

Л. РЫЧКОВА

ПОЎНАТЭКСТАВЫЯ БАЗЫ ДАДЗЕННЫХ НА МАТЭРЫЯЛЕ ДРАМАТУРГІЧНЫХ ТВОРАЎ ЯНКІ КУПАЛЫ¹

Тэндэнцыя да камп'ютэрызацыі лінгвістычных даследаванняў, якую разглядаюць як агульную для сучаснага мовазнаўства², пакуль яшчэ мала закрунула айчыннае мовазнаўства, і тэрміны *камп'ютэрная* ці *корпусная лінгвістыка* ўсё яшчэ не ўспрымаюцца як тыя, што маюць дачыненне да даследавання беларускай мовы.

Экспертным шляхам вызначана³, што пераход да аўтаматызаваных спосабаў збору і апрацоўкі першаснай моўнай інфармацыі павысіць эфектыўнасць работы кожнага даследчыка, скараціць тэрміны і кошт распрацовак у два разы, пашырыць зыходную базу лінгвістычных даследаванняў у некалькі разоў. Аснова аўтаматызацыі лінгвістычных даследаванняў — стварэнне камп'ютэрных масіваў спецыяльным шляхам падрыхтаваных тэкстаў (камп'ютэрных корпусаў) і на іх матэрыяле — тэкставых лінгвістычных баз дадзеных, якія разглядаюцца як «сукупнасць інфармацыі, размешчанай так, каб спрасціць увод і пошук»⁴. Методыка фарміравання баз дадзеных увогуле і тэкставых у прыватнасці знаходзіцца ў стадыі станаўлення. Відавочна, што аб'ектыўна яна павінна базіравацца на патэнцыяльных напрамках аналізу, якія склаліся ў межах лінгвістычных прадстаўленняў аб тэксце.

Тэрмін *лінгвістыка тэксту*, які шырока ўжываецца ў сучасным мовазнаўстве, адлюстроўвае статус тэксту як асобнага аб'екта лінгвістычнага аналізу побач з традыцыйнымі *сказам/выказваннем, словам/словаўжываннем, марфемай/морфам*. Атрымаўшы статус самастойнага лінгвістычнага аб'екта, тэкст па-ранейшаму застаецца «моўным матэрыялам» (тэрміналогія Л. У. Шчэрбы), які служыць мэтам даследавання іншых аб'ектаў (любых лінейных адзінак, з якіх складаецца тэкст).

Тэкст таксама застаецца аб'ектам даследавання традыцыйнай тэксталагіі, накіраванай на вывучэнне моўных сродкаў, якія адлюстроўваюць кампазіцыйна-стылёвыя асаблівасці жанру, творчую манеру пісьменніка, стыль аўтара.

¹ У артыкуле выкарыстаны вынікі, атрыманыя ў межах выканання навукова-даследчых тэм № 28-96 і № 30-99, якія фінансаваліся Міністэрствам адукацыі РБ, а таксама тэма № Г98-385, якая фінансуецца Беларускай рэспубліканскай Фондам фундаментальных даследаванняў.

² Бондарко Л. В. Фонетический фонд языков России // Возрождение культуры России: Язык и этнос. С.-Петербург, 1995. С. 29.

³ Использование ЭВМ в лингвистических исследованиях // Г. А. Грязнухина и др. Киев, 1990. С. 12.

⁴ Kennedy G. An Introduction to Corpus Linguistics — N. Y., 1998. — P. 57; пераклад наш — Л. Р.

«Шматграннасць» і поліаспектнасць тэксту як аб'екта і матэрыяла лінгвістычнага аналізу ўскладняюць апрыёрнае вылучэнне патэнцыяльных тыпаў аб'ектаў, непазбежнае на этапе праектавання базы дадзеных, прызначанай для мэтаў шматаспектнага лінгвістычнага аналізу пэўнага тыпу тэкстаў. Відавочна, што пры вывучэнні спецыфічных рыс індывідуальна-аўтарскага стылю тэксту неабходна ўключаць у апрыёрны набор параметраў аналізу першасныя параметры, значэнні якіх выяўляюцца непасрэдна ў тэксце/тэкстах, які/якія падлягаюць аналізу.

Пры пабудове баз дадзеных, прызначаных для лінгвістычнага аналізу, неабходна ўлічваць як «рэальныя параметры», так і «характарыстыкі баз дадзеных, якія будуць неабходны для пастаноўкі задач на ЭВМ у бліжэйшы перыяд часу»⁵. Адсюль — важнасць праблемы дадатковага лінгвістычнага анатавання зыходных камп'ютэрных масіваў (корпусаў) тэкстаў, што набліжае «прыкметавую» натуральную мову да «аб'ектнай» мовы, зразумелай камп'ютэру⁶.

Выбар у якасці першакрыніц для пабудовы поўнатэкставых баз дадзеных драматургічных твораў не выпадковы. Як правіла, пры стварэнні камп'ютэрных корпусаў мастацкіх тэкстаў абіраюць тыя творы, дзе роля і памеры дыялагічнай мовы абмежаваны, таму прынцыпы і метадыка фарміравання базы дадзеных на матэрыяле драматургічных твораў патрабуюць спецыяльнага даследавання і з'яўляюцца новымі не толькі для беларускай прыкладной лінгвістыкі. Ускладняюць пастаўленую задачу і таму робяць яе больш цікавай у навуковым плане адносная складанасць тэксту драматургічнага твора і яго спецыфічная структураванасць.

Вядома, што тэксты, якія належаць да аднаго жанру («функцыянальнага тыпу»), «звычайна падобныя ў моўным афармленні і ў некаторых асаблівасцях сэнсавай і фармальнай структуры»⁷. Асноўнай асаблівасцю драматургічных твораў можна лічыць іх спецыфічную кампазіцыю, якая дазваляе дакладна адрозніваць мову аўтара ад мовы персанажаў і вылучаць мову кожнага персанажа. Гэта, безумоўна, дае дадатковыя магчымасці для правядзення тэксталагічнага аналізу і супастаўляльных лінгвістычных даследаванняў.

Вывучэнне літаратуры, прысвечанай аналізу драматургічных твораў, дазволіла вызначыць пэўныя напрамкі аналізу тэкстаў гэтага жанру, сярод якіх можна назваць:

1) аналіз моўных сродкаў, якія служаць для характарыстыкі персанажаў:

- «моўныя характарыстыкі персанажаў»⁸;
- «індывідуалізацыя» і «тыпізацыя» мовы персанажаў⁹;

⁵ Использование ЭВМ в лингвистических исследованиях // Г. А. Грязнухина и др. Киев, 1990. С. 16.

⁶ Естественный язык, искусственные языки и информационные процессы в современном обществе // Р. Г. Котов и др. М., 1988. С. 157.

⁷ Ванников Ю. В. Типы научных и технических текстов и их лингвистические особенности. Методическое пособие. Ч. 1. М., 1984. С. 4.

⁸ Щербина А. А. О речевой характеристике сатирических персонажей русской советской комедии (некоторые специфические средства): Автореф. дисс. ... канд. филол. наук. Киев, 1958. 18с. і інш.

⁹ Новиков Ю. С. К вопросу об индивидуализации речи персонажей в реалистической драме // Проблемы реализма и художественной правды. Львов, 1961. Вып. 1. С. 150—160; Долгих А. И. К вопросу об индивидуализации и типизации речи персонажей // Русский язык в школе. 1959. № 2. С. 46—51.

- 2) аналіз «арганізацыі» аўтарскіх рэмарак¹⁰;
- 3) аналіз асаблівасцей дыялагічнай мовы або «драматургічнага дыялога»¹¹;
- 4) аналіз сродкаў дасягнення «камічнага эфекту»¹², «сродкаў гумару і сатыры»¹³;
- 5) аналіз ролі, якую выконваюць у драматургічных творах уласныя імёны¹⁴;
- 6) аналіз разнастайных сродкаў, уласцівых мове пэўнага драматурга, такіх, як гіперсінтаксічныя — «паўтараемыя моўныя сродкі»¹⁵; сінтаксічныя — «асаблівасці сінтаксісу»¹⁶, «словы-сказы»¹⁷, «устаўныя канструкцыі»¹⁸ і іншыя; лексіка-сінтаксічныя — роля выклічнікаў¹⁹, «функцыянальная роля азначэнняў»²⁰; лексіка-фразеалагічныя — «фальклорная фразеалогія»²¹, лексіка-фразеалагічная сінаніміка»²², «кантэкстуальная

¹⁰ Винник Н. Ф. Диалог и функции авторских ремарок в пьесах «Дачники», «Дети солнца», «Вирвары» М. Горького. Днепропетровск, 1958. 27.; Асанов К. К. Некоторые особенности языковой организации авторских ремарок в драматургическом тексте в связи с фактором адреса / Моск. гос. лингв. ун-т. М., 1991. 17 с. Рук. деп. в ИНИОН № 44724 от 10.06.91.

¹¹ «Борисова М. Б. О строении диалога в драматургии М. Горького // Учен. зап. Са-рат. ун-та. 1957а. Т. 56. С. 456—473 і інш

¹² Михлина М. Л. о некоторых языковых приемах создания комического эффекта: (На материале советской драматургии (30—40 гг.)) // Учен. зап. Душанб. пед. ин-та. 1962. Т. 31. Вып. 14. Филол. сер. С. 3—14; Мурзина М. Н. Языковые средства комического в советских комедиях 20-х — 30-х годов // Тез. докл. и сообщ. на итоговой науч. конф. Луганского пед. ин-та за 1964 г. Сер. филол. наук. Луганск, 1965. С. 53—57.

¹³ Удовиченко Г. М. Языковые средства юмора в украинской советской комедии: Автореф. дисс. ... канд. филол. наук. Киев, 1953. 14 с.; Касьянов А. В. О речевых средствах юмора и сатиры в комедии Н. В. Гоголя «Ревизор» // Учен. зап. Армавирск. пед. ин-та. 1962. Т. 4. Вып. 2. С. 3—19.

¹⁴ Волкова Л. П. Литературная ономастика и ее особенности в драматургии первой трети XIX столетия (на материале пьес И. А. Крылова и А. С. Грибоедова). // Питання стилістики української мови в її взаємозв'язку з іншими слов'янськими мовами. Тез. доп. міжвузівської наук. конф. Чернівці, 1963. С. 89—91; Шейнина Е. П. Семантика собственных имен в пьесах А. Н. Островского // Вопросы лексической и фразеологической семантики. Ростов н/Д., 1979. 126 с.; Иезуитов С. А. Имена и фамилии в пьесах Горького «Егор Булычев и другие», «Достигаев и другие», «Васса Железнова» // Горьковские чтения. Горький, 1990. С. 53—58; Ryczkova L. Имена персонажей драматургических произведений янки Купалы как центры культурно-тематических полей текста // Studia Slawistyczne i Nazewnictwo na pograniczach etniczno-jezykowych: Materiały Międzynarodowej Konferencji Onomastycznej Białystok-Suprasl 26—27 X 1999/ Uniwersytet w Białymstoku Instytut Filologii Wschodniosłowiańskiej. С. 250—256.

¹⁵ Кубарев Е. М. Композиционные функции повторяющихся речевых средств в комедии Н. В. Гоголя «Ревизор»: (Сопоставительно-текстологический анализ) // Науч. тр. Куйбышев. пед. ин-та. 1979. Т. 213. С. 18—27.

¹⁶ Басенко К. В. Некоторые особенности синтаксиса пьес А. Н. Островского // Учен. зап. Ровен. пед. ин-та. 1963. Т. 8. С. 49—87.

¹⁷ Аронова Ф. А. Грамматико-стилистическая характеристика слов-предложений в драмах А. М. Горького // Науч. зап. Харьк. пед. ин-та. 1957. Т. 28. С. 93—118 і інш.

¹⁸ Аникин А. Ветастные конструкции в «Русском лесе» // Большой мир. Статьи о творчестве Л. Леонова. М., 1972. С. 451—463.

¹⁹ Ломов А. Г. Структурно-грамматические особенности междометий: (На материале пьес А. Н. Островского). // Исследования по русскому и славянскому языкознанию. Самарканд. 1980. С. 65—77.

²⁰ Кузнецова Н. И., Цепова С. П. Функциональная роль определений в драмах М. Горького // Исследование языка художественных произведений. Куйбышев, 1975. С. 68—72.

²¹ Ломов А. Г. Фольклорная фразеология в пьесах А. Н. Островского // Сб. науч. тр. Самарканд. гос. ун-та. 1980. № 15. С. 166—176.

²² Митрофанов Г. Ф. Лексико-фразеологическая синонимика пьесы М. Горького «Последние» // Учен. зап. Томск. ун-та. 1963. № 45. С. 127—142.

этымалагізацыя фразеалагізмаў²³; уласна лексічныя — запазычаная лексіка²⁴; асаблівасці антонімаў²⁵.

Мадэль поўнатэкставай базы дадзеных, прызначанай для аўтаматызацыі лінгвістычнага аналізу драматургічных твораў, павінна ўлічваць не толькі ўзроўні ўласна тэксту і жанру, але і ўзровень мовы пісьменніка, індывідуальна-аўтарскія асаблівасці канкрэтнага тэксту.

Тэкст, «створаны канкрэтным аўтарам у канкрэтны час, у канкрэтнай сітуацыі»²⁶, разглядаецца як аб'ект стылеметрыі, у той час як вывучэннем індывідуальна-аўтарскіх асаблівасцей займаецца дыягностыка. Як стылеметрыя, так і дыягностыка выкарыстоўваюць лінгвістычныя метадыкі аналізу, што таксама было ўлічана пры распрацоўцы мадэлі лінгвістычнай базы дадзеных і пры стварэнні самой базы.

Аналіз навуковай літаратуры, прывешчанай даследаванням мовы Янкі Купалы, дазваляе вылучыць шэраг параметраў, якія характарызуюць як манеру пісьменніка²⁷, так і індывідуальна-аўтарскія асаблівасці канкрэтных твораў²⁸. Асаблівую ўвагу даследчыкаў выклікалі такія з'явы, як фальклорныя «традыцыі» або «матывы»²⁹ ў творчасці Янкі Купалы, шматлікія лексічныя асаблівасці яго твораў³⁰. У меншай ступені даследаваны асаблівасці семантыка-сіntaxічнай структуры драматургічных твораў Янкі Купалы³¹.

Вылучэнне патэнцыяльных напрамкаў аналізу пэўнага корпусу тэкстаў, як ужо адзначалася раней, з'яўляецца неабходнай умовай фарміравання канкрэтнай лінгвістычнай базы дадзеных, прызначанай для шматаспектнага аналізу такога корпусу. Па-першае, гэта неабходна для распрацоўкі спецыяльнага праграмнага забеспячэння; а па-другое — для фарміравання правілаў уводу тэксту, якія не павінны ствараць перашкод і «не павінны патрабаваць грувацкіх праграмісцкіх рашэнняў»³².

²³ Ломов А. Г. Контекстуальная этымалогія фразеологізмаў: (На матэрыяле пьес А. Н. Островскага) // Сб. науч. тр. Самарканд. гос. ун-та. 1979. № 12. С. 17—26.

²⁴ Грабчыкаў С. М. З назіранняў над русізмамі ў творах В. Дуніна-Марцінкевіча // Весн. Бел. дзярж. ун-та. 1971. № 3. С. 42—48 і інш.

²⁵ Куритдинов А. Стилистические особенности антонимов в пьесах А. Н. Островского // Межвуз. сб. науч. тр. Ташкент. ун-та. 1979. № 596. С. 27—31.

²⁶ Мартыненко Г. Я., Чебанов С. В. Стилеметрия // Прикладное языкознание. СПб, 1996. С. 422; пераклад наш — Л. Р.

²⁷ Гульман Р. І. Тэксталагія твораў Янкі Купалы: Пытанні вызначэння аўтэнтчнага тэксту. Мн.: Навука і тэхніка, 1971. 174 с. і інш.

²⁸ Жыркевіч М. Аб мове камедыі Я. Купалы «Паўлінка» // Даслед. па беларус. і рус. мовах. Мн., 1958. С. 69—80; Лепешаў І. Мова «Тутэйшых» Янкі Купалы // Польшыя. Мн., 1998. С. 241—248 і інш.

²⁹ Ленсу Е. Я. Фольклорныя традыцыі в дооктябрьском творчестве Янки Купалы. Мн.: Выш. школа, 1965. 122 с.; Макарэвіч А. Фальклорныя матывы ў драматургіі Янкі Купалы. Мн.: Навука і тэхніка, 1969. 99 с.

³⁰ Лепешаў І. Фразеалагізмы ў камедыі Я. Купалы «Паўлінка» // Купалава і Коласова слова. Мн., 1981. С. 15—30; Пальцаў Г. В. Лексічныя элементы польскага паходжання ў мове «Паўлінкі» Я. Купалы // Творчая спадчына Я. Купалы і Я. Коласа і развіццё славянскіх моў і літаратур. Мн., 1982. С. 168 і інш.

³¹ Бурак Л. І. Функцыянальныя аналагі сказаў у драматургічных творах Янкі Купалы // Весн. Беларус. ун-та. Сер. 4. Філалогія, журналістыка, педагогіка, псіхалогія. 1982. № 1. С. 39—44.

³² Аверина С. А., Алексеева Е. Л., Герд А. С. Автоматизация обработки древних текстов // Прикладное языкознание. СПб., 1996. С. 510; пераклад наш — Л. Р.

У залежнасці ад сістэмы лінгвістычнага анатавання, якая ўжываецца ў пэўнай базе дадзеных, прынята вылучаць агульнамэтавыя тэкставыя базы дадзеных і базы дадзеных спецыяльнага прызначэння³³. Для пабудовы агульнамэтавых баз дадзеных выкарыстоўваюць сістэму лінгвістычнага анатавання, дастатковую для вырашэння задач структурызацыі тэкстаў і размежавання нятоесных словаўжыванняў. Аднак ужо гэты этап праектавання лінгвістычнай базы дадзеных дазваляе рэалізаваць такія напрамкі фармальнага аналізу, як пошук часцін мовы (з улікам словаформаў і асаблівасцей кантэксту), пошук лексічных і праграматычных амонімаў (у тым ліку ў кантэксце), пошук тоесных словаформаў у кантэксте сказаў; вылучэнне часткі тэксту па любых фармальных прыкметах; вылучэнне сказаў па пэўнай прыкмеце або іх камбінацыі і інш.

Істотна пашырае спектр патэнцыяльных напрамкаў аналізу ўзбагачэнне сістэмы лінгвістычнага анатавання з мэтай эксплікацыі выражэння аналізу. Працэс апрыёрнага выяўлення параметраў аналізу, і асабліва іх прыкмет, дастаткова складаны. Па-першае, сярод прыкмет могуць быць «шумавыя»³⁴, адлюстраванне якіх у базе дадзеных можа прывесці да зніжэння дакладнасці аналізу; па-другое, павелічэнне сістэмы дадатковых памет у корпусе тэкстаў вядзе да «засмечвання» базы дадзеных і ўскладняе камп'ютэрную апрацоўку. Да таго ж, паколькі лінгвістычнае анатаванне корпусу тэкстаў пры стварэнні базы дадзеных ажыццяўляецца, як правіла, уручную, неабходнасць унясення ў тэкст мноства памет непазбежна вядзе да памылак уводу і патрабуе вельмі цяжкай і дарагой працэдуры праверкі.

Калі параметры аналізу разглядаць з пункту гледжання аналізіруемых аб'ектаў, а прыкметамі параметраў, якія патрабуюць фармальнага выражэння ў тэкставай базе дадзеных, лічыць прыкметы гэтых аб'ектаў, то такі корпус тэкстаў можа разглядацца як інфармацыйна-пошукавы масіў, дзе можна весці пошук любых тэкставых адзінак — аб'ектаў, якія валодаюць пэўнымі значэннямі прыкмет, набораў прыкмет і іх камбінацый.

З улікам усяго вышэйсказанага, на матэрыяле аднаго камп'ютэрнага корпусу тэкстаў драматургічных твораў Янкі Купалы замест адной базы дадзеных быў створаны комплекс ўзаемазвязаных лінгвістычных баз як інфармацыйна-пошукавых падмасіваў. Цэнтрам комплексу з'яўляецца агульнамэтавая поўнатэкставая база дадзеных, на падставе якой фармулююцца і праз якую аб'ядноўваюцца ў адзіны інфармацыйна-пошукавы масіў поўнатэкставыя базы дадзеных спецыяльнага прызначэння. Мадэль комплексу баз дадзеных з'яўляецца адкрытай і дазваляе ўключаць у комплекс дадатковыя базы дадзеных.

Сістэма лінгвістычнага анатавання, прынятая для пабудовы агульнамэтавай базы дадзеных, дазваляе:

— у дастатковай ступені адназначна і экспліцытна структураваць корпус тэкстаў і адпаведна вылучаць тэкст асобнага твору, тэкст аўтарскіх рэ-

³³ Фрэнсіс У. Н. Проблемы формирования и машинного представления большого корпуса текстов // Новое в зарубежной лингвистике. Вып. XIV. Проблемы и методы лексикографии. М.: Прогресс, 1983. С. 334—352.

³⁴ Марусенко М. А. Атрибуция анонимных и псевдонимных текстов методами прикладной лингвистики // Прикладное языкознание. СПб, 1996. С. 473.

марак, тэкст рэплік кожнага з персанажаў (як ва ўсім творы, так і ў межах акта ці з'явы), сказ/выказванне, словаўжыванне;

- вылучаць часціны мовы;
- адлюстроўваць аманімію;
- ідэнтыфікаваць уласныя імёны ў пачатку сказа;
- вылучаць аналітычныя граматычныя формы (з улікам магчымасці лінейнага «разрыву» гэтых форм).

Першая са створаных баз дадзеных спецыяльнага прызначэння мадэлюе лексічныя асаблівасці драматургічных твораў Янкі Купалы, бо прынятая пры яе пабудове сістэма дадатковага лінгвістычнага анатавання дазваляе асобна вылучаць наступныя аб'екты аналізу:

- неаднаслоўныя намінацыі (не толькі з улікам магчымасці іх «разрыву», але і з азначэннем моўных сродкаў, якія могуць выступаць у такой функцыі);
- фразеалагізмы, прыказкі, прымаўкі, выслоў'і, параўнанні (таксама з улікам магчымасці іх «разрыву» і з захаваннем «унутранай» індэксацыі аб'ектаў);
- ідэнтыфікаваць розныя падвіды лексічных адзінак: аказіяналізмы любой структуры, варварызмы, макаранізмы, дыялектызмы.

Другую базу дадзеных спецыяльнага прызначэння можна лічыць «граматычнай», бо яна мадэлюе сістэму граматычных катэгорый беларускай мовы³⁵. У якасці асноўнага аб'екта мадэлявання выступае агульная парадыгма катэгарыяльнага ўзроўню, якая ў большасці выпадкаў адпавядае пэўнай часціне мовы. У межах граматычнай катэгорыі такога ўзроўню паслядоўна вылучаюцца і задаюцца ў выглядзе набору прыкмет і іх пэўных значэнняў такія традыцыйныя для лінгвістычнага апісання сістэм словазмянення адзінкі, як прыватная і канкрэтная парадыгмы³⁶, парадыгматычны рад, граматычная форма слова, маркер словаформы і іншыя.

Трэцяя база дадзеных спецыяльнага прызначэння плануецца як сінтаксічная, дзе асноўным аб'ектам аналізу будзе з'яўляцца сінтаксічная структура сказа.

Апрацоўка дадзеных у базах, выбар і выдача тых лінгвістычных аб'ектаў, якія характарызуюцца патрэбным наборам прыкмет, ажыццяўляецца аўтаматычна на падставе тыповых і індывідуальных запытаў. Форма выдачы дадзеных можа прадугледжваць кантэкст рознага памеру, упарадкаванне аб'ектаў па падабенстве прыкмет і абумоўлівацца далейшымі мэтамі лінгвістычнага аналізу.

Поўнатэкставыя базы дадзеных можна разглядаць як спецыфічны матэрыял для лінгвістычных даследаванняў³⁷, падрыхтаваны такім чынам, каб, захаваўшы ўсе якасці зыходнага тэксту, скараціць час і палегчыць працу лінгвістаў пры адборы з тэксту адзінак аналізу. У адказ на свой запыт лінгвіст атрымлівае ўпарадкаваную «картатэку», гатовую для правядзення ўласнага аналізу. Асабліва важна, што пры апісанні неад-

³⁵ Лінгвістычнае анатаванне «лексічнай» базы дадзеных і аб'ектна-прыкметавое мадэляванне граматычных аб'ектаў праведзена з непасрэдным удзелам супрацоўнікаў Інстытута мовазнаўства НАН РБ Лукашанца А. А., Кунцэвіч Л. П., Нікалаевы В. М., Кандрасені І. У.

³⁶ Откупщикова М. И. Общая парадигматика в морфологии. СПб, 1995. 88 с.

³⁷ Рычкова Л. В. Полнотекстовые базы данных как материал для лингвистических исследований // Е. Карский и современное языкознание. Материалы шестых научных чтений. Ч. 1. Гродно, 1996. С. 227–230.

наслоўных адзінак аналізу можна атрымаць усе выпадкі іх узаемадзеяння ў тэксце, якія не абмяжоўваюцца традыцыйнай для «ручнага» аналізу спалучаль-насьцю адзінак, а ўключаюць таксама звесткі аб накладанні, уключэнні ці перасячэнні аб'ектаў аналізу.

Структура комплексу поўнатэкставых баз дадзеных на матэрыяле драматургічных твораў Янкі Купалы дазваляе таксама праводзіць аўтаматычны многамерны лінгвістычны аналіз.

Гродна

Summary

A complex of full-text linguistic data-bases built on the same corpora of texts of Yanka Kupala drama works devoted to multiparametric linguistic analysis is described. Objects analysis and their parameters in the texts have been ordered in the all-purpose and several special data-bases. The parametric modeling of data-bases was made regarding the existing trends in investigations of other drama texts and those by Yanka Kupala.