міркування пояснюють доцільність створення Браунського корпусу української мови й слугують дороговказом для його укладачів.

Спираючись на досвід «Браунської родини корпусів» й пристосовуючи їх до особливостей побутування української мови, окреслімо принципи наповнення Українського Браунського корпусу. Попри те, що бажано дотримуватися структури оригінального Браунського корпусу, вважаємо, що варто допустити певні відхилення, як це зробили, наприклад, болгарські лінгвісти. Сформулюймо вимоги до текстів у корпусі.

1. Оригінальні (тобто неперекладні) твори.
2. Твори, створені й опубліковані за відносно короткий проміжок часу (до 10 років).
3. За змоги тексти мають бути зредаговані, окрім тих, щодо яких цю ознаку в принципі неможливо встановити.
4. Підбір текстів за категоріями й підкатегоріями має відповідати первісному Браунському корпусу. Виняток становить категорія F, у якій жанр «вестерн» можна замінити іншою пригодницькою літературою, представленою в Україні.
5. Корпус складається з 500 фрагментів довжиною 2000 слів плюс залишок до кінця речення.
6. Фрагмент повинен бути витяжкою з одного тексту, окрім випадків, коли фрагмент складається з кількох коротких текстів кількох авторів (короткі новини).
7. Тексти мають бути підібрані згідно з класифікацією за типом (інформативні й художні тексти), категорією, підкатегорією. Браунський корпус має 15 категорій, поділених на різну кількість підкатегорій [2].

Порівняння Браунського корпусу й ЛОБ засвідчує, що немає потреби (а іноді й змоги) копіювати структуру взірцевого корпусу до найменших дрібниць. Наприклад, співвідношення між книжками й періодичними виданнями в межах підкатегорії чи щоденними й тижневими періодичними виданнями (в категоріях A–C) може різнитися.

Отже, створення Українського Браунського корпусу за вказаними параметрами дасть змогу створити корисний дослідницький і навчальний ресурс. Інші принципи побудови корпусу варто розглянути окремо.

Література

1. http://dcl.bas.bg/Corpus/home_bg.html – режим доступу: 19.01.2012 р.
2. http://icame.uib.no/brown/bcm.html – режим доступу: 19.01.2012 р

Богдан Шуневич
Львівський державний університет безпеки життєдіяльності

Українсько-англійський комп'ютерний словник пожежно-технічних термінів: лексичні матеріали, програмне забезпечення

Зарубіжний і вітчизняний ринок програмного забезпечення пропонує велику різноманітність комп'ютерних словників. Серед відомих українських комп'ютерних

словників можна назвати, наприклад, інтегровану лексикографічну систему «Словники України» Інституту мовно-інформаційних досліджень НАН України, систему електронних навчальних словників «ГЛОСА» та ін. Лабораторії комп'ютерної лінгвістики Київського національного лінгвістичного університету, електронний багатотематичний тлумачний словник MultiLock галузевого Нормативно-термінологічного центру нафтогазового комплексу, системи PolyDic v. 1.0, PolyDic ML 3.0 Технічного комітету стандартизації науково-технічної термінології Держспоживстандарту та Міністерства освіти і науки, молоді та спорту України.

Комп'ютерним словником називають «словник, процедури укладання якого здійснює комп'ютер» [1].

Мета доповіді – провести порівняльний аналіз програмного забезпечення PolyDic v. 1.0 [2], за допомогою якого укладається «Англійсько-український комп'ютерний словник з робототехніки» [3], і лінгвістичної бази даних Військового інституту Київського національного університету імені Тараса Шевченка [4], яка запропонована викладачам кафедри іноземних мов та технічного перекладу Львівського державного університету безпеки життєдіяльності для укладання «Українсько-англійського комп'ютерного словника пожежно-технічних термінів» [5].

Для укладання «Англійсько-українського комп'ютерного словника з робототехніки" нами використовується програмне забезпечення з вільним кодом (open source), а саме PolyDic, версія 1.0, яке укладено під керівництвом Романа Мисака (Національний університет «Львівська політехніка») [2].

Система укладання та перегляду електронних словників PolyDic складається з двох програм: PolyDic Editor – для набирання, форматування та редагування словникових баз даних та PolyDic Viewer – для перегляду електронних словників.

Пошук терміна у PolyDic Editor відбувається за першими уведеними літерами в інтерактивному режимі у вікні пошуку. Програма запам'ятовує історію (почерговість) переглянутих статей, в якій зберігається послідовність до десяти переглянутих статей. Особливістю PolyDic та її істотною перевагою над іншими електронним словниками є механізм фільтрів, який можна застосувати до термінів, тексту статей або до них обох одночасно. Наприклад, можна залишити видимими слова, які починаються на "абр", або слова, що мають закінчення "ан". У статтях, у разі потреби, встановлюються зв'язки з іншими статтями, які можна переглянути просто натиснувши курсором миші по відмітці зв'язку. Переклади або тлумачення супроводжуються короткими поясненнями, наприклад, щодо галузі застосування.

У програмі PolyDic Editor передбачено можливість набору словникової бази частинами з подальшим злиттям цих частин в єдину базу. Ця функція корисна під час розподілення праці з уведення інформації у базу між різними операторами-користувачами.

Система PolyDic v. 1.0 розроблена для формування комп'ютерних словників, паперові версії яких уже були укладено або видано, і їх макро- та мікроструктура повністю відповідала паперовим версіям. До недоліків цієї системи можна віднести: обмежену кількість мов перекладних словників (дві); вбудований в систему шрифт не підтримує Unicode і не дає змогу вводити літери з діакритичними знаками; система не підтримує мультимедійні об'єкти.

«Українсько-англійський комп'ютерний словник пожежно-технічних термінів» заплановано створити в рамках лінгвістичної бази даних Військового інституту Київського національного університету імені Тараса Шевченка.

Метою розроблення бази даних є створення загальнодоступного і високоякісного порталу з військової і, в тому числі, пожежно-технічної термінології, а також уніфікація і стандартизація військово-технічних термінів.

На порталі представлена вся необхідна інформація про термін (переклад, пояснення, фото та відео матеріали). Користувачі можуть не тільки користуватися термінологічною базою, а й одночасно приймати активну участь в її розширенні та поліпшенні шляхом додавання відсутньої інформації.

Проект передбачає створення багатомовної термінологічної бази даних військово-технічних термінів, поки що англійською, німецькою, французькою. російською та українською мовами.

Досвід укладання комп'ютерних словників дасть можливість апробувати вітчизняне програмне забезпечення для укладання комп'ютерних словників, а також порівняти різні параметри цих словників, вибрати кращий варіант програмного забезпечення для подальшої словникової роботи в нашому та інших університетах.

Література

1. Карпіловська Є. А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика. – Донецьк: ТОВ «Юго-Восток, Лтд», 2006. – 188 с.
2. Система укладання та перегляду електронних словників PolyDic v.1.0. – Режим доступу до Веб-сторінки: www.lp.lviv.ua. – Заголовок з екрана, 2011.
3. Шуневич Б. Проект англійсько-українського комп'ютерного словника з робототехніки / Б. Шуневич, В. Голтвян, М. Маляр // Лінгвістичні проблеми та інноваційні підходи до викладання чужоземних мов у вищих навчальних закладах, м. Львів, 28-30 жовтня 2010 р. – Львів: ЛДУ БЖД. – С. 83.
4. Лінгвістична база даних Військового інституту Київського національного університету імені Тараса Шевченка. – Режим доступу до Веб-сторінки: http://www.mildic.com/admin. – Заголовок з екрана, 2011.
5. Короткий українсько-англійський словник зі сфери надзвичайних ситуацій / Вовчаста Н.Я., Бугайська О.В. та ін. (За ред. Коваля М., Шуневича Б.). – Львів: Вид-во ЛДУ БЖД, 2010. – 184 с.

Катерина Яковенко
Київський національний університет імені Тараса Шевченка

Створення лінгвістичного корпусу у міжмовних експериментально-фонетичних дослідженнях

Питання створення лінгвістичного корпусу постає перед кожним дослідником на одному з перших етапів його мовознавчих студій. Правильний підбір корпусу даних є запорукою успішності експерименту, адже це матеріал, на основі якого базуватимуться всі подальші висновки і від якого залежатиме істинність отриманих

кінцевих результатів, а також можливість їх використання як достовірного джерела для майбутніх досліджень.

Інформаційна база даних, що слугує основою для експериментів у галузі експериментальної фонетики, суттєво різниться від лінгвістичного корпусу для аналізу граматичних явищ, функціональних аспектів мовних одиниць або вивчення статистичних параметрів мови. Підставою для цього є не лише якісно інший об'єкт і предмет дослідження, але й чинники, що мають бути враховані при його аналізі. Так, якщо зазвичай під поняттям «лінгвістичний корпус» розуміють масиви текстів, то у фонетиці мова йде про масив слів, об'єднаних відповідним до мети фонетичного дослідження принципом.

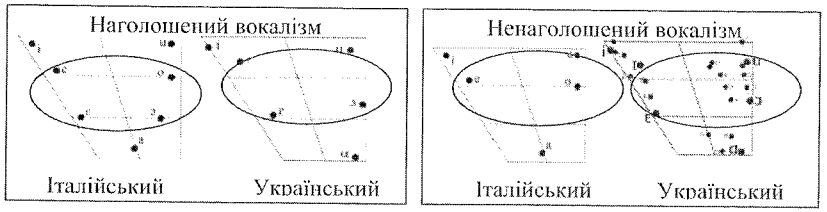
Детальний опис особливостей проведення фонетичних досліджень, практичні поради щодо створення лінгвістичного корпусу, вибору мовців, систем аудіозапису та подальшого аналізу даних подає американський фонетист Пітер Ладефогед у праці «Аналіз фонетичних даних: вступ до експериментально-фонетичних досліджень та інструментальних технологій».

Корпус для міжмовних фонетичних студій і вивчення явища іншомовного акценту має свою специфіку. Так, у ході експериментально-фонетичного дослідження італійського мовлення українців з метою вивчення особливостей засвоєння чужомовного вокалізму, були проаналізовані критерії, які необхідно враховувати при виборі мовного матеріалу для акустичного запису. У їхній основі лежить суть самого явища іноземного акценту, що полягає в накладанні фонетичної системи рідної мови на іноземну за таким алгоритмом:

- звуки та їхні опозиції в іноземній мові (М2), ідентичні або «схожі» на звуки рідної (М1), будуть замінені звуками рідної мови (М1):
$[a_1 \approx a_2 \rightarrow a_1]$, $[a_1{:}b_1 \approx a_2{:}b_2 \rightarrow a_1{:}b_1]$;
- опозиції звуків іноземної мови (М2), відсутні або «нові» для рідної (М1), не дотримуватимуться: $[a_2{:}b_2 \neq M1 \rightarrow ?]$;
- опозиції звуків рідної мови (М1), відсутні в іноземній (М2), все одно зберігатимуться в іноземній мові (М2): $[a_1{:}b_1 \neq M2 \rightarrow a_1{:}b_1]$.

Важливо наголосити на спостереженні за процесом засвоєння звуків, позначених як «нові» і «схожі»: якісно нові звуки іноземної мови засвоюються неносіями мови значно швидше, ніж схожі. Це пояснюється складністю перцепції тонкої межі між акустично подібними чужомовними і вже існуючими у фонетичній базі мовця фонемами, а часом і відсутністю їхнього графічного розрізнення на письмі.

Так, наприклад, досліджуючи засвоєння італійського вокалізму українцями, спочатку варто проаналізувати самі системи вокалізму обох мов у їхніх наголошених та ненаголошених позиціях (використання символів МФА).



НАВЧАЛЬНІ СЛОВНИКИ,
УКЛАДЕНІ В ЛАБОРАТОРІЇ КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ
КИЇВСЬКОГО НАЦІОНАЛЬНОГО ЛІНГВІСТИЧНОГО УНІВЕРСИТЕТУ

МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЛІНГВІСТИЧНИЙ УНІВЕРСИТЕТ

«КОМП'ЮТЕРНА ЛІНГВІСТИКА: СУЧАСНЕ ТА МАЙБУТНЄ»

Матеріали міжнародної науково-практичної конференції

КИЇВ 2012

ОРГАНІЗАТОРИ: Київський національний лінгвістичний університет
Лабораторія комп'ютерної лінгвістики

РЕКОМЕНДОВАНО ДО ВИДАННЯ рішенням Вченої ради КНЛУ
(Протокол № 6 від 30 січня 2012 р.)

Технічний редактор, комп'ютерна верстка Льон Олександр Васильович

Комп'ютерна лінгвістика: сучасне та майбутнє. Матеріали Міжнародної науково-практичної конференції – К.: КНЛУ, 2012.– 52 с.

Збірник сформовано за матеріалами Міжнародної науково-практичної конференції: «Комп'ютерна лінгвістика: сучасне та майбутнє» (23-24 лютого 2012 р., Київський національний лінгвістичний університет м. Київ).

Валентина Перебийніс, Тетяна Бобкова
Київський національний лінгвістичний університет

Історія лабораторії комп'ютерної лінгвістики КНЛУ

Лабораторія комп'ютерної лінгвістики КНЛУ була створена у 2002 році з ініціативи Сергія Микитовича Назарова, тодішнього проректора з наукових питань КДЛУ. Напрямок досліджень лабораторії – комп'ютерна навчальна лексикографія, укладання перекладних навчальних словників. За 10 років у лабораторії укладено серію англо-українських й україно-англійських навчальних словників для початкового та середнього рівня викладання англійської мови – всього 4 словники, при цьому словники першого рівня витримали три видання у паперовому форматі, а в комп'ютерному форматі англо-український словник має звуковий супровід.

Крім перекладних словників у лабораторії укладено словники-довідники «Труднощі англійського слововживання для українців» та «Морфологія англійського дієслова: система та функціонування», де наводяться дані про словозмінну парадигму англійського дієслова та про вживаність кожної словозмінної форми біля 300 найуживаніших дієслів у сучасній художній прозі, драмі, наукових та суспільно-політичних текстах. Всі ці словники можна побачити на книжковій виставці, організованій бібліотекою КНЛУ. Ще однією фундаментальною працею лабораторії є навчально-методичний комплекс, який складається з англо-українського та україно-англійського навчальних словників з методичними коментарями для другого рівня навчання англійської мови. Він існує в комп'ютерному форматі, планується його видання у паперовому форматі. В лабораторії укладаються не лише навчальні словники. Так, на замовлення московського видавництва «Астрель» укладено «Російсько-український розмовник» і «Російсько-український словник» на 210 тисяч статей. Т.В. Бобкова уклала англо-українсько-російський словник термінів з комп'ютерної лінгвістики, що містить переклад і тлумачення близько 1000 термінів.

Значну увагу приділяють в лабораторії використанню та створенню корпусів текстів. Так, на матеріалі мільйонного корпусу документів НАТО укладено ЧС на замовлення Міністерства освіти і науки України (2007 р.). Розроблено кілька корпусів: Багатомовний корпус субтитрів до кінофільмів – розробник К.М. Лебедєв; В.О. Коломієць працює над створенням навчального корпусу англійських текстів, написаних носіями української мови (UCLE); В. Орел створює корпус анотацій до англійських статей з комп'ютерної лінгвістики.

За останні п'ять років співробітники лабораторії виступили з доповідями на 15 конференціях, зокрема на дистанційному занятті-семінарі з дистанційного навчання іноземним мовам, організованим Львівським університетом безпеки життєдіяльності.

Існує ще одна сфера діяльності лабораторії комп'ютерної лінгвістики – забезпечення навчального процесу на відділенні прикладної (комп'ютерної) лінгвістики, яке відкрилося, знову з ініціативи Сергія Микитовича Назарова майже одночасно з лабораторією. Ця робота була пов'язана з великими труднощами, оскільки в Україні склалася дуже несприятлива ситуація для структурно-математичної лінгвістики. В середині минулого століття ця галузь мовознавчих наук набула значного розвитку в колишньому СРСР, у тому числі й в Українській РСР. Працювали

НАВЧАЛЬНІ ПОСІБНИКИ, УКЛАДЕНІ В ЛАБОРАТОРІЇ КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ КИЇВСЬКОГО НАЦІОНАЛЬНОГО ЛІНГВІСТИЧНОГО УНІВЕРСИТЕТУ