

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 particular genre of text in English that is not a novel (e.g., music lyrics, poems, forum entries, tweets) and (ii) the same genre in a different language than English. Compare to which degree (i) and (ii) conform to the law, and to which extent they behave the same/differently and why. 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).

After the report has been evaluated, you will be asked to answer 2-3 short questions on it. This counts for the outstanding 20% of your grade for this exercise. Even if working in groups, everyone is expected to understand their report fully and be able to answer any questions about it.

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 (why this data? 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)? How do they compare? 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 (20%: 15% replicability/5% clarity)

Make your code publicly available (hosted on, e.g., [OSF](#) or [github](#)). Remember to extensively comment it. 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.