

КОРПУСНЕ МОДЕЛЮВАННЯ ЕВОЛЮЦІЇ МОВНОГО СТИЛЮ В ПАСТИРСЬКИХ ПОСЛАННЯХ АНДРЕЯ ШЕПТИЦЬКОГО

Гулик Юрій Володимирович

аспірант кафедри прикладної лінгвістики
Національний університет "Львівська політехніка", Україна

Науковий керівник: Левченко Олена Петрівна

доктор філологічних наук, професор, завідувачка кафедри прикладної лінгвістики
Національний університет "Львівська політехніка", Україна

Корпусне моделювання еволюції мовного стилю в пастирських посланнях Андрея Шептицького є актуальним і цікавим напрямком досліджень з багатьма потенційними перевагами. Розуміння змін у мовному стилі відображає не лише історичний контекст, але й культурні, соціальні та релігійні аспекти [2]. Дослідження такого роду може допомогти краще зрозуміти, як змінювалася мовна практика та сприйняття релігійних текстів протягом часу [1].

Аналіз стилістики пастирських послань Андрея Шептицького через корпусне моделювання надає можливість виявити тенденції, патерни та впливи, що могли впливати на формулювання його думок і ідей. Це дослідження може бути корисним як для лінгвістичних аналітиків, так і для істориків, релігієзнавців та культурологів.

Подальший аналіз еволюції мовного стилю в пастирських посланнях Андрея Шептицького може сприяти глибшому розумінню релігійного дискурсу, його ролі в суспільстві та змінам у сприйнятті церковної мови протягом часу. Такі дослідження можуть мати практичне значення для вивчення історії церковної культури, формування діалогу між різними релігійними та культурними групами, а також для вдосконалення методів викладання релігійної літератури та культурології.

Корпусне моделювання є потужним інструментом для аналізу еволюції мовного стилю в текстах різних жанрів та авторів, включаючи пастирські послання Андрея Шептицького [4]. Цей метод дослідження базується на зборі та аналізі великих обсягів текстових матеріалів з метою виявлення стилістичних та мовних особливостей, а також змін, які відбуваються зі зміною часу.

Для проведення корпусного аналізу пастирських послань Андрея Шептицького спершу був створений корпус текстів, який включав у себе всі доступні послання. Цей корпус був оброблений та підготовлений для подальшого аналізу з використанням мови програмування Python. Python використовувався для автоматизації процесу обробки та аналізу текстів, що значно полегшило роботу дослідників.

Одним із ключових етапів було роззначення тексту з використанням програми на Python, яка дозволила ідентифікувати та виділяти основні риси стилю та мови, зокрема цитати з Біблії. Для цього були використані певні техніки обробки тексту та алгоритми пошуку, що дозволили автоматично розпізнавати та відзначати цитати з Біблії у текстах пастирських послань.

Наступним кроком було тренування нейронної мережі на підготовленому корпусі текстів з метою виявлення цитат з Біблії та визначення їх вірогідності. Тренування нейронної мережі відбувалося з врахуванням специфіки структури текстів Біблії та пастирських послань, що дозволило побудувати модель, здатну аналізувати та класифікувати тексти за наявністю цитат.

Отримана після тренування нейронна мережа представляє собою інструмент для автоматичного виявлення цитат з Біблії у пастирських посланнях Андрея

Шептицького з вказанням ступеня вірогідності. Це дозволяє швидко та ефективно аналізувати великі обсяги текстів з релігійного контексту та досліджувати їх мовні та стилістичні особливості. Такий підхід відкриває нові можливості для дослідження еволюції мовного стилю в релігійних текстах та їх зв'язку з Біблійними джерелами.

У нашому дослідженні ми використовували пастирські послання Андрея Шептицького для тренування нейронної мережі на датасеті, що містить цитати з Біблії. Стилїстика Біблії є вкрай специфічною, тому наявність великого обсягу даних дозволяє мережі виявляти у тексті фрагменти, що ймовірно є цитатами з Біблії, а також встановлювати числові характеристики вірогідності цитати.

Для ефективної обробки текстів ми використовували токенизацію, що дозволяє розділити текст на логічні частини. Нормалізація тексту, яка включає стандартизацію реєстру слів та видалення знаків пунктуації, була важливим етапом обробки. Крім того, застосували стеммізацію та лемматизацію для виокремлення основ слів.

Для аналізу та обробки текстів також використовувалася концепція корпусу, що представляє собою набір текстів для лінгвістичного аналізу. «Стоп-слова», що не несуть семантичного навантаження, були використані для усунення шуму з тексту. POS-тегування та статистичне моделювання мови використовувалися для визначення частин мови та побудови статистичних моделей мови [3].

Ми використовували готовий токенизатор та модель BERT, яка є нейронною мережею з композицією кодувальників трансформера. BERT навчався на текстових даних корпусу BooksCorpus та текстах англomовної Вікіпедії, а якість моделі оцінювалась на наборі завдань GLUE.

У процесі тренування ми розділили датасет на тренувальну та тестову вибірки, використовуючи бібліотеку scikit-learn для оцінки роботи моделі. Частина алгоритмів була реалізована через мову програмування Python для покращення продуктивності.

Наш підхід дав можливість ефективно виявляти цитати з Біблії у пастирських посланнях Андрея Шептицького та оцінювати їх вірогідність. Використання нейронних мереж та сучасних методів обробки тексту відкриває нові можливості для дослідження еволюції мовного стилю та релігійних текстів.

Корпусне моделювання еволюції мовного стилю в пастирських посланнях Андрея Шептицького є важливим і актуальним напрямком досліджень, яке може принести значний внесок у розуміння не лише лінгвістичних аспектів, але й культурних, релігійних та історичних контекстів. Аналіз мовного стилю відомого духовного лідера через призму корпусного підходу дозволяє виявити тенденції та зміни у сприйнятті релігійних текстів протягом часу, що може мати важливе значення для розуміння ролі церковної мови в суспільстві та формування діалогу між різними релігійними та культурними групами.

Список використаних джерел:

1. Бобкова Т. Корпус текстів: основні аспекти визначення.
URL: http://www.mova.info/corpus_papers/bobkova-corpus.pdf (дата звернення: 10.02.2024).
2. Дарчук Н. П. Дослідницький корпус української мови: основні засади і перспективи. Мовознавство. Фольклористика. – Київ : Видавничо-поліграфічний центр “Київський університет”, 2010. № 21. с. 45-49.
3. Демська-Кульчицька О. Що таке корпус текстів?
URL: <https://ekmair.ukma.edu.ua/server/api/core/bitstreams/f53cab32-377b-4349-9114-3e336a865af5/content> (дата звернення: 10.02.2024).
4. Шептицький А. Твори митрополита Андрея Шептицького. Пастирські послання до духовенства й вірних Станиславівської Епархії (1899-1904) (1935). URL: <http://irbis-nbuv.gov.ua/ulib/item/UKR0007499> (дата звернення: 10.02.2024).