

UOT:81.

**ЛИНГВОСТАТИСТИКА, ИЛИ КВАНТАТИВНАЯ ЛИНГВИСТИКА:
ЗАКОН ЦИПФА, ЗАКОН МЕНЗЕРАТА – АЛЬТМАННА
ПРЕИМУЩЕСТВА ЛИНГВОСТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЙ**

Халилова Гамида Азизага гызы
МГУ имени М.В. Ломоносова в г. Баку
докторант Института языкознания им. Насими
Национальной Академии Наук Азербайджана
hamida.khalilova@mail.ru
Orcid ID: 0000-0002-7423-8850

Аннотация: В данной статье даются основные определения лингвостатистики (лингвистической статистики), или, иначе, квантативной лингвистики. В статье рассматриваются методы лингвостатистики, отмечается преимущество лингвостатистических исследований. Среди методов лингвостатистики особо рассматриваются закон Ципфа и закон Мензерата – Альтманна. В статье также говорится, что благодаря методам лингвостатистики можно определить разнообразие речи, применяя коэффициенты разнообразия речи, также возможно выявить частотное соотношение тех или иных единиц, к примеру, слов или словоформ в тексте, подсчитать количество токенов в тексте или словоформ, а также можно определить особенности языка того или иного писателя, специфику художественного текста на языке оригинала и при переводе, характерные особенности языка художественных текстов. Это также позволит создать частотные словари, а также корпуса, к примеру, корпуса языка того или иного автора.

Ключевые слова: лингвостатистика, квантативная лингвистика, корпусная лингвистика, конкорданс, корпус, закон Ципфа, токен, художественный текст, частотность

Açar sözlər: linqvostatistika, kvantativ linqvistika, korpus linqvistikası, konkordans, korpus, Zipf qanunu, token, bədii mətn, tezlik

Keywords: linguistic statistics, quantitative linguistics, corpus linguistics, concordance, corpus, Zipf's law, token, literary text, frequency

Начиная со второй половины XX века в науке стали зарождаться новые области, которые являлись смежными, т.е. возникли на стыке двух и более наук. К примеру: корпусная лингвистика, компьютерная лингвистика, а также квантативная, или, другими словами, лингвистическая статистика (лингвостатистика).

Лингвостатистика (лингвистическая статистика) – это «1) в широком смысле: область применения статистических методов в языкознании; 2) в узком смысле: изучение некоторых математических проблем, связанных с лингвистическим материалом, главным образом с типами статистических распределений языковых единиц в тексте» [5].

Следует отметить, что лингвостатистику также называют и квантативной лингвистикой: «В западноевропейской литературе термином «квантативная лингвистика» обозначается та область науки, которую в отечественном языкознании называют лингвистическая статистика или лингвостатистика, статистика речи, статистика текста» [9]. Из вышесказанного следует, что чаще на Западе ученые и исследователи называют «лингвистическую статистику (лингвостатистику) «квантативной лингвистикой».

Рассмотрим некоторые определения исследователей относительно данного понятия и термина. Так, Моисеева отмечает, что «квантитативная лингвистика – это термин, широко используемый в современной научной литературе для обозначения междисциплинарного направления прикладных исследований, в котором количественные или статистические методы анализа служат основным инструментом изучения языка и речи» [7, стр. 10].

А.В. Кашеева же отмечает следующее: «Популярность квантитативных методов позволила со временем сформироваться направлению в лингвистике, называемому квантитативной лингвистикой (quantitative linguistics). Это одно из направлений прикладной лингвистики, которое занимается изучением языка с помощью статистических методов» [2]. А.В. Кашеева, как и я [9], а также и многие другие исследователи, подчеркивает, что в зарождении данной науки сыграло важную роль применение методов др. наук. Помимо этого, также важное значение в формировании смежных наук имеет нынешнее техническое развитие.

Хотелось бы отметить важность использования лингвостатистических методов. Однако до того, как перейти к этому, нужно понять, а что же представляют собой лингвостатистические методы.

Итак, лингвостатистический метод – это собой «совокупность статистических методик, приёмов и процедур, используемых для получения количественных данных о языковых явлениях, с целью раскрытия статистических закономерностей функционирования единиц языка и речи, а также для установления статистических закономерностей построения текстов» [4, стр. 574]. И.Ю. Моисеева же отмечает, что благодаря методам, известным в математической статистике возможно изучать законы и структуру языка и изменения в системе [7, стр. 10].

Итак, рассмотрим некоторые методы лингвостатистики.

Наиболее распространенным является метод анализа, который основан на так называемом законе Ципфа, сводящемся к уравнению вида: $F \times i = \text{const}$, где F — частота слова в частотном словаре, а i — ранг этого слова, т. е. номер в списке слов, упорядоченном по уменьшающейся частоте. С поправочными коэффициентами Б. Мандельброта закон подтверждается на многих и разнообразных текстах. Таким образом, закон Ципфа выступает как модель описания распределения слов по частоте, однако вводимые на каждый случай поправочные коэффициенты в значительной мере лишают его предсказательной силы [5].

Существуют и др. методы, например: закон Мензерата. Закон Мензерата (Menzerath's law), или закон Мензерата – Альтманна (Menzerath–Altmann law), гласит следующее: согласно этому закону «с увеличением размера языковой единицы уменьшаются размеры его компонентов и наоборот» [1]. Как известно, первый вариант закона предложил П. Мензерат, он сформировал этот закон на основе данных слогов слов немецкого языка. Было определено: чем длиннее слова, чем короче оказались составляющие их слоги. Далее же Г. Альтманн «придал этому закону строгую математическую форму» [1], показав тем самым, что данный закон можно применить и на материале и других структур языка (например, предложений и клаузов, клаузов и составляющих их слов). Как отмечает Галиева, «при этом обязательной монотонности не требуется, средний размер составляющих является функцией от размера конструктора» [Галиева]. Итак, формула, которую предложил Г. Альтманн: $y(x) = a \cdot x^b \cdot e^{cx}$, где y – средняя длина конститuentов, x – длина конструктора, a , b , c – параметры модели [10].

Итак, выше были рассмотрены самые основные методы лингвостатистики.

Я считаю, что использование лингвостатистических методов является наиболее продуктивным, особенно при исследовании художественных текстов. Таким образом, важность и преимущество лингвостатистических исследований, в частности же художественных текстов, заключается в том, что такой подход позволит оценить «характер использования привлеченных в текст языковых средств, которое осуществляется при помощи лингвостатистических методов. К примеру, для определения разнообразия речи применяются так называемые коэффициенты разнообразия речи» [9], также благодаря методам лингвостатистики можно выявить частотное соотношение тех или иных единиц, к примеру, слов или словоформ в тексте, подсчитать количество токенов в тексте или словоформ, а также можно определить особенности языка того или иного писателя, специфику художественного текста на языке оригинала и при переводе, характерные особенности языка художественных текстов. Это в свою очередь упростит создание корпусов, к примеру, корпуса языка того или иного автора.

Библиография

1. Галиева А.М. Закон Мензерата – Альтманна: эксперименты с текстами на татарском языке, Ученые записки казанского университета\ Серия «Гуманитарные науки», Т. 1, 163, кн. 1, 180–189 с.
2. Кашеева А. В. Квантитативные и качественные методы исследования в прикладной лингвистике [Электронный ресурс]. – URL: <https://cyberleninka.ru/article/n/kvantitativnye-i-kachestvennye-metody-issledovaniya-v-prikladnoy-lingvistike>
3. Квантитативная лингвистика [Электронный ресурс]. – URL: https://studwood.net/1639714/informatika/ponyatie_kvantitativnoy_lingvistiki_otlichie_matematiceskoy_lingvistiki
4. Комарова З. И. Методология, метод, методика и технология научных исследований в лингвистике: учебное пособие. 2012
5. Лингвистический энциклопедический словарь [Электронный ресурс]. – URL: <https://tapemark.narod.ru/les/269b.html>
6. Лингвостатистика и вычислительная лингвистика Труды по лингвостатистике; Тарту, 1982. – 169 с.
7. Моисеева И. Ю. Квантитативная лингвистика и новые информационные технологии: учебное пособие / И. Ю. Моисеева; Оренбургский гос. ун-т. Оренбург: ОГУ, 2017. 101с.
8. Понятие квантитативной лингвистики и ее отличие от математической лингвистики [Электронный ресурс]. URL: https://studwood.net/1639714/informatika/ponyatie_kvantitativnoy_lingvistiki_otlichie_matematiceskoy_lingvistiki
9. Халилова Г.А. Компьютерная лингвистика и применение лингвостатистических методов при исследовании художественных текстов BİLİMSEL ARAŞTIRMALAR ELMİ ARAŞDIRMALAR / Tezis və məqalələr toplusu, p. 33-35
10. Altmann G. Prolegomena to Menzerath's law // Glottometrika. – 1980. – No 2. – P. 1–10
11. Dornyei Z. Research Methods in Applied Linguistics. OUP. 2007
12. Lazaraton A. Quantitative and qualitative approaches to discourse analysis // Annual Review of Applied Linguistics. 2002. № 22.
13. Cramer I. The parameters of the Altmann–Menzerath law // J. Quant. Linguist. – 2005. – V. 12, No 1. – P. 41–52.
14. Xəlilova H.Ə. Mirzə Ələkbər Sabirin satiralarının korpus tədqiqatı. Sabirin mətn korpusunun yaradılması Filologiya məsələləri, dekabr 2023.