

АЛГАРЫТМІЗАЦЫЯ АДБОРУ МАТЭРЫЯЛУ ДА СЛОЎНІКА СПАЛУЧАЛЬНАСЦІ ПРЫМЕТНІКА Ё БЕЛАРУСКАЙ МОВЕ

Аб'ектам беларускай лексікаграфіі мінулага стагоддзя як правіла становіліся ізаляваныя моўныя адзінкі. Такі падыход да мовы, калі перавага аддаецца не вывучэнню сувязей паміж элементамі сістэмы, але вывучэнню саміх элементаў, рэалізуецца ў большасці выданняў XX ст. (пяцітомны «Тлумачальны слоўнік беларускай мовы» пад агульнай рэд. К. К. Атраховіча; «Слоўнік іншамовных слоў» А. М. Булыкі; «Слоўнік новых слоў беларускай мовы» В. І. Уласевіча, Н. М. Даўгулевіча). У сучаснай лінгвістыцы ўсё большая ўвага надаецца даследаванню сістэмных адносін у мове. Гэта выражаецца, перад усім, у выхадзе з друку беларускіх ідэаграфічных слоўнікаў: «Асацыятыўны слоўнік беларускай мовы» А. І. Цітовай, «Славянскі асацыятыўны слоўнік: рускі, беларускі, балгарскі, украінскі» Н. В. Уфімцавай і інш.

Адной з перспектывных мэт нашай навуковай працы з'яўляецца стварэнне беларускага слоўніка спалучальнасці прыметніка. Слоўнік спалучальнасці, незаменны ў працэсе як выкладання, так і вывучэння пэўнай мовы як замежнай, да сёння не складзены ў айчынай лінгвістыцы. Зразумела, што комплекснае лексікаграфічнае апісанне моўных элементаў – надзвычай складаная задача нават для калектыву аўтараў. Даследаванне, у якім разглядаюцца спалучальнасныя ўласцівасці беларускага прыметніка, у сваю чаргу, магло б стаць першым крокам на шляху да вырашэння гэтай праблемы. Варта таксама адзначыць, што выбар сінтаксічных канструкцый з прыметнікам ў якасці прадмета вывучэння абумоўлены, у тым ліку, недастатковай распрацаванасцю такіх спалучэнняў у параўнанні, напрыклад, з камбінацыямі дзеяслова з рознымі часціцамі мовы. Як піша Н. Юдзіна, «гэта, верагодна, выклікана большай зручнасцю апісання сінтагматычных магчымасцей дзеяслова са структуральных пазіцый» [2, с. 3-4].

Як вядома, прафесійныя лексікаграфы пры адборы моўнага матэрыялу кіруюцца не ўласнай лінгвістычнай інтуіцыяй, а, перш за ўсё, дакладным алгарытмам, пэўным наборам крытэрыяў. Суадносна, намі былі вылучаны тры асноўныя этапы працы над артыкулам будучага слоўніка спалучальнасці: а) пошук кантэкстных ужыванняў прыметнікаў асобнай тэматычнай групы ў мастацкай літаратуры; б) параўнанне адабраных камбінацый з ілюстрацыйнымі прыкладамі слоўнікавых артыкулаў ўжо існуючых лексікаграфічных выданняў (у выпадку супадзення з ілюстрацыйным матэрыялам канструкцыя ўключаецца ў картатэку слоўніка спалучальнасці); в) адбор найбольш частотных словаспалучэнняў з ліку тых, якія не былі зафіксаваны акадэмічнымі слоўнікамі (такія канструкцыі таксама папаўняюць картатэку).

Для праверкі карэктнасці працы складзенага алгарытма на пачатковым этапе даследавання намі былі ўзяты прыметнікі ядра адной з базавых тэматычных груп беларускай мовы – шэсць прыметнікаў смаку (*салодкі, кіслы, горкі, салёны, востры, прэсны*). Акрамя таго, сярод асноўных беларускіх лінгвістычных слоўнікаў былі выбраны тыя, якія ў большай ці меншай ступені адлюстроўваюць спалучальнасныя ўласцівасці прыметніка як граматычнага класа, а менавіта:

тлумачальныя, перакладныя, сінанімічныя, антанімічныя і нек. інш. Было выяўлена, што найбольш поўна валентнасныя магчымасці атрыбутава раскрываюцца ў найважнейшым філалагічным даведніку беларускай літаратурнай мовы 20-га ст. – у «Тлумачальным слоўніку беларускай мовы», дзе, акрамя дэфініцыі, для кожнай лексемы абавязкова падаецца **слоўны выраз**, прыклад ужывання (**смачныя яблыкі** ‘прыемны на смак’/ **сон** ‘які дае задавальненне; прыемны’ / **выраз** ‘сакавіты, каларытны’).

Другой асноўнай крыніцай рэлевантных звестак аб натуральным ужыванні атрыбутыўных канструкцый (акрамя слоўнікавага артыкула) з’яўляецца мастацкі тэкст. Кіруючыся распрацаваным алгарытмам для пошуку кантэкстных ужыванняў прыметнікаў (у дадзеным выпадку прыметнікаў смаку) намі былі прааналізаваны тэксты паэм Я. Коласа («Новая зямля», «Сымон-музыка») і Я. Купалы («Адплата кахання», «Бандароўна», «Гарыслава», «За што», «Зімою», «Калека», «Курган», «Магіла льва», «На Дзяды», «На Куццю», «Нікому», «У Піліпаўку», «Яна і Я» і інш.). Менавіта ўласнай творчай працай класікі айчынай літаратуры заклалі падмурак для далейшага фарміравання пісьмовага ўнармаванага беларускага слова, паказалі яго патэнцыял у стварэнні высокамастацкіх тэкстаў. Аналіз паказаў, што ў паэмах выкарыстоўваецца: 5 прыметнікаў з ліку Коласавых твораў і 3 прыметнікі з ліку твораў Купалавых. Акрамя таго, сярод адабраных кантэкстных ужыванняў было выяўлена 12 канструкцый, не адлюстраваных у беларускай лексікаграфіі, гэта значыць адзінак, якія належаць маўленню. Найбольш частотнымі з іх ў паэмах сталі словаспалучэнні *салодкі сон і горкая доля*.

Паколькі заяўлены намі алгарытм прадугледжвае працу з вялікай колькасцю мастацкіх тэкстаў, матэрыялы, атрыманыя ў працэсе аналізу твораў Я. Коласа і Я. Купалы, неабходна параўнаць з канструкцыямі, выкарыстанымі іншымі аўтарамі. Корпус тэкстаў для супастаўлення, такім чынам, склалі эталонныя ўзоры беларускай прозы ХХ ст. Прыцягненне атрыбутыўных словаспалучэнняў з твораў В. Быкава, І. Мележа, І. Шамякіна, Н. Гілевіча, Р. Барадуліна, У. Караткевіча дазваляе максімальна раскрыць спалучальнасны патэнцыял лексем разглядаемай групы. Пошук кантэкстных ужыванняў, напрыклад, для прыметніка **горкі** паказвае, што яго спалучэнне з назоўнікам **сляза** з’яўляецца тыповым: такі варыянт сустракаем як у Я. Коласа (*Непрыветна сустрачала [поле] / Тых, хто скрыўджанай душой / Смутак нёс свой, сваё гора / У свет гэты, гэту даль / Каб рассяць на прасторы / Слёзы горкія і жаль...*; *Страх і жаласць агарнулі / Плача горкімі слязьмі*// *Людзі гэты плач пачулі / З лесу едучы з дрыўмі...* «Сымон-музыка»), так і ў мастацкіх тэкстах І. Шамякіна (*Каб яна пачала спачуваць, я, мабыць, расплакаўся б, бо слёзы былі блізка, комам стаялі ў горле, горкія, салёныя.* («Трывожнае шчасце»)), В. Быкава (*І ён [начальнік], абліваючыся горкімі слязьмі ў душы, кінуў яе на малую Волечку, якая цэлы дзень то плакала, то тоненькім дзіцячым галаском няяла маці ўлюбёную ёю «Вутачку» («Аблава»)).* Акрамя дадзенай канструкцыі, прыметнік **горкі**, як высветлілася, узаемадзеінічае з 44-ма іншымі назоўнікамі: *горкае дзіця* (≈ 9 % ад агульнай колькасці), *горкія слёзы* (≈ 7 %), *горкі вопыт* (≈ 4 %), *горкія яблыкі* (≈ 4 %) і інш. Як бачна, зыходным словаспалучэннем **горкая сляза** валентнасныя магчымасці атрыбутава не

вычэрпваюцца. У перспектыве, пры складанні слоўнікавага артыкула для адпаведнага прыметніка тэматычнай групы смаку дастаткова будзе толькі адабраць найбольш частотныя варыянты словаўжывання для кожнага з яго значэнняў, пазначаных у тлумачальным слоўніку.

Распрацаваны намі алгарытм, у выніку, дазволіў правесці аб'ектыўны аналіз спалучальных уласцівасцей прыметнікаў асобнай тэматычнай групы, ажыццявіць сістэмны падыход да выбару лінгвістычнага матэрыялу (былі задзейнічаны элементы мовы/маўлення), а таксама рэгламентаваў адбор атрыбутыўных канструкцый у рэестр слоўніка спалучальнасці.

Літаратура

1. Карповские научные чтения. Вып. 6 : сб. науч. ст. : в 2 ч. / редкол.: А. И. Головня (отв. ред.) [и др.]. – Мінск., 2012. – Ч. 2.
2. Тлумачальны слоўнік беларускай мовы: у 5 т. / рэдкал.: К. Атраховіч (гал. рэд.) [і інш.]. – Мінск, 1977.
3. Юдина, Н. В. Лексическая сочетаемость в когнитивном аспекте (на материале конструкции «прилагательное + существительное») : автореф. дис. д-ра фил. наук : 28.02.2006 / Н. В. Юдина. – М., 2006. – 40 с.

М.І. Краскоўскі

ПРАБЛЕМЫ МАШЫННАГА ПЕРАКЛАДУ БЕЛАРУСКІХ ДЗЕЯСЛОВАЎ ПЕРЫФЕРЫ ЛСГ УСПРЫМАННЯ НА РУСКУЮ МОВУ (НА МАТЭРЫЯЛЕ МАСТАЦКІХ ТЭКСТАЎ Я. КОЛАСА)

З ліку задач, якія вырашаюцца пры сістэмна-функцыянальным вывучэнні дзеяслоўнай лексікі рускай і беларускай моў, можна вылучыць як асноўныя, звязаныя з супастаўляльным аналізам унутранай арганізацыі і ўзаемадзеяння розных лексіка-семантычных груп, так і факультатыўныя, сярод якіх асобнае месца займае распрацоўка праблемы машыннага перакладу фраз з дзеясловамі *перыферыі* базавых ЛСГ. Некаторыя аспекты дадзенага пытання ўжо разглядаліся намі раней, калі на матэрыяле названых адзінак паралельных груп рускіх і беларускіх дзеясловаў успрымання былі апісаны асаблівасці працы з «анлайн-перакладчыкам Google»¹.

У апошнія дзесяцігоддзе ў айчынным мовазнаўстве значна павялічылася колькасць даследаванняў, прысвечаных *статыстычным*² машынным перакладчыкам. Трэба адзначыць, што найчасцей увагу лінгвістаў прыцягвае менавіта вэб-сэрвіс кампаніі Google. Па сукупнасці характарыстык прадукт сапраўды не мае аналагаў: па-першае, гэта адзіны на сённяшні дзень статыстычны машынны перакладчык высокай якасці, які падтрымлівае *беларускую мову*; па-другое, гэта адзіны анлайн-перакладчык такога маштабу, які функцыянуе ў глабальным секце ў *вольным* доступе. Калі, напрыклад, замежны студэнт актыўна

¹Гл.: Карповские научные чтения. Вып. 6 : сб. науч. ст. : в 2 ч. / редкол.: А. И. Головня (отв. ред.) [и др.]. – Мн., 2012. – Ч. 2.

²Пра адрозненні *традыцыйных* машынных перакладчыкаў, выкарыстоўваючых тэхналогію, заснаваную на ўжыванні алгарытмаў (Rule-based machine translation), ад *статыстычных*, у аснову якіх пакладзена тэхналогія вылічэння верагоднасці тэкставых супадзенняў (Statistical-based machine translation), таксама пісалася ў папярэдніх даследаваннях.