

Опыт квантитативного анализа социально-политической повестки в дневниках сталинского периода (по материалам электронного корпуса дневников "Прожито")

Научный руководитель – Бонч-Осмоловская Анастасия Александровна

Крюков Артем Игоревич

Студент (магистр)

Национальный исследовательский университет «Высшая школа экономики», Факультет гуманитарных наук, Москва, Россия

E-mail: kryukov.ai39@gmail.com

В тоталитарных странах, таких как сталинский СССР, граница между личным и общественным крайне размытая, политическая повестка, пропагандистские лозунги проникают во все сферы жизни человека. В этом исследовании предпринята попытка проанализировать влияние советского социально-политического дискурса на дневниковые записи граждан СССР с конца 1920-х до 1953 г. Этот период включает в себя формирование культа личности Сталина, периоды политических репрессий, ВОВ и множество внешнеполитических событий (от конфликтов на Дальнем Востоке до гражданской войны в Испании), в которых СССР принимал активное участие. Все эти события активно обсуждались в советской прессе и обществе.

Работа находится в методологической рамке “дальнего чтения” филолога Ф. Моретти [3]. Этот метод был придуман для изучения неканонической художественной литературы, но как показывают последние исследования [1, 4], он применим и к ряду других источников, в том числе к дневникам [2]. Дневники тоже имеют свой канон, который изучают исследователи (дневник Анны Франк, Екатерины Великой, Пришвина), но большинство дневников интересны только родственникам авторов, редко печатаются и становятся предметом отдельного научного исследования. Поэтому для введения этих исторических источников в научный оборот стоит прибегнуть к “дальному чтению” и сопутствующим ему методам компьютерной лингвистики.

Аналізу подвергся корпус дневников “Прожито” [5]. Записи заданного периода составляют почти треть корпуса, это 119394 записи 858 авторов. Для сравнения и выявления закономерностей был также изучен публицистический подкорпус НКРЯ. Из методов «дальнего чтения» были выбраны тематическое моделирование и поиск коллокаций, которые были реализованы методами Python-библиотек для обработки естественного языка NLTK и машинного обучения sklearn. Корпус был предварительно лемматизирован с помощью библиотеки `rumorphy2` и очищен от пунктуации. Тематическое моделирование делалось на 20 тем по 10 компонент по каждому году с 1928 по 1953 год.

Из полученных топиков мы отобрали несколько биграмм, которые можно отнести к общественно-политической лексике. Некоторые из них повторялись из года в год, например, такие базовые для лексикона советского человека слова как «советская власть» (выделяется как компонента разных тем в 1928, 1929, 1930, 1935, 1937, 1941, 1942 гг.), «классовый враг»/«классовая борьба» (наблюдаются в моделировании за 1929, 1930, 1932, 1935 год), «социалистическое строительство» (1928, 1929), «враг народа» (1937, 1938), «диктатура пролетариата» (1930, 1935, 1937) и др. Был проанализирован контекст некоторых неоднократно встречающихся топиков.

Судя по корпусу, по крайней мере пропагандистские кампании против врагов народа нашли отклик у обычных советских граждан. Пики их упоминаний в дневниках и

публицистике по времени совпадают. Однако, если учитывать тот факт, что корпус публицистики в разы меньше корпуса дневников, равенство упоминаний терминов говорит о по-настоящему масштабной кампании (см. рис. 1).

Одной из самых распространенных биграмм (721 вхождение в корпус) оказалась «советская власть». На графике видно два скачка её упоминаний - в 1937 году в НКРЯ (в рамках кампании против «врагов народа» акцентировался их вред «советской власти») и в годы ВОВ. Резкий скачок в годы войны можно попробовать объяснить оккупацией многих регионов СССР немецкими войсками, чей режим постоянно сравнивался с положением при советской власти (см. рис. 2).

Можно предположить, что внешней политике в советском обществе уделяли повышенное внимание. Например, значительный рост упоминаний «Испании» и «испанского народа» пришёлся на 1936-1937 гг., когда там разворачивалась гражданская война и в которой СССР открыто поддерживало республиканцев. (см. рис. 3, 4).

Интересна эволюция топики «гражданская война». Помимо пика в 1936-1937 гг. из-за упомянутой выше войны в Испании, на графике виден ещё больший пик, пришедшийся на годы ВОВ, который никак не коррелирует с упоминаниями в прессе. Помимо упоминаний этой темы в контексте школьной программы или описании чьего-то боевого опыта, в годы войны авторы, зная, какие лишения их ждут, сравнивают войну с Германией с другим травматичным личным опытом - опытом Гражданской войны. (см. рис. 5)

Отдельно стоит остановиться на аномальном пике биграмм «северный полюс», приходящийся на 1937 год (см. рис. 6). Она относится к топику об экспедиции папанинцев и открытием первой в мире дрейфующей станции «Северный полюс». Отчасти такое большое количество записей по этой теме можно объяснить тем, что дневники всех членов экспедиции есть в корпусе. Но за новостями об экспедиции пристально следили и обычные советские граждане, встречая новость о посадке в Арктике с восхищением и гордостью.

Тема сталинской конституции предсказуемо активно освещалась в прессе, но, как ни странно, пик упоминаний её в дневниках приходится на год-два раньше (см. рис. 7). Проект новой конституции, формально значительно расширяющей права советских граждан, был воспринят авторами дневников в основном позитивно. Некоторые подвергали критике «раболопный» тон прессы и партийной номенклатуры в адрес конституции и её инициатора.

Методы тематического моделирования и поиска коллокаций позволили нам выделить среди других те события общественной жизни, которые нашли наибольший отклик у авторов дневников. Среди них - борьба с врагами народа во время «ежовщины», экспедиция папанинцев, принятие новой сталинской конституции. Внешнеполитическая повестка также оказалась небезразлична - люди следили за войной в Испании, противоборством коммунистов и националистов в Китае, применением американцами атомной бомбы. Безусловно, стоит отметить события Великой Отечественной войны. Хотя темы 1941-1945 годов были переполнены топиками, посвящённым военным действиям, голоду, смертям и другим спутникам войны, из рефлексии людей нам удалось вытащить неочевидное сопоставление опыта войны с немцами с опытом войны гражданской как наиболее похожего по масштабам бедствия.

Источники и литература

- 1) Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using Topic Modeling Methods for ShortText Data: A Comparative Analysis. *Front. Artif. Intell.*, 3, 42.
- 2) Blevins, C. (2010). Topic modeling Martha Ballard's diary. Cameron Blevins.
- 3) Moretti, F. (2000). Conjectures on world literature. *New left review*, 54-68.

- 4) Riddell, A. B. (2014). How to read 22,198 journal articles: Studying the history of German studies with topic models. *Distant Readings: Topologies of German culture in the long nineteenth century*, 91-114.
- 5) Прожито.орг: <https://prozhito.org/>

Иллюстрации



Рис. 1. График частотности упоминания биграмммы 'враг народа'



Рис. 2. График частотности упоминания биграммы 'советский власть'



Рис. 3. График частотности упоминания униграммы 'испания'



Рис. 4. График частотности упоминания биграммы 'испанский народ'



Рис. 5. График частотности упоминания биграммы 'гражданский война'



Рис. 6. График частотности упоминания биграммы 'северный полюс'



Рис. 7. График частотности упоминания униграммы 'конституция'