

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

ISSN 0320-3077

УКРАЇНСЬКЕ МОВО- ЗНАВ- СТВО

Міжвідомчий
науковий
збірник

Засновано у 1973 р.

Випуск 43

У збірнику вміщені дослідження з актуальних проблем комп'ютерної лінгвістики.
Для викладачів, науковців, учителів, студентів.

This volume presents investigations on topical issues in Computational Linguistics.
It is intended for lecturers, researchers, teachers, and students.

РЕДАКЦІЙНА КОЛЕГІЯ	А. К. Мойсієнко, д-р філол. наук, проф. (відп. ред.); І. М. Арібжанова, канд. філол. наук, доц. (відп. секр.); О. В. Бас-Кононенко, канд. філол. наук, доц.; П. І. Білоусенко, д-р філол. наук, проф.; Л. П. Гнатюк, д-р філол. наук, доц.; Н. В. Гуїванюк, д-р філол. наук, проф.; С. Я. Єрмоленко, д-р філол. наук, проф.; А. П. Загнітко, д-р філол. наук, проф.; І. В. Козленко, канд. філол. наук, доц. (заст. відп. ред.); Ю. Л. Мосенкіс, д-р філол. наук, проф.; Ю. Ф. Прадід, д-р філол. наук, проф.; О. С. Снитко, д-р філол. наук, проф.; В. Ф. Чемес, канд. філол. наук, доц.
Адреса редколегії	01601, Київ-601, б-р Т. Шевченка, 14, Інститут філології; ☎ (38044) 239 33 49, 239 33 54
Рекомендовано	Ученою радою Інституту філології 06.10.2012 року (протокол № 2)
Зареєстровано	Постановою президії ВАК України від 14.04.2010, пр. № 1-05/3
Засновник і видавець	Київський національний університет імені Тараса Шевченка Видавничо-поліграфічний центр "Київський університет" Свідоцтво внесено до Державного реєстру ДК № 1103 від 31.10.02

Автори опублікованих матеріалів несуть повну відповідальність за підбір, точність наведених фактів, цитат, економіко-статистичних даних, власних імен та інших відомостей. Редколегія залишає за собою право скорочувати й редагувати подані матеріали. Рукописи та дискети не повертаються.

© Київський національний університет імені Тараса Шевченка,
Видавничо-поліграфічний центр "Київський університет", 2013

УДК 81'322

*Марія Шведова,
Дмитро Січінава*
Інститут російської мови Російської академії наук

КОРПУСНА ЛІНГВІСТИКА ТА ЛЕКСИКО-ГРАМАТИЧНА ТИПОЛОГІЯ¹

Наводяться приклади застосування паралельних корпусів Національного корпусу російської мови для розв'язання завдань зіставлення граматики та лексики у перекладних текстах.

Ключові слова: корпусна лінгвістика, паралельний корпус, граматична типологія, лексична типологія, семантика.

Моделі та стимули перекладу

Є багато можливостей практичного застосування Національного корпусу російської мови (НКРМ, <http://ruscorpora.ru>) в галузі зіставної граматичної та лексичної семантики. Базою для міжмовних (типологічних і контрастивних) досліджень у НКРМ є передусім паралельний корпус – корпус, який складається з оригінальних та перекладних текстів, що вирівняні по реченнях. До складу Національного корпусу російської мови за станом на січень 2013 р. входять 8 двомовних паралельних корпусів: англійсько-російський (та, відповідно, російсько-англійський), німецько-російський, французько-російський, іспансько-російський, італійсько-російський, польсько-російський, українсько-російський та білорусько-російський (для кожної мовної пари подані також тексти, що перекладені у зворотному напрямку) [Добровольський, Кретов, Шаров 2005], [Андреева, Касевич 2005], [Добровольський 2009, Січінава, Шведова, Тищенко-Монастирська 2011], [Sitchinava 2012]. Сукупний обсяг цих корпусів перевищує

¹ Стаття підготована за підтримки програми Президії РАН №36-П “Корпусна лінгвістика” та гранту РФФІ №12-06-33038 “Контрастивні корпусні дослідження російських та французьких дієслівних категорій у поліваріантних паралельних текстах”.

37 мільйонів слововживань, і розвиток їх продовжується. На стадії проекту знаходяться ще кілька двомовних корпусів за участю російського (вірменський, латиський, угорський, татарський та інші). Кожна мова паралельного корпусу має морфологічну розмітку, що дозволяє здійснювати точний пошук за словоформою, пошук за лексемою та/або набором граем, а також за різними комбінаціями цих параметрів.

Крім того, існує багатомовний корпус на базі 25 мов, здебільшого слов'янських, який містить 9 паралельних текстів (граматична розмітка є вже не для кожної мови); його обсяг – 5 мільйонів слововживань. Заплановано створення поліваріантних корпусів, до яких входить кілька різних перекладів тексту на одну мову; нині триває робота над створенням російсько-французького поліваріантного паралельного корпусу.

Тексти вирівнюються по реченнях; основою виділення співвідносних частин є поділ на речення в оригінальному тексті (у перекладі він часто буває іншим). Слід зауважити, що за радянської доби художній переклад із російської мови на мови народів СРСР і навпаки (на сьогодні саме перекладів цього часу найбільше в українсько-російському та білорусько-російському паралельних корпусах) був досить приблизним на різних рівнях, при наповненні та аналізі паралельного корпусу це доводиться постійно враховувати. Причини цього є різними: загальна настанова певних перекладачів, роль автора у створенні авторизованого перекладу (зокрема, він і сам може бути перекладачем), а нерідко й цензурні вимоги. Довільність та різні помилки фіксуються і в перекладах за участю інших мов, хоча й менш систематично. Для зазначення тієї чи іншої неточності в перекладі – пропуску, додавання, заміни тексту зі зміною змісту – у корпусі використовується спеціальний рівень розмітки – атрибут **loose**. Наприклад, тег **omit** використовується на позначення значного пропуску в перекладі (приклад із повісті Л. Толстого “Смерть Івана Ільича” у французькому перекладі М. Оффмана; не поданий у перекладі фрагмент виділений).

<para id="353">

<se lang="ru">Было лучшее общество, и Иван Ильич танцевал с княгиней Труфоновой, сестрою той, которая известна учреждением общества "Унеси ты мое горе".</se>

<se lang="fr" loose="omit">Ivan Illitch dansa avec la princesse Troufonova, la propre sœur de l'illustre fondatrice de la société "Foin de peines !".</se>

</para>

Використання паралельних корпусів у порівняльних дослідженнях є предметом багатьох праць, назвемо, передусім, спеціальний номер часопису STUF за редакцією М. Сісоува та Б. Вельхлі [Cysouw, Wälchli (eds.) 2007]; див також зазначені вище статті [Андреева, Касевич 2005] і [Добровольский 2009]. При перекладі лексики, як і граматичних конструкцій (чітку межу між ними не завжди можна провести), суттєву роль відіграють **моделі перекладу та стимули перекладу**. Зазвичай для конструкції А у мові В існують (реально представлені в текстах) кілька відповідників, які повторюються не випадково, – моделей перекладу (translation patterns; пор. [Hasselgård, Oksefjell (eds.) 1999]). Практично можна стверджувати, що, коли є значний за обсягом корпус перекладних текстів, можна виокремити кілька моделей перекладу конструкції мови А у мові В.

Стимулом перекладу назвемо конструкції в мові оригінальних текстів, що відповідають певним конструкціям у мові перекладу ("що стимулює появу конструкції Х у перекладі на мову В?").

Відповідники граматичних конструкцій

У праці [Сичинава 2011] представлено результати статистичного дослідження стимулів перекладу в англійських та німецьких текстах для конструкції з часткою *было*, типу *пошёл было, да вернулся*. Було виявлено цікаві стійкі відповідники, причому не тільки і не стільки граматичні, скільки лексичні. Наприклад, у німецькій – частки *eben* та *gleich* 'рівно', в англійській – прислівники нетривалої дії типу *for a while* 'деякий час'. З огляду на ці відповідники можна чимало сказати про семантику російської конструкції (яка не має точного відповідника в англійській та німецькій мовах і переважно позбавлена будь-якого лексичного

стимулу при перекладі). Те ж саме стосується і відповідників іншомовних граматичних категорій у російській мові. Так, перфект прогресива в англійській мові (*I have been doing*) передається російським недоконаним видом (теперішнього або минулого часу) в актуально-тривалому значенні. Разом із цим зазначено такий лексичний засіб для передавання перфекта прогресива, як дієслова типу *начать*, *решить*, які вказують на актуальність для теперішнього моменту початкової межі ситуації:

Perhaps people *have been celebrating* Bonfire Night early – it's not until next week, folks! [Joanne Kathleen Rowling. Harry Potter and the Sorcerer's Stone (1997)]

Кажется, народ уже начал *праздновать* день Порохового Заговора – рановато, господа, он будет только на следующей неделе! [Дж. К. Роулинг. Гарри Поттер и Волшебный камень / М. Спивак].

Well, it's a new thing the boss *has been trying*. [Stephen King. The Lawnmower Man (1975)]

Это новая методика, которую решил *испробовать* наш босс. [Стивен Кинг. Газонокосильщик / пер. не указанных]

Відповідники французького часу *passé antérieur* у російських текстах послуговуються додатковими лексичними засобами: *лишь, немедленно, сразу*:

Quand Passepartout *eut fini*, il se trouva calme et comme soulagé. [Jules Verne. Le tour du monde en quatre-vingt jours (1872)]

Отведя душу, Паспарту сразу успокоился и почувствовал большое облегчение. [Жюль Верн. Вокруг света за восемьдесят дней / Н. Габинский, Я. Лесюк]

За допомогою паралельного корпусу можна виявити особливості вживання подібних граматичних конструкцій у близькоспоріднених мовах. Так, для російської мови не характерна конструкція “*было* + модальне дієслово” (вирази *могло было, следовало было* замість *могло бы, следовало бы* є помилковими). Натомість сполуки з дієсловом *могли* у давноминулому часі, який етимологічно відповідає російській конструкції, нерідко можна

зустріти в західноукраїнських дореволюційних письменників – І. Франка, О. Мартовича, О. Кобилянської. Наприклад:

І нікого не було тут, кому б *міг був* повірити свій жаль. [Ольга Кобилянська. Земля (1901)]

И никого не было тут, кому он *мог бы* поверить своё горе. [Ольга Кобилянская. Земля / пер. не вказаний].

Із дієсловами повинності (“долженствования”) така конструкція в українській мові представлена набагато ширше:

Настал день, когда электронный скальпель *должен был* коснуться разумного мозга... [Б. В. Зубков, Е. С. Муслин. Исповедь после смерти (1966)]

Настав день, коли електронний скальпель *мав був* торкнутися розумного мозку... [Б. В. Зубков, Е. С. Муслин. Німа сповідь / пер. не вказаний]

...в остаточній редакції твою одчайдушну телеграму *мусили були* прийняти без цих двох слів. [Оксана Забужко. Польові дослідження з українського сексу (1996)]

...в окончательной редакции твою отчаянную телеграмму *должны были* принять без этих двух слов [Оксана Забужко. Полевые исследования украинского секса / Е. Мариничева].

У мовленні письменника-білінгва Г. Квітки-Основ'яненка є цікавий приклад граматичної інтерференції. Квітка сам переклав російською мовою свою повість “Конотопська відьма” і відповідно до українського давноминулого часу вжив модальну конструкцію з *было* (в російській літературній мові зустрічалася у XVIII ст.), замінивши лише дистантне положення показника контактним:

Ище *было* уп'ятерить *подобало* за таковое злодіяніе... [Конотопська відьма (1833)].

Еще *было подобало* упятерить за таковое злодеяние...

Приклади подібної інтерференції між оманливо близькими східнослов'янськими мовами не такі рідкісні, і в неавторському

перекладі також (зокрема, помітний вплив білоруських конструкцій та лексики проглядається в деяких радянських перекладах із білоруської на російську).

Лексичні відповідники

Дослідження моделей та стимулів перекладу в реальному корпусі є, зокрема, інструментом верифікації тверджень про “мовну специфіку” тих чи інших лексем, їх “перекладуваність” тощо. Очевидно, найбільш “специфічний” із семантичного погляду лексичні кожної мови відповідатиме найбільший діапазон моделей і стимулів перекладу – оскільки “головний” відповідник, який підходить до всіх або майже всіх контекстів, в іншій мові відсутній. Так, лексеми, що їх деякі дослідники відносять до “російської мовної картини світу”, такі, як *тоска*, *душа* або *удаль*, мають досить широкий діапазон варіантів перекладу на англійську мову.

Наприклад, кожен із наведених англійських перекладів слова *удаль* у паралельному корпусі поки що трапився по одному разу: *abandonment, assurance, violence, effrontery, jockeying, bravado, daring, fearless confidence*. При цьому характерно, що в перекладах з англійської на російську слово *удаль* та його похідні майже не вживаються. Англійське слово *soul* перекладається практично однозначно як російське *душа* (крім фразеологізмів типу *poor soul*), а російське *душа* може перекладатися по-різному. Так, у контекстах типу *в глубине души* це слово передається англійським *mind*. Цікавим є приклад із *soul* в англійському тексті і *сознание* в російському:

And, all through that silent drive back to Green Street, the *souls* of both of them revolved a single thought: ‘Why, oh! why should I have to expose my misfortune to the public like this? [John Galsworthy. In *Chancery* (1920)]

И все время, пока они молча ехали на Грин-стрит, у обоих *в глубине сознания* неотступно вертелась одна и та же мысль: “Почему, ах, почему мне приходится вот так выставлять напоказ мои невзгоды? [Джон Голсуорси. В петле / М. Богословская]

Перекладні моделі та стимули для слова *тоска* надзвичайно різноманітні (наводимо переклади не тільки за допомогою імен-

ника: *gloom, dismay, grief, agony, excruciating feelings, distracted moods, anguish, ache, despair, woe, loneliness, distress, frustration, desperation, blue, misery, nuisance, sadness, nostalgia, yearning, longing, lust, missing, glumly, sulkily, wistfully*).

You must be as *blue* as I am about our nation's great unvictory. [Kurt Vonnegut. *Hocus Pocus* (1990)]

Значит, на вас нагнала таку ж *тоску*, как и на меня, великая пиррова победа нашего народа. [Курт Воннегут. *Фокус-покус* / М. Ковалева]

А. Вежбицька вважає найбільш прийнятним, хоч і приблизним, до російського *тоска* англійський відповідник *yearning*. Водночас у корпусі “спектр” відповідників англійського слова *yearning* виявився зміщеним у бік семантики палкого бажання (*желание, жажда, тянуть, томление, томный, жадность, томило желание*).

Так само значна міжмовна варіативність моделей перекладу спостерігається й між близькоспорідненими мовами, наприклад російською та українською. Незважаючи на те, що обсяг паралельного українсько-російського корпусу майже вдвічі менший за англійсько-російський, для російського слова *тоска* в українській знаходимо також значну кількість еквівалентів-іменників (за винятком застарілих): *туга, жаль, жалощі, смуток, нудьга, нудота, сум, журба, скорбота, гризота, зневага, тоскність* (в 11-томному “Словнику української мови” має ремарку “рідковживане”), *скруха*. Крім того, відповідниками до російського *тоска* у предикативній функції виступають українські прислівники на -о: *тоскно, тужливо, тужно, журно, нудно*.

За допомогою корпусу можна дати хронологічну оцінку таким синонімічним рядам. Деякі українські еквіваленти до російського *тоска*, які нерідко зустрічаються в ранніх літературних текстах, не вживані в сучасних: це “застаріле” (за СУМ) *зануда*, діалектизми *жель* (не поданий в СУМі, хоч у західноукраїнському мовленні функціонує досі), *жура, туск*, а також слова *печаль, тоска*, які графічно однакові з російськими. Зауважимо, що у творах визначного прозаїка 1920–1930-х років, опонента

русифікації Миколи Хвильового основним засобом вираження цього значення є саме українське слово *тоска*.

Щодо слова *душа*, – яке є і в російській, і в українській, – то у перекладі фразеологізмів міжмовна варіативність досить відчутна: *душой стремился – прагнув, болит душа – уболіває, кто болеет душой – кому болить; горела душа – пекло; Оленька-душа – Оленька-серце, душа моя – серце; до глубины души – доглибно, от души – щиро, з щирим почуттям, в глубине своей собачьей души – в глибині свого собачого ества*.

Корпус можна використати у процесі дослідження синонімів. Наприклад, російські слова *влажный* і *мокрый* та їхні англійські відповідники розрізняються відповідно до означуваного об'єкта:

(долоні, руки) *damp, wet, moist* (від поту) => *влажный*
(трава, земля, пісок) *damp, humid, wet, soaking* => *влажный, мокрый*
(одяг) *damp, humid, soaking* => *мокрый*.

Семантична несумісність синонімів із означуваним об'єктом зумовлена авторським баченням перекладу:

Trees, forced by the *damp heat*, found too little soil for full growth, fell early and decayed: creepers cradled them, and new saplings searched away up. [William Golding. *Lord of the Flies* (1954)]

Мокрой жарой деревья толкало в рост, но на хилом слое почвы они не заживались, рано валились и гнили; их обнимали лианы, сквозь них пробивались новые ростки. [Уильям Голдинг. Повелитель мух / Е. Суриц]

У російській мові *мокрая жара* – сполука на межі допустимості; у 230-мільйонному одномовному корпусі НКРМ вона зустрілася лише один раз у контексті *ню-йоркская “мокрая жара”*, де вона оформлена як елемент “чужої мови”.

Таким чином, за допомогою паралельного корпусу можна швидко отримувати велику кількість реальних перекладацьких рішень, що були прийняті носіями мови при створенні перекладу, й аналізувати виявлені відповідники в лексиці та граматиці. Завдання їх інтерпретації (оцінка достовірності відповідників,

інтерференції та помилок, аналіз контексту, виокремлення різних значень тощо), звичайно, залишається справою дослідника. Однак обсяг попередньої підготовчої роботи за допомогою паралельного корпусу може бути значно скорочений.

1. Андреева Е. Г. Грамматика и лексика (на материале англо-русского корпуса параллельных текстов) / Е. Г. Андреева, В. Б. Касевич // Национальный корпус русского языка: 2003–2005. – М., 2005. – С. 297–307.
2. Добровольский Д. О. Корпус параллельных текстов: архитектура и возможности использования / Д. О. Добровольский, А. А. Кретов, С. А Шаров // Национальный корпус русского языка: 2003–2005. – М., 2005. – С. 263–296.
3. Добровольский Д. О. Корпус параллельных текстов в исследовании культурно-специфичной лексики / Д. О. Добровольский // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. – СПб., 2009. – С. 383–401.
4. Сичинава Д. В. Комплексное исследование одноязычного и параллельного корпусов в грамматических исследованиях / Д. В. Сичинава // Труды Международной конференции “Корпусная лингвистика–2011”. – СПб. – С. 316–332.
5. Сичинава Д. В. Паралельні українсько-російський та російсько-український корпуси / Д. В. Сичинава, О. О. Тищенко-Монастирська, М. О. Шведова // Лексикографічний бюлетень, 20. – К., 2011. – С. 35–38.
6. Parallel Texts. Using Translational Equivalents in Linguistic Typology. Theme issue in Sprachtypologie & Universalienforschung STUF 60.2. / eds. Cysouw M., Wälchli B. – 2007.
7. Out of Corpora: Studies in Honour of Stig Johansson / H. Hasselgård, S. Oksefjell. – Amsterdam ; Atlanta, 1999.
8. Sitchinava D. Parallel corpora within the Russian National Copus / D. Sitchinava // Prace Filologiczne. – LXIII. – 2012. – S. 271–278.

Приводятся примеры применения параллельных корпусов Национального корпуса русского языка для решения задач сопоставления грамматики и лексики в переводных текстах.

Ключевые слова: корпусная лингвистика, параллельный корпус, грамматическая типология, лексическая типология, семантика.

The paper deals with some examples of using the parallel corpora within the Russian National Corpus for the cross-linguistic analysis of grammar and lexicon in translated texts.

Keywords: corpus linguistics, parallel corpus, grammatical typology, lexical typology, semantics.

Стаття надійшла до редакції 1.09.2012