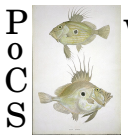

Principles of Complex Systems, Vols. 1, 2, & 3D
CSYS/MATH 6701, 6713, & a pretend number
University of Vermont, Fall 2023
Solutions to Assignment 21
... but maybe the printing press was heavier than the siege weapon

Name: Kevin Motia

Conspirators: Chris O'Neil

---

1. (3 points each)

   Using your text of choice, generate word shifts comparing two "interesting" regions of text.

   Use the Python package described in Ref. [?].

   (Various Matlab versions made by the Unreliable Deliverator do exist and need to be shared on Gitplaces.)

   Links to paper versions (arXiv is always best), Github repository, and an exhilarating Twitter feed can be found here: https://pdodds.w3.uvm.edu/research/papers/gallagher2021a/.

   "Interesting" is anything you find interesting. Could be books 3 and 12 in a series, second half of a book compared to the first half, season 4 of a show versus all seasons, etc.

   Aim to find two texts that are both reasonably large (more than 10^4 words) and fairly different in average happiness scores (though even the same scores can be meaningfully explored with word shifts).

   Let's call the two texts $T^{(1)}$ and $T^{(2)}$. In your plots, you should label them meaningfully based on your choices).

   Use a reasonable exclusion lens of your choice, e.g., [4, 6] or [3, 7].

   (a) Produce a word shift comparing text $T^{(2)}$ relative to text $T^{(1)}$. Use the average happiness of text $T^{(1)}$ as the baseline.

   (b) Interpret the word shift. Does what you see make sense? Are there any surprises? Are some words being used in what the average person might not think is their primary meaning? For example, "crying" in Moby Dick means yelling, and "sick" can mean "awesome."

(c) Produce a word shift comparing text $T^{(1)}$ relative to text $T^{(2)}$. Use the average happiness of text $T^{(2)}$ as the baseline.

(d) Comment on any asymmetries you see (the basic word shifts we use are asymmetric).

(e) Produce a word shift comparing text $T^{(1)}$ relative to text $T^{(2)}$. Now use 5 as the baseline reference score (neutral on the happiness-sadness spectrum of 1–9).

(f) Compared to your first word shift, how interpretable is this one?

Solution:

□

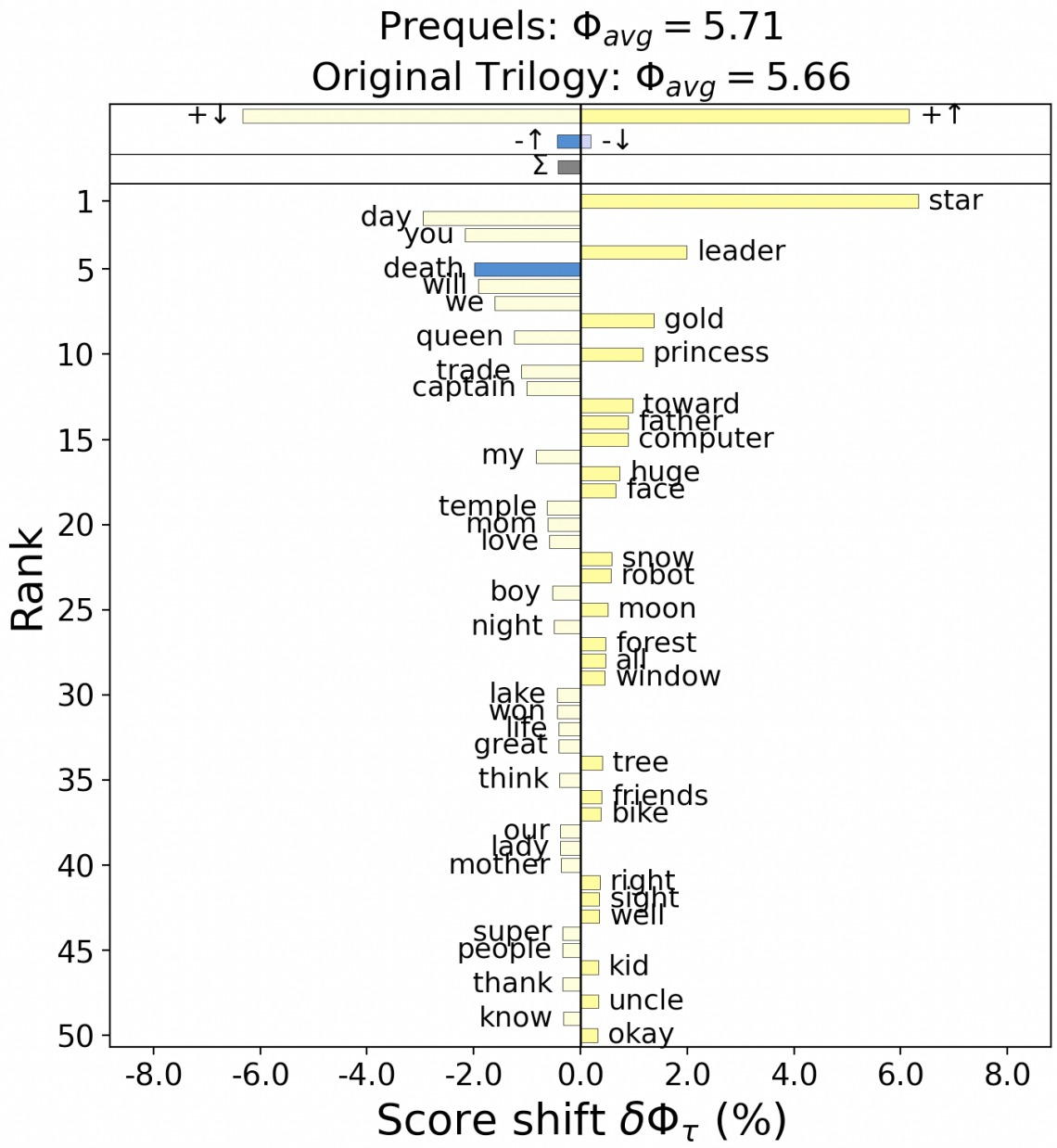

Figure 1: "Super" is a high-frequency word in the 6 Star Wars movies observed. Notably, the word is generally used when referring to enemy droids and ships such as "super battle droids" and "super star destroyers"; it does not take on the meaning in traditional modern language that is synonymous with "swell". Similarly, we can observe the word "temple" to understand how a trend can be visualized. In the prequel trilogy, the Jedi temple is a common setting. Its presence in the text declines because it is not a location that is ever visited in the original trilogy.

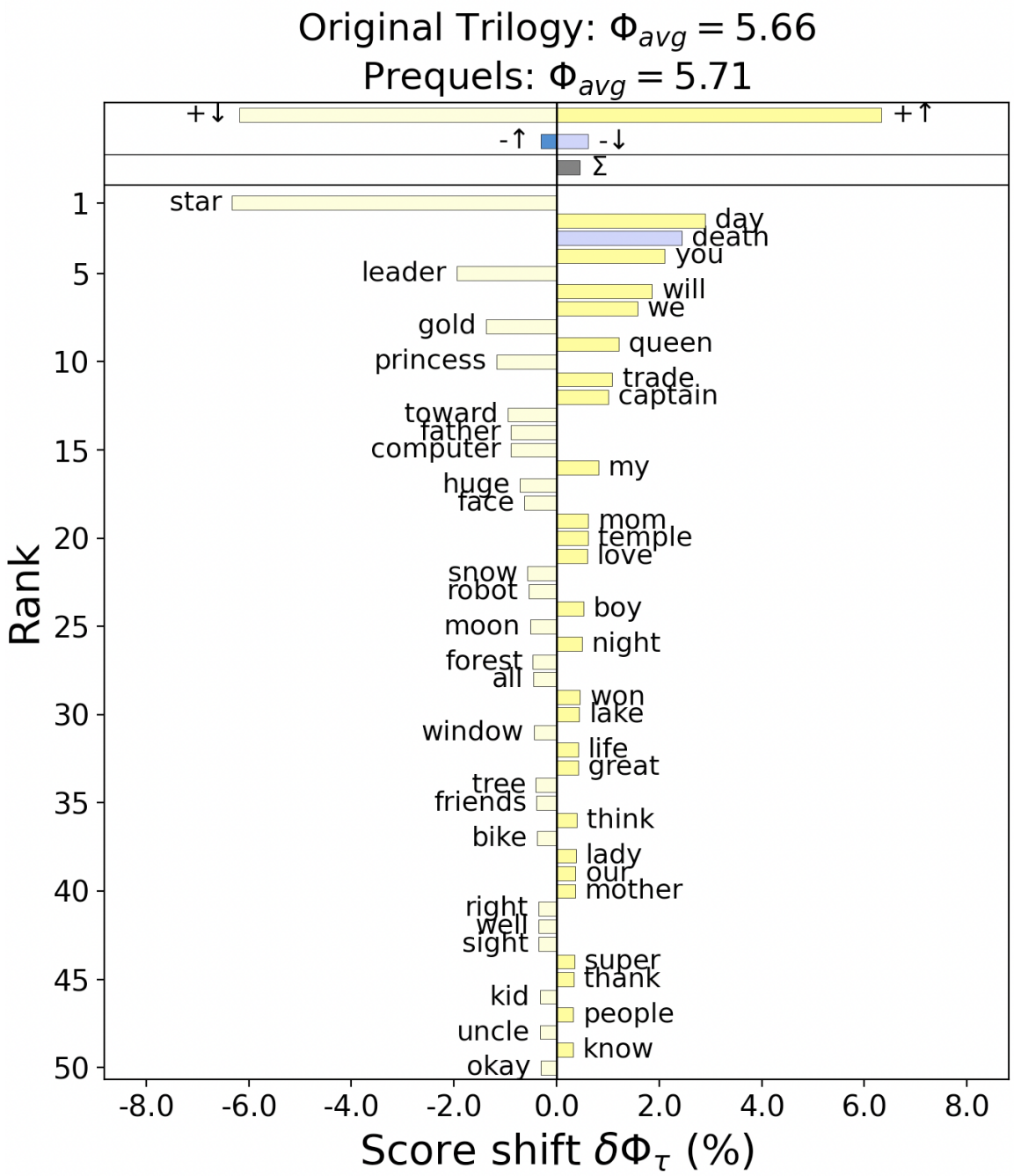

Figure 2: The word 'death' is the only negative word that appears in the top 50 ranked words, and it is a negative word that is being used less in the prequel trilogy relative to the original trilogy. It's also notable that there are no words which increase in prevalence in the prequels to the extent that the use of 'star' decreases.

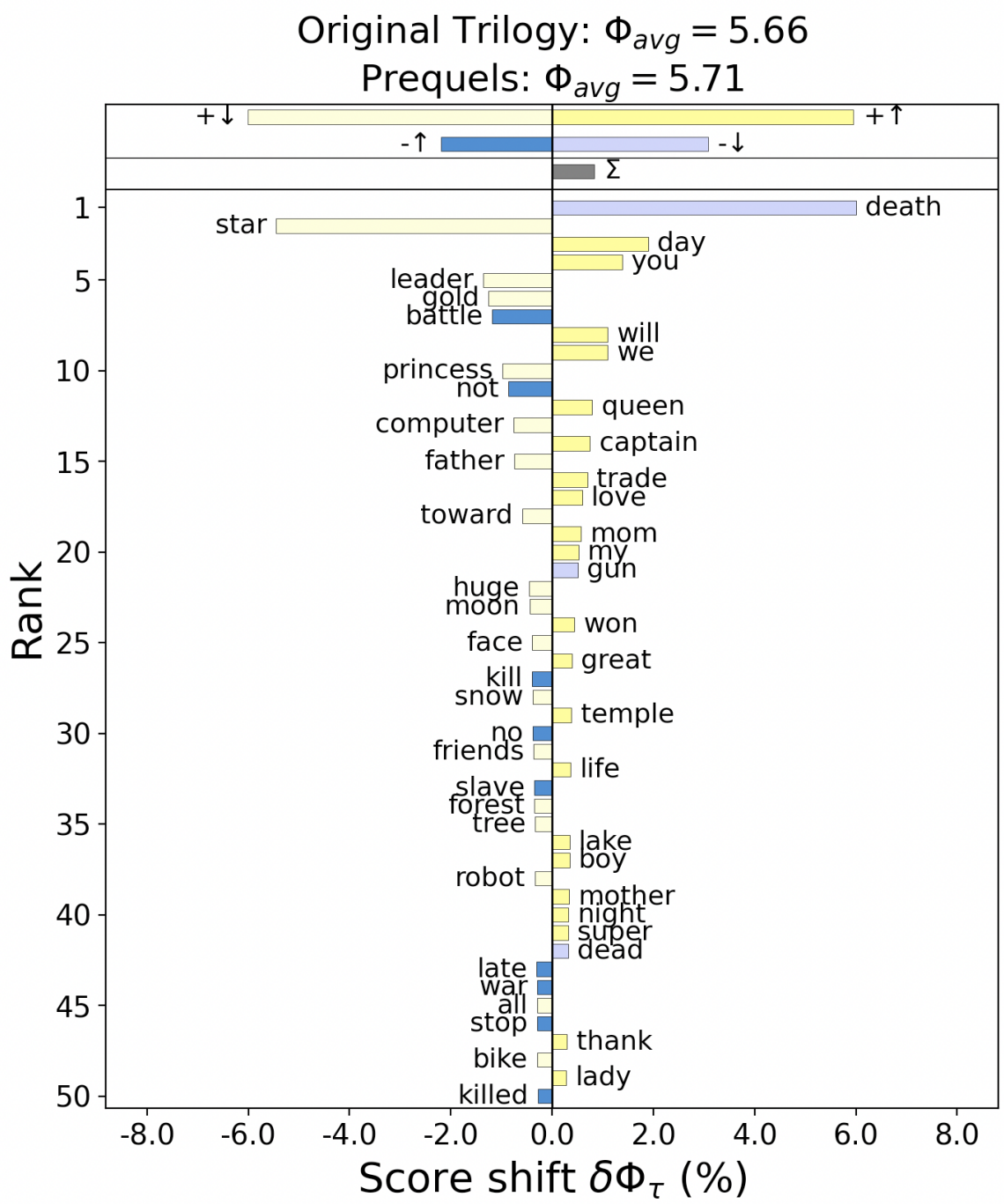

Figure 3: This figure is more interpretable than Figure 1 in that it helps visualize word shifts for a more diverse population of words. In the first word shift, we are able to see the word shift for only a single negative word, whereas here we use a neutral baseline happiness score to visualize positive and negative word frequency without bias to either side of the happiness spectrum.