

Создание корпуса художественных текстов азербайджанского языка

Халилова Гамида Азизага гызы
Сотрудник

Бакинский филиал Московского государственного университета имени М.В.Ломоносова,
Баку, Азербайджан
E-mail: hamida.khalilova@mail.ru

Корпусная лингвистика представляет собой одну из бурно развивающихся областей науки. «Корпусная лингвистика – раздел компьютерной лингвистики, занимающийся разработкой общих принципов построения и использования лингвистических корпусов с использованием компьютерных технологий» [Захаров 2005: 3]. Процесс ее развития в эпоху информационных технологий необратим, в силу чего корпусная лингвистика получила статус самостоятельной науки [Халилова 2024].

Основным понятием корпусной лингвистики является корпус текстов. «Под названием лингвистический, или языковой, корпус текстов понимается большой, представленный в электронном виде, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач» [Захаров 2005: 6]. Корпус текстов – это то, что стремится создать корпусная лингвистика. На самом же деле, корпус текстов выступает, «с одной стороны, как ее объект, т.е. является исходным речевым материалом для корпусной лингвистики и для других лингвистических дисциплин» [Бойкова]; с другой же стороны, корпус текстов – результат деятельности корпусной лингвистики [Бойкова].

Создание корпуса национального языка – это стратегический вопрос государственного значения. Корпус языка важен тем, что он способствует сохранению языка, его лексического состава, при помощи данного корпуса возможно выявить изменения, происходящие в языке на протяжении времени и т. д. [Халилова 2024]. Благодаря корпусу текстов исследователь может работать с массивом текстов, имея возможность извлечь оттуда интересующую его лексическую единицу, то есть может проанализировать, выявить поведение данной единицы в естественной языковой среде, не теряя при этом время на сбор материала, разметку и т. д. [Халилова 2024].

Актуальность и новизна исследования обусловлены значимостью создания корпусов текстов азербайджанского языка, который на данный момент не представлен корпусами текстов. Поскольку в текстах одной личности язык проявляется как целостная система, создание авторских корпусов имеет важное значение для развития корпусной лингвистики, что особенно актуально для азербайджанского языкоznания. Следует особо подчеркнуть, что создание корпусов текстов отдельных азербайджанских авторов, корпуса художественных текстов не менее важно и для создания Национального корпуса (НК) азербайджанского языка. НК включают в себя все стили и жанры. Так, в НК входят и произведения художественной литературы, так как именно произведения художественной литературы и устного народного творчества имеют культурную значимость, в них также заключены прецедентные феномены.

В значительной степени влияет на актуальность данной работы и выбор материала – художественных текстов М. Физули, М.А. Сабира, Н. Гянджеви, которые не являлись ранее объектом корпусного исследования. Выбор материалов также не случаен. Интерес к личности и творческому наследию Н. Гянджеви, М. Физули и М.А. Сабира не теряет своей актуальности. Важность изучения творческого наследия Низами подтверждает и контракт, подписанный Центром Н. Гянджеви и Оксфордским университетом при содействии и участии вице-президента НАНА, ректора филиала МГУ в г. Баку, академика

Конференция «Международная площадка конференции "Ломоносов-2025" в Азербайджане»

Наргиз Пашаевой. Язык и стиль, которыми писал основатель азербайджанской сатирической поэзии М.А. Сабир, также значимы для исследования: он писал просто, сочетая живую народную речь с литературным языком, создавая этим новый стиль в поэзии. М. Физули является основоположником литературного *азербайджанского языка*, что весьма значимо для исследования [Халилова 2024].

Результаты корпусного исследования:

- Создан корпус художественных текстов, включающий на данный момент тексты М.А. Сабира, «Хамсе» Н. Гянджеви, произведения М. Физули, написанные на азербайджанском языке;
- Созданы авторские корпуса: Корпус текстов М.А. Сабира, Корпус текстов «Хамсе» Н. Гянджеви (переводы на азербайджанский язык), Поэтический корпус текстов М. Физули,
- Составлены частотные списки;
- Выявлены самые частотные единицы в текстах по-отдельности и по созданным корпусам;
- Определены количественные соотношения токенов и словоформ;
- Определено частотное соотношение определенных лексических единиц;
- Выявлены некоторые особенности употреблений языка М.А. Сабира. Например: а) употребление арабизмов: *əzz edəsiz*, *əvəti*; б) употребление персизмов: *bərəngixtə rərxas*; *Yeksər* и др.
- Выявлены некоторые особенности языка М. Физули. Например: а) обильное использование вспомогательных глаголов *eylətək* (делать) və *olmaq* (быть), *etmək* (делать); б) словоупотребление *ol* (быть в форме третьего лица ед. ч.) располагается рядом с личным местоимением 3-го лица ед. ч.

Таким образом, данная работа является пионерской в создании корпусов текстов азербайджанского языка. Более того, это исследование имеет существенное значение для развития корпусной лингвистики и улучшения качества лексической и синтаксической обработки языка. Результаты же вносят важный вклад в дальнейшие лексикографические исследования и служат основой для новых научных изысканий в области языка, перевода и других направлений.

Источники и литература

- 1) Бойкова С.Н. Введение в корпусную лингвистику: [Электронный ресурс] / URL: <http://core.ac.uk/download/pdf/75999967.pdf>
- 2) Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие. – СПб., 1 2005. – 48 с.
- 3) Халилова Г.А. «Развитие азербайджанской корпусной лингвистики. Создание корпуса текстов азербайджанских авторов (поэтов и писателей) и национального корпуса азербайджанского языка»: [Электронный ресурс] // Вестник Дагестанского государственного университета. Серия 2: Гуманитарные науки 2024 / 2. URL: <https://cyberleninka.ru/article/n/razvitie-azerbaydzhanskoy-korpusnoy-lingvistik-i-sozdanie-korpusa-tekstov-azerbaydzhanskih-avtorov-poetov-i-pisatelyey-i>