English Influence on Gujarati: A Statistical Examination of News Headlines
When cultures come into contact with one another, their languages may not always be similar. Many times, foreign words must be incorporated into a different language, especially in instances when no direct translation can be found. For example, 'taco' in English is borrowed from Spanish, and 'chutney' in English is borrowed from Hindi. Different words are borrowed into English. This study examines what English words are borrowed into Gujarati.

Gujarati is an Indo-European language spoken by approximately 55 million people [3]. It is primarily spoken in the state of Gujarat in India. In the US, it was one of the fastest growing foreign language groups between 2010 and 2017 [1].

Using a public dataset of headlines taken from 3 categories (technology, entertainment, or business) of the News18 Gujarati Website, a database of headlines with 5026 total characters was built [2]. A human rater classified each word as borrowed (defined as having a sound and intended meaning that is closely related to an English word) or a native Gujarati word. These words were then manually classified by their part of speech (noun, adjective/adverb, verb).

The research questions of interest were:
Is there a difference in the number of borrowed words seen in different headline categories?
Is there a relationship between the part of speech and the headline category of borrowed words?

We found that there was no difference between the number of borrowed words across the three headline categories or the three parts of speech, Kruskal-Wallis $\chi^2(2, N = 97) = 0.875$, $p = 0.64$. There was also no relationship between part of speech and headline category in terms of number of borrowed English words, $\chi^2(2, N = 189) = 1.058$, $p = 0.59$, $V = 0.08$.

Although the results were not statistically significant, there were still some interesting patterns shown in the data. In some cases, words that had no Gujarati equivalent were borrowed, such as 'February.' Even though India typically follows a lunar calendar, sometimes months must be given in the western format. Conversely, a word such as 'budget' has an equivalent concept in Gujarati, but it is possible that even native speakers are more familiar with the English word rather than the original Gujarati one. The question of why 'budget' and other words may be borrowed could have something to do with influences of British occupation on India, and also may be due to increases in global communication and connection.

One other interesting pattern involved instances where sound and meaning borrowing occurred with a Gujarati morphological modification. For instance, the word 'center' was borrowed and written out in Gujarati script to sound like 'seṇṭar.' It was written in a plural form, but instead of using the English plural marker 's', which would have been 'centers,' it was written using the Gujarati plural marker 'o' as 'seṇṭaro.' This may indicate that some words in these headlines underwent a two-step borrowing process first involving the English sound and meaning and then undergoing native Gujarati morphological rules.

References

[1]   Bhattacharya, A. (2018, September). *America's fastest growing foreign language is from south India*. Quartz India.
        https://qz.com/india/1399825/telugu-is-uss-fastest-growing-foreign-language/

[2]   Gaurav. (2019, December). Gujarati News Dataset, Version 1. Retrieved from
        https://www.kaggle.com/disisbig/gujarati-news-dataset

[3]   *Gujarati: The language spoken by more than 55 million people*. (2017, January). The Straits Times.
        https://www.straitstimes.com/singapore/gujarati-the-language-spoken-by-more-than-55-million-people