

АНАЛИЗ ДЛИН ПРЕДЛОЖЕНИЙ В ТЕКСТАХ ПУБЛИЧНЫХ ВЫСТУПЛЕНИЙ НА МАТЕРИАЛЕ ЕЖЕГОДНЫХ ВЫСТУПЛЕНИЙ ЛИДЕРОВ ПАРТИЙ ВЕЛИКОБРИТАНИИ

Цижмовска Н.Л.¹, Мартюшев Л.М.¹

¹⁾ Уральский федеральный университет имени первого Президента России Б.Н. Ельцина, г. Екатеринбург, Россия

E-mail: nata-cizmov@yandex.ru

ANALYSIS OF SENTENCE LENGTHS IN PUBLIC SPEAKING BASED ON THE MATERIAL OF THE ANNUAL SPEECHES PARTY LEADERS OF THE UK

Tsizhmovska N.L.¹, Martyushev L.M.¹

¹⁾ Ural Federal University, Yekaterinburg, Russia

We have two main results under this study. The first result is linear decrease in the length of sentences (with the slope coefficient is 0.14). The second result is that the distribution of sentence length is best described by the Weibull distribution.

С середины XX века был достигнут большой прогресс в исследовании языка при помощи математических методов, и, в частности, в области статистического анализа длин предложений (см. обзоры [1,2]). Полученные результаты имеют не только теоретическое, но и прикладное значение, в частности для классификации текстов по темам, анализа тональности текста, создания чат-ботов, установления авторства текстов и многое другое.

Целенаправленное изучение зависимости длины слов и предложений от времени (эпохи) было начато А.В. Мельниковой [1], которое было продолжено и существенно расширено [2]. Анализировалась выборка инаугурационных речей президентов США из 58 текстов в период с 1789 до 2017 года. Используя статистический анализ, было определено, что за 228 лет средняя длина предложений в тексте линейно уменьшается (коэффициент угла наклона составляет 0.14 ± 0.03). Так вначале изученного периода средняя длина равна 45 словам, а в конце 14.

Для проверки того насколько полученный результат в [2] имеет универсальный характер для англоязычных выступлений в данной работе были проанализированы ежегодные выступления лидеров партий Великобритании, которые были взяты из электронного архива [3]. Было проанализировано 224 текста в период с 1895 до 2018. Средняя длина предложений в тексте также линейно уменьшается (коэффициент угла наклона составляет 0.14 ± 0.02). Так вначале изученного периода средняя длина предложения равна 31 слову, а в конце 14. Медиана и максимальное количество слов в предложении аналогично линейно уменьшаются. Интересным фактом стало то, что среднее количество предложений в текстах со временем линейно увеличивается (в 1985 году – 144 предложения, в 2018 – 378), при

этом среднее значение количества слов в текстах со временем не изменяется, а лишь колеблется в определенных значениях.

Сильным отличием Великобритании от США является большое количество текстов, являющихся дву- и более модальными. Так если в США было 14 из 48 текстов, то в Великобритании – 138 из 224 текстов. При этом подтверждения влияния модальности на другие статистические параметры не было выявлено.

Распределения длин предложений проверялись на соответствие к дву- и однопараметрическим распределениям таким, как Вейбулла, логнормальное, нормальное, полунормальное, Рэлея. Ранжирование вероятности принадлежности выборки к распределению проводилось по критерию Колмогорова-Смирнова. С большой вероятностью можно сказать, что рассмотренные речи распределены по Вейбуллу. Так на первом месте Вейбулл оказывается 147 раз, на втором – 72, на 3 – 4 раза, то есть во всех речах он занимает с 1 по 3 место. Только в 32 текстах уровень значимости ниже 5%. Средний же уровень значимости для 224 текстов для Вейбулла составил 0,25, для США он составил 0,51.

Вышеперечисленное подтверждает и расширяет не только полученные ранее результаты, но и их универсальный характер для других англоязычных выступлений [1,2].

1. А. В. Мельникова, ВКР магистра (руководитель Мартюшев Л.М.), Екатеринбург: УРФУ (2019).
2. N. L. Tsizhmovska and L. M. Martyushev, AIP Conf. Proc. 2313, 070029 (2020).
3. British Political Speech <http://britishpoliticalspeech.org/index.htm>