

18. Burtkopa *ſch* ir daudznozīmīga, tādēļ vispirms jānomaina lielākas burtu kombinācijas, leksēmas vai leksēmu daļas, kurās burtkopai *ſch* mūsdienās atbilst līdzskanis *ž* (piemēram, *allaſch→allaž*, *daſch→daž*, *daſch[]→dažs*, *grieſch→griež*, *ļauſch→ļauž*, *muiſch→muiž*, *mūſch→mūž*, *ſchēl→žēl*).

19. Jāveic vēl dažas maiņas ar *ſch*: *tſch→č*, *ſchki→šķi*.

20. Jāveic maiņa *ſch→š*.

21. Jāveic maiņa *dſ→dz*.

22. Jāveic maiņa *ſ→s*.

23. Jāveic maiņas *Schk→šķ*, *Sch→š*.

24. Jāveic maiņa *S→s*.

25. Jānomaina *s→z* vārda sākuma zilbē (piemēram, *[]ais→aiz*, *[]bes→bez*, *[]is→iz*, *[]mas→maz*, *[]neis→neiz*, *[]us→uz*, *[]ūs→ūz*).

26. Jāveic maiņas *Schk→šķ*, *Sch→ž*. Reģistrjutība 1, jo maiņa attiecas tikai uz lielo *S*.

27. Jāveic maiņa *S→z*. Reģistrjutība 1, jo attiecas uz lielo *S*. (Pirms tam gan ar aizstājējzīmi *Sk→%k*, *Sl→%l*, *Sm→%m*, *Sp→%p*, *St→%t*. Kad veikta maiņa *S→z*, tad *%→s*.

28. Pēc tam jāveic pārceltā teksta pārbaude, atsevišķu gadījumu pārveide. Papildu likumi katrā avotā ir atšķirīgi.

Avoti

Avotu saīsinājumus skat. https://senie.korpuss.lv/

Noderīgi resursi

Unicode 17.0 Character Code Charts. https://www.unicode.org/charts/

Projekta grupas publikācijas

Andronova, Everita; Frīdenberga, Anna; Pretkalniņa, Lauma; Siliņa-Piņķe, Renāte; Skrūzmane, Elga; Trumpa, Anta; Vanags, Pēteris. User-friendly Search Possibilities for Early Latvian Texts: Challenges Posed by Automatic Conversion. *DHNB 2022 Conference 'Digital Humanities in the Nordic and Baltic Countries': 6th Conference, Uppsala, Sweden, 15–18 March, 2022: Proceedings* / eds.: Karl Berglund, Matti La Mela, Inge Zwart (CEUR Workshop Proceedings, Vol. 3158), pp. 168–176. Pieejams: dhnb.eu/wp-content/uploads/2022/03/DHNB2022_book_of_abstracts.pdf [skatīts 03.11.2025.].

Andronova, Everita; Frīdenberga, Anna; Pretkalniņa, Lauma; Siliņa-Piņķe, Renāte; Skrūzmane Elga; Trumpa, Anta; Vanags, Pēteris. Latviešu valodas senāko rakstu pieminekļu konvertācija mūsdienu rakstībā: iepriekšējā pieredze un automatizācijas mēģinājumi = The Conversion of Early Written Latvian Texts into Modern Spelling: Previous Experience and Automatization Attempts. *Aktuālas problēmas literatūras un kultūras pētniecībā : rakstu krājums = Current Issues in Research of Literature and Culture : Scientific Journal* / atb. red. Anita Helviga; Liepājas Universitāte. Humanitāro un mākslas zinātņu fakultāte, Kurzemes Humanitārais institūts. LU Literatūras, folkloras un mākslas institūts. Lietuviešu literatūras un folkloras institūts, Vītauta Dižā universitāte. Liepāja: LiePA, 2022, 27. sēj., 346.–358. lpp. Pieejams: https://dom.lndb.lv/data/obj/1035006.html [skatīts 03.11.2025.].

Andronova, Everita; Frīdenberga, Anna; Pretkalniņa, Lauma; Siliņa-Piņķe, Renāte; Skrūzmane, Elga; Trumpa, Anta; Vanags, Pēteris. New possibilities for exploring early Latvian texts: switching to the *NoSketchEngine*. *Baltic Journal of Modern Computing*, Vol. 12, N 4 (2024), pp. 548–559. Pieejams: https://www.bjmc.lu.lv/fileadmin/user_upload/lu_portal/projekti/bjmc/Contents/12_4_16_Andronova.pdf [skatīts 03.11.2025.].

Andronova, Everita; Frīdenberga, Anna; Siliņa-Piņķe, Renāte; Skrūzmane, Elga; Trumpa, Anta; Vanags, Pēteris. Variantums kā konvertācijas izaicinājums: Georga Manceļa tekstu atveide mūsdienu rakstībā. *Letonica*, Nr. 47 (2022), Rīga: Latvijas Universitātes Literatūras, folkloras un mākslas institūts, 2022, 188.–207.lpp. Pieejams: lulfmi.lv/storage/files/letonica-47.pdf [skatīts 03.11.2025.].

Vanags, Pēteris; Frīdenberga, Anna; Trumpa, Anta. Georga Manceļa rakstības principi: darbības pirmais posms (līdz 1631. gadam). *Baltistica*, 58 (2) (2023), Vilnius: Vilniaus universitetas, 2023, 329.–353. lpp. Pieejams: https://www.baltistica.lt/index.php/baltistica/article/view/2529/2428 [skatīts 03.11.2025.].

Citas publikācijas

Bollmann, Marcel; Petran, Florian; Dipper, Stefanie (2011). Rule-based normalization of historical texts. *Proceedings of the Workshop on Language Technologies for Digital Humanities and Cultural Heritage*. Hissar, Bulgaria,

September, pp. 34–42. Pieejams: https://aclanthology.org/W11-41.pdf [skatīts 03.11.2025.].

Bollmann, Marcel; Petran, Florian; Dipper, Stefanie (2014). Applying Rule-Based Normalization to Different Types of Historical Texts — An Evaluation. *Human Language Technology Challenges for Computer Science and Linguistics Lecture Notes in Computer Science*. Springer International Publishing, pp. 166–177. Pieejams: https://marcel.bollmann.me/pub/ltc11.pdf [skatīts 03.11.2025.].

Bollmann, Marcel (2018). *Normalization of Historical Texts with Neural Network Models*. Bochum: Stefanie Dipper. Pieejams: https://www.linguistics.rub.de/forschung/arbeitsberichte/22.pdf [skatīts 03.11.2025.].

Domingo, Miquel; Casacuberta, Francisco (2018). Spelling Normalization of Historical Documents by Using a Machine Translation Approach. *Proceedings of the 21st Annual Conference of the European Association for Machine Translation.* Alacant, Spain, May 2018, pp. 129–137. Pieejams: https://aclanthology.org/2018.eamt-main.13.pdf [skatīts 03.11.2025.].

Oravecz, Csaba; Sass, Bálint; Simon, Eszter (2010). Semi-automatic normalization of Old Hungarian codices. *Proceedings of the ECAI Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*. Faculty of Science, University of Lisbon Lisbon, Portugal, August, pp. 55–59.

Paikens, Pēteris; Pretkalniņa, Lauma; Rituma, Laura (2024). A Computational Model of Latvian Morphology. *Proceedings of Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING)*, Turin, Italy, pp. 221–232. Pieejams: https://aclanthology.org/2024.lrec-main.20 [skatīts 03.11.2025.].

Paikens, Pēteris; Rituma, Laura; Pretkalniņa, Lauma (2013). Morphological analysis with limited resources: Latvian example. *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA)*. Oslo, Norway, pp. 267–278. Pieejams: https://aclanthology.org/W13-5624.pdf [skatīts 03.11.2025.].

Pettersson, Eva; Megyesi, Beáta; Nivre, Joakim (2012a). Parsing the Past – Identification of Verb Constructions in Historical Text. In *Proceedings of the 6th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pp. 65–74, Avignon, France. Association for Computational Linguistics. Pieejams: https://aclanthology.org/W12-1010.pdf [skatīts 03.11.2025.].

Pettersson, Eva; Megyesi, Beáta; Nivre, Joakim (2012b). Rule-based normalisation of historical text – a diachronic study. *Proceedings of KONVENS 2012* (LThist 2012 workshop), pp. 333–341. Pieejams: https://www.oegai.at/konvens2012/proceedings/50_pettersson12w/ [skatīts 03.11.2025.].

Pettersson, Eva; Megyesi, Beáta; Nivre, Joakim (2013). Normalisation of Historical Text Using Context-Sensitive Weighted Levenshtein Distance and Compound Splitting. *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013)*. Oslo, Norway, Sweden: Linköping University

Electronic Press, pp. 163–179. Pieejams: https://aclanthology.org/W13-5617/ [skatīts 03.11.2025.].

Piotrowski, Michael (2022). *Natural language processing for historical texts* (Revised ed.). Switzerland: Springer Nature. DOI: https://doi.org/10.1007/978-3-031-02146-6 [skatīts 03.11.2025.].

Porta, Jordi; Sancho, José-Luis; Gómez, Javier (2013). Edit transducers for spelling variation in Old Spanish. *Proceedings of the workshop on computational historical linguistics at NODALIDA 2013*. NEALT Proceedings Series 18 / Linköping Electronic Conference Proceedings 87, pp. 70–79. Pieejams: https://ep.liu.se/ecp/087/006/ecp1387006.pdf [skatīts 03.11.2025.].

Rayson, Paul; Dawn Archer; Smith Nicholas (2005). VARD versus Word – A comparison of the UCREL variant detector and modern spell checkers on English historical corpora. *Proceedings from the Corpus Linguistics Conference Series, online e-journal*, Vol. 1, Birmingham, UK, July. Pieejams: https://www.researchgate.net/publication/237394000_VARD_versus_Word [skatīts 03.11.2025.].

Sánchez-Marco; Cristina, Boleda; Gemma, Padró, Lluís (2011). Extending the tool, or how to an notate historical language varieties. *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*. Portland, OR, USA: Association for Computational Linguistics, pp. 1–9. Pieejams: https://aclanthology.org/W11-1501/ [skatīts 03.11.2025.].

Dokuments apspriests 2025. gada 5. novembrī.