

DOI <https://doi.org/10.30525/978-9934-26-227-2-86>

**ON THE PROBLEM OF PARALLEL CORPORA DATABASE
(PUBLICISTIC STYLE)**

**ДО ПРОБЛЕМИ БАЗИ ДАНИХ ПАРАЛЕЛЬНИХ КОРПУСІВ
ТЕКСТІВ (ПУБЛІЦИСТИЧНИЙ СТИЛЬ)**

Safonova N. M.

*Candidate of Philological Sciences,
Associate Professor,
Associate Professor at the Department
of Applied Language Studies,
Comparative Linguistics
and Translation
National Pedagogical Dragomanov
University
Kyiv, Ukraine*

Сафонова Н. М.

*кандидат філологічних наук,
доцент,
доцент кафедри прикладної лінгвістики,
порівняльного мовознавства
та перекладу
Національний педагогічний університет
імені М. П. Драгоманова
м. Київ, Україна*

У лінгвістиці ступінь вагомості результатів дослідження напряму залежить від репрезентативності матеріалу дослідження, а саме: що більший обсяг фактажу, то більша вірогідність правильності шуканого результату, що більше мовних фактів вибрано, систематизовано, класифіковано й проаналізовано, то достовірніші спостережувані закономірності. Але традиційна форма збирання і систематизації інформації – загальновідома лексична паперова картотека – сьогодні не задовольняє потреб дослідників-філологів і через доступ до таких картотек, і через трудомісткість роботи з ними. А коли йдеться про двомовну чи багатомовну картотеку як основу або базу знань для паралельних корпусів, то паперовий аналог узагалі не розглядається. І тоді постає перед дослідником чи укладачем паралельного корпусу першорядне завдання оцифрування друкованих текстів, особливо якщо дослідник вивчає яесь мовне / мовленнєве явище діахронно, або використання текстів, які вже в однаковій мірі переведені в електронний формат.

Щодо другого способу сучасні комунікаційні ресурси спростили доступ до електронних текстів. Це поки що стосується не всіх функціональних стилів мовлення, але окремі з них отримати таку прерогативу. І вона очевидна.

Мета нашої наукової розвідки й конкретні дослідницькі завдання опосередковано полягають у вивченні ситуації стилістичного вияву

текстової інформації, але більшою мірою націлені на опис інформаційної галузі, яка творить масиви співвідносної в різних мовах інформації (власне текстів-оригіналів і їхніх перекладів), і вивчити спроможність таких текстових масивів ставати основою паралельних корпусів текстів.

Коли йдеться про корпус текстів, зокрема й паралельний, важливо тримати баланс стилістичної співвіднесеності текстів, залучених у його структуру. Це так звана збалансованість [1, с. 129] як конститутивна ознака корпусу.

І якщо ще донедавна вважали, що відсоток художніх текстів, тобто текстів художнього стилю, мусить бути чисельно більшим, то віднедавна (а йдеться про початок епохи постмодернізму) говорити про це вже не доводиться аж так однозначно, оскільки на перший план виходить уже не якийсь уявний, здебільшого вигаданий мовець (автор, ліричний герой, персонаж тощо), який мовить і діє у вигаданому світі, а конкретний мовець, жорстко прив'язаний до ситуації мовлення. Тобто, ми є свідками зміни корпусної текстової домінанти – з художнього стилю на інший.

Цим іншим стилем, за нашим спостереженням, стає публіцистичний стиль. Цей стиль не є одновимірним ні щодо жанрів, ні щодо підстилів, ні щодо форм і способів передачі інформації. Він має розгалужену систему форм (усна, писемна), способів / форматів (текст, відеоформат, аудіоформат), жанрів (стаття, замітка, допис, звіт, нарис, інтерв'ю, репортаж, рецензія, огляд, есе, лист, реклама, сатира) вияву.

Тексти, які формують цей стиль, за кількісним виявом можна структурувати за ядерно-периферійним принципом: найбільша кількість текстів окремого жанру / форми (допис, замітка, репортаж) становитимуть ядро стилю, менші кількісно – напівпериферію (стаття, огляд, інтерв'ю, есе, нарис, портрет, рецензія), а ті, кількість яких найменша, – периферію (реклама, лист, фейлетон, сатира). І ядром цього стилю є допис / репортаж, або новина, про подію, яка відбулася в конкретному темпорально-локально-мовному континуумі.

Таке спостереження є цілком логічним, позаяк у сучасному динамічному темпі подій, швидкоплинності новини як явища оперативність доведення інформації до загалу є визначальною, а тому так звані «короткі жанри» в пріоритеті. Інше спостереження стосується вже тріади «час – місце – мова»: ці складники взаємозалежні у трактуванні події, яка стала об'єктом репортажу. Але не лише вони вплинуть на остаточний зміст повідомлення. Втім, такі ознаки, як мета, наратив повідомлення залишаємо за межами нашої розвідки, як такі, що

вимагають іншого, більш глибокого аналізу із застосуванням методів і методик психолінгвістики і доробку когнітології.

Між мовою і корпусом текстів завжди стоїть посередник. Цей посередник – база даних, ресурс, який слугує джерелом наповнення майбутнього корпусу. Для паралельного корпусу текстів потрібна така база, яка одночасно, буквально в режимі реального часу, подаватиме тексти кількома мовами. При цьому, логічно, що оригінальний текст буде однією мовою, а решта текстів – перекладами цього ж тексту на інші мови із збереженням національно-мовної, граматичної й етнокультурної специфіки як мови-оригіналу так і з урахуванням перелічених характеристик мов-перекладу. Очевидно, що текст перекладу (переклад як продукт) – це певний баланс між тим, що написано мовою оригіналу, і тим, що читатимуть (мусять прочитати) не носії мови оригіналу, а ті, кому переклад призначено.

Далі поділимося ще одним спостереженням. Оскільки багато-мовний новинний сайт призначений для користувачів різних мов, які, очевидно, не володіють / не достатньо володіють мовою оригіналу, то переклади мусять враховувати не лише граматику / лексику / семантику кожної мови перекладу, а й передавати квант наративності тексту-оригіналу, культурної релятивності [2, с. 273].

Новинні сайти, що пропонують новини кількома мовами, наприклад, Українське національне інформаційне агентство «Укрінформ» <https://www.ukrinform.ua/info/contacts.html>, яке публікує новини в реальному часі вісьмома мовами, може бути непоганою базою для подальшого укладання паралельних корпусів текстів із тим лише застереженням, що цей корпус матиме назву «українсько-російський, або українсько-англійський, або українсько-іспанський, або українсько-німецький, або українсько-французький, або українсько-японський, або українсько-польський», із першим компонентом «українсько-», оскільки новинна стрічка формується спершу українською мовою, а потім відбувається літературний переклад на кожну із названих мов. При цьому не важко помітити, що неукраїнськомовні сторінки містять не всі новини, які подано українською мовою. Крім того, відсоток таких новин у різних мовах перекладу різний – від понад 20–35% перекладу оригінальних текстів англійською мовою, 10–20% німецькою, лише до 2–5% японською. Але оскільки будь-який корпус регулюється принципом відібраності [1, с. 129] як конститутивної ознаки, то цілком можливо використовувати й таку кількість для паралельного корпусу текстів.

Зупинимося на перевагах і недоліках таких баз для творення паралельних корпусів текстів. Незаперечні плюси, на наш погляд, більш вагомі, ніж мінуси, тому перелічимо спочатку їх:

1) такі бази даних можуть бути одночасно основою як двомовних паралельних корпусів текстів, так і багатомовних (у нашому випадку – аж восьмимовних) паралельних корпусів;

2) тексти на таких сайтах є сучасними, злободенними, і на їхньому прикладі легко простежуються мовотворчі процеси, поява нових слів і значень, виникнення нових словотвірних типів і моделей, розширення чи навпаки звуження семантики полісемантів, евфемізація чи дисфемізація тем / понять, виникнення фразеологізмів, розвиток семантики прецедентних феноменів тощо; така собі синхронія в реальному часі;

3) майже безкінечна можливість поповнення бази все новими й новими текстами;

4) широка тематика текстів бази (від політики й спорту до культури й науки), а отже, тяжіння до максимального охоплення найрізноманітніших лексико-граматичних ресурсів;

5) уже готовий цифровий (електронний) текстовий ресурс, який значно легше піддівати розмітці.

До недоліків таких ресурсів відносимо:

1) несистемність і спорадичність у виборі текстів перекладів;

2) повнота бази даних для багатомовних паралельних корпусів текстів залежить від мови, якою перекладено найменше текстів оригіналу;

3) фізична неможливість вивчення певного мовного явища в діячій діячій;

4) тексти великих жанрів (звична всім аналітика) публіцистичного стилю рідко коли підлягають перекладу.

Вивчення кількомовних інформаційних сайтів наштовхує на ще один висновок: усі тексти без винятку є гіпертекстово зв'язаними і мають активні відсилки до інших взаємопов'язаних текстів. А це ще один новий ракурс дослідження таких баз даних: здатність / нездатність текстів-оригіналів і текстів-перекладів симетрично відносити до інших текстів.

Отже, інформаційні новинні ресурси є прекрасним джерелом наповнення баз даних текстів публіцистичного стилю для паралельних корпусів текстів, а багатомовні сайти можуть якісно оптимізувати роботу дослідника.

Література:

1. Демська О. Вступ до лексикографії. Київ : Вид. дім «Києво-Могилянська академія, 2010. 266 с.
2. Селіванова О. О. Сучасна лінгвістика : термінологічна енциклопедія. Полтава : Довкілля-К, 2006. 716 с.

DOI <https://doi.org/10.30525/978-9934-26-227-2-87>

**REALITIES IN UKRAINIAN TRANSLATIONS
OF POLISH FAIRY TALE TEXTS**

**РЕАЛІЇ В УКРАЇНСЬКИХ ПЕРЕКЛАДАХ
ПОЛЬСЬКИХ КАЗКОВИХ ТЕКСТІВ**

Stanislavova L. L.

*Candidate of Philological Sciences,
Associate Professor at the Department of
Slavic Philology
Khmelnyskiy National University*

Станіславова Л. Л.

*кандидат філологічних наук,
доцент кафедри слов'янської філології
Хмельницький національний
університет*

Serkova Yu. I.

*Senior Teacher at the Department of
Slavic Philology
Khmelnyskiy National University,
Khmelnyskiy, Ukraine*

Серкова Ю. І.

*старший викладач кафедри
слов'янської філології
Хмельницький національний
університет
м. Хмельницький, Україна*

Казки народів світу мають багато спільних рис – на рівні тем, мотивів казок, їх дидактичної мети. Це наслідок того, що бажання, надії і мрії усіх народів досить схожі. Казка – це розповідь про вигадані, фантастичні чи авантюрні події. Кінцівка є переважно оптимістичною: добро перемагає зло. Але казки кожного народу відбивають і певні особливості історії, природних умов життя, праці і побуту, менталітету цього народу. Тому оригінал казкового тексту зазвичай містить національний колорит, який формується за рахунок використання національних реалій. Реалії є мовними одиницями, які позначають елементи «чужої» культури, мають національне, історичне, місцеве або побутове забарвлення і не мають еквівалентів в інших мовах і культурах.