

### Summary

The theme under analysis belongs to the field of the applied translation. Translation ambiguity in the process of terms translation is not only linguistic problem. It was and is also the problem in the sphere of theological investigations.

The given article deals with the problems of translation ambiguity and faithfulness of translation of two greek trinitarian terms “οὐσία” and “ὑπόστασις”.

УДК 811.111'33

## ЗАСТОСУВАННЯ КОРПУСНОГО ПІДХОДУ У ДОСЛІДЖЕННІ МОВИ

Приц О.М.

*ДВНЗ «Ужгородський національний університет»*

На сучасному етапі все частіше науковці говорять про вихід мовознавства з кола суто гуманітарних наук.

Корпусна лінгвістика – розділ комп'ютерної лінгвістики, який займається розробкою загальних принципів побудови та використання лінгвістичних корпусів (корпусів текстів) з використанням комп'ютерних технологій. Корпусна лінгвістика базується на використанні корпусу, тобто великого об'єму живого мовного матеріалу, взятого із різних джерел і введеного до комп'ютера. Вона досліджує по-новому як мовлення, так і мову і, в результаті цього, об'єктивним шляхом отримуються нові лінгвістичні дані. Перевагою цього напрямку дослідження є те, що він виключає суб'єктивізм, який властивий традиційному мовознавству, і керується об'єктивними результатами [2, с. 5].

За останнє десятиліття корпусна лінгвістика перейняла досвід, накопичений в інших достатньо розвинених областях, що об'єднуються терміном АОР (автоматична обробка тексту), або ж АРТ (автоматичне розуміння тексту). Сучасна Корпусна лінгвістика має дві лінії розвитку — лінгвістичний та інформаційний аналіз тексту.

Інформаційний аналіз реалізується в системах типу ІПС (інформаційно-пошукові системи), а лінгвістичний аналіз має на меті інтерпретацію, з'ясування

сенсу тексту, при цьому лінгвістичний аналіз характеризується локальністю, тобто проходить в рамках однієї пропозиції. Інформаційний аналіз, навпаки, «бачить» текст як ціле. Ці два напрями розвивалися майже паралельно, практично ніяк не стикаючись, унаслідок чого не отримали достатнього розвитку. Лінгвістичний аналіз не забезпечив конкретний переклад, оскільки одна невірно інтерпретована фраза може в контексті набувати абсолютно іншого значення.

У наш час у лінгвістиці з'являються дуже близькі поняття: «електронні бібліотеки», «масив текстів», «колекція текстів», «електронний архів», «повнотекстова база даних». Серед них виділяються лінгвістичні корпуси, або мовні корпуси. Корпус текстів — це вид корпусу даних, одиницями якого є тексти або уривки з текстів, що включають, наприклад, певні фрагменти макроструктури текстів даної проблемної області.

Отже, відповідно до твердження О. Демської-Кульчицької [1, с. 56-58] обов'язковими корпусними ознаками є: підібраність, репрезентативність, збалансованість, машиночитаність, стандартність.

Цінність корпусу вбачається в наступному:

- одного разу укладений корпус може багато разів використовуватися;
- корпус показує мовні дані в їх реальному оточенні, що дозволяє досліджувати лексичну і граматичну структуру мови, а також безперервні процеси мовних змін, що відбуваються у мові впродовж певного відрізка часу;
- корпус характеризується репрезентативністю, або збалансованим складом текстів, що дозволяє використовувати його для тестування пошукових машин, машинних морфологій, систем перекладу, а також використовувати його в різних лінгвістичних дослідженнях;
- корпус має важливе значення для викладання мови, оскільки за допомогою корпусу можна швидко і ефективно перевірити особливості вживання незнайомого слова або граматичної форми.

Робота з корпусами, тобто з масивами текстів, представленими в електронному вигляді, давно вже стала одним з основних методів лінгвістичних досліджень. Так, ще в 1960-і роки створився Браунівський корпус (США) [7], який включає 1 млн. слів; у 1970-і : LOB корпус (Великобританія, Норвегія), який також включає 1 млн. слів. У 1980-і роки створилися такі корпуси, як: Машинний Фонд російської мови, Уппсальський корпус російської мови (Швеція), 1 млн. слів, The Bank of English, Birmingham, 20 млн. слів. У 1990-і створений British National Corpus [6], який включає 100 млн. слів, а також інші національні корпуси (угорський, італійський, хорватський, чеський, японський) обсягом 100 млн. слів. В цей же час створювався The Bank of English, Birmingham, що включає 600 млн. слів. На початку XXI ст. створювалися такі корпуси, як American National Corpus [4], 100 млн. слів і Gigaword corpora (англійський, арабський, китайський), що включає 1 млрд. слів. Особливу увагу слід звернути на російський національний корпус, над яким працюють лінгвісти Москви і Санкт-Петербурга.

Існування корпусів текстів дає можливість значно розширити і автоматизувати аналіз мовного матеріалу, який є найважливішою базою будь-якого лінгвістичного дослідження. Чим більше матеріалів аналізується, тим вища значущість висновків і рівень їх достовірності.

Сучасні комп'ютерні програми дозволяють знаходити потрібні приклади з корпусів текстів, які зберігаються в електронному вигляді на комп'ютері. Це економить значну кількість часу у порівнянні з традиційною технологією збору прикладів вручну.

Стаття присвячена вивченню застосування корпусного підходу у дослідженні мови. Об'єктом дослідження є застосування корпусів у лексикографії та граматиці.

Досягнення поставленої мети припускає розв'язання наступних завдань: 1) розглянути один з перспективних напрямків комп'ютерної лінгвістики – комп'ютерна лексикографія; 2) описати основний метод аналізу текстів на базі корпусів (Corpus-Based Approach).

Говорячи про застосування корпусів у лексикографії, варто зазначити, що мова йде про один з перспективних напрямків комп'ютерної лінгвістики, яка є в свою чергу складовою частиною прикладної лінгвістики, — *комп'ютерної лексикографії*. У центрі уваги цього напрямку – створення електронних словників і енциклопедій, які розробляються сьогодні як автономні і мережеві програмні продукти. На відміну від лексикології – теоретичної дисципліни, що займається розробкою методів опису значення слів (у тому числі і методів їх тлумачення), проблематика лексикографічних робіт лежить у сфері словарної практики. Центр її інтересу – способи організації словникової статті, принцип роботи словників і технологія їх створення. Лінгвісти беруть участь як у розробці лінгвістичних (філологічних) словників, так і в створенні енциклопедичних видань. Сучасна лексикографія істотно розширила і підсилила свій інструментарій комп'ютерними технологіями створення і експлуатації словників. Цей напрям прикладної лінгвістики одержав назву комп'ютерної лексикографії. Спеціальні програми – бази даних, комп'ютерні картотеки, програми обробки тексту – дозволяють в автоматичному режимі формувати словникові статті, зберігати інформацію і обробляти її. Безліч різних комп'ютерних лексикографічних програм можна розділити на дві великі групи: програми підтримки лексикографічних робіт і автоматичні словники різних типів, що включають лексикографічні бази даних [3, с. 1].

Важливу роль комп'ютерна лексикографія відіграє при вивченні іноземної мови, так як у процесі навчання використовуються різні словники. Раніше це були тільки статистичні друковані на папері томи, при використанні яких витрачалось багато часу на пошук тієї чи іншої лексичної одиниці. З розвитком прикладної лінгвістики і такого з її напрямків, як комп'ютерна лексикографія, почали з'являтися перші автоматизовані словники, з допомогою яких можна було висвітлити ту чи іншу лексичну одиницю і її переклад або значення на екрані монітора. З розвитком комп'ютерних технологій цей процес став більш швидким і

зручним. Нові мультимедійні технології дали можливість швидко показувати не тільки значення або переклад, а навіть приклади вживання і почути вимову його. Відкриваючи словникову статтю, перед нами постає візуальна картинка із зображенням цього слова. Все це допомагає нам краще зрозуміти значення, переклад і вживання лексичних одиниць в динаміці. Поява електронних словників в інтернеті і на різних носіях є результатом сумісної роботи лінгвістів, програмістів, психологів і різних технічних спеціалістів.

Корпусна лінгвістика дає матеріал для дослідження мови та її варіантів і визначає основний метод аналізу текстів на базі корпусів (Corpus-Based Approach) [5, с. 12]. Однією з важливих особливостей цього методу аналізу текстів є дослідження не тільки власне лінгвістичних явищ (наприклад, граматичне чи лексичне функціонування слів та їх зв'язок з іншими лексемами), але і таких явищ як частотність вживання тих чи інших лексем або граматичних явищ у різних жанрах та діалектах.

З кінця 80-х років корпуси текстів почали використовувати для підготовки та випуску граматик. Найвідомішим проектом, який представляв особливості англійської граматики, був проект повної англійської граматики на основі Лондонсько-Лундського корпусу (1985). З того часу більшість британських та американських видавництв видають словники та граматики, побудовані на матеріалах корпусів, Британського та Американського національних корпусів англійської мови[6].

Традиційні підручники з граматики часто акцентували свою увагу на граматичних структурах, описуючи їх форму та значення, однак не враховуючи як ці граматичні конструкції вживаються у контексті, дискурсі, стилі мовлення.

Корпусні дослідження показали, що в сучасній англійській мові у розмовному мовленні частіше зустрічається Present Tense, в той час як у писемному мовленні Past Tense [8, с. 54]. У розмовному мовленні форма дієслова TO BE у Past Tense зустрічається не так часто, ніж у писемному мовленні, в той час як дієслово TO GET у Past Tense навпаки частіше зустрічається у писемному мовленні, ніж у розмовному мовленні [9, с. 86].

Таким чином, інформація про граматичну конструкцію у згаданих вище граматиках подається у вигляді розгорнутого опису, пояснення її структури, значення, синтаксичних зв'язків дискурсі, з частотними діаграмами та графіками, які висвітлюють частоту використання цієї конструкції у розмовному чи писемному мовленні.

Підсумовуючи викладене вище, зазначимо, що корпусний підхід, або метод лінгвістичного дослідження, який базується на корпусах текстів, зорієнтований на прикладне вивчення мови, його функціонування у живому мовленні.

## Література

1. Демська-Кульчицька О.М. Основи національного корпусу української мови. – К.: Наукове видання Ін-ту укр. мови НАН України, 2005. – 219 с.
2. Магомедова Адигат Нурахмагджиевна. Корпусная лингвистика и контекстное разрешение лексической многозначности слов : Дис. ... канд. филол. наук : 10.02.19 : Махачкала, 2004.- 203 с.
3. Хоменко Ф.В. Комп'ютерна лексикографія при вивченні іноземної мови, Національний авіаційний університет, м. Київ  
[//http://ev.nuos.edu.ua/en/content/komp%E2%80%99yuternaleksikograf%D1%96ya-pri-vivchenn%D1%96-%D1%96nozemnoi-movi](http://ev.nuos.edu.ua/en/content/komp%E2%80%99yuternaleksikograf%D1%96ya-pri-vivchenn%D1%96-%D1%96nozemnoi-movi)
4. American National Corpus// <http://americannationalcorpus.org/>
5. Biber D., Conrad S., Reppen R. Corpus Linguistics: Investigating language structure and use. - Cambridge University Press, 2001. – 312 p.
6. British National Corpus // <http://www.natcorp.ox.ac.uk/>
7. BROWN CORPUS MANUAL// <http://khnt.aksis.uib.no/icame/manuals/brown/>
8. Longman Grammar of Written and Spoken English. London, 1999. -298 p.
9. Mindt D. An Empirical Grammar of the English Verb System. Berlin, 2000. - 250 p.

## Summary

The article deals with the corpus-based approach as one of the linguistic methods based on the text corpora which is orientated on the applied language studying, its functioning in the speech. The author focuses on the computer lexicography which is examined as one of the perspective methods of the computer linguistics and on the corpus-based approach which is described as the main method of the text analyses based on the text corpora.