

Модель корпусу текстів Наддністрянщини

Ігор Кульчицький

Кафедра прикладної лінгвістики, Національний університет “Львівська політехніка”, УКРАЇНА, м. Львів, вул. С. Бандери, 12,
E-mail: bis.kim@gmail.com (стиль Affiliation)

Abstract – The article deals with individual aspects of creating a corpus of texts which according to the place of creation or publication can be attributed to Naddnistrianska Ukraine. The time frames of texts of such corpus are outlined.

Ключові слова — лінгвістика, корпус текстів, Наддністрянщина, модель, тег.

Для сучасної лінгвістики корпуси — провідний технологічний інструментарій мовознавчих студій [4, 5]. Тому останні десятиліття відзначились укладанням корпусів текстів для основних мов світу (наприклад, American National Corpus, British National Corpus, Brown Corpus, Český národní korpus, Esti kirjakeele korpus, Korpus Slovenskega Jezika, Hrvatski elektronski tekstovni arhiv, International Corpus of English, Trésor de la langue française тощо) [2; 4].

В Україні корпусна лінгвістика ще не має великих здобутків і перебуває на етапі свого становлення. На сьогодні найбільшим є Український національний лінгвістичний корпус Українського мовно-інформаційного фонду НАНУ (наразі закритого типу) [4, 21]. У відкритому доступі (див.: <http://www.mova.info/>) на лінгвістичному порталі «Mova.info» Інституту філології Київського університету імені Тараса Шевченка представлено Дослідницький корпус сучасної української мови обсягом у 3 млн. слівформ, побудований як інформаційно-довідкова система [2, 19]. Отже, актуальним на сьогоднішній день залишається залучення відповідних науковців до створення українськомовних корпусів.

Таку спробу зроблено на кафедрі прикладної лінгвістики НУЛП. Вирішено створити корпус текстів, написаних чи виданих на Наддністрянській Україні. Обумовлено це тим, що на сьогодні все частіше увагу укладачів корпусів привертають специфічні варіанти мови. У свій час Святослав Караванський зазначав, що в Галичині, починаючи з першої половини XIX ст., виник власний, базований на живому мовленні і дещо відмінний від наддніпрянського, варіант української мови [3, 102]. На нашу думку, корпус текстів, написаних такою мовою, стане належною частиною загальноукраїнського лінгвістичного корпусу.

Укладання будь-якого корпусу передбачає попереднє вирішення ряду завдань, серед яких: вибір текстів відповідно до типу створюваного корпусу — синхронічного чи діахронічного, одномовного чи багатомовного тощо; побудова електронної бібліотеки — сукупності

творів різних стилів із повним бібліографічним описом; визначення відповідно до міжнародних стандартів набору тегів для маркування текстів тощо [4, 411]. На сьогоднішній день до корпусу вирішено залучати тексти, створені у період від 1816 до 1944 року. Нижня межа пов'язана зі створенням о. Іваном Могильницьким у цьому році у м. Перемишлі «Товариства галицьких греко-католицьких священників для поширення письмами просвіти і культури серед вірних». Верхня межа обумовлена остаточною окупацією більшовиками Західної України, після якої галицький варіант української мови занепав, а поширена, але відмінна від наддніпрянської, лексика отримала позначку «діал.».

Моделлю для випрацювання принципів добору та маркування текстів через стильове розмаїття обрано Літературно-науковий вісник, який почав виходити з 1889 року, коли Наукове товариство імені Шевченка вирішило зреформувати часпис «Зоря» та створити доповнене і розширене щомісячне видання з таким рубриками: оригінальна белетристика і переклади з чужих літератур; оригінальні і перекладні статті про найважливіші здобутки світової та своєї науки; критично-літературні студії з українського і чужого письменства; хроніка літератури і культурного життя українсько-руського народу в Австрії і Росії; огляди літератури і культурного життя в слов'янським і загальнолюдським світі; бібліографія. Під такою назвою часопис виходив до 1932 року. Публікація тимчасово припинялась: серпень 1914–1916, 1920–1921. У різні періоди вісник видавали у Києві та Львові за редакцією Володимира Гнатюка, Михайла Грушевського, Івана Франка, Дмитра Донцова. Від 1932 до 1939 року вісник мав назву «Місячник літератури, мистецтва, науки й громадського життя». Окрім Наукового товариства імені Шевченка у Львові часопис видавали: Українсько-руська видавнича спілка у Львові; Українська видавнича спілка у Києві; Українська видавнича спілка у Львові. Зараз доступні томи вісника перетворюють у цифрову форму та визначається набір тегів для метатекстової та структурної розмітки текстів.

Література

1. Барвінський О. Історія української літератури. Ч. 2 / О. Барвінський. — Львів: Друкарня НТШ, 1921. — 429 с.
2. Демська О. М. Текстовий корпус: ідея іншої форми / О. М. Демська. — К.: ВПЦ НАУКМА, 2011. — 282 с.
3. Караванський С. Секрети української мови / С. Караванський. — К.: УКСП «Кобза», 1994. — 152 с.
4. Корпусна лінгвістика: Моногр. / В. А. Широков, О. В. Бугаков, Т. О. Грязнухіна, О. М. Костишин, М. Ю. Кригін; НАН України, Укр. мов.-інформ. фонд. — К.: Довіра, 2005. — 472 с.