

Exercise 1: Zipf's Law of Abbreviation

This exercise counts 25% toward your final grade. Your task is to write a brief report on whether Zipf's Law of Abbreviation holds for (i) a different language than English and (ii) a particular genre of text in English that is not a novel (e.g., music lyrics, poems, forum entries, tweets). The report should be a PDF document; no longer than 2 pages. It should include the sections described below. The 2-page limit does not include references (optional); supplementary material (optional); and the list of contributions (only if working in groups).

Introduction (10%)

What is Zipf's Law of Abbreviation? What language will you test and why? What genre will you test and why? Explain your expectations prior to conducting the analysis.

Material and methods (20%)

Describe your approach to answer the question whether Zipf's Law of Abbreviation holds for (i) and (ii), as well as the data you will be using (where did you get it? how did you process it?). In particular, explain how you tokenized the data.

Results (40%)

Your main results. Does the Law hold for (i) and (ii)? Why (not)? Were your expectations met? Do you think your results would generalize to other data of the same kind? Why (not)? Discuss limitations of your data and methods.

Code (30%: 20% replicability/10% clarity)

Add a link to the code (hosted on, e.g., [OSF](#) or [github](#)) or send it in with your submission. Remember to extensively comment your code. Mention the dependencies that need to be fulfilled to run the code.

List of contributions (unlimited space / only if working in groups)

Who did what in your group. You can use the [CRediT](#) system or a variant thereof.