

В. Ю. Бойка (Беларусь, Мазырскі дзяржаўны педагагічны ўніверсітэт імя І. П. Шамякіна)

РОЛЯ СТАТЫСТЫЧНАГА МЕТАДУ Ў ВЫВУЧЭННІ СТЫЛІСТЫЧНЫХ ПАРАМЕТРАЎ ЛЕКСІКІ БЛОГАЎ

У працэсе даследавання тэкстаў такога інтэрнэт-жанру, як блог, перад лінгвістам паўстае праблема аб'ектыўнасці атрыманых тэарэтычных абагульненняў. У прыватнасці, гэта датычыцца вывучэння стылістычных рэсурсаў лексікі. У класічных падручніках нярэдка адзначаецца, што ў тэкстах таго ці іншага стылю часта ўжываюцца лексічныя адзінкі пэўных разрадаў. Напрыклад, гутарковыя тэксты звычайна змяшчаюць прастамоўныя адзінкі, жарганізмы, аргатызмы, навуковыя – спецыфічныя тэрміны, юрыдычныя – “канцылярызмы”. Падобныя вывады робяцца, відавочна, на падставе супастаўляльнага аналізу тэкстаў, створаных у пэўнай сферы сацыяльных адносін. “Лакмусавай паперкай”, якая дазваляе выявіць стылістычна афарбаваныя словы, з'яўляецца так званая нейтральная лексіка.

На першы погляд, гэты падыход можа быць плённым і ў працэсе вывучэння стылістычных параметраў блогаў. Разглядаючы стылістычна

нейтральную літаратурную лексіку як пэўны эталон, лінгвіст можа вылучыць у тэкстах інтэрнэт-дзённікаў асобныя групы слоў, што не адпавядаюць літаратурнай норме і маюць яркую эмацыянальна-экспрэсіўную афарбоўку. Аднак гэтыя адзінкі зусім не абавязкова з'яўляюцца агульнымі стылістычнымі маркерамі дадзенага жанру. Справа ў тым, што кожны стыль і жанр маюць свае экстралінгвістычныя ўмовы, што ўплываюць на адбор лексічных сродкаў. Напрыклад, тэксты навуковага і афіцыйна-справавога стылю ствараюцца ва ўмовах даволі строгай рэгламентацыі. Гэта абумоўлівае лексічнае падабенства тэкстаў розных жанраў, але аднаго і таго ж стылю. Пэўная рэгламентацыя, не выражаная экспліцытна, прысутнічае і ў гутарковым стылі з яго ўяўнай адвольнасцю выбару моўных сродкаў. У камунікатыўнай сітуацыі нязмушанай размовы прадзэнт тэксту падсвядома імкнецца выкарыстоўваць сродкі зніжанай экспрэсіі, прастамоўныя словы і выразы, паколькі выключна літаратурная лексіка не адпавядае чаканням рэцыпіента. Такім чынам, мы выкарыстоўваем у жанры вуснай гутаркі спецыфічную “гутарковую” лексіку не столькі ў сілу стылістычнай “уседазволенасці”, колькі праз неабходнасць “падстройвацца” пад сітуацыю зносін і камунікатыўныя інтэнцыі суразмоўцы.

Тэксты блогаў прадуюцца ў іншых камунікатыўных умовах. Па-першае, запісы ў інтэрнэт-дзённікаў не толькі адрасаваны патэнцыйнаму партнёру па камунікацыі, але і з'яўляюцца формай аўтакамунікацыі. Асоба чытача, а таксама прадметна-побытавы кантэкст аказваюць нязначны ўплыў на працяканне зносін. Таму аўтар блога мае большую свабоду выбару лексічных сродкаў, чым удзельнік “жывой” размовы. Ён у значна меншай ступені “падстройваецца” пад патэнцыйнага камуніканта (апошні, дарэчы, можа і ўхіліцца ад зносін, не пакідаючы каментарыя). Хаця перад даследчыкам, які сустракае ў асобна ўзятым блогу элементы гутарковага маўлення, узнікае “спакуса” аднесці гэты жанр да гутарковага стылю. Але блогі маюць вялікую тэматычную разнастайнасць, у іх часта можна вылучыць “паджанры”. Так, многія аўтары выкарыстоўваюць персанальную інтэрнэт-старонку для публікацыі ўласных вершаў, падарожных нататак, філасофскіх эсэ. Жанравая неаднароднасць часта ўласціва нават асобным запісам у межах аднаго блога. Акрамя таго, на форму і змест выказвання ўплывае шэраг суб'ектыўных фактараў: узрост аўтара, узровень яго адукаванасці, характар, псіхічны стан і г. д. Калі адны карыстальнікі актыўна ўжываюць у інтэрнэт-маўленні гутарковыя элементы, то іншыя могуць аддаваць перавагу стылістычна нейтральным. Таму мы не можам катэгарычна сцвярджаць, што тыя ці іншыя лексічныя сродкі, выяўленыя пры аналізе асобных блогавых тэкстаў, уласцівыя жанру як такому.

Для таго, каб пазбегнуць суб'ектывізму ў працэсе стылістычнага аналізу лексікі блога, мэтазгодна выкарыстоўваць статыстычны метада

даследавання. У прыватнасці, гаворка ідзе пра камп'ютэрны падлік частотнасці слоў. У дадзеным выпадку мы можам дакладна вызначыць наступныя параметры:

1. Працэнтныя суадносіны нарматыўна-літаратурнай і нелітаратурнай лексікі;
2. Марфалагічны састаў лексікі (колькасць ужытых назоўнікаў, прыметніка, дзеясловаў і г. д.);
3. Найбольш частотныя лексічныя адзінкі;
4. Працэнтныя суадносіны слоў пэўнай групы і агульнай колькасці лексічных адзінак.

Маючы дадзеныя камп'ютэрнага аналізу частотнасці слоў, мы можам вызначыць “удзельную вагу” слоў пэўных груп у агульным корпусе тэкстаў. Гэта дапаможа пазбегнуць суб'ектыўнасці і стварыць эмпірычна абгрунтаванае апісанне лексічных рэсурсаў жанру блога. Дзеля таго, каб атрымаць найбольш аб'ектыўныя вынікі, неабходна выканаць шэраг умоў:

1. У агульным корпусе тэкстаў, прызначаных для камп'ютэрнага аналізу, павінны быць прадстаўлены фрагменты пісьмовага маўлення дастаткова вялікай колькасці карыстальнікаў (100-200);
2. Фрагменты пісьмовага маўлення павінны адбірацца шляхам суцэльнай выбаркі (за выключэннем запазычаных тэкстаў);
3. Маўленне розных блогераў павінна быць прадстаўлена прыблізна эквівалентнымі аб'ёмам тэкстаў.

Такім чынам, ужываючы сучасныя статыстычныя метады апрацоўкі лінгвістычных дадзеных, мы атрымліваем аб'ектыўныя дадзеныя аб лексічным саставе тэкстаў беларускамоўных блогаў. Але падлік частотнасці слоў пры дапамозе камп'ютэрных праграм мае некаторыя хібы, якія неабходна ўлічваць. Так, пры аўтаматычным стварэнні спісу ўжытай лексікі не ўлічваецца яе магчымая шматзначнасць і аманімія. Ігнараванне гэтых з'яў можа прывесці да скажэння выніковых дадзеных, асабліва ў дачыненні да элементаў гутарковага маўлення. Напрыклад, назоўнік *тормаз* у значэнні ‘прыстасаванне для спынення або запавольвання руху машыны, цягніка’ і яго амонім *тормаз* як ‘марудлівы чалавек’ будуць прадстаўлены як адна лексема. Але відавочна, што яны маюць розную стылістычную афарбоўку.

Другім фактарам, што ўскладняе інтэрпрэтацыю вынікаў камп'ютэрнага падліку частотнасці слоў, з'яўляецца правапісная неаднароднасць. Шэраг блогераў карыстаецца “альтэрнатыўнай” арфаграфіяй (т. зв. “тарашкевіцай”). Адпаведна *снег* і *сьнег* камп'ютэрная праграма распазнае як розныя словы.

Пэўныя складанасці для даследчыка стварае і той факт, што даступнае праграмае забеспячэнне не дазваляе распазнаваць граматычныя формы слоў тыпу *хадзіць*, *хаджу*, *ходзяць*, а таксама *брат*, *брата*, *братоў* і г. д. Некаторыя з праграм некарэктна працуюць з беларускімі літарамі,

паколькі першапачаткова былі створаны для аналізу рускамоўных або англамоўных тэкстаў.

Аднак нават з улікам усіх пагрэшнасьцяў, якія ўласцівы вынікам частотнага аналізу слоў, статыстычны метада павінен стаць, на нашу думку, адным з асноўных інструментаў вывучэння стылістычных параметраў лексікі. На дадзены момант толькі частотны падлік можа даць найбольш аб'ектыўнае ўяўленне аб асаблівасцях лексічнага саставу тэкстаў блогаў.