

ЧАСТОТНЫЙ АНАЛИЗ КОРПУСА ТЕКСТОВ КАК МЕТОД ИСТОРИЧЕСКОГО ИССЛЕДОВАНИЯ: ОПЫТ АНАЛИЗА ГАЗЕТЫ «ПРАВДА», 1953 – 1991 ГГ.

Ахмеров Д.А. (Университет ИТМО)

Научный руководитель – кандидат социологических наук

Колозарики П.В. (Университет ИТМО)

Введение. Частотный анализ текста широко применяется в гуманитарных, политических и социальных науках. Чаще всего, в фокусе исследователей находятся несколько документов, в которых подсчитывается вхождение определенных понятий, однако корпуса текстов, насчитывающие сотни и даже тысячи документов игнорируются из-за трудоемкости анализа, недостаточного финансирования и отсутствия «рабочих рук». В частности, об этом во введении к «Словарю основных исторических понятий» писал Райнхарт Козеллек, один из отцов основателей и идеологов истории понятий. Несмотря на развитие методов анализа данных и их доступность, до сих пор историками не был описан опыт исследования большого массива исторических источников с помощью частотного анализа, который был бы основан на актуальных компьютерных методах анализа данных. Более всего эта лакуна заметна в диахронических исследованиях на материалах периодических изданий и эго-документов.

В центре этого доклада находится следующий вопрос: каким образом историки могут использовать частотный анализ для изучения большого корпуса исторических источников. Данная проблема рассматривается на примере анализа корпуса газеты «Правда» за 1953-й — 1991-й гг.

Основная часть. Применение метода частотного анализа корпуса текстов позволяет историкам зафиксировать, в какие моменты некоторое понятие появляется или наиболее активно используется в дискурсах определенной эпохи. В свою очередь, оперирование этими понятиями отпечатывается в исторических источниках, например в газетах, публичных выступлениях, письмах и дневниках.

Наиболее перспективным, на наш взгляд, является использование данного аналитического метода для исследований в рамках истории политических и социальных понятий, поскольку подсчет встречаемости слов в большом массиве исторических источников позволяет зафиксировать факты употребления понятий, а затем рассмотреть их в диахронической перспективе с применением «точечных» аналитических методов, например дискурс-анализа. Так, на примере подсчета количества вхождений слов «крестьянин», «крестьянство» и производных от них были выявлены периоды и моменты, в которые советская власть активно обращалась к этим понятиям, — показательно, что частота вхождений в выпуски газеты, посвященные началу посевных работ, а также реорганизации машинно-тракторных станций в 1958-м году, на несколько порядков выше по сравнению с прочими выпусками за те же годы. В паре с методом дискурс-анализа возможно определить, какое место занимали «крестьяне» и «крестьянство» в рамках советского авторитетного дискурса в период позднего социализма, а также проследить, каким образом менялось употребление этих понятий в диахронической перспективе. Обращаясь к корпусам оцифрованных документов за этот же период и сопоставляя их друг с другом появляется возможность реконструировать дискурсы советской власти, существовавшие в исследуемый период.

Области применения частотного анализа корпуса текстов ограничиваются лишь фантазией учёного. К примеру, Марк Альфано в одном из своих исследований использует частотный анализ для определения трудов Фридриха Ницше, в которых философ обращается к понятиям добродетель, стимул и инстинкт (*virtue, drive, instinct*). Кроме того, Альфано прослеживает, каким образом меняется их употребление в работах Ницше.

Выводы. Данное исследование помогает понять, каким образом специалисты в области гуманитарных, социальных и политических наук могут использовать метод частотного анализа корпусов текстов, насчитывающих сотни и тысячи документов, в своих исследованиях. Кроме того, настоящая работа очерчивает возможности применения этого аналитического метода как исследовательского метода *per se*, так и как вспомогательного инструмента для настройки исследовательской оптики, а также обращает внимание на трудности и ограничения применения частотного анализа текстов.

Список использованных источников:

1. Alfano M. Digital Humanities for History of Philosophy: A Case Study on Nietzsche // Research Methods for the Digital Humanities / eds. L. Levenberg, T. Neilson, D. Rheams. — Cham. Palgrave Macmillan, 2018. P. 85–99.
2. Белоусов К. И., Зелянская Н. Л. Образ мира политика (по данным частотного анализа лексики) // Политическая лингвистика. 2012. №2. С. 60–65.
3. Горожане в процессах потребления массовой информации и пропаганды // Грушин Б. А. Четыре жизни России в зеркале общественного мнения. Очерки массового сознания россиян времен Хрущева, Брежнева, Горбачева и Ельцина в 4-х книгах. Жизнь 2-я. Эпоха Брежнева (часть 2-я). — М.: Прогресс-Традиция, 2006. С. 529–610.
4. Словарь основных исторических понятий: Избранные статьи в 2-х т. Т. 1 / Пер. с немецкого К. Левинсон; сост. Ю. Зарецкий; К. Левинсон, И. Ширле; научн. ред. перевода Ю. Арнаутова. — М.: Новое литературное обозрение, 2014. С. 36–37.
5. Шумилина Т. В. Применение частотного анализа текстов СМИ для оптимизации процесса коммуникации // Вестник Московского университета. Серия 10. Журналистика. 2017. №2. С. 67–79.