

ЭЛЕКТРОННАЯ СІСТЭМА РАСПРАЦОЎКІ ЛАТЫШСКА-ЛАТГАЛЬСКАГА СЛОЎНІКА

Гісторыя

Нягледзячы на шматвяковую пісьмовую традыцыю і прызнанне яе аўтарытэтнымі і палітычна неангажыраванамі балтыстамі асобнай мовай (напрыклад, Уладзімір Мікалаевіч Топараў у гутарцы з аўтарам у кастрычніку 2005 года, за два месяцы да сваёй смерці, падкрэсліў, што лічыць латгальскую асобнай балтскай мовай), латгальская мова у сілу палітычных прычын у Латвіі лічыцца дыялектам, а з 2000 года афіцыйна вызначаны яе статус у арт. 3.4 Закону аб дзяржаўнай мове: «Дзяржава забяспечвае захаванне, абарону і развіццё латгальскай пісьмовай мовы як гістарычнага падвіда латышскай мовы.»¹ [1] На першы погляд артыкул глядзіцца добра, але, на жаль, гэта ўражанне зманліва: паняцце падвіда мовы нідзе не вызначана і не апісана, таму латгальская мова *de facto* ў Латвіі апыняецца «нічым»: ёй нішто не належыць (ні ў прававым, ні ў матэрыяльным плане), з дзяржаўнага бюджэту ніякія сродкі наогул не адводзяцца. Фактычна, мова рэгіянальнай карэннай народнасці знаходзіцца ў горшым становішчы чым мовы нацыянальных меншасцяў. Аналагічныя мовы ў іншых краінах Еўропы звычайна маюць статус рэгіянальнай мовы, які ў прававым полі ЕС надзяляе іх пэўнымі правамі, але дзяржаўныя органы Латвійскай рэспублікі пытанні аб прысваенні латгальскай мове статусу рэгіянальнай ходу не даюць.

Праўда, латгальскія гаворкі даследуюцца дыялекталагамі дзяржаўных навуковых устаноў нароўні з латышскімі, што, безумоўна, дае вялікі ўклад і ў развіццё латгальскай мовы (адным з найбольш яркіх прыкладаў з'яўляецца двухтомная слоўнік колупскай гаворкі [2]), але гэта ўклад, усё-такі не прамы, а ускосны.

У дадзенай сітуацыі адзінай рухаючай сілай у галіне развіцця латгальскай мовы з'яўляюцца прыватныя асобы і недзяржаўныя арганізацыі. Прычым, так як нават на конкурснай аснове ў розных дзяржаўных і недзяржаўных праграмах атрымаць сродкі пад праекты даследавання і развіцця латгальскай мовы практычна немагчыма, то ў асноўным гэта работа праводзіцца на асновах добраахвотнай працы.

З сярэдзіны 90-х гадоў складаннем і напісаннем (першыя гады – ў тэкставым файле) латышка-латгальскага слоўніка займаецца Армад Коцінь. У 1999 годзе Анс Атаол Берзінь стварыў першую версію электроннага слоўніка. Так як слоўнік быў аднаўзроўневым і толькі часткова структураваным, то даволі хутка стала ясна, што ён не задавальняе патрабаванням стварэння паўнаважнага пераводнага слоўніка, таму ў 2003 годзе Берзінем была распрацавана сістэма распрацоўкі цалкам структураванага пераводнага слоўніка, аб якой распавядзем больш падрабязна.

Канцэпцыя

На пачатковым этапе распрацоўкі слоўніка была пастаўлена галоўная задача: слоўнік павінен быць універсальным, прычым па розных прыкметах.

Па-першае, ён павінен быць і электронным, і «папяровым», г.зн., з адных і тых жа слоўнікавых дадзеных праграма павінна мець магчымасць згенераваць вывад артыкулу і ў выглядзе, прынятым у традыцыйных друкаваных слоўніках, і ў розных відах слоўнікаў электронных, у тым ліку інтэрактыўных. Па-другое, з яго павінны аўтаматычна вылучацца спецыялізаваныя слоўнікі, напрыклад, тэрміналагічны слоўнік пэўнай вобласці (г.зн., павінна падтрымлівацца фільтраванне). Па-трэцяе, так як у працэсе распрацоўкі слоўніка прапрацоўваецца ў тым ліку і матэрыял, які не патрэбен для пераводнага слоўніка, але ў той жа самы час з'яўляецца патрэбным і каштоўным для слоўнікаў іншага прызначэння (тлумачальнага, дыялектнага, сінонімаў), то было-б пажадана мець магчымасць ўводзіць і гэтую інфармацыю, каб у перспектыве мець магчымасць аўтаматычна вылучыць і слоўнікі вышэйзгаданых відаў.

Відавочна, што неабходнай умовай такой ўніверсальнасці вываду з'яўляецца поўная структураванне дадзеных, г.зн., дадзеныя артыкулаў павінны быць пабіты на драбнючкі семантычныя складнікі, і кожны складнік павінен быць адпаведна апісан. Так як артыкулы друкаваных слоўнікаў звычайна ўтрымліваюць больш падрабязную інфармацыю, то мы прынялі рашэнне распрацоўваць структуру на іх аснове. Намі быў праведзены падрабязны аналіз структуры артыкулаў больш за дзесятак розных перакладных слоўнікаў, выдадзеных у другой палове XX стагоддзя, і ў выніку была створана фармальна структура, якой можна было апісаць любы артыкул любога з прааналізаваных слоўнікаў.

Структура і тэхнічная рэалізацыя

Многія распрацоўшчыкі для захоўвання слоўнікавых артыкулаў аддаюць перавагу карыстацца XML фарматам, але наш выбар спыніўся на рэляцыйнай базе дадзеных, так як рэалізацыю нашай структуры ў выглядзе табліц мы палічылі больш нагляднай. Акрамя таго, у нас былі падставы меркаваць, што пры абранай наме сістэме пошуку і вываду база дадзеных SQL будзе працаваць хутчэй XML (хоць мы гэтага строга не правяралі).²

У дадзеным артыкуле мы не ставім задачу распісаць структуру цалкам, таму толькі апішам найбольш істотныя яе складнікі. Для апісання артыкулаў пераводнага слоўніка мы карыстаемся чатырма табліцамі: слоў, значэнняў,

¹ Фармулёўка «пісьмова мова» выкарыстоўваецца спецыяльна, каб не называць яе (паўнаважнай) мовай; прыдуманая латышскімі лінгвістамі другой паловы XX стагоддзя.

² У агульным выпадку немагчыма гаварыць аб тым, што хутчэй – XML або рэляцыйная БД, усё залежыць ад структуры і прымянення дадзеных. [3]

перакладаў і прыкладаў³. Кожны слоўнікавы артыкул апісваецца такой колькасцю запісаў табліцы слоў, колькі ў дадзенага слова⁴ амонімоў і амографоў. Кожнаму амоніму/амографу могуць падпарадкоўвацца неабмежаваная колькасць значэнняў. Кожнае значэнне альбо змяшчае спасылку-пераход на значэнне нейкага іншага слова, альбо ўтрымлівае неабмежаваную колькасць перакладаў і прыкладаў. Уражанне прастаты структуры, што робяцца пры такім, спрошчаным яе апісанні, зманліва, так як: 1) ва ўсіх табліцах вялікая колькасць розных і розна тыповых палёў, падрабязнае апісанне якіх заняло-б некалькі старонак; 2) структура трохузроўневая, што для слоўніка нямала – большасць электронных слоўнікаў наогул аднаўзроўневых. Увогуле-то, наша задача і была цалкам структуравана апісаць артыкула любога папяровага слоўніка пры найменшай магчымай колькасці узроўняў базы дадзеных, і мы пераканаліся, што менш трох іх не можа быць ніякім чынам. Фактычна, мы маглі вынесці амонімы/амографы ў асобны ўзровень пад словам, але, так як большасць слоў валодаюць толькі адным амонімам/амографам, а колькасць агульных палёў розных амонімаў/амографоў можна палічыць на пальцах адной рукі, пры чым усе яны - аўтаматычна ствараная тэхнічная інфармацыя (напрыклад, час ўводу), то гэта б толькі павялічыла гравасткасць структуры і выбару дадзеных.

Змесціва

Так як слоўнік цалкам структураваны і, у выніку таго – фільтруем, то перад намі не ўставала класічнае лексікаграфічнае пытанне абмежавання ўтрымання, г.зн., што ўключаць, што не ўключаць: наша заданне сабраць і ўвесці як мага больш інфармацыі, пры гэтым яе цалкам структурае і маркіруе, каб у выпадку неабходнасці мы маглі аўтаматызаванай выбіраць толькі патрэбную яе частка.

Так як складанне слоўніка вядзецца па прыватнай ініцыятыве і практычна не мае фінансавання, то і работа вядзецца нястала, з прычыны чаго словы дадаюцца не па пэўнай сістэме, а нерэгулярна, у залежнасці ад сабраных намі ці трапілі да нас моўных дадзеных (вусная размова, матэрыялы экспедыцый, фальклор, літаратура, прэса цяперашніх і мінулых гадоў і т.п.).

За 8 гадоў намі ў слоўнік ўведзены ўжо больш за 14 000 артыкулаў, якія паралельна рэдагуюцца і дапрацоўваюцца, г.зн., на дадзены момант не ўсе з іх скончаныя.

У 2010 годзе пры невялікі падтрымцы Дзяржаўнага Фонду культурнага капіталу Латвіі намі быў створаны публічны доступ да слоўніка па адрасе <http://vuordineica.lv/>.⁵ На дадзеным сайце карыстальнікі могуць не толькі праводзіць пошук па ўсяму слоўніку і чытаць яго артыкулы, але і іх каментавать, такім чынам перадаючы аўтарам дадатковую лінгвістычную інфармацыю, якая інакш, магчыма, не была-бы сабраная.⁶

Перспектывы развіцця

У будучыні, вядома, плануем працягваць дадаваць і дапрацоўваць слоўнікавыя артыкулы. Таксама хочам распрацаваць новую версію праграмага забеспячэння, палепшыўшы інтэрфейс распрацоўніка слоўніка і дадаўшы ў структуру новыя падначаленыя дадзеныя, напрыклад, дыялектныя формы аднаго і таго ж перакладу. Слоўнік, вядома, не з'яўляецца сіметрычным, г.зн., з латышка-латгальскага немагчыма аўтаматычна атрымаць паўнаватасны латгальска-латышскі, але ўсё ж у нейкай ступені гэта актуальнае пытанне⁷ можна вырашыць, напрыклад, шляхам дадання пошуку па перакладах.

Акрамя таго, мы зацікаўлены ў выкарыстанні нашай сістэмы распрацоўкі слоўнікаў для іншых моў, таму, калі ў вас паўстаў цікавасць, то смела выказвайце свае прапановы.

Літаратура

1. Valsts valodas likums. – Rīga: LR Saeima, 1999.
<http://www.likumi.lv/doc.php?id=14740>
2. Reķēna, A. Kalupes izloksnes vārdnīca, I-II sēj. – Rīga: LVI, 1998.
3. Nicola, M. XML versus Relational Database Performance. – San Jose, 2010.
<http://nativexmlatabase.com/2010/08/22/xml-versus-relational-database-performance/>

³ Вядома, база дадзеных утрымлівае і іншыя табліцы для захоўвання дадатковай змястоўнай (напрыклад, аб крыніцах) і службова-тэхнічнай інфармацыі. (Строга кажучы, службова-тэхнічную інфармацыю, якая адносіцца да дадзенага запісу, утрымоўваюць ўсе, у тым ліку і вышэйзгаданыя асноўныя табліцы.)

⁴ У дадзеным выпадку мы пад словам разумеем нейкую чараду, г.зн., спарадкаваную сукупнасць, графем.

⁵ Мы палічылі, што слоўнік, хоць і знаходзіцца ў стадыі распрацоўкі, усё ж ўжо ўтрымлівае дастаткова інфармацыі, каб быць людзям карысным, асабліва ўлічваючы адсутнасць ніякага слоўніка латгальскай мовы і ўзрастаючую ролю латгальскай мовы ў грамадстве.

⁶ Гэта асабліва важна, улічваючы вялікую ролю дыялектаў і ненармаванасць латгальскай мовы.

⁷ Яно сапраўды актуальна: многія карыстальнікі цікавяцца магчымасцю перакладу ў іншым накірунку.