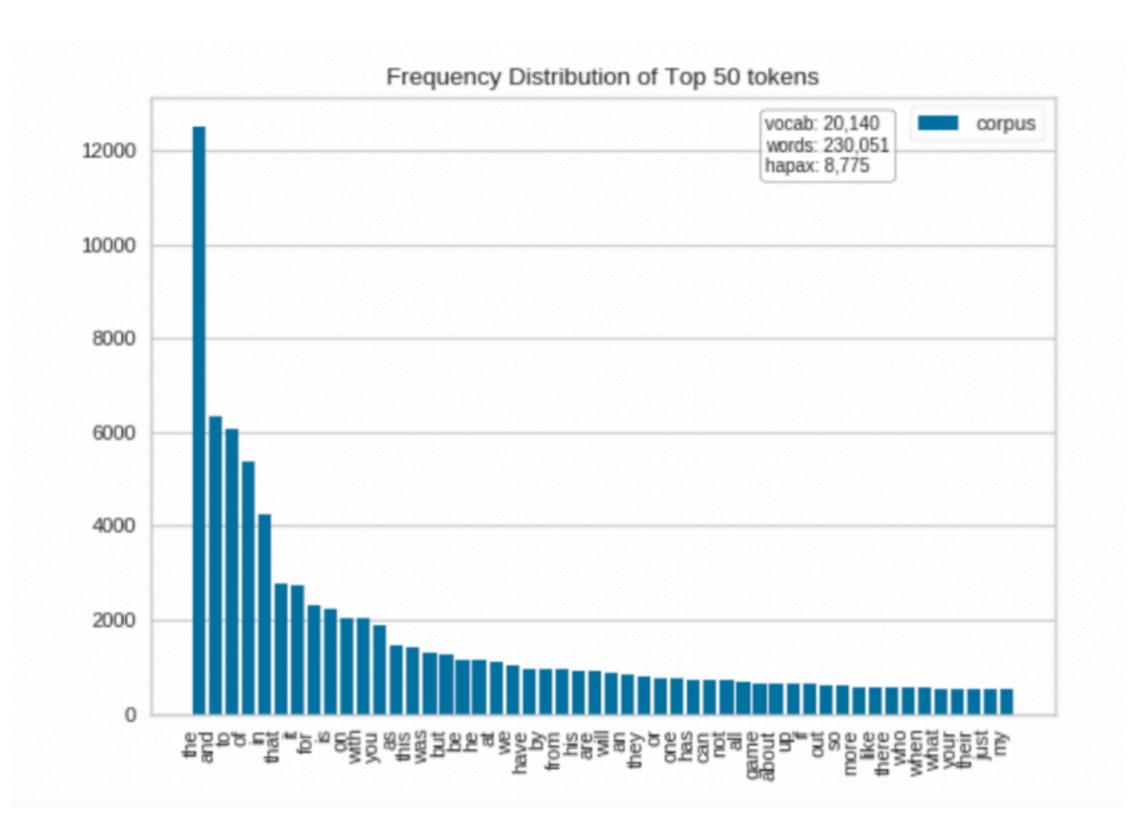
Why is modeling human language hard?

Non-uniform word distribution



The frequency of the k^{th} most frequent word is ~ proportional to 1/k

- Zipf, 1949

Compositionality

Language is **compositional**: units such as words can combine to create phrases, which can combine by the very same principles to create larger phrases. For example, a **noun phrase** can be created by combining a smaller noun phrase with a **prepositional phrase**, as in *the whiteness of the whale*. The prepositional phrase is created by combining a preposition (in this case, of) with another noun phrase (the whale). In this way, it is possible to create arbitrarily long phrases, such as,

(1.1) ...huge globular pieces of the whale of the bigness of a human head.²

The meaning of such a phrase must be analyzed in accord with the underlying hierarchical structure. In this case, huge globular pieces of the whale acts as a single noun phrase, which is conjoined with the prepositional phrase of the bigness of a human head. The interpretation would be different if instead, huge globular pieces were conjoined with the prepositional phrase of the whale of the bigness of a human head — implying a disappointingly small whale. Even though text appears as a sequence, machine learning methods must account for its implicit recursive structure.

- Eisenstein, 2019, Chp 1

Why is conversation is hard?

World Knowledge

Domain knowledge

Discourse knowledge

What flights to SFO are there tomorrow afternoon?

I found three flights from WAS to SFO tomorrow afternoon:



- \$199 DCA-SFO 2:15-3:44pm
- \$377 BMI-SFO 4:45-5:15pm
- \$232 IAD-SFO 6:00-6:30pm

Ok book the last expensive one for two adults and 1 infnt

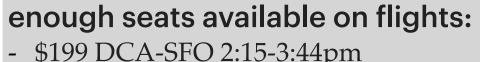


Ok, you want to book:

\$232 IAD-SFO 6:00-6:30pm Right?

No, the DCA flight

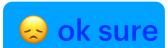




Unfortunately there are not

- \$232 IAD-SFO 6:00-6:30pm Would you like to book this flight?

\$377 BMI-SFO 4:45-5:15pm





Ok, first I'll need the name and ID of the first adult passenger.