

ПРЕДМЕТНА ГАЛУЗЬ ЗАГАЛЬНОМОВНОГО КОРПУСУ: ПИТАННЯ ПРО СУРЖИК

Статтю присвячено загальній проблемі репрезентації у стандартному електронному текстовому корпусі національної мови, зокрема її літературного, територіального та соціолектного рівнів з урахуванням явища суржику.

Ключові слова: корпус, літературна мова, територіальний діалект, соціальний діалект, суржик.

Однією з найважливіших якісних ознак будь-якого корпусу є його здатність коректно відображати предметну галузь, чи той мовний рівень, що його задекларовано корпусом. *Предметна галузь* для корпусу може мати різний обсяг (від надзвичайно великого до дуже малого), оскільки її визначає, з одного боку, обраний рівень (субрівень) опису, а з іншого – комплекс дослідницьких завдань. Так, якщо йдеться про авторський корпус і лінгвістичний опис стосується авторської мови, то предметною галуззю буде авторська мова, репрезентована творами досліджуваного автора, інтерв'ю з ним, можливо, записами його усного мовлення.

Корпус сучасної української мови як певна теоретична модель для втілення детермінована такими базовими характеристиками: дослідницький (зорієнтований на широкий спектр лінгвістичних завдань), загальнономовний (його предметну галузь становить сучасна українська мова), фрагментний (збудований із текстових фрагментів), мішаний (містить писемні й усні текстові фрагменти), динамічний (постійно поповнюється текстами), синхронний (охоплює рівень сучасної української мови з урахуванням територіальної специфіки як в Україні, так і поза її межами на історичних етнічних землях та в діаспорі) й одномовний (є результатом мовної діяльності носіїв сучасної української мови), які далі детермінують предметну галузь такого електронного текстового корпусу, а саме: сучасна українська загальнонаціональна мова у таких формах її існування, як літературна мова, територіальний діалект і частково соціальний діалект. Центральним ідіомом у такому корпусі є літературна мова – «унормована мова суспільного спілкування, загальноприйнята в писемній та усній практиці» [4, с. 318]. Найважливіша ознака якої, крім низки інших («писемна фіксація, унормованість, загальнообов'язковість норм і їхня кодифікація, розвинута функціонально-стильова система і стилістична диференціація засобів вираження, взаємодія писемного (книжного) і розмовного

мовлення» [7, с. 54]), що перетворює її на центр у корпусній репрезентації. За Л. Масенко, це «ідеальний зразок мови, яка реалізується в більш чи менш досконалій формі у мовленні та письмі освічених людей» [7, с. 41], тобто йдеться про трактування літературної мови як «ідеального зразка». Крім того, літературна мова, на що вказували ще представники Празької школи, є «складною багаторівневою структурою, поліфункціональним утворенням, на відміну від народних говорів, що у принципі виконують тільки комунікативну функцію. Безумовно, саме багаторівневості і поліфункціональності літературної мови творить і визначає різноманітність відношень літературної мови з іншими ідіомами (стратами), і [...] навпаки, багато у чому ці співвідношення визначають характер літературної мови» [10, с. 12]. І як такий ідіом (страта), літературна мова у корпусному відображенні може бути організована як окремих корпус літературної мови або ж як сегмент генерального корпусу і мати у межах цього корпусу найбільші кількісні характеристики, що відповідає статусу літературної мови і її реальним числовим параметрам у мові.

У межах загальнонаціональної мови літературна мова стоїть в опозиції до, так би мовити, нелітературної, чи, за Л. Масенко, «протистоїть територіальним і соціальним діалектам, різним типам койне, просторіччю й суржику – як вища, опрацьована форма існування мови» [7, с. 41]. Виникнувши на діалектній основі, сучасна українська мова засвідчує зв'язок зі своїми територіальними діалектами, тому годі ставити під сумнів потребу введення до загальнономовного корпусу діалектних текстів, оскільки, як слушно зауважує Н. Дзюбишина-Мельник, «жива природна мова, якими є її територіальні різновиди, залишається невичерпним джерелом для подальшого розвитку мови у будь-якій її формі: чи то штучного страту (ідіома), а власне, стандартної (тобто літературної) мови, чи то природних, але соціально зумовлених різновидів мови, або корпоративних субмов» [3, с. 25]. Крім того, діалек-

тна лексика, репрезентована діалектними текстами, є частиною лексики сучасного усного мовлення, однією з синхронних форм реалізації мовної системи, і якщо йдеться про репрезентативність дослідницького корпусу загальнонаціональної мови, то введення діалектного мовлення, діалектних текстів до його структури – не просто бажана, а важлива передумова досягнення репрезентативності та вичерпності текстового корпусу, на що також додатково впливає й такий чинник, як, за О. Гердтом, його функціонування у синхронії та одночасна належність до діахронії [див.: 1, с. 71–72]. Саме одночасна належність до синхронії та діахронії, крім іншого, є найпершим аргументом на користь уведення діалектного матеріалу до загального стандартного корпусу сучасної української мови.

Не таким однозначним є питання з приводу соціальних діалектів у корпусі сучасної української мови, яке пов'язано, по-перше, з певними екстралінгвістичними чинниками, і, по-друге, з проблемою суржику. Так, уведення до корпусу сучасної української мови текстів, які б репрезентували соціальні діалекти, зумовлені, з одного боку, загальною лінгвістичною тенденцією до «термінологічної універсалізації соціально маркованої лексики» [8, с. 47], а з іншого – із завданням оптимальної репрезентації та вичерпності мовних даних у загальномовному дослідницькому корпусі, а також із розширенням в українському мовознавстві дескриптивного (на противагу прескриптивному, наскрізно домінуючому в нашій традиції) підходу до наукового вивчення мови. Прескриптивний підхід, як слушно зазначає Л. Ставицька, «можна розглядати у контексті нехтування антропоцентричною парадигмою в українських лінгвістичних студіях, ігнорування самої мовної особистості, людини-носія мови, а отже, й дегуманізації самого поняття мови» [8, с. 15], не кажучи вже про спотворення реальної сучасної українськомовної картини світу. Соціальний діалект, чи соціолект, трактуємо як «напівавтономну або неавтономну форму існування загальнонародної мови національного періоду, функціонально закріплену за певним професійно-корпоративним соціумом, із властивою йому специфічною лексичною системою та варіативним за якістю й кількістю інвентарем граматичних і фонетичних особливостей, зумовлених соціолінгвістичними характеристиками його носіїв» [5, с. 382]. До соціальних діалектів, за Л. Ставицькою, зараховуємо аргю (особлива мова певної відокремленої професійної чи соціальної групи, яка складається з видозмінених елементів однієї або двох природних мов [8, с. 22]), жаргон (напіввідкрита лексико-фразеологічна підсистема, яку застосовує та чи та соціальна група з метою відособлення від решти мовної спільноти [6,

с. 33]) і сленг (різновид розмовної мови, яку суспільство оцінює як підкреслено неофіційну («побутова», «фамільярна», «довірлива» [8, с. 40]). У межах жаргону додатково диференціюємо професійний жаргон і професійну розмовну мову. Таким чином, якщо розглядати соціолект, репрезентований аргю, жаргоном і сленгом, як складову (особливо на рівні лексики) загальнонаціональної мови, то наявність соціолектних текстів у корпусі сучасної української мови є передумовою репрезентативності корпусу, повноти відображення предметної галузі. Однак очевидно, що соціолектні тексти не можуть кількісно переважати над діалектними текстами, текстами літературної мови, вони мають у кількісному відношенні відтворювати баланс субрівнів сучасної української літературної мови і, крім того, віддзеркалювати традицію пріоритетності літературної мови, чого загалом нескладно досягнути, пам'ятаючи про усну природу соціолектних текстів.

Важливим аспектом, що зумовлено необхідністю релевантного відображення у корпусі сучасної мовної картини світу українців, є доцільність уведення суржику до корпусу сучасної української мови. Стратегія уведення / неуведення суржикових текстів до корпусу безпосередньо мотивована рівнем наукового осмислення явища суржику, найбільше його статусу і місця у парадигмі національної мови. Найпоєднаніше виступає проти належності суржику до «тіла мови»¹ Н. Дзюбишина-Мельник: «Непоодинокі випадки кваліфікації суржика – мови, пересиченої дублетами, зайвими запозиченнями з усіх рівнів російської мови в українську – як прояву просторіччя і, отже, складової національної мови. Фактично ж має йтися про український піджин, що характеризується рядом специфічних ознак порівняно з традиційно виокремлюваними піджинами» [3, с. 26] і далі «Український піджин / креольська мова (інакше – суржик), як і об'єктна лексика (разом з усіма своїми дериватами та новорозгалуженою модальністю) до тіла національної мови не належить» [3, с. 27]. А в іншій своїй праці Н. Дзюбишина-Мельник взагалі кваліфікує суржик як патотекст, зокрема: «Саме перетворення *патотексту*, яким є суржик, в ноосферу України й спонукало нас простежити динаміку використання суржику й суржикізмів як стилістичного засобу в художніх текстах з кінця XIX до початку XXI ст.» [2, с. 17] Однак саме те, що, попри невизнання суржику сегментом тіла національної мови, дослідник усе-таки звертає свою увагу на суржикізми і суржикові тексти загалом, говорячи про чотири етапи «у використанні суржику та суржикізмів (тобто запозичень-інтерферентів із російської мови) у художніх тек-

¹ Термін запропонований Н. Дзюбишиною-Мельник [див.: 3, с. 24–27].

стах, які ми пов'язуємо не лише зі стилістичними ресурсами цього мовного явища, а й з різним трактуванням персонажа-мовця, що зазнає чужомовного, а разом чужокультурного тиску» [2, с. 17] і, що цікаво, в аналізі Н. Дзюбишиної-Мельник привертає увагу специфіка четвертого етапу – поч. ХХІ ст., – коли уже «з'являється твір (принаймні нам відомий лише один), де зроблено спробу відтворити суржиком буття персонажів достатньо високого інтелектуального рівня з акцентом на дорослій людині» [2, с. 19].

Однак в українському мовознавстві існують інші погляди на позиціонування суржику щодо загальнонаціональної мови. Як певне межове явище пропонують трактувати суржик Л. Ставицька і В. Труб: «суржик – це не кодифікована просторозмовна мова (курсив мій. – О. Д.), яка виникла як наслідок масової тривалої контактної українсько-російської двомовності в її диглосній формі. Суржик постає в результаті системної інтерференції на фонетичному, морфологічному, лексичному, синтаксичному рівнях; він репрезентований цільнооформленими лексемами – суржикізмами, які накладаються на українську або російську мовну основу; виявляє себе на базі регіональних різновидів української мови як мовний код в осіб з різним типом мовної компетенції, у різнотипних соціально-корпоративних та комунікативних сферах» [9, с. 77]. Цю ж думку розвиває і Л. Масенко, зокрема, говорячи про потребу наукового вивчення суржику: «До головних проблем у вивченні суржику належить визначення місця в системі усних форм побутування української мови, встановлення критеріїв, згідно з якими той чи інший різновид усного мовлення слід кваліфікувати як суржик, а це, своєю чергою, передбачає необхідність чіткого розмежування інтерферованих і мішаних форм мовлення, встановлення інтерференційного порогу, за яким однобічний вплив контактної російської мови на українську переростає в конвергенцію двох мов» [6, с. 53–54], де, з одного боку, маємо екстраполяцію на систему усних форм побутування української мови, а з іншого – на конвергенцію мов. Окрім того, розглядаючи суржик у діячності, Л. Масенко схильна трактувати останній як соціолект української мови: «суржик, вживаний у ХVІІІ–ХІХ ст., можна, очевидно, кваліфікувати як соціолект окремих груп українського етносу – прошарку місцевого чиновництва, яке пристосувало своє мовлення до загальноімперського офіціозу, та ідіолектів селян, зросійщених під час військової служби. [...] Розширення соціальної бази суржику, початкові процеси витіснення ним питомих, здавна сформованих говірок української мови слід пов'язувати із соціально-економічними й демографічними змінами кінця ХІХ – початку ХХ століття, з початком індустріалізації, розвитком

промисловості і капіталістичних відносин, створенням єдиного товарного ринку» [6, с. 16], однозначно вказуючи на негативний характер його поширення, що «перериває зв'язок української мови зі своїми питомими джерелами, передусім діалектними, відтак [суржик] є явищем глибоко деструктивним» [7, с. 80].

Аналізуючи праці українських соціолінгвістів, доходимо висновку, що, попри загальну послідовність трактування суржику, сьогодні усе ще триває процес його наукового вивчення або як цілісного окремого утворення, або як частини сучасної української мови – субутворення чи окремих його аспектів, зокрема: фонетичного морфологічного, особливо лексичного і синтаксичного. Проте як слушно зауважує Л. Масенко: «Попри поступове накопичення наукових розвідок, присвячених суржику, актуальним завданням лишається створення бази польових досліджень для його лінгвістичного опису. Тільки наявність солідної бази зразків мішаного українсько-російського мовлення з різних регіонів України дасть достатні підстави для уточнення дефініцій суржику та його лінгвістичних характеристик» [7, с. 72].

Із розглядом суржику за межами системи сучасної української мови зникає потреба уводити його до загальномовного корпусу, оскільки цей сегмент текстів лежить поза предметною галуззю корпусу. Натомість, дотримуючись іншого погляду, така потреба виникає, бо, відповідно, суржик потрапляє у поле описуваної предметної галузі. Двоаякість ситуації потребуватиме компромісного рішення щодо представлення суржику до залучення в корпусі. Цей компроміс схильні шукати в архітектурі корпусу, пропонуючи її як субкорпусну, тобто генеральний корпус має передбачати у своїй будові існування автономних субкорпусів, одним з яких може бути і субкорпус суржику. Але тут маємо пам'ятати, що архітектура загальномовного корпусу може позначитися на процесах усамостійнення суржику. Однак будь-які рішення щодо цієї гіпотези можливі лише після побудови самого корпусу і підтвердження або заперечення її.

Таким чином, якщо корпус сучасної української мови – це зібрання електронних текстів, що репрезентують національну мову на певному етапі її існування (сучасна українська мова) в усьому різноманітті жанрів, стилів, територіальних і соціальних варіантів, то його предметну галузь становитиме сучасна українська загальнонаціональна мова у таких формах існування, як літературна мова, територіальний діалект і частково соціальний діалект (арго, жаргон, сленг), а суржик винесено за межі цієї предметної галузі, оскільки запропоновано субкорпус суржику, де саме це мовне явище й становитиме предметну галузь.

Список літератури

1. Гухман М. М. Исторические и методологические основы структурализма / М. М. Гухман // Основные направления структурализма / [отв. ред. М. М. Гухман, В. Н. Ярцева]. – М. : Наука, 1964. – С. 5–45.
2. Дзюбишина-Мельник Н. Суржик і суржикізма : стилістичні ресурси / Н. Дзюбишина-Мельник // Наукові записки НАУКМА. Філологічні науки. – 2010. – Т. 111. – С. 16–20.
3. Дзюбишина-Мельник Н. Тіло національної мови / Н. Дзюбишина-Мельник // Магістеріум. Мовознавчі студії. – 2009. – Вип. 37. – С. 24–27.
4. Єрмоленко С. Я. Літературна мова / С. Я. Єрмоленко // Українська мова : енциклопедія / [редкол.: В. М. Русанівський, О. О. Тараненко, М. П. Зяблюк та ін. – 2-ге вид., випр. і допов.]. – К. : Укр. енциклопедія ім. М. П. Бажана, 2004. – С. 318–320.
5. Коровушкин В. П. Инвективная лексика и фразеология в английском и русском военных подъязыках : контрастивно-социолингвистический аспект / В. П. Коровушкин // Злая
- лая матерная... : [сб. ст. / под ред. В. И. Жельвиса]. – М. : Эдиториал, 2005. – С. 376–420.
6. Масенко Л. Суржик : між мовою і язиком / Л. Масенко. – К. : Вид. дім «Києво-Могилян. акад.», 2011. – 135 с.
7. Масенко Л. Нариси з соціолінгвістики : [навч. посіб.] / Л. Масенко. – К. : Вид. дім «Києво-Могилян. акад.», 2010. – 243 с.
8. Ставицька Л. Арго, жаргон, сленг : соціальна диференція української мови / Л. Ставицька. – К. : Критика, 2005. – 464 с.
9. Ставицька Л. Суржик : міф, мова, комунікація / Л. Ставицька, В. Труб // Українсько-російська двомовність. Лінгво-соціокультурні аспекти : [зб. наук. праць]. – 2007. – № 1. – С. 31–120.
10. Толстой Н. И. Славянские литературные языки и их отношение к другим языковым идиомам (стратам) (Опыт сравнительного рассмотрения) / Н. И. Толстой // Основные направления структурализма / [отв. ред. М. М. Гухман, В. Н. Ярцева]. – М. : Наука, 1964. – С. 9–24.

O. Demska

THE CONTENT OF THE GENERAL CORPUS: WORD ABOUT THE SURZHYK

The article deal with the general problem of the representation standard language, territorial and social dialects levels in the modern computer text corpus of the national language, with especially regard to the surzhyk.

Keywords: corpus, standard language, territorial dialect, social dialect, surzhyk.

Матеріал надійшов 11.06.2012