

**РАЗМЕЖАВАННЕ АМАФОРМАЎ ДЛЯ ЎКЛАДАННЯ  
СЛОЎНІКА МОВЫ ЯКУБА КОЛАСА: ПРАКТЫЧНЫ АСПЕКТ**

Ключавую ролю пры падрыхтоўцы слоўнікаў розных тыпаў, у тым ліку і слоўнікаў мовы пісьменнікаў, адыгрываюць спосабы і прыёмы аўтаматычнай апрацоўкі тэкстаў. Адна з галін мовазнаўства – корпусная лінгвістыка – звязана са стварэннем і развіццём корпусаў тэкстаў, іх прымяненнем у якасці інструмента лінгвістычных даследаванняў. Корпус тэкстаў, якія складаюць пісьмовую спадчыну Якуба Коласа, найбольш поўна прадстаўлены ў 20-томным акадэмічным выданні збору твораў пісьменніка (2007–2012). Наяўнасць электроннага варыянту дадзеных тэкстаў дазваляе вучоным працаваць над укладаннем слоўніка мовы класіка, выкарыстоўваючы шырокія магчымасці іх аўтаматычнай апрацоўкі.

Аўтаматычная апрацоўка тэксту дзеліцца на некалькі этапаў, адным з якіх з'яўляецца марфалагічны. Для поўнага адлюстравання лексікону класіка неабходна ўлічваць усе выкарыстаныя ім словы, а таксама формы слоў – суадносіць іх з пэўнай парадыгмай словазмянення. На дадзеным этапе для кожнага словаўжывання вызначаюцца марфалагічныя характарыстыкі (род, склон, трыванне і г.д.) і пачатковая форма слова, якая называецца лемай. Аптымізаваць працэс можна дзякуючы выкарыстанню электроннай граматычнай базы, створанай спецыялістамі Інстытута мовазнаўства імя Якуба Коласа НАН Беларусі. З'яўляючыся надзейнай асновай марфалагічнай разметкі тэкстаў, яна дапаўняецца спецыфічнымі аўтарскімі лексемамі і формамі, абумоўленымі індывідуальным мастацкім стылем пісьменніка і тагачасным станам развіцця літаратурнай мовы.

Задача марфалагічнай разметкі ўскладняецца аманіміяй, якая ў беларускай мове выступае на лексічным і граматычным узроўнях. Апошні ў найбольшай ступені выклікае распаўсюджанасць аманімічных форм. Суаднясенне аманімічнай формы з парадыгмай змянення слова нават пры аўтаматычнай апрацоўцы адбываецца ўручную: не існуе надзейных алгарытмаў, якія дазволілі б аўтаматызаваць гэты працэс. Для тэкстаў на англійскай мове існуюць метады зняцця аманіміі, заснаваныя на верагоднаснай мадэлі, якія даюць высокую дакладнасць. Распрацоўка алгарытмаў для рускай і беларускай моў, якія адрозніваюцца свабодным размяшчэннем слоў у сказе, выклікае цяжкасці [1]. І ўсё ж некаторыя правілы спалучальнасці слоў могуць часткова вырашыць гэтае пытанне.

У аснову класіфікацыі амонімаў на ўзроўні слоўнікавых форм кладуцца два асноўныя крытэрыі: а) аднесенасць да адной і той жа ці розных часцін моў; б) поўнае або частковае супадзенне парадыгм словазмянення. На гэтай падставе выдзяляюць поўныя і няпоўныя лексічныя, поўныя і няпоўныя лексіка-граматычныя амонімы [2, с. 227]. Напрыклад, да поўных лексіка-граматычных амонімаў адносяцца словы нязменных часцін мовы. З аналізуемага корпусу бачым, што лема *блізка* можа рэпрэзентаваць дзве розныя нязменныя часціны мовы. У адным выпадку гэта прыслоўе: “*Не бядуі, што восень блізка...*”. У другім – прыназоўнік: “*Па-другое – замацавацца на зямлі ніколі не пашкодзіць, асабліва, калі яна блізка горада*”.

За межамі пералічаных груп застаюцца аманімічныя з’явы, якія працягваюцца на ўзроўні граматычных форм, – амаформы. Амаформы ў межах парадыгмы аднаго і таго ж слова – гэта супадзенне ў гучанні і напісанні граматычных форм скланяльных часцін мовы. Так, у межах лемы *лес* прадстаўлена аманімічная форма *лесу*, якая ўжываецца як: 1. форма роднага склону адзіночнага ліку назоўніка мужчынскага роду: “*А лесу шум і песня поля // Сатруць з душы нягоды след*”; 2. форма давальнага склону: “*Сонца апускалася ўжо к лесу, і ціха, і хораша свяціліся яго верхавіны ў залатым бляску апошніх лучоў*”.

Аманімічныя формы ў межах адной і той жа часціны мовы ў лексіконе Якуба Коласа маюць розную ступень распаўсюджанасці. Так, сярод 112 выпадкаў ужывання формы *птушкі* абсалютная большасць прыпадае на форму множнага ліку, у той час як колькасць ужыванняў формы роднага склону адзіночнага ліку складае толькі 7 фіксацый, або 6,25% ад агульнай колькасці. Статыстычныя даныя адносна аманімічнай формы *лесу* наступныя: 440 фіксацый роднага склону і толькі 10 – давальнага (2,2% ад агульнай колькасці ўжыванняў). Што тычыцца фармальных прыкмет, якія маглі б дазволіць карэктнае аўтаматычнае вызначэнне патрэбнай формы ў пэўным кантэксце, то варта адзначыць наступнае. Спалучэнне “прыназоўнік *к* + форма *лесу*” ў 100% выпадкаў будзе рэпрэзентаваць давальны склон. Ужыванне іншых прыназоўнікаў (*да, каля, узбоч, ад* і інш.) у аналагічным спалучэнні ўказвае выключна на родны склон.

Амаформы ў межах парадыгм розных слоў паяўляюцца ў выніку выпадковага супадзення ў адным вымаўленні і напісанні граматычных форм слоў, якія адносяцца да адной і той жа або да розных часцін мовы. Тут магчымы разнастайныя камбінацыі, напрыклад, “*нязменная часціна мовы + форма зменнай часціны мовы*” (прыслоўе і назоўнік, часціца і дзеяслоў і інш.). Так, нязменнае слова *бывала* з граматычнага боку варта разглядаць як часціцу, што выкарыстоўваецца ў значэнні пабочнага слова і паказвае на дзеянне, якое нерэгулярна паўтаралася ў

мінулым: “Збяруцца, **бывала**, і люлькі закураць, // І так углыбляцца ў размовы яны, // Што лбы свае зморшчаць і бровы нахмураць...”. У той жа час адна з форм лемы *бываць* мае ідэнтычнае напісанне і гучанне, рэпрэзентуючы пры гэтым зусім іншыя граматычныя характарыстыкі (дзеяслоў, незакончанае трыванне, прошлы час, жаночы род, адзіночны лік): “Узышоўшы на ганак школы, зноў прыпыняліся, і тут часамі **бывала** між імі невялікая спрэчка, бо, як праўдзівыя хрысціяне, яны ўступалі першае месца адзін другому”.

Часам да з’явы граматычнай аманіміі далучаецца з’ява аманіміі графічнай, калі пры аднолькавым напісанні словы ці іх формы адрозніваюцца пастаноўкай націску. Пры машыннай апрацоўцы тэкстаў такія словы таксама патрабуюць адпаведнай граматычнай ідэнтыфікацыі. Возьмем да прыкладу форму *стала*. Найбольш пашыранае яе ўжыванне ў тэкстах Якуба Коласа – форма прошлага часу ніякага роду дзеяслова закончанага трывання: “Не бядуў, што цяжка **стала** // Жыць у вечнай цемнаце”. Такая ж форма з’яўляецца адзіна магчымай для нязменнага слова – прыслоўя: “Так за ёмкім павуком цягнуцца ўсе павуцінкі, калі ён, намеціўшы сваю забытаную ахвяру, не спяшаючыся, **стала** настаўляе на яе свае чэпкія лапы”. У той жа час як аманімічны варыянт ужываецца форма роднага склону адзіночнага ліку назоўніка мужчынскага роду *стол*: “Потым крошкі з **стала** // Чыста сцёрла-змяла”. Аўтаматычна ідэнтыфікаваць апошняю форму дапамагае кантэкст “прыназоўнік + *стала*”: *каля стала, да стала* і інш.

Размежаванне граматычнай аманіміі пры аўтаматычнай апрацоўцы тэкстаў для ўкладання слоўніка мовы Якуба Коласа – працаёмкі працэс, аднак ён неабходны для карэктнага складання рээстра загаловачных слоў і адлюстравання парадыгм зменных лексем. У працэсе працы над корпусам можна вылучыць фармальныя прыкметы, якія дазваляюць ідэнтыфікаваць пэўныя аманімічныя формы, аднак надзейных алгарытмаў для гэтага пакуль не існуе.

## ЛІТАРАТУРА

1. Порохнин, А. А. Анализ статистических методов снятия омонимии в текстах на русском языке / А. А. Порохнин // Вестник Астраханского государственного технического университета. Серия: управление, вычислительная техника и информатика. – 2013. – № 2. – С. 168–174.

2. Лексікалогія сучаснай беларускай літаратурнай мовы / пад рэд. А. Я. Баханькова. – Мінск : Навука і тэхніка, 1994. – 463 с.