

Корпусное исследование лексических трендов, связанных с COVID-19, в
англоязычном медиа- и интернет-дискурсе.

Вадим С. Малецкий

Национальный исследовательский университет «Высшая школа экономики»,
Москва, Россия, vsmaletskiy@edu.hse.ru

Пандемия COVID-19 стала одним из главных событий последних лет, значительно повлиявшим на многие сферы жизни, в том числе – на язык (в частности – английский), где ее след можно наблюдать уже сейчас. Однако большинство работ, исследующих влияние пандемии на лексику английского языка, фокусируются на неологизмах: анализируют способы их создания и применяют к ним различные классификации (см., например, [Абрамова 2020; Катермина, Липириди 2020; Селезнева 2020; Akut 2020; Asif и др. 2020]) не уделяя внимания другим важным, на наш взгляд, изменениям, которые происходят в языке в период подобных событий – например, сдвигам в семантике и узусе существующих лексических единиц. Целью данной работы является исследование трендов в лексике английского языка, вызванных продолжающейся пандемией COVID-19; для этой цели были использованы как количественные (корпусный и статистический), так и качественные (описательный, аналитический и сравнительный) методы исследования.

На первоначальном этапе работы был произведен обзор интернет-ресурсов, посвященных обсуждению лексики, характерной для периода пандемии¹. Из них для последующего анализа была отобрана 41 языковая единица (например: *aerosol*, *face covering*, *hand gel*, *mask*, *respirator* и др.) – слова и словосочетания, про которые утверждалось, что частота их употребления возросла или же у них сформировались новые значения. Таким образом, анализ был разделен на две части: анализ трендов частотности и анализ сочетаемости. Для того, чтобы проанализировать тренды, для каждой из анализируемых единиц были собраны данные о частотности их употребления за два периода: (1) до начала пандемии (2018-2019 гг.) и (2) в период пандемии (январь 2020 г. - апрель 2021 г.). Данные за первый период собирались за весь интервал времени по корпусу Google Books; данные за второй период – ежемесячно, на основе The Coronavirus Corpus² (к такому разделению пришлось прибегнуть из-за того, что

¹ <https://goo.su/7R4i>; <https://goo.su/7R4j>; <https://goo.su/7r4K>.

² <https://www.english-corpora.org/corona/>.

последний не содержит в себе данных до начала пандемии³). Все данные о частотностях употребления были масштабированы (размещены в интервале $[0, 1]$), после чего изменения в частоте встречаемости рассматриваемых единиц были визуализированы с помощью графиков. Примеры наиболее интересных наблюдений представлены на Рис. 1-2.

Так, при сравнении графиков частотности употребления слов “epidemic” и “pandemic” (Рис. 1) можно отметить их зеркальность. Это соотносится с историей распространения COVID-19: до марта 2020 года ситуация с вирусом считалась эпидемией, однако, 11 марта ВОЗ объявила пандемию коронавирусной инфекции⁴.

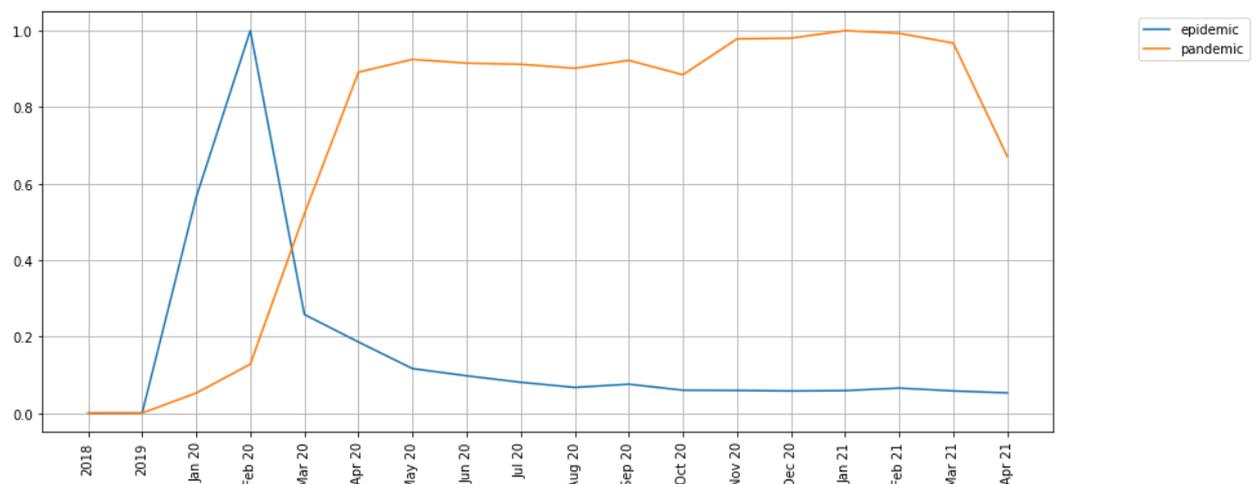


Рисунок 1. Сравнение графиков трендов частотности epidemic и pandemic

Графики слова “zoom”, словосочетания “work from home” и его аббревиатуры “wfh” (Рис. 2) почти идентичны. Такая схожесть не случайна, так как во время карантина многие предприятия перевели своих сотрудников на удаленную работу и именно приложение Zoom стало самым популярным для коммуникации внутри компаний⁵.

³ Два этих корпуса несколько отличаются по жанровому составу, что могло, некоторым образом, отразиться на результатах; в последующем планируется оценить степень этого влияния, сравнив данные The Coronavirus Corpus с данными Timestamped JSI Web Corpus (см. ниже).

⁴ <https://goo.su/7R9q>.

⁵ <https://goo.su/7R9s>.

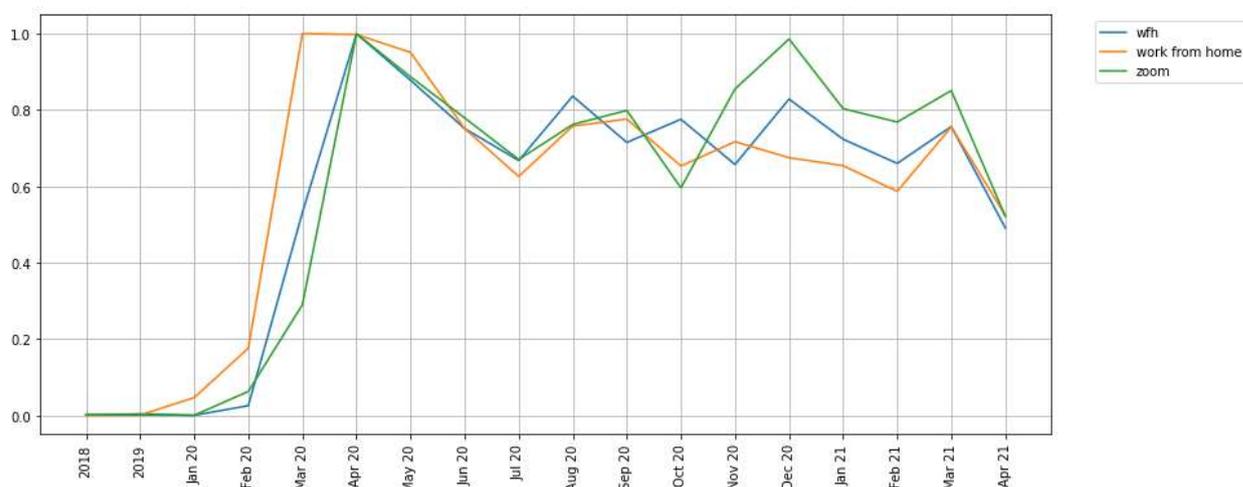


Рисунок 2. Сравнение графиков трендов частотности WFH, work from home и Zoom

Для анализа изменений в узусе рассматриваемых единиц в ходе двух выделенных периодов были проанализированы данные о паттернах их совместной встречаемости с другими лексическими единицами. Для этих целей с помощью поискового интерфейса инструмента SketchEngine⁶ из корпуса Timestamped JSI Web Corpus были отобраны по 2000 произвольных контекстов их употребления в каждом из периодов и выделены их наиболее статистически значимые коллокации (юниграммы и биграммы), с помощью статистической метрики ΔP [Ellis, Ferreira-Junior 2009]. Один из наиболее показательных случаев изменения коннотаций и коллокаций представлен в Табл. 1:

Mask	2018-2019 гг.	2020 г.
Unigrams	identity; sleep; hockey; helmet; performance; Halloween; villain; cream; hood; batman; victim; skin	coronavirus; covid-19; sanitizer; pandemic; mandatory; enforce; refuse; healthcare; supply; compulsory; shortage
Bigrams	cover face; hide identity; Korean beauty	social distancing; protective equipment; hand sanitizer; wash hand; public health; face covering; disease control; prevent spread

Таблица 1. Примеры наиболее частотных сочетаемостей для mask

⁶ <https://www.sketchengine.eu/>.

Видно, что все рассматриваемые единицы в 2020 г. изменили контекст своего использования по сравнению с периодом 2018-19 гг. В 2020 г. Все они в основном употреблялись вместе с лексикой, связанной с пандемией COVID-19, однако встречались и другие тематики, например, политическая, экономическая, технологическая и социальная. В период 2018-2019 гг. темы лексикона были более разнообразны: к перечисленным выше можно добавить бытовую, религиозную, военную, экологическую и спортивную.

В результате были сделаны следующие выводы:

1. Такие глобальные события, как пандемия коронавирусной инфекции, не могут не оставить свой след в языке — помимо провоцирования создания неологизмов, частотность употребления ряда уже имеющихся в языке слов испытывает резкий рост.

2. Этот взрывной рост, в свою очередь, неизбежно влияет на семантику языковых единиц, что отражается в их коллокациях и контексте использования.

3. Существующая экстралингвистическая обстановка находит свое отражение в языке, и для периода пандемии можно выделить несколько основных мотивов, которые оставили наиболее явный след. Так, выделяются экономические проблемы: нехватка средств защиты от коронавируса из-за недостаточных объемов производства. К ним можно добавить возросшую активность государственных структур, которые были обязаны вводить меры для ограничения распространения вируса и следить за их соблюдением.

Список литературы

1. Абрамова 2020 – Абрамова Г. С. Особенности английских неологизмов понятийно-тематической области COVID-19 // Филологические Науки Вопросы Теории И Практики. 2020. Т. 13. № 9. С. 172–176.

2. Катермина, Липириди 2020 – Катермина В. В., Липириди С. Х. Особенности отображения пандемии коронавируса в лексике медицинского дискурса (на основе английских неологизмов) // Известия Волгоградского Государственного Педагогического Университета. 2020. Т. 147. № 4. С. 170–175.

3. Селезнева 2020 – Селезнева О. Н. Влияние COVID-19 на словарный состав современного английского языка // Филология, лингвистика, журналистика: актуальные вопросы и современные аспекты. Пенза, 2020. С. 24–26.

4. Akut 2020 – Akut K. B. Morphological Analysis of the Neologisms during the COVID-19 Pandemic // Int. J. Engl. Lang. Stud. 2020. Т. 2. № 3. С. 01–07.

5. Asif и др. 2020 – Asif M. и др. Linguistic Analysis of Neologism Related to Coronavirus (COVID-19) // SSRN Electron. J. 2020.

6. Ellis, Ferreira-Junior 2009 – Ellis N. C., Ferreira-Junior F. Constructions and their acquisition: Islands and the distinctiveness of their occupancy // Annu. Rev. Cogn. Linguist. 2009. Т. 7. № 1. С. 188–221.