Week 2

- Representing words and meanings
- Language modeling

Agenda

-
-
-

Models of Natural Language (NL)

What is an NL model?

How do we build an NL model?

Why should we care about NL models?

. . . let's focus on the **why** aspect first.

Why do we care about NL models?

Let's consider tokenization, a core task to any natural language processing analysis.

Given a sentence s

Naive tokenizer

NLTK tokenizers

N-grams & tokenizers

Modeling natural language: Main challenges

Processing raw text intelligently is difficult:

- it's common for words that look completely different to mean almost the same thing
- the same words in a different order can mean something completely different
- most words are rare
- even splitting text into useful word-like units can be difficult in many languages (see Japanese)

Source is spaCy website

Pre-DL models of the language

Post-DL models of the language

The architecture of Natural Language Models