

КОЛИЧЕСТВЕННЫЕ МЕТОДЫ ИЗУЧЕНИЯ ЯЗЫКА

Сущность количественных методов заключается в использовании подсчетов и измерений при изучении языковых явлений. Применение математического аппарата при изучении языка связано со стремлением использовать исследовательские приемы точных наук, чтобы сблизить с ними языкознание. Когда количественные методы опираются на математическую статистику, они могут быть названы статистическими методами. В силу этого иногда количественные методы разделяют на количественные и статистические, используя общий для них термин «квантитативный метод».

Применение количественных методов к изучению языка началось с использования статистики в лингвистических исследованиях. Русский математик А. А. Марков в 1913 г. издал работу «Опыт статистического исследования над текстом «Евгения Онегина», иллюстрирующий связь испытаний в цепи», в которой установил частотность повторения гласных и согласных звуков в тексте романа А. С. Пушкина, опираясь на теорию вероятностей.

С помощью подсчетов распределения букв в текстах иногда можно увидеть специфику того или иного языка. Например, при статистических подсчетах частотности букв в белорусских текстах выяснилось, что буквы *а*, *я* покрывают около 18 % употреблений всех букв, что значительно отличает белорусские тексты, например от русских текстов. В этой цифре отражается такая существенная фонетическая особенность белорусского языка, как аканье и яканье, которая в белорусском языке отражается на письме.

Статистическое измерение текста, выявление его статистической структуры интересуют современную теорию информации. С точки зрения теории информации частота появления в тексте сигнала, или символа, (это будет вероятность) и сочетаний сигналов (условная вероятность) представляет собой статистическую структуру текста. При многократном повторении сигнала в тексте важным становится определение количества информации, передаваемой сигналом. Количество этой информации приравнивается к мере недостающей информации, то есть к величине неопределенности. Количественная мера неопределенности ситуации получила в теории информации название энтропии, которая вычисляется по известной формуле.

Соединение статистических методов с приемами дистрибутивного метода легло в основу дистрибутивно-статистического анализа. С его помощью описывают структуру текста и языка, на котором он создан, без каких-либо сведений о семантике. Единицы языка и отношения между ними в этом случае выделяются в процессе анализа, а не используются как готовый материал, что очень важно при дешифровке древних рукописей, написанных на неизвестных языках.

С помощью количественных методов описывается поведение различных языковых единиц – фонем, букв, морфем, слов – и частота различных фактов, зафиксированных в текстах: частота употребления единиц, их распределения в текстах разного жанра, сочетаемости с другими единицами и т.д. Например, на основе изучения количества употреблений всех лексических единиц, встречающихся в определенной совокупности текстов, создаются частотные словари разных языков. Согласно данным частотных словарей, на 1000 самых употребительных слов приходится приблизительно 85% всех использованных в текстах лексических единиц. Эти данные имеют первостепенное значение при составлении словарей-минимумов и отборе лексических явлений для активного и пассивного усвоения при изучении иностранных языков.

Статистические выкладки используются при установлении авторства произведения на основе подсчета особенностей его языка, при определении приблизительной датировки произведений и хронологии произведений одного и того же автора. Польский этнограф Ян Чекановский в 1927 г. и американский языковед Моррис Сводеш с 1948 г. стали применять лексикостатистику в историческом и сравнительном языкознании, где на основе количественных показателей словарного состава стремились определить наличие родства между языками и время их вычленения из единого праязыка.

Количественные методы предполагают исследование обширных массивов текстов, поэтому хранение и обработка исходных данных, многократное обращение к ним дают наибольший эффект с применением ЭВМ. Количественные методы создают перспективу превращения обычной структурной модели языка в структурно-вероятностную модель, в которой единицы языка обладают каждая своим «весом», измеряемыми оказываются их связи и отношения.

Количественное изучение языка тесно связано с его качественным анализом и в идеале не должно проводиться в отрыве от него, поскольку сами по себе количественные данные не всегда имеют научную значимость.