

Як бачым, большасць варыянтных прыметнікаў у сучаснай беларускай мове ўзніклі ў выніку ўзаемадзеяння высокапрадуктыўных у сістэме прыметнікаў суфіксаў *-н-*, *-ск-* (*-цк-*), *-ов-* (*-ав-*), якія, далучаючыся да асноў назоўнікаў, выражаюць адноснасць спосабу маніфестацыі значэння.

Літаратура

1. Бережан С.Г. К проблеме лексической вариантности // Вариантность как свойство языковой системы. – М.: Наука, 1982. – Ч. I. – С. 36.
2. Климкова Л.А. Притяжательные прилагательные в современном белорусском и русском литературном языке: Автореф. дис. ...канд. филол. Наук. – Минск, 1987. - 24 с.
3. Пилипович Т.П. Словообразовательные варианты прилагательных в современном белорусском литературном языке: Автореф. дис. ...канд. филол. Наук. – Минск, 1987. - 22 с.
4. Шкраба І.Р. Варыянтнасць у сучаснай беларускай мове. – Мінск: Асар, 2004. - 240 с.

А.В. Гюнтар¹

СТВАРЭННЕ ЛІНГВІСТЫЧНЫХ БАЗ ВЕДАЎ ДЛЯ СІНТЭЗАТАРА БЕЛАРУСКАГА МАЎЛЕННЯ

Сінтэзатар маўлення па тэксце для беларускай мовы можа выкарыстоўвацца ў розных сферах і для вырашэння розных задач: у call-цэнтрах і аўтаінфармацыйных сістэмах, у разнастайных прыстасаваннях для нямых і людзей са слабым зрокам, пры вывучэнні замежных моў (слоўнікі і перакладчыкі на аснове тэхналогіі сінтэзу маўлення). Акрамя таго, выкарыстанне сінтэзатара беларускага маўлення можна прапанаваць выкладчыкам у адукацыйным працэсе любой дысцыпліны. Але для эфектыўнага выкарыстання сінтэзатара маўлення неабходна забяспечыць максімальную правільнасць яго работы.

На сённяшні дзень існуюць тры рэалізацыі сінтэзатара маўлення па тэксце для беларускай мовы. Гэта сістэма сінтэзу маўлення на славянскіх мовах “Мультифон” [2, ст. 180], сістэма сінтэзу маўлення для мабільных прыстасаванняў [3, ст. 83] і інтэрнэт-версія (<http://www.corpus.by/tts3>) [1, с. 375].

Стварэнне лінгвістычных баз дадзеных для фанемна-алафоннага працэсара сінтэзатара беларускага маўлення прызначана для праверкі работы стацыянарнага сінтэзатара “Мультифон” і інтэрнэт-версіі сінтэзатара, распрацаваных у лабараторыі распазнавання і сінтэзу маўлення Аб’яднанага інстытута праблем інфарматыкі НАН Беларусі (АІПІ НАН Беларусі).

Непасрэдна перад тым, як прааналізаваць работу стацыянарнага сінтэзатара “Мультифон” і інтэрнэт-сінтэзатара, неабходна сабраць матэрыял, на аснове якога будзе праводзіцца іх тэставанне. Для гэтага неабходна стварыць тэставыя мноствы слоў, якія будуць групувацца па тэмах. Праца па стварэнні такіх тэставых мностваў вялася на падставе адзіных табліц правілаў пераўтварэння “Фанема-Алафон” для рускай і беларускай моў, распрацаваных у лабараторыі распазнавання і сінтэзу маўлення Аб’яднанага інстытута праблем інфарматыкі НАН Беларусі.

¹ Магістрантка кафедры сучаснай беларускай мовы, навуковы кіраўнік дацэнт Г.К. Чахоўскі.

У адзіных табліцах правілаў пераўтварэння “Фанема-Алафон” для рускай і беларускай моў алафон пазначаецца імем фанемы і наступнымі за ім трыма лікавымі індэксамі, якія абазначаюць: для галосных – націскны, часткованаціскны ці ненаціскны гук, а таксама левае і правае асяроддзе; для зычных – адзіночны ці падвоены алафон, левае і правае асяроддзе. Алафоны, якія існуюць толькі ў рускай мове пазначаны **тлустым шрыфтам**, а алафоны, якія існуюць толькі ў беларускай мове пазначаны **тлустым курсівам**, астатнія – агульныя для двух моў. Улічваючы гэта, а таксама пэўныя асаблівасці мовы, дадзеныя табліцы для рускай мовы налічваюць 759 алафонаў, а для беларускай мовы – 761 алафон. Гэта колькасць не з’яўляецца дакладнай, паколькі былі ўлічаны розныя магчымыя пазіцыі для кожнага канкрэтнага алафона, за выключэннем тых, якія відавочна не існуюць для рускай ці беларускай мовы, але гэтыя дадзеныя не былі правераны на практыцы.

Для стварэння базы, якая б дапамагала ў пошуку памылак аналізуемых сінтэзатараў неабходна ўлічваць тое, што ў мове, а тым больш у жывым маўленні дастаткова часта сустракаюцца выключэнні з правілаў, розныя варыянты. Менавіта таму для стварэння тэставых мностваў трэба карыстацца ведамі, атрыманымі эмпірычным шляхам.

Для атрымання такіх ведаў неабходна знайсці прыклад на кожны алафон, пазначаны ў адзіных табліцах правілаў пераўтварэння “Фанема-Алафон”, гэта дазволіць стварыць лінгвістычную базу ведаў, якая забяспечыць правільную работу разглядаемых сінтэзатараў.

Апішам працэс збору матэрыялу для стварэння тэставых мностваў больш падрабязна.

Першым крокам у зборы інфармацыі для фарміравання тэставых мностваў быў экспертны падбор з арфаграфічнага слоўніка адносна адзіных табліц правілаў пераўтварэння “Фанема-Алафон” для рускай і беларускай моў. Гэта значыць, што на базе гэтых табліц праводзіўся адвольны пошук слоў на любыя алафоны па слоўніку і пры дапамозе праграмы VocEditor.

Праграма VocEditor значна палягчае працу лінгвіста ў пошуку слоў. Напрыклад, трэба знайсці слова з алафонам L000 (адзіночны зычны, перад якім стаіць любая фанема ці дапаможны сімвал, а пасля яго – паўза). Каб пачаць пошук у праграме, неабходна выбраць патрэбную мову (у нашым выпадку беларускую), у акне Find увесці сімвалы *л, дзе зорачка – любая колькасць любых літар і націснуць на кнопку Go! У выніку машына выдае спіс слоў, згодна з запытам (словы, якія заканчваюцца на л).

Гэта была першая, “ручная” стадыя збору інфармацыі, праведзеная для азнаямлення з асаблівасцямі работы над пастаўленай задачай. Для палягчэння запланаванай працы былі скарыстаны розныя інструменты, якія будуць апісаны ніжэй, здольныя па магчымасці аўтаматызаваць рашэнне задачы.

Наступныя дзеянні былі звязаны з апрацоўкай пачатковых напрацовак тэматычных мностваў. Першапачатковай базай для пакрыцця вызначаных алафонаў з’яўляліся спецыяльна распрацаваныя ў лабараторыі распазнавання і сінтэза маўлення АПП НАН Беларусі тэматыкі слоў (“прафесіі”, “ежа і напоі”,

“геаграфічныя назвы” і г.д.). Для пераўтварэння іх у алафонны выгляд быў скарыстаны сэрвіс “Text-to-speech synthesis”², у які на ўваход неабходна падаваць арфаграфічны тэкст з расстаўленымі асноўнымі і пабочнымі націскамі.

Наступны крок у падрыхтоўцы матэрыялу для пакрыцця наяўнага спіса алафонаў – іх пошук і сарціроўка. Дадзены працэс таксама аўтаматызаваны: “Service of Searching and Sorting Allophones/Сэрвіс пошуку і сарціроўкі алафонаў”¹ дазваляе атрымаць спіс алафонаў з прыкладамі слоў у алфавітным парадку і па колькасці сустракання ў зададзеным тэксце. Дадзены сэрвіс быў распрацаваны для вузка-прафесійнага кола карыстальнікаў, якія займаюцца ўдасканаленнем працы беларускамоўнага СМТ. Яго мэтай з’яўляецца складанне ўпарадкаваных (па алфавіце альбо па колькасці) спісаў алафонаў па тэксце, пададзеным на ўваход у алафонным выглядзе.

Знойдзеныя алафоны пазначаны колерам, што палягчае іх пошук у кантэксце (слове ці спалучэнні слоў). Таксама дадзены сэрвіс зручны тым, што пры апрацоўцы вялікіх электронных тэкстаў адразу відаць колькасць знойдзеных алафонаў (агульная колькасць і колькасць унікальных алафонаў).

Напрыклад, слова “Кра+каў”. Пасля апрацоўкі сэрвісам “Text-to-speech synthesis” атрымліваем алафонны выгляд:

K004,R022,A022,K004,A331,W010,/,#P4.

Далей уносім гэтае слова ў “Service of Searching and Sorting Allophones”. Пад кнопкай “Get allophones!/Знайсі алафоны!” размешчаны патрэбныя нам дадзеныя:

The total number of elements is 6.

Different elements are 5

Адсюль відаць, што агульная колькасць алафонаў – 6, пры гэтым унікальных пяць, паколькі алафон *K004* сустракаецца два разы.

Пасля таго, як сабраныя тэматыкі апрацавалі праз “Service of Searching and Sorting Allophones”¹ стала зразумела, што дадзенага матэрыялу недастаткова для поўнага пакрыцця алафонаў беларускай мовы. Таму з электроннай версіі арфаграфічнага слоўніка беларускай мовы былі ўзяты ўсе пачатковыя формы назоўнікаў і апрацаваны пры дапамозе таго ж сэрвіса. Па выніках апрацоўкі дадзенага матэрыяла пакрыццё алафонаў склала толькі 561 адзінка. Многія пазіцыі алафонаў сустракаюцца на стыку слоў, таму неабходна было ўлічваць не толькі адзінкавыя словы, але і розныя спалучэнні слоў.

Улічваючы вышэйсказанае для апрацоўкі быў узяты падручнік “Грамадазнаўства” за 10 клас, які папярэдне быў апрацаваны праграмай VocEditor. Гэта праграма дазволіла знайсці ўсе новыя словы для электроннага слоўніка сінтэзатара маўлення “Мультифон” і расстаноўкі адназначных націскаў у словах-амографах падручніка “Грамадазнаўства”. Дадзеная кніга пасля апрацоўкі праграмай склала 40 раздзелаў, якія адзін за адным паступова былі пратэставаны.

² Сэрвіс, створаны на аснове інтэрнэт-версіі сінтэзатара маўлення <http://www.corpus.by/tts3>.

Нягледзячы на тое, што была апрацавана вялікая колькасць матэрыялу, пакрыццё алафонаў складала толькі 589 адзінак.

На пляцоўцы www.corpus.by быў распрацаваны яшчэ адзін сэрвіс, тэставанне якога дапамагала ў фарміраванні тэставых мностваў для сінтэзатараў. Гэты сэрвіс атрымаў назву “Orthoepic Dictionary Generator” і пры яго стварэнні ставілася задача зрабіць такі сэрвіс, пры дапамозе якога было б магчыма атрыманне нарматыўнага вымаўлення слоў у графічным выглядзе (іх транскрыпцыі).

Для таго, каб мець магчымасць пастаяннага паляпшэння работы сэрвіса, вызначана “адваротная сувязь”, пры якой карыстальнік, які зафіксаваў памылкі, можа даслаць ліст распрацоўшчыкам на электронны адрас з апісаннем праблемы.

Паколькі пасля ўсёй праведзенай працы пакрыццё алафонаў не перавышала мяжу 604 адзінкі, пошук алафонаў, для якіх яшчэ не знойдзены адпаведнікі, праводзіўся “ўручную”.

Пры правядзенні пошуку “ўручную” браўся пэўны алафон са спісу, аналізавалася яго асяроддзе і пры дапамозе арфаграфічнага слоўніка, праграмы VocEditor, а часта і шляхам самастойнага падбора шукаліся неабходныя прыклады.

Пасля збору матэрыялу было неабходна яго апрацаваць і сістэматызаваць. Быў створаны лінгвістычны рэсурс – база ведаў (БВ), у якой пазначаўся арфаграфічны выгляд слова/спалучэння слоў, іх трансфармацыя ў алафонны выгляд стацыянарным і інтэрнэт-версіяй сінтэзатара маўлення. Пасля гэтага пазначаўся эталонны (правільны) варыянт слова і далей адбывалася яго суаднясенне з атрыманым выглядам пасля апрацоўкі сінтэзатарамі, пасля чаго прапісваліся заўвагі наконт знойдзеных памылак у такой трансфармацыі (малюнак 1).

A	B	C	D	E	F	G	H
Бегемот.	B'004,E343,GH'004,E241,M002,O012,T000,/#P4,	B'004,E343,GH'004,E241,M002,O012,T000,/#P4,	>B'004,E343,>,GH'004,E241,>,M002,O012,T000,>,>				
Бутсел.	B002,U013,S'002,E341,L000,/#P4,	B002,U013,S'002,E341,L000,/#P4,	>B002,U013,>,S'002,E341,L000,>,>				
Бу+йвал.	B002,U013,J'013,V002,A311,L000,/#P4,	B002,U013,J'013,V002,A311,L000,/#P4,	>B002,U013,J'013,V002,A311,L000,>,>				
Вавёрка.	V002,A213,V'011,O042,R003,K004,A330,/#P4,	V002,A213,V'011,O042,R003,K004,A330,/#P4,	>V002,A213,>,V'011,O042,R003,>,K004,A330,>,>				
Верабей.	V'002,E342,R002,A223,B'002,E043,J'010,/#P4,	V'002,E342,R002,A223,B'002,E043,J'010,/#P4,	>V'002,E342,R002,A223,>,B'002,E043,J'010,>,>				
Во+жык.	V001,O012,ZH004,Y322,K000,/#P4,	V001,O012,ZH004,Y322,K000,/#P4,	>V001,O012,>,ZH004,Y322,K000,>,>				
Во+ўк.	V001,O011,W013,K000,/#P4,	V001,O011,W013,K000,/#P4,	>V001,O011,W013,K000,>,>				
Газзаль.	GH004,A232,Z002,E023,L'000,/#P4,	GH004,A232,Z002,E023,L'000,/#P4,	>GH004,A232,>,Z002,E023,L'000,>,>				
Гарнастаў.	GH004,A332,R001,N004,A222,S002,T001,A023,J'01	GH004,A332,R001,N004,A222,S002,T001,A023,J'01	>GH004,A332,>,R001,N004,A222,>,S002,T001,A023,J'010,>,>				
Гепард.	GH'004,E241,P001,A012,R003,T000,/#P4,	GH'004,E241,P001,A012,R003,T000,/#P4,	>GH'004,E241,>,P001,A012,R003,T000,>,>				
Дзік.	DZ'002,I042,K000,/#P4,	DZ'002,I042,K000,/#P4,	>DZ'002,I042,K000,>,>				
Дзяцел.	DZ'002,A043,C'002,E341,L000,/#P4,	DZ'002,A043,C'002,E341,L000,/#P4,	>DZ'002,A043,>,C'002,E341,L000,>,>				
Дро+зд.	D001,R022,O022,S002,T000,/#P4,	D001,R022,O022,Z003,T000,/#P4,	>D001,R022,O022,S002,T000,>,>	U corpus.by няма асіміляцыі па глухасці			
Журавель.	ZH004,U322,R002,A223,V'011,E043,L'000,/#P4,	ZH004,U322,R002,A223,V'011,E043,L'000,/#P4,	>ZH004,U322,>,R002,A223,>,V'011,E043,L'000,>,>				
Затяц.	Z002,A023,J'012,A342,C000,/#P4,	Z002,A023,J'012,A342,C000,/#P4,	>Z002,A023,J'012,A342,C000,>,>				
Зе+бра.	Z'002,E041,B001,R012,A320,/#P4,	Z'002,E041,B001,R012,A320,/#P4,	>Z'002,E041,>,B001,R012,A320,>,>				
Зу+бр.	Z002,U021,B001,R010,/#P4,	Z002,U021,B001,R010,/#P4,	>Z002,U021,B001,R010,>,>				
Зяю+ля.	Z'004,A243,Z'002,U043,L'004,A340,/#P4,	Z'004,A243,Z'002,U043,L'004,A340,/#P4,	>Z'004,A243,>,Z'002,U043,L'004,A340,>,>				

Малюнак 1: Фрагмент базы ведаў «Арфаграфічны выгляд – алафонны адпаведнік»

Пасля таго, як была сфарміравана дадзеная БВ, сабраныя словы былі

прапушчаны праз “Service of Searching and Sorting Allophones” і паслядоўна атрыманы дадзеныя ўносіліся ў спецыяльны файл з поўным спісам алафонаў для беларускай мовы ў алфавітным парадку, дзе каля пэўнага алафона пазначаўся алафонны выгляд слова, пасля - арфаграфічны выгляд, а таксама тэматыка (малюнак 2).

	A	B	C	D	E	F	G
1	Алафон	Алафонны выгляд	Слова	Тэматыка			
2	A000						
3	A001						
4	A002						
5	A003						
6	A010	T002,U222,R001,M002,A010,/#P4	Турма+	Месцы і ўстановы			
7	A011	P002,L002,A011,W013,K'002,I340,/#P4,	Пла+ўкі.	Вопратка, яе элементы і абугак			
8	A012	S'002,P'002,A241,V011,A012,K000,/#P4,	Спява+к.	Прафесіі			
9	A013	P002,R012,A322,T002,P002,R012,Y221,M002,A013	Прадпрыма+льнік.	Прафесіі			
10	A020	M'004,A242,ZH002,A020,/#P4,	Мяжа+.	Месцы і ўстановы			
11	A021	K004,A332,S002,M004,A212,N002,A021,W013,T0	Касмана+ўт.	Прафесіі			
12	A022	R022,E221,A022,L000,/#P4,	Рэа+д.	Грошы			
13	A023	D004,Y222,Z002,A023,J'013,N'004,E342,R000,/#P	Дыза+йнер.	Прафесіі			
14	A030	R022,U222,K002,A030,/#P4,	Рука+.	Часткі цела			
15	A031	K002,A031,V012,A310,/#P4,	Ка+ва.	Ежа і напоі			
16	A032	M004,U312,Z004,Y222,K002,A032,N003,T000,/#P	Музыка+нт.	Прафесіі			
17	A033	GH002,A033,L'003,SH002,T002,U322,K000,/#P4,	Га+льштук	Вопратка, яе элементы і абугак			
18	A040	S'002,V'002,I243,N'102,A040,/#P4,	Свіння+.	Жывёлы			

Малюнак 2: Фрагмент базы ведаў для пакрыцця алафонаў беларускай мовы

У працэсе даследавання ўзніклі пэўныя складанасці. Напрыклад, дастаткова цяжка было знайсці прыклады на падвоеныя алафоны (за выключэннем гукаў [з'], [с'], [н'], [л'], [дз'], [ц'], [ж], [ш], [ч], якія ў інтэрвакальным становішчы падаўжаюцца). Для таго, каб сінтэзатар згенераваў, напрыклад, падвоены мяккі зычны [л'], пасля якога ідзе звонкі зычны (L'101), было знойдзена спалучэнне слоў, паміж якімі прыйшлося паставіць сімвал Ь для абазначэння злітнасці вымаўлення (паку+ль Ьлье+).

Пры падборы прыкладаў на кожны алафон было вырашана ўключыць і тыя прыклады, якія з'яўляюцца, так скажам, памылковымі, гэта значыць, што ў сучаснай беларускай мове не было знойдзена адпаведнага слова/словазлучэння, але тэарэтычна дадзены алафон у пэўным асяроддзі можа сустрацца (напрыклад, у назвах лекаў, запазычаных словах, прозвішчах і г.д.). Тым больш, што сінтэзатар генеруе не мову, а маўленне, і карыстальнік можа ўвесці абсалютна любы тэкст, які наўмысна ці выпадкова ўтрымлівае ў сабе памылкі. Таму ўключэнне такіх прыкладаў - неабходнасць для карэктнай работы сінтэзатара.

У выніку выкананай працы пакрыццё алафонаў складала 704 адзінкі, пры гэтым колькасць алафонаў для беларускай мовы прапануецца паменшыць, бо створаныя адзіныя табліцы правілаў пераўтварэння “Фанема-Алафон” улічваюць абедзве мовы (рускую і беларускую), а таму ў іх сустракаюцца пэўныя недакладнасці.

Такім чынам, стварэнне лінгвістычных баз ведаў для тэставання фанемна-

алафоннага працэсара сінтэзатара беларускага маўлення з’яўляецца важнай і разам з тым складанай задачай. Магчымасць аўтаматычнай апрацоўкі некаторых дадзеных дазваляе значна палегчыць працу лінгвіста па вырашэнні дадзенай праблемы.

Літаратура

1. Гецэвіч, Ю.С. Кампаненты для розных платформаў сінтэзатара маўлення па тэксце для інтэлектуальных сістэм / Ю.С. Гецэвіч, Д.А. Пакладок, Д.В. Брэк // Открытые семантические технологии проектирования интеллектуальных систем = OpenSemanticTechnologiesforIntelligentSystems (OSTIS–2013) : материалы III Междунар. науч.-техн. конф. (Минск, 21–23 февраля 2013 года) / редкол. : В.В. Голенков (отв. ред.) [и др.]. – Минск: БГУИР, 2013 г. – С. 375–382.)

2. Лобанов, Б.М. Компьютерный синтез и клонирование речи // Б.М. Лобанов, Л.И. Цирульник / Минск, Белорусская Наука, 2008.

3. Цирульник, Л.И. Система синтеза речи по тексту для мобильных телефонов / Л.И. Цирульник, Д.А. Покладок // Речевые технологии. – 2010. – № 1. – С. 83–91.

І.А. Купцэвіч³

“АБЫ НЕ БЫЛО ВАЙНЫ”: СВЕТАПОГЛЯД БЕЛАРУСАЎ І ЎКРАЌНЦАЎ, АДЛЮСТРАВАНЫ Ў ПАРЭМІЯХ

Успрыманне чалавекам навакольнай прыроднай і сацыяльнай рэчаіснасці, свайго месца ў свеце, стаўленне да сябе і іншых людзей, ацэнкі, маральныя нормы і каштоўнасныя арыенціры – усё гэта складае светапогляд чалавека. Сукупнасць людзей, аб’яднаных на пэўнай тэрыторыі адным дзяржаўным утварэннем ці яго аналагамі і агульнай мовай, культурай, гістарычным мінулым – гэта ўжо народ, ці нацыя (паняцці ў гэтым кантэксце сінанімічныя). І пра светапогляд можна меркаваць у дачыненні не толькі да чалавека, але і да нацыі. Пры гэтым светапогляд нацыі не будзе з’яўляцца сумай ці сярэднім арыфметычным светапоглядаў яе прадстаўнікоў (якія могуць істотна адрознівацца), а будзе з’яўляцца наборам нейкіх тыповых характарыстык, уласцівых большасці прадстаўнікоў нацыі, характарыстык, якія прышчапляюцца людзям з маленства ў працэсе несвядомага пераймання калектыўнага досведу папярэдніх пакаленняў, што адбываецца само сабой праз любыя кантакты з асяроддзем, калі чалавек выходзіць сярод прадстаўнікоў пэўнага народа. Называецца ж гэты агульнанародны светапогляд, як вядома, менталітэтам. Менталітэт абумоўлены звычайна гістарычным лёсам народа і вызначае нацыянальную культуру, а разам з ёй адлюстроўваецца ў нацыянальнай мове, фарміруючы моўную карціну свету. Натуральна, што моўныя карціны свету розных нацый, як і іх менталітэты, істотна адрозніваюцца, паколькі яны фарміраваліся ў розных гістарычных абставінах. Моўная карціна свету народа праяўляецца, напрыклад, у безэквівалентнай лексіцы, адметных канатацыях слоў, фразеалагізмах, стылістычным ладзе мовы, у міфалогіі і

³ Студэнтка 6 курса завочнай формы навучання спецыяльнасці «Беларуская філалогія», навуковы кіраўнік дацэнт Т.Л. Чахоўская.