

## РАСПРАЦОЎКА ПРАГРАМНЫХ ІНСТРУМЕНТАЎ ДЛЯ РАШЭННЯ ЗАДАЧ ЛІНГВІСТЫКІ

Краўцова В.С.

Беларускі дзяржаўны ўніверсітэт інфарматыкі і радыёэлектронікі  
г. Мінск, Рэспубліка Беларусь

Пятрова Н.Я. – к. філал. н., дацэнт

Разглядаюцца асаблівасці выкарыстання сучасных інфармацыйных тэхналогій у лінгвістыцы. Мэтай работы з'яўляецца правядзенне аналітычнага агляду выкарыстання інфармацыйных тэхналогій і рэсурсаў інфарматызацыі лінгвістычных ведаў. Практычная значнасць аналітычнага агляду заключаецца ў тым, што разгледжаныя інфармацыйныя тэхналогіі і рэсурсы інтэрнэт могуць быць выкарыстаны для паскарэння і палягчэння навуковых даследаванняў лінгвістаў.

Сёння выкарыстанне інфармацыйных тэхналогій (ІТ) значна павялічвае эфектыўнасць функцыянавання любой сферы чалавечай дзейнасці. Лінгвістыка не з'яўляецца выключэннем. Якімі б далёкімі галінамі, на першы погляд, ні здаваліся ІТ і лінгвістыка, неаспрэчная сувязь паміж імі пацвярджаецца хоць бы фактам існавання камп'ютарнай лінгвістыкі. Больш за тое, калі ўлічыць, што галоўным прадметам даследавання ў лінгвістыцы з'яўляюцца тэксты і што менавіта сучасныя ІТ адкрываюць новыя магчымасці для апрацоўкі і аналізу тэкстаў і прадастаўляюць разнастайныя сродкі стварэння, распаўсюджвання, пошуку і ўліку тэкставай інфармацыі, сувязь і ўзаемадзеянне гэтых дзвюх сфер становяцца яшчэ больш відавочнымі. Сказанае вышэй тлумачыць актуальнасць тэмы дадзенай працы [1].

Лінгвістыка – гэта навука аб чалавечай мове наогул і аб усіх мовах свету як яго індывідуалізаваных прадстаўніках, якая даследуе сутнасць і прыроду мовы, праблему яе паходжання і агульныя законы яе развіцця і функцыянавання.

Камп'ютарная лінгвістыка – напрамак ў прыкладной лінгвістыцы, арыентаваны на выкарыстанне камп'ютарных інструментаў: праграм, камп'ютарных тэхналогій арганізацыі і апрацоўкі дадзеных для мадэлявання функцыянавання мовы ў тых ці іншых умовах, сітуацыях, праблемных сферах і г. д., а таксама ўсёй сферы прымянення камп'ютарных мадэляў мовы ў лінгвістыцы і сумежных дысцыплінах [2].

Прыкладная лінгвістыка – кірунак у мовазнаўстве, які займаецца распрацоўкай метадаў рашэння практычных задач, звязаных з выкарыстаннем мовы. Разглядаючы вызначэнне паняцця «інфармацыйныя тэхналогіі» па адносінах да лінгвістыкі, можна сказаць, што тэхналогіі ў лінгвістыцы – гэта сукупнасць законаў, метадаў і сродкаў атрымання, захоўвання, перадачы, распаўсюджвання інфармацыі аб мове і законах яе функцыянавання з дапамогай камп'ютараў. Калі суаднесці гэта вызначэнне з тымі задачамі, якія вырашае сучасная прыкладная лінгвістыка, то можна адзначыць, што паняцце "інфармацыйныя тэхналогіі" ў лінгвістыцы суадносіцца ў асноўным з задачамі прыкладной лінгвістыкі. Да іх ліку можна аднесці:

- стварэнне сістэм штучнага інтэлекту;
- стварэнне сістэм аўтаматычнага перакладу;
- стварэнне сістэм спараджэння тэкстаў;
- стварэнне сістэм навучання мове;
- стварэнне сістэм разумення вуснага маўлення;
- стварэнне аўтаматызаваных інфармацыйна-пошукавых сістэм;
- распрацоўка розных баз дадзеных (слоўнікаў, картак, каталогаў, рэестраў і т. п.) для гуманітарных навук;
- распрацоўка рознага тыпу аўтаматычных слоўнікаў;
- распрацоўка сістэм перадачы інфармацыі ў сетцы інтэрнэт і г. д. [3].

Цяпер звернемся да праграм, звязаных з аналізам тэкстаў і вылічальнай лінгвістыкай, а таксама адпаведных рэсурсаў, даступных сёння ў глабальнай сетцы інтэрнэт.

Варта адзначыць, што ў сучасных умовах дэтальнае вывучэнне якой-небудзь навуковай тэмы пачынаецца з азнаямлення з вялікай колькасцю літаратурных крыніц, недаступных у друкаваным выглядзе. Адзіным выйсцем у дадзеным выпадку з'яўляецца выкарыстанне міжнароднай сеткі інтэрнэт і пошук у ёй матэрыялаў па кірунках даследавання.

Як наступства пошуку навуковых матэрыялаў у сетцы інтэрнэт (значэнне якой у сучасных навуковых даследаваннях немагчыма недаацаніць) узнікае неабходнасць выкарыстання тэкставых рэдактараў – Microsoft Word, WordPerfect, WordPro, праграм Acrobat Reader (пры чытанні і стварэнні электронных тэкставых дакументаў), ABBYY FineReader (пры іх распазнанні падчас сканавання) і шэрагу іншых.

Важна адзначыць і неабходнасць валодання навыкамі працы ў такой праграме стварэння прэзентацый для візуалізацыі дакладаў на канферэнцыях, як Microsoft PowerPoint [4].

Для аўтаматычнай карэктуры тэкстаў існуе праграма, якая носіць назву AfterScan. AfterScan можна выкарыстаць для карэкціроўкі тэксту, уведзенага ўручную, тэксту пасля распазнавання, для

апрацоўкі тэкстаў на старажытнай мове, а таксама тэкстаў з інтэрнэта. Для працы ў гэтых рэжымах створаны спецыяльныя версіі праграмы: Webmaster – для тэкстаў з інтэрнэта і Antique – для старажытных тэкстаў.

AfterScan на дадзены момант з'яўляецца адзінай праграмай, якая можа аўтаматычна выпраўляць памылкі, якія ўзнікаюць пры сканаванні і распазнаванні тэксту [5].

Таксама можна ўзгадаць праграмы кантэнт-аналізу, якія грунтуюцца на статыстычных метадах аналізу тэкстаў. Напрыклад, узгадаем сярод іх наступныя сродкі:

– "ТАСТ" – адна з самых "старажытных" праграм, якая працуе яшчэ пад сістэмай DOS. Яна была створана ў адным з універсітэтаў Канады, русіфікавана айчыннымі навукоўцамі і выкарыстоўваецца досыць часта гісторыкамі пры вывучэнні спецыфічных тэкстаў. Нягледзячы на "сур'ёзны ўзрост", мае досыць шмат функцый, аднак часам з-за праблем русіфікавання дае збоі ў працы;

– "VAAL" ("ВААЛ") – адна з самых вядомых айчынных праграм кантэнт-аналізу. Гэта экспертная сістэма лексічнага і кантэнт-аналізу тэкстаў; прагназуе эффект неўсвядомленага ўздзеяння тэксту на масавую аўдыторыю, аналізуе тэксты з пункту гледжання такога ўздзеяння, выяўляе асобна-псіхалагічныя якасці аўтара тэксту.

Сярод праграм для аўтаматычнай апрацоўкі і пераўтварэння тэкстаў можна вылучыць наступныя:

– Textar – рэвалюцыйная праграма для візуалізацыі і даследавання тэкстаў, якая з'яўляецца сапраўдным творам мастацтва. Тэкст прайграецца на экране камп'ютара ў выглядзе галактыкі, у якой словы атрымліваюць ролю зорак. Словы, якія часцей сустракаюцца, свецяцца ярка, а рэдкія – зусім не бачны. Радкі тэксту выводзяцца ў выглядзе закручанай спіралі па мяжы "галактыкі" кропкавым шрыфтам, вышынёй у адзін піксаль. Праграма дазваляе бачыць структуру тэксту, узаемасувязі паміж словамі і кантэкстам, праглядаць частотны і алфавітны індэкс слоў. Пры руху курсора па галактыцы словы ўспыхваюць і загарваюцца прамянямі (паказальнікамі на кантэкст), а таксама гучаць у рознай танальнасці;

– TextTransformer v 1.4.1 – магутная праграма, прызначаная для выканання калі заўгодна складаных пераўтварэнняў над мноствам тэкставых файлаў. Дазваляе аналізаваць, змяняць, ствараць і выдаляць зададзеныя элементы тэкстаў, якія апрацоўваюцца;

– LeoBilingua – праграма, якая дазваляе генераваць білінгва-тэкст, ці тэкст з двух сінхронных частак на розных мовах.

Сярод шматлікіх слоўнікаў і тэзаўрусаў асабліваю цікавасць для аўтара ўяўляюць (у сувязі з даследаваннем англамоўных мастацтвазнаўчых тэкстаў) British National Corpus (калекцыя больш за 100 мільёнаў слоў сучаснай англійскай пісьмовай і вуснай мовы), Acronym Finder (поўны слоўнік акронімаў і скарачэнняў англійскай мовы, змяшчае больш за 173 тысячы элементаў), WordNet (электронны тэзаўрус, які адлюстроўвае ўсе магчымыя тлумачэнні слоў англійскай мовы і які паказвае ўзаемасувязі паміж імі; для кожнага слова можна знайсці сінонімы, аб'яднаныя ў сэнсавыя групы і атрымаць спасылкі на родавыя або вытворныя паняцці), тлумачальны слоўнік Merriam Webster (on-line версія знакамітага Тлумачальнага слоўніка англійскай мовы) і інш.

Звяртаюць на сябе праграмы сінтэзу і распазнавання прамовы, напрыклад: Sacrament Text-to-Speech Engine v2. 0 (сістэма сінтэзу рускай мовы, створаная мінскай кампаніяй "Сакрамент"; дазваляе прамаўляць тэксты мужчынскім і жаночым голасам, робячы пры гэтым інтанацыйныя паўзы, змяняючы па жаданні карыстальніка тон і тэмбр маўлення); Govorka (невялікая праграма чытання тэксту голасам для Windows; наладжваецца на розныя мовы і галасы; выкарыстоўвае рэдагуемыя слоўнікі вымаўленняў; дазваляе запісаць сінтэзаваную гаворку ў MP3-файл); CSLU Toolkit (набор інструментальных сродкаў для стварэння дадаткаў, якія апрацоўваюць маўленчую і гукавую інфармацыю; ажыццяўляе сінтэз і распазнаванне прамовы, агучванне тэкстаў, апрацоўку спектраграм і гукавых сігналаў) і іншыя [6].

Такім чынам, мы пераканаліся ў існаванні велізарнага мноства разнастайных праграм і рэсурсаў, якія могуць значна палегчыць працу з лінгвістычным матэрыялам. Лінгвісту застаецца толькі выбраць прыдатныя для даследавання праграмы і выкарыстаць іх у сваёй прафесійнай дзейнасці.

#### **Спіс выкарыстаных крыніц:**

1. Артеменок, Г. А. *Современные направления в лингвистике: методические рекомендации* / Г. А. Артеменок. – Витебск: ВГУ имени П.М. Машерова, 2017. – 51 с.
2. Зубов, А. В. *Информационные технологии в лингвистике: учеб. пособие для студ. лингв. фак-тов высш. уч. учреждений* / А. В. Зубов, И. И. Зубова. – М.: Академия, 2004. – 208 с.
3. Потапова, К. Р. *Новые информационные технологии и лингвистика: учебное пособие* / К. Р. Потапова – М.: КомКнига, 2005. – 368 с.
4. Всеволодова, А. В. *Компьютерная обработка лингвистических данных: учеб. самоучитель* / А. В. Всеволодова – Ярославль: Литература, 2005. – 96 с.
5. Логичев, С. В. *Каталог лингвистических программ и ресурсов в сети* / С. В. Логичев // Русская виртуальная библиотека [Электронный ресурс]. – Режим доступа: <http://www.rvb.ru/soft/catalogue/index.html>. – Дата доступа: 27.11.2019.
6. *Компьютерная лингвистика* // Онлайн-энциклопедия "Кругосвет" [Электронный ресурс]. – Режим доступа: [http://www.krugosvet.ru/enc/gumanitarnye\\_nauki/lingvistika/KOMPYUTERNAYA\\_LINGVISTIKA.html](http://www.krugosvet.ru/enc/gumanitarnye_nauki/lingvistika/KOMPYUTERNAYA_LINGVISTIKA.html). – Дата доступа: 27.11.2019.