использовать иностранный язык для установления контактов и достижения взаимопонимания в социальной и профессиональной сфере с зарубежными партнерами, отводится коммуникативным технологиям [1]. В этом контексте имеют значение знание и соблюдение характеристик и особенностей видов речевой деятельности:

– диалогическая речь: ситуативная обусловленность, ролевое поведение, смена инициативности, эмоциональность и темп речи, клише начала, поддержания и завершения разговора, фонетическое и интонационное оформление реплик;

– монологическая речь: ситуативная обусловленность, логичность, связность, широкая обращённость, правильность оформления речи, приёмы привлечения внимания, клише;

– аудирование: мотивированность, прогнозирование на уровне слова, предложения, текста, достаточный объем оперативной памяти в опыт слухового восприятия**,** определение главной и второстепенной информации.

– чтение: определение вила текста, выбор способа чтения и степени понимания, самостоятельная семантизация незнакомых лексических единиц, определение главной и второстепенной информации, определение ключевых слов и элементов связи предложений и частей текста, анализ полученной информации, её интерпретация.

Полагаем, что реализация учебных технологий в процессе изучения иностранного языка повысит эффективность овладения им, что позволит будущим специалистам достичь успеха в коммуникации и откроет путь к самореализации.

СПИСОК ЛИТЕРАТУРЫ

1. Булатова, Д. В. Иностранный язык как средство профессиональной подготовки студентов неязыковых вузов / Д. В. Булатова // Профессиональное образование. – 1996. – № 1. – С. 78–87.

2. Котунова, М. Н. Современные технологии обучения иностранному языку [Электронный ресурс]. – Режим доступа: https://gigabaza.ru/doc/75847.html. – Дата доступа: 12.04.2020.

І.А. ШВЕД, І.В. ПОЎХ
Брэст, БрДУ імя А.С. Пушкіна

ПЕРСПЕКТЫВЫ ВЫКАРЫСТАННЯ МЕТАДАЎ АЎТАМАТЫЧНАЙ КЛАСІФІКАЦЫІ НАРАТЫЎНЫХ ЖАНРАЎ

У СУЧАСНАЙ ФАЛЬКЛАРЫСТЫЦЫ[1]

Адной з найбольш актуальных праблем сучаснай фальклярыстычнай жанралогіі выступае недастатковая акрэсленасць шэрагу жанраў, найперш наратыўных, і адсутнасць дакладнага размежавання паміж імі. Лічбавізацыя фальклорных наратываў і іх далейшая аўтаматычная класіфікацыя дазваляюць зрабіць больш зручным пошук патрэбных наратываў (напрыклад, праз адбор вынікаў пошуку па жанры ці ранжыраванне знойдзеных дакументаў) і вызначыць ступень унікальнасці кожнага жанру. Адпаведна замежныя фальклярысты даволі актыўна пачалі распрацоўваюць метады аўтаматычнай класіфікацыі наратыўных жанраў. Аўтаматычная класіфікацыя жанраў разлядаецца як дадатковы інструмент, які, як слушна адзначаюць ірландскія навукоўцы Айдан Фін (*Aidan Finn*) і Нікалас Кушмерык (*Nicholas Kushmerick*), дапаўняе актуальныя тэхналогіі і забяспечвае атрыманне больш дакладных вынікаў. Адпаведнасць таго ці іншага дакумента пэўнаму запыту залежыць ад карыстальніка, які склаў запыт. Жанравая і / або стылёвая прыналежнасць тэксту дае каштоўную дадатковую інфармацыю, якая дазваляе вызначыць, якія дакументы ў найбольшай ступені адпавядаюць таму ці іншаму запыту [1].

Разам з пашырэннем тэкставых баз дадзеных, даступных карыстальнікам, жанр набывае ўсё большую значнасць для даследчыкаў, нароўні з тэматычным і структурным прынцыпамі класіфікацыі. Адным з найбольш распаўсюджаных метадаў класіфікацыі і лінгвістычнага аналізу тэкстаў, у тым ліку фальклорных, выступае метад “знешніх сімвалаў” (*surface cues*). Фармальныя знешнія элементы сінтаксічнага, прасадычнага і семантычнага ўзроўняў зліваюцца, спалучаюцца ці супярэчаць адзін аднаму, увасабляючы тыя ці іншыя сэнсы альбо выконваючы пэўныя функцыі. Амерыканскія лінгвісты Брэт Кеслер (*Brett Kessler*) і Джэфры Нанберг (*Geoffrey Nunberg*), а таксама нямецкі лінгвіст Хайнрых Шутцэ (*Hinrich Schütze*) разглядаюць жанр як сукупнасць кампанентаў, якія адпавядаюць тым ці іншым знешнім сімвалам, сцвярджаючы, што эфектыўнасць і дакладнасць катэгарызацыі жанраў з дапамогай знешніх сімвалаў не саступае іншым, больш глыбокім структурным падыходам. Даследчыкі вылучаюць чатыры асноўныя групы знешніх сімвалаў, якія вызначаюць жанравую прыналежнасць таго ці іншага твора: структурныя, лексічныя, знакавыя і вытворныя сімвалы, – тры апошнія з якіх прыдатныя для аўтаматычнай класіфікацыі. Лексічныя сімвалы ахопліваюць формулы

[1] Праца выканана пры фінансавай падтрымцы Беларускага рэспубліканскага фонда фундаментальных даследаванняў у межах задання “Беларуская фальклярыстыка ў сучасным свеце: метадалагічны, праблемна-тэматычны дыяпазон, тэарэтычныя навацыі”; дамова № Г20У-004, нумар дзяржаўнай рэгістрацыі 20201112 ад 26.06.2020.

звароту, стылёва маркіраваныя афіксы, лексемы са значэннем часу і інш. Да знакавых сімвалаў адносяцца пераважна знакі прыпынку і іншыя сродкі падзелу і размежавання фраз, сказаў і частак сказа. Да групы вытворных сімвалаў належаць каэфіцыэнты (сума іншых сімвалаў, атрыманая шляхам ператварэння вынікаў усіх вылічэнняў у натуральныя лагарыфмы) і варыятыўныя меры (ступень варыятыўнасці пэўнага колькаснага сімвала ў тэксце – напрыклад, працягласці сказаў) [2].

Сярод іншых метадаў аўтаматычнай класіфікацыі наратыўных жанраў, якія знайшлі ўвасабленне ў сучасных фальклярыстычных даследаваннях, варта адзначыць аналіз тэксту як набору слоў (*bag-of-words analysis*), калі наратыў зашыфроўваецца як вектар прыкмет (*feature vector*), што паказвае на наяўнасць ці адсутнасць у ім пэўнага слова. Гэты падыход да класіфікацыі тэкстаў, як адзначаецца ў літаратуры пытання [1], спрацоўвае найлепш, калі асновай класіфікацыі выступае тэма наратыву; адпаведна, жанравы класіфікатар, які выкарыстоўвае адзначаны метад, будзе найбольш эфектыўным для твораў адной тэматыкі. Сучасныя лінгвісты актыўна ўжываюць і статыстычныя метады на аснове вектару прыкмет, у прыватнасці статыстычны аналіз часцін мовы (*POS Statistics*) і статыстычны аналіз тэксту (*Text Statistics*). Даследчыкі разглядаюць статыстычны аналіз часцін мовы як эфектыўны механізм адлюстравання моўнага стылю, дастатковы для распрацоўкі алгарытма размежавання жанраў. У межах гэтага аналізу наратыў разглядаецца як вектар, які складаецца з 36 прыкмет часцін мовы – па адной на кожнае ўмоўнае абазначэнне той ці іншай прыкметы, – прадстаўленых у выглядзе адсотку ад агульнай колькасці слоў у тэксце. Выкарыстанне адзначанага метаду прадугледжвае засяроджванне ўвагі на схеме ўжывання пэўных часцін мовы ў наратыве (а не на ўласна лексічным складзе тэксту як наборы слоў) і на тыпе тэксту (а не яго тэме), што, у сваю чаргу, дазваляе вызначыць жанр наратыва незалежна ад яго тэматыкі. Нарэшце, статыстычны аналіз тэксту ахоплівае такія параметры як сярэднюю даўжыню сказа, размеркаванне доўгіх слоў і сярэднюю даўжыню слова, а таксама частотнасць ужывання функцыянальных слоў і знакаў прыпынку.

Шырока распаўсюджаны ў еўрапейскіх лінгвістычных даследаваннях таксама метад аналізу тэкстаў на аснове шматэлементных (шматслоўных, шматзнакавых, сінтаксічных і г.д.) паслядоўнасцяў (*n-gram*). Традыцыйна ў якасці элементаў могуць выступаць словы, знакі, прыкметы часцін мовы і іншыя элементы, якія ўжываюцца ў тэксце адзін за адным. (Літара *n* паказвае на колькасць элементаў у паслядоўнасці) [3].

Аўтаматычная класіфікацыя тэкстаў і іх элементаў актыўна выкарыстоўваецца еўрапейскімі даследчыкамі ў дачыненні да фальклорных наратываў. Распрацаваная калектывам галандскіх навукоўцаў [4] сістэма

аўтаматычнай класіфікацыі фальклорных наратыўных жанраў ахоплівае ўсе прыведзеныя вышэй метады. Аб'ектам даследавання выступіў корпус фальклорных наратываў, сабраных на тэрыторыі Нідэрландаў з XVI стагоддзя да сучаснасці: казкі, легенды (асобна вылучаюцца легенды пра святых і гарадскія легенды), асабістыя наратывы, загадкі, сітуацыйныя задачы, жарты і песні. Даследчыкі вылучылі чатыры асноўныя групы знешніх сімвалаў (*surface cues*): лексічныя прыкметы, стылёвыя і структурныя прыкметы, інфармацыю пра асяроддзе і метададзеныя.

Групу лексічных прыкмет склалі ўніграмы (аднаэлементныя паслядоўнасці) і шматзнакавыя паслядоўнасці (*n-gram*) даўжынёй ад двух да пяці сімвалаў, у тым ліку знакі прыпынку і прабелы. Стылёвыя і сінтаксічныя прыкметы ахопліваюць уніграмы і біграмы (двухэлементныя паслядоўнасці) часцін мовы, пунктуацыю, пустыя радкі і статыстыку тэксту. Адбор уніграм і біграм праводзіўся з выкарыстаннем модульнага морфасінтаксічнага аналізатара і маркіравальніка для галандскай мовы *TADPOLE*, распрацаванага для апрацоўкі вялікіх (ад некалькіх мільёнаў да мільярду слоў) корпусаў тэкстаў, створаных у выніку лічбавізацыі як сучасных матэрыялаў, так і архіўных дадзеных. Сярод бясспрэчных пераваг адзначанай тэхналогіі яе стваральнікі называюць дакладнасць, высокую лінейную хуткасць апрацоўкі і выкарыстанне невялікага аб'ёму памяці [5]. Пунктуацыя паказвае колькасць знакаў прыпынку з папраўкай на агульную колькасць знакаў у тэксце. Пустыя радкі – колькасць пустых радкоў з папраўкай на агульную колькасць радкоў у тэксце (выкарыстоўваецца як прыкмета песенных тэкстаў). Нарэшце, у статыстыку тэксту ўваходзяць працягласць тэксту, сярэдняе значэнне і стандартнае адхіленне даўжыні сказаў, колькасць слоў у сказе і даўжыня слоў.

Інфармацыя пра асяроддзе выступае вызначальнай прыкметай легенды як жанру і выражаецца ў колькасці аўтаматычна маркіраваных найменных існасцей (тапонімы, антрапонімы, назвы арганізацый, мерапрыемстваў і г.д.), падлічанай з дапамогай апісанай вышэй тэхналогіі *TADPOLE*. Кожны тып найменных існасцей улічваецца як асобная прыкмета. Варта адзначыць, што аналіз метададзеных (ключавыя словы, найменныя існасці, не маркіраваныя аўтаматычна, кароткі змест і год стварэння ці выканання) праводзіўся ўручную, таму вынік аўтаматычнай класіфікацыі жанраў падводзіўся як з улікам, так і без уліку адпаведнай групы.

У выніку праведзенага галандскімі навукоўцамі даследавання было высветлена, што найбольш эфектыўным паказальнікам аўтаматычнай класіфікацыі жанраў выступаюць шматзнакавыя паслядоўнасці (*n-gram*), здольныя, з аднаго боку, улічваць пунктуацыю і канчаткі слоў, а з іншага – ігнараваць памылкі правапісу і варыянтныя напісанні [4].

Ранжыраванне як метад аўтаматычнай класіфікацыі наратыўных жанраў таксама актыўна выкарыстоўваецца еўрапейскімі фальклярыстамі, у першую чаргу, калі крытэрыем размежавання выступае тып аповеду – сюжэтна-тэматычнае падабенства групы наратываў, якое адрознівае яе ад іншых груп. Аўтаматычнае вызначэнне сюжэтнага тыпу таго ці іншага наратыву метадам ранжыравання адбываецца шляхам надання найбольш адпаведнаму тыпу сюжэту найвышэйшага рангу. Такі падыход, як сцвярджаюць даследчыкі, не толькі дазваляе ідэнтыфікаваць тып сюжэту, але і ўсталёўвае новыя сувязі паміж сюжэтамі, а таксама дапамагае выявіць тэкставыя запазычанні і перыфразы [6]. Тэкставыя падабенствы праяўляюцца на розных узроўнях, ад поўнай ідэнтычнасці да сюжэтнага падабенства. Тэкставыя запазычанні ахопліваюць пашырэнне тэксту, яго змяненне ці звужэнне. Сюжэтнае падабенства можа вынікаць з агульнай першакрыніцы. Тым не менш, тэкставыя падабенствы заўжды выходзяць за межы лексічных альбо сюжэтных сыходжанняў: у іх аснову кладуцца падзеі, матывы (наратыўныя элементы) і персанажы. У адрозненне ад тэкставых запазычанняў, прыналежнасць сюжэтаў да аднаго і таго ж тыпу ўсталёўваецца на больш абстрактным узроўні, чым лексічны (напрыклад, месца дзеяння не заўжды супадае дакладна).

Пры вызначэнні тыпу сюжэту шляхам ранжыравання сістэма аўтаматычна складае рэйтынг магчымых тыпаў сюжэтаў для пэўнага наратыву. Даследаванне складаецца з наступных этапаў: стварэнне зыходнай выбаркі сюжэтаў метадам *BM25* (прынцып найбольшага супадзення), адбор 50 найбольш адпаведных варыянтаў на аснове супадзенняў знешніх прыкмет, выдаленне паўтораў і складанне канчатковага рэйтынгу тыпаў сюжэтаў. З усяго дыяпазону знешніх прыкмет, на аснове якіх ствараецца вузкая выбарка, даследчыкі вылучаюць тры асноўныя групы: лексічныя падабенствы (на ўзроўні ўніграм, біграм (двухэлементных паслядоўнасцяў), шматзнакавых паслядоўнасцяў (2–5), блокаў, найменных існасцей, паказчыкаў часу і прасторы), тэкставыя сыходжанні (увесь тэкст – толькі назоўнікі – толькі дзеясловы) і трайныя сінтаксічныя сыходжанні (дзейнік-выказнік-дапаўненне).

Ацэньваючы эфектыўнасць метаду ранжыравання, даследчыкі адзначаюць магчымасць яго выкарыстання ў дачыненні да розных тыпаў тэкстаў – як вусных, так і пісьмовых, а таксама з мэтай вызначэння сыходжанняў паміж дыялектнымі ці гістарычнымі варыянтамі наратываў [6].

Праведзены аналіз некаторых замежных даследаванняў у галіне распрацоўкі метадаў аўтаматычнай класіфікацыі лінгвістычных элементаў і яе выкарыстання ў дачыненні да фальклорных наратываў выступае сведчаннем эфектыўнасці аўтаматычнай жанравай класіфікацыі наратываў, а таксама ўсталявання ўзаемасувязяў і запазычанняў паміж тэкстамі.

Аўтаматызацыя класіфікацыі і ранжыравання тэкстаў спрашчае і ўдасканальвае працэс стварэння і выкарыстання рэгіянальных і нацыянальных баз дадзеных фальклорных матэрыялаў ва ўмовах іх лічбавізацыі, дазваляе распрацаваць алгарытмы іх інтэрпрэтацыі і ўвядзення ў кантэкст міжнародных фальклярыстычных даследаванняў, найперш кампаратыўных.

СПІС ЛІТАРАТУРЫ

1. Finn, A. Learning to Classify Documents According to Genre / A. Finn, N. Kushmerick // Journal of the American Society for Information Science and Technology. – 2006. – Vol. 57, № 11. – P. 1506–1518.
2. Kessler, B. Automatic Detection of Text Genre / B. Kessler, G. Nunberg, H. Schütze [Electronic resource]. – Mode of access: https://www.researchgate.net/publication/1783012_Automatic_Detection_of_Text_Genre. – Date of access: 23.06.2021.
3. Sidorov, G. Syntactic N-grams as Machine Learning Features for Natural Language Processing / G. Sidorov, F. Velasquez, E. Stamatatos, A. Gelbukh // Expert Systems with Applications. – 2014. – Vol. 41, № 3. – P. 853–860.
4. Nguyen, D, Automatic Classification of Folk Narrative Genres / D. Nguyen, D. Trieschnigg, Th. Meder, M. Theune // Proceedings of KONVENS 2012 (LThist 2012 workshop). – Vienna : September 21, 2012. – P. 378–382.
5. Bosch van den, A. An Efficient Memory-Based Morphosyntactic Tagger and Parser for Dutch / A. van den Bosch, B. Busser, S. Canisius, W. Daelemans [Electronic resource]: Computational Linguistics in the Netherlands. – Mode of access: https://www.researchgate.net/publication/313005307_An_efficient_memory-based_morphosyntactic_tagger_and_parser_for_Dutch. – Date of access: 25.06. 2021.
6. Nguyen, D. Folktale Classification Using Learning to Rank / D. Nguyen, D. Trieschnigg, M. Theune [Electronic resource]: Proceedings of the 35th European conference on Advances in Information Retrieval. – Mode of access: https://www.researchgate.net/publication/262215479_Folktale_Classification_Using_Learning_to_Rank. – Date of access: 27.06. 2021.

Е.И. ТАРАШКЕВИЧ
Минск, ВА РБ

СИСТЕМА УПРАЖНЕНИЙ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ КУРСАНТОВ И СЛУШАТЕЛЕЙ ПО ОВЛАДЕНИЮ ПРОФЕССИОНАЛЬНОЙ ЛЕКСИКОЙ