

Natural Language Processing

CS 174 Section B TTh 1100-1230

CTC214

Intro to NLP

CS 174 Section B TTh 1100-1230

CTC214

Natural Language Processing?



lan·guage

/ˈlɑNGgwij/

noun

1. the method of human communication, either spoken or written, consisting of the use of words in a structured and conventional way.
"a study of the way children learn language"
2. a system of communication used by a particular country or community.
"the book was translated into twenty-five languages"

Similar:

tongue

speech

mother tongue

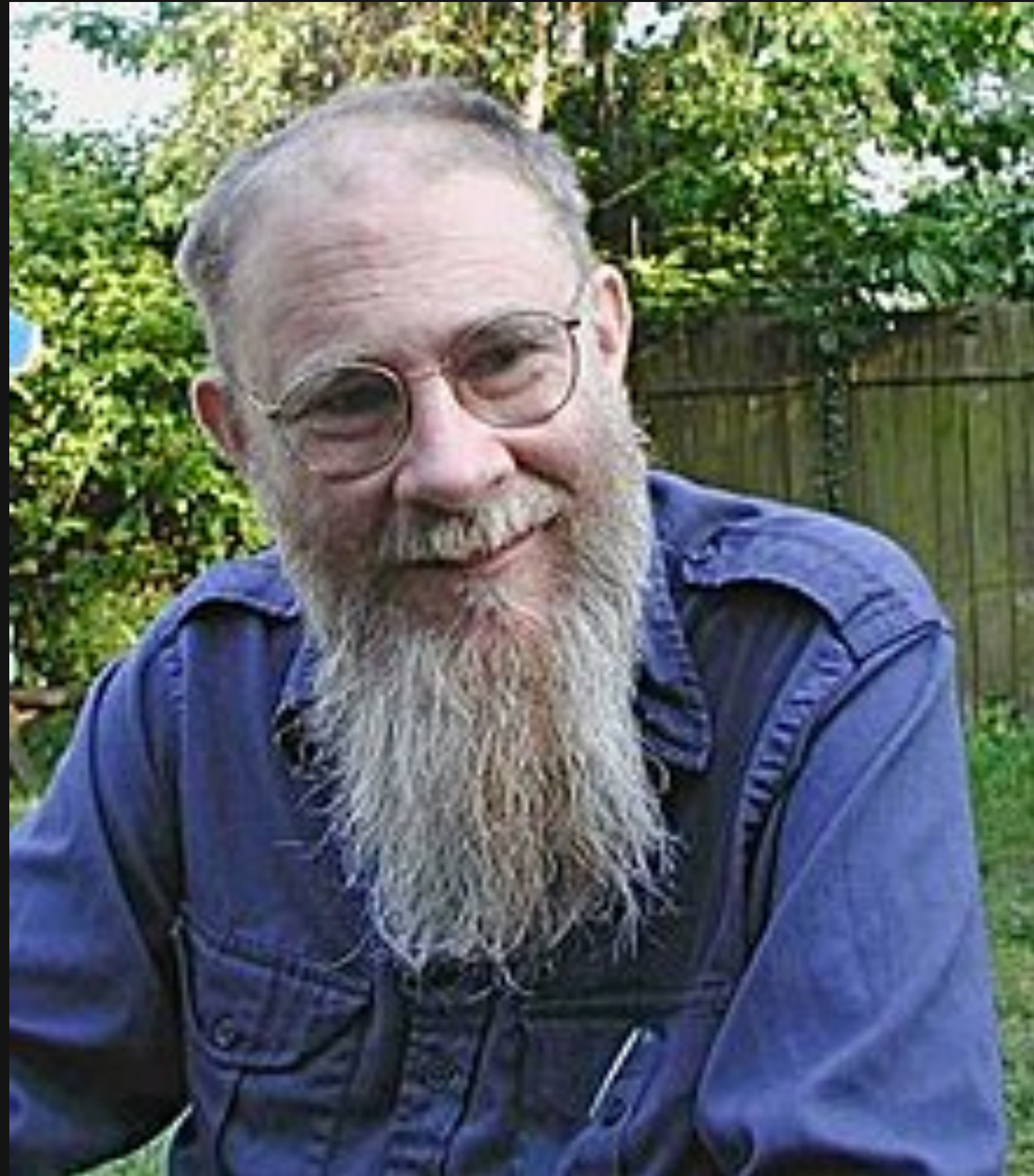
native tongue

dialect



Defining Meaning

Wooo Philosophy

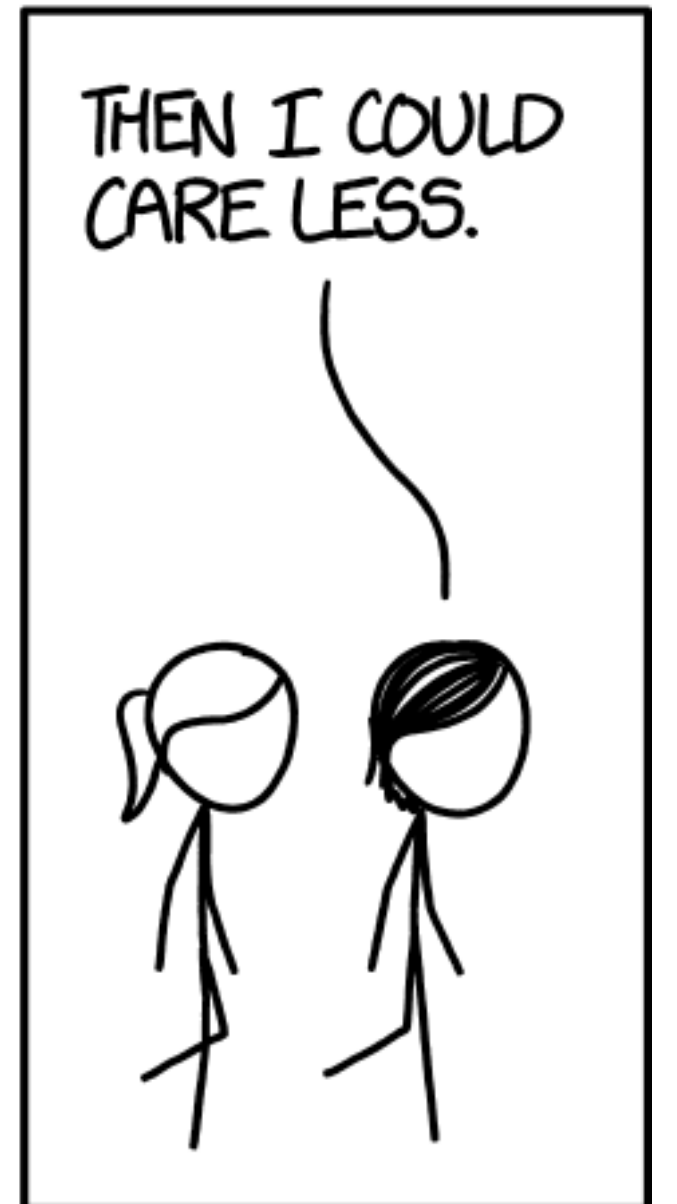
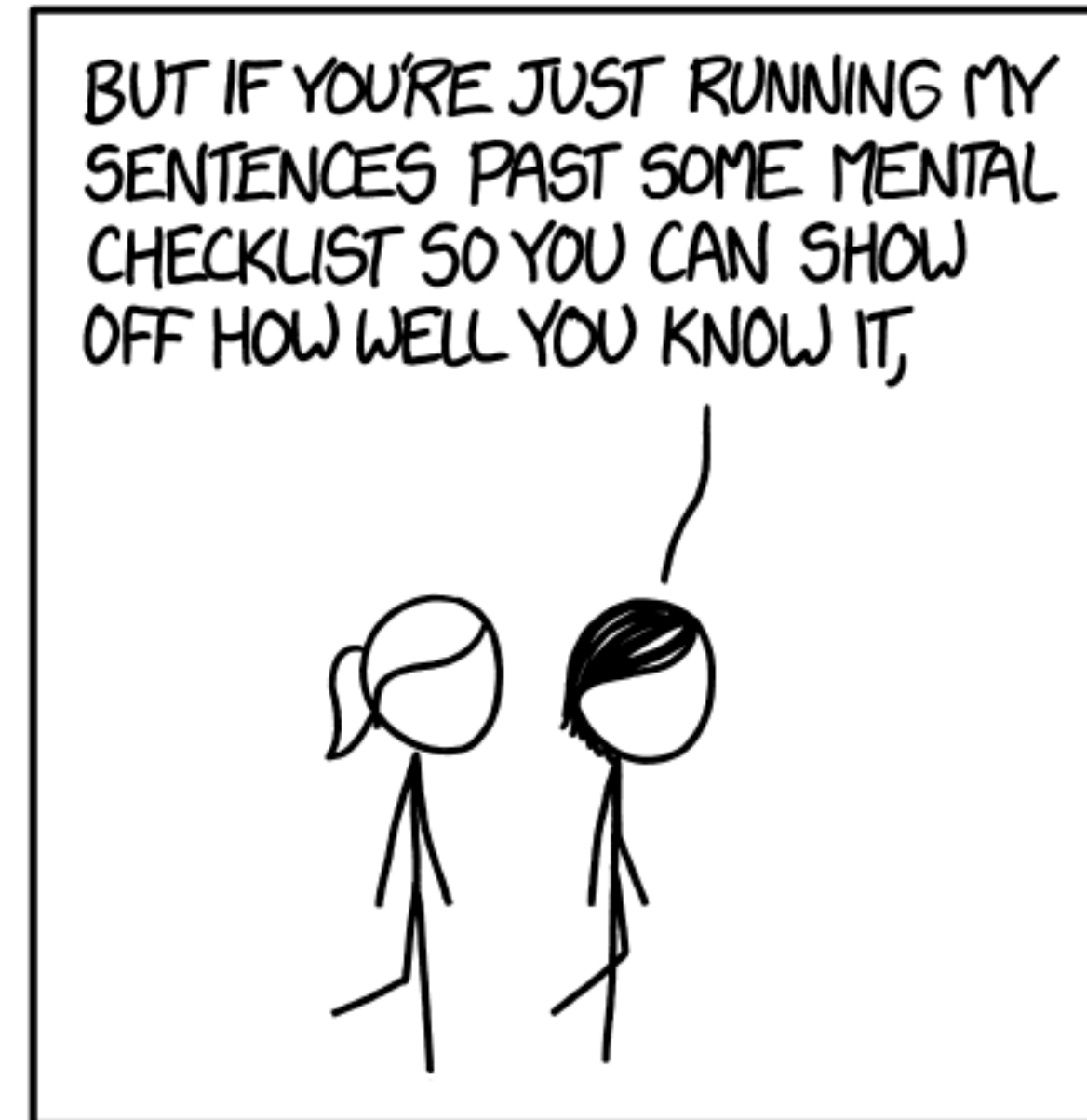
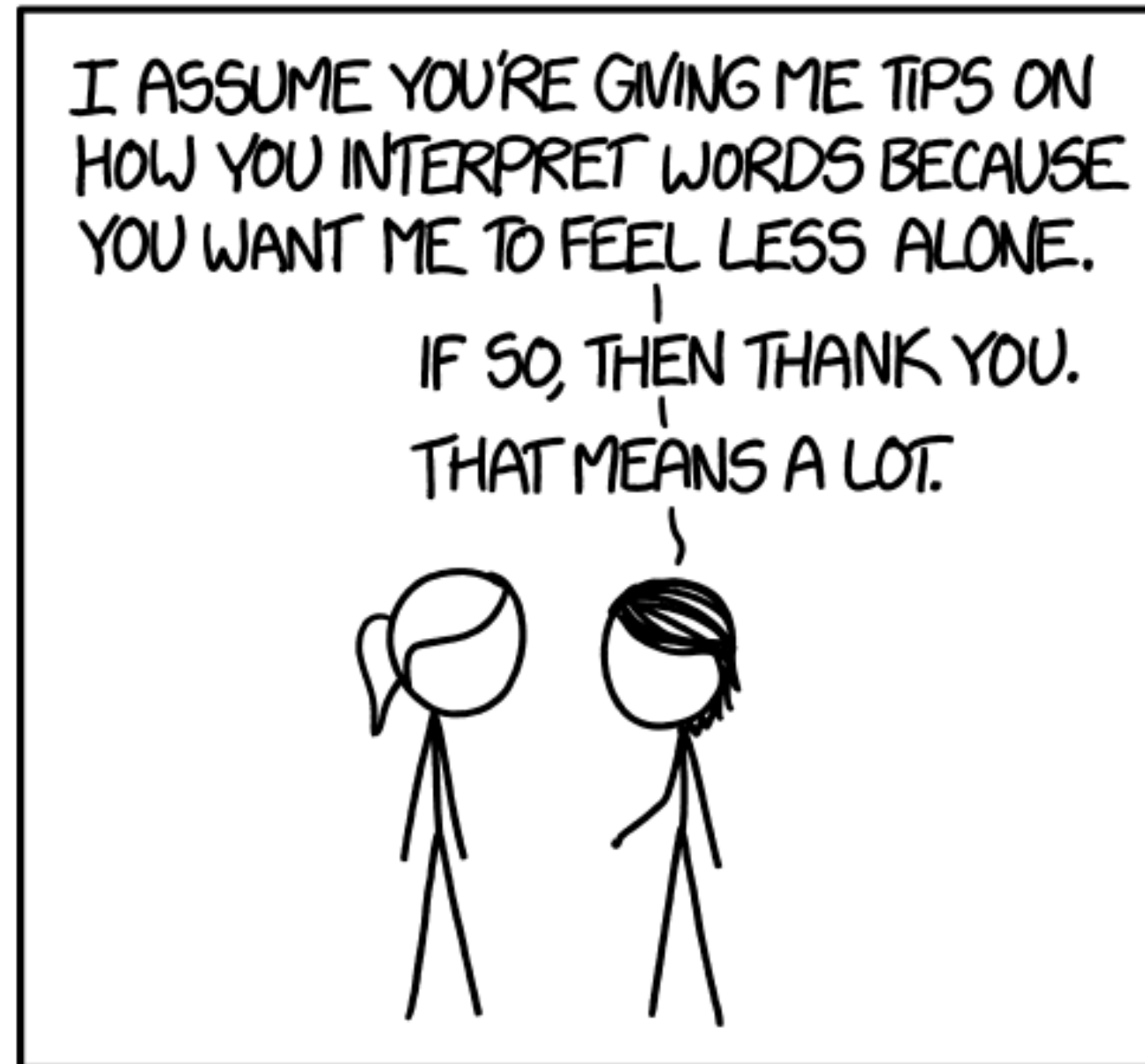
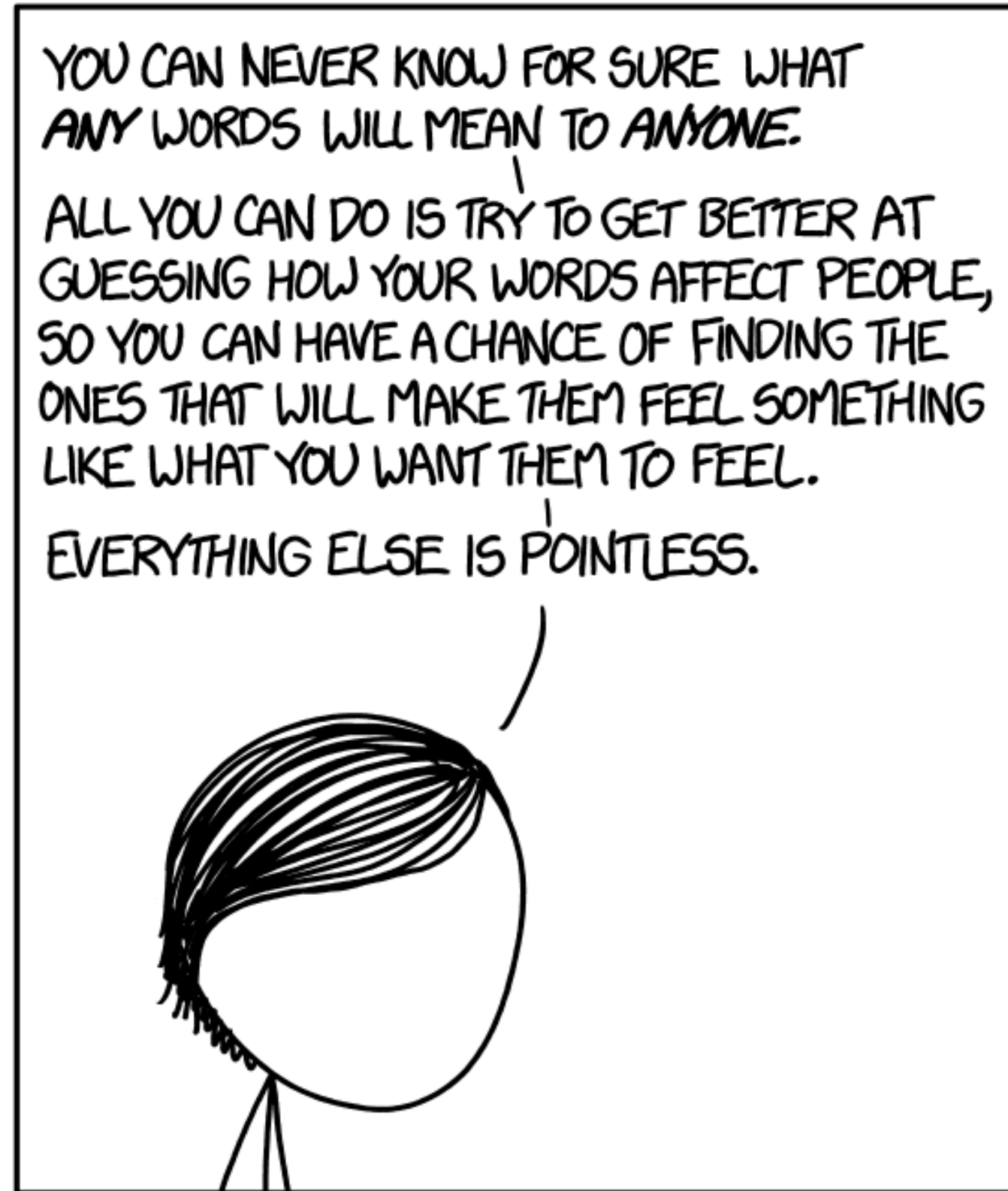
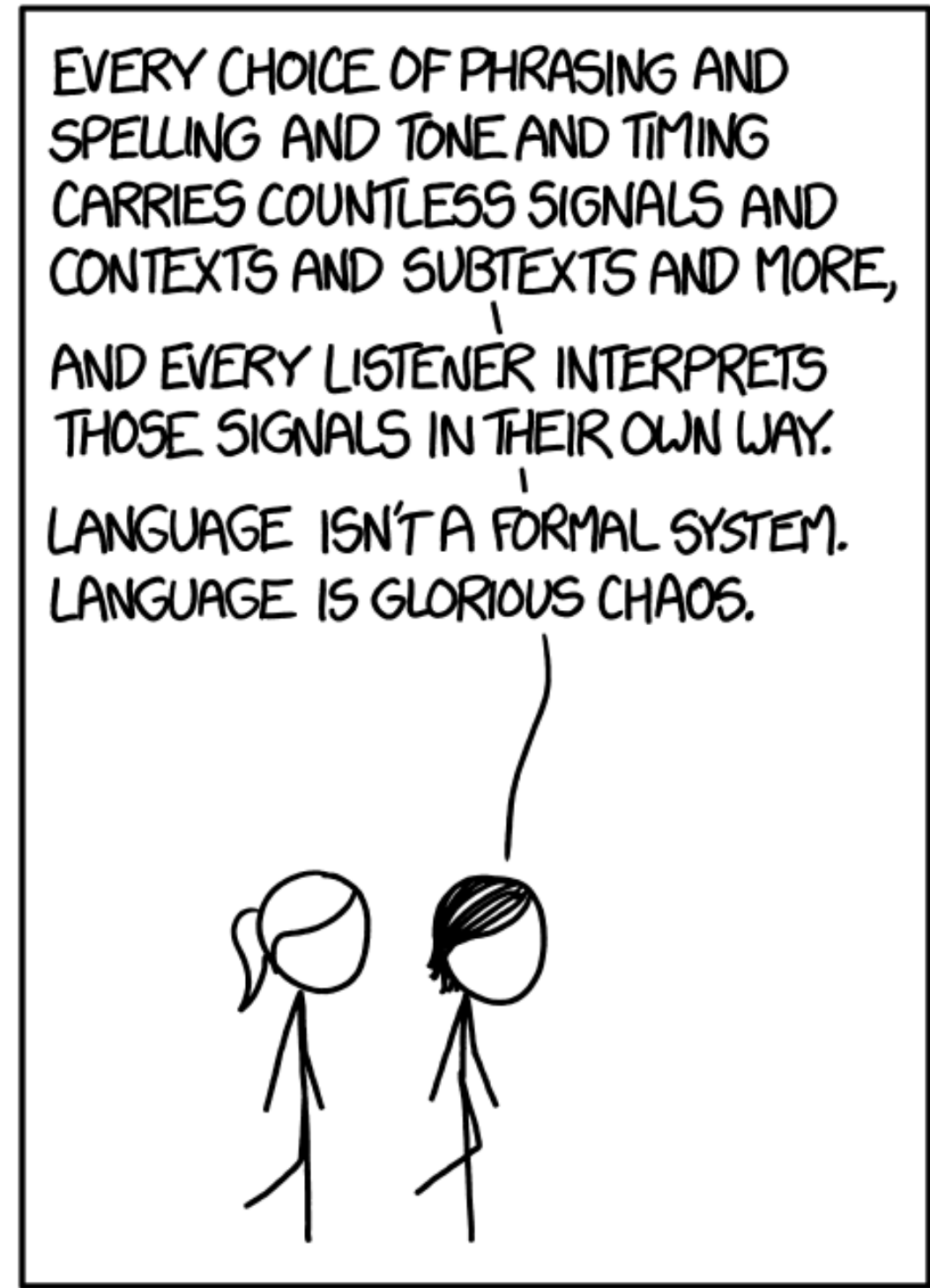
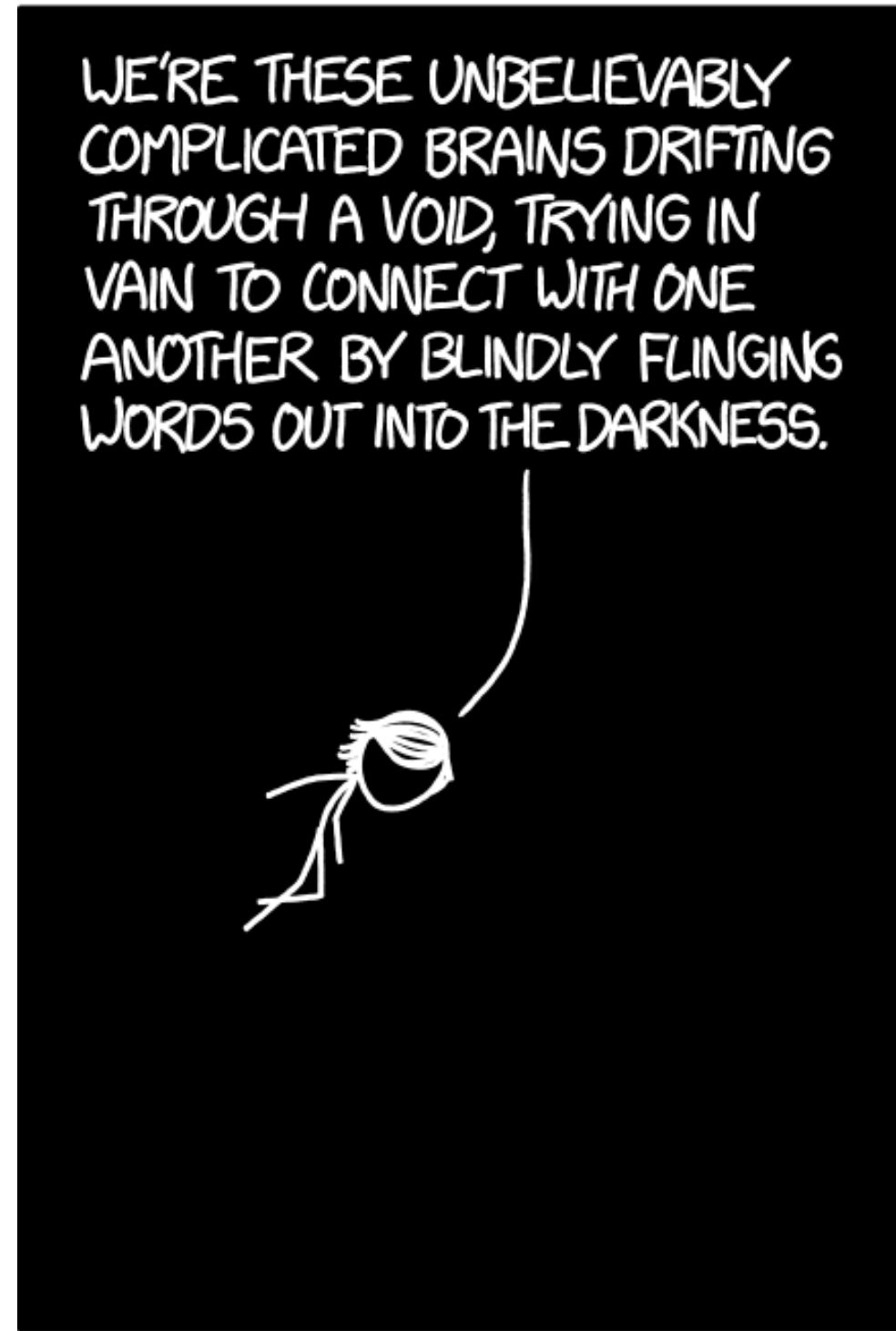
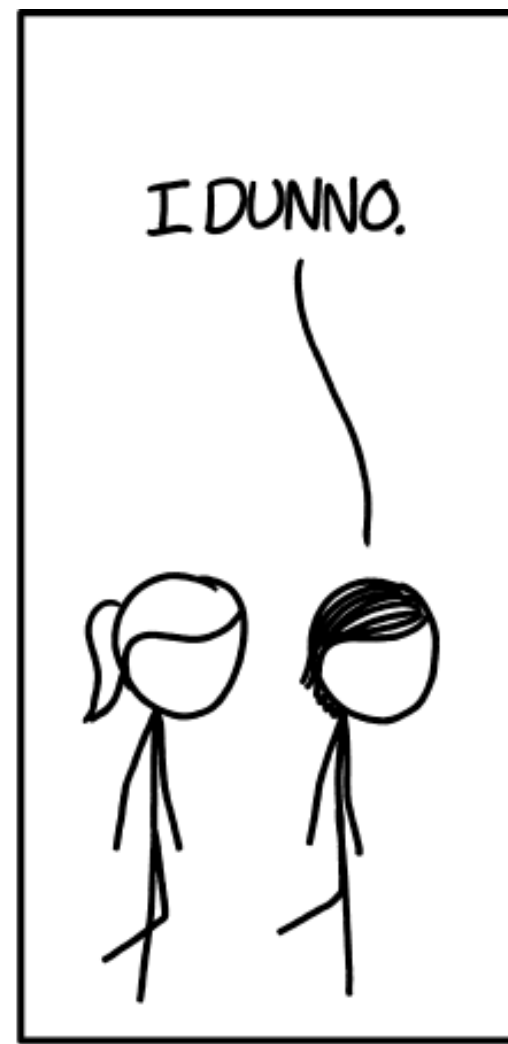
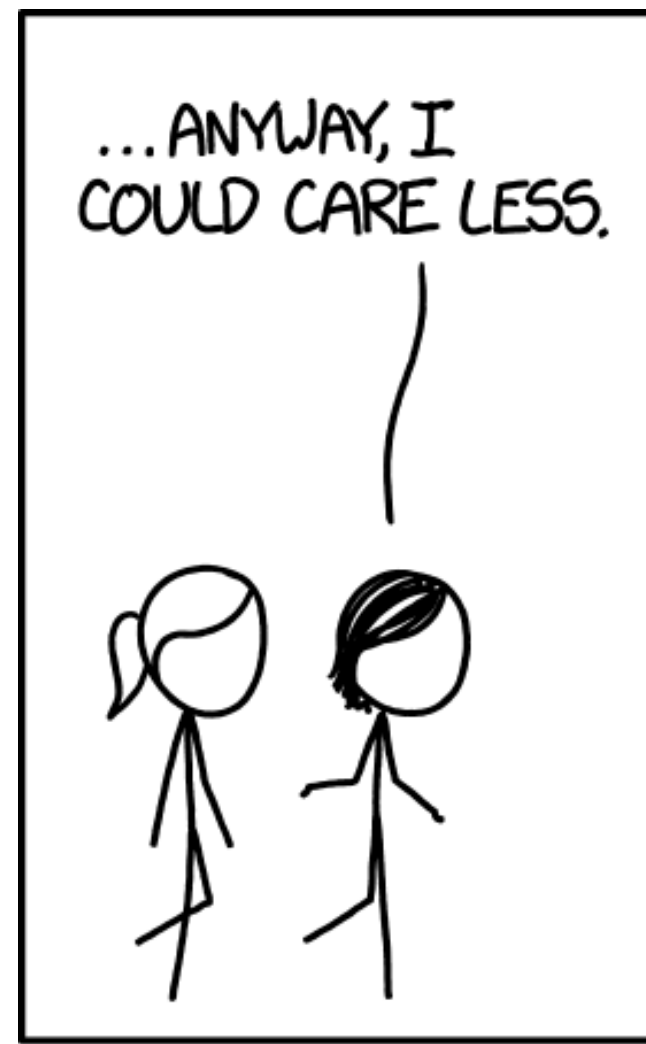


David Lewis

I distinguish two topics: first, the description of possible languages or grammars as abstract semantic systems whereby symbols are associated with aspects of the world; and, second, the description of the psychological and sociological facts whereby a particular one of these abstract semantic systems is the one used by a person or population. Only confusion comes of mixing these two topics. (Lewis 1970: 19)

What is the meaning of a symbol for a person/group?

In what scenarios regarding the person/
group does that symbol have that
particular meaning?



Syntax. Semantics.
Discourse.

Syntax

A grammar does not define the meaning of
a string, only their form.

Formal Grammar

Basic foundations of language

At its most simple, a grammar defines
what combination of characters is
accepted by a language.

Formal Grammar

Basic foundations of language

Identifying the root component of a word.

Lemmatization/Stemming

Basic foundations of language

Removal of prefixes, suffixes, modifiers
for plural, tense, etc.

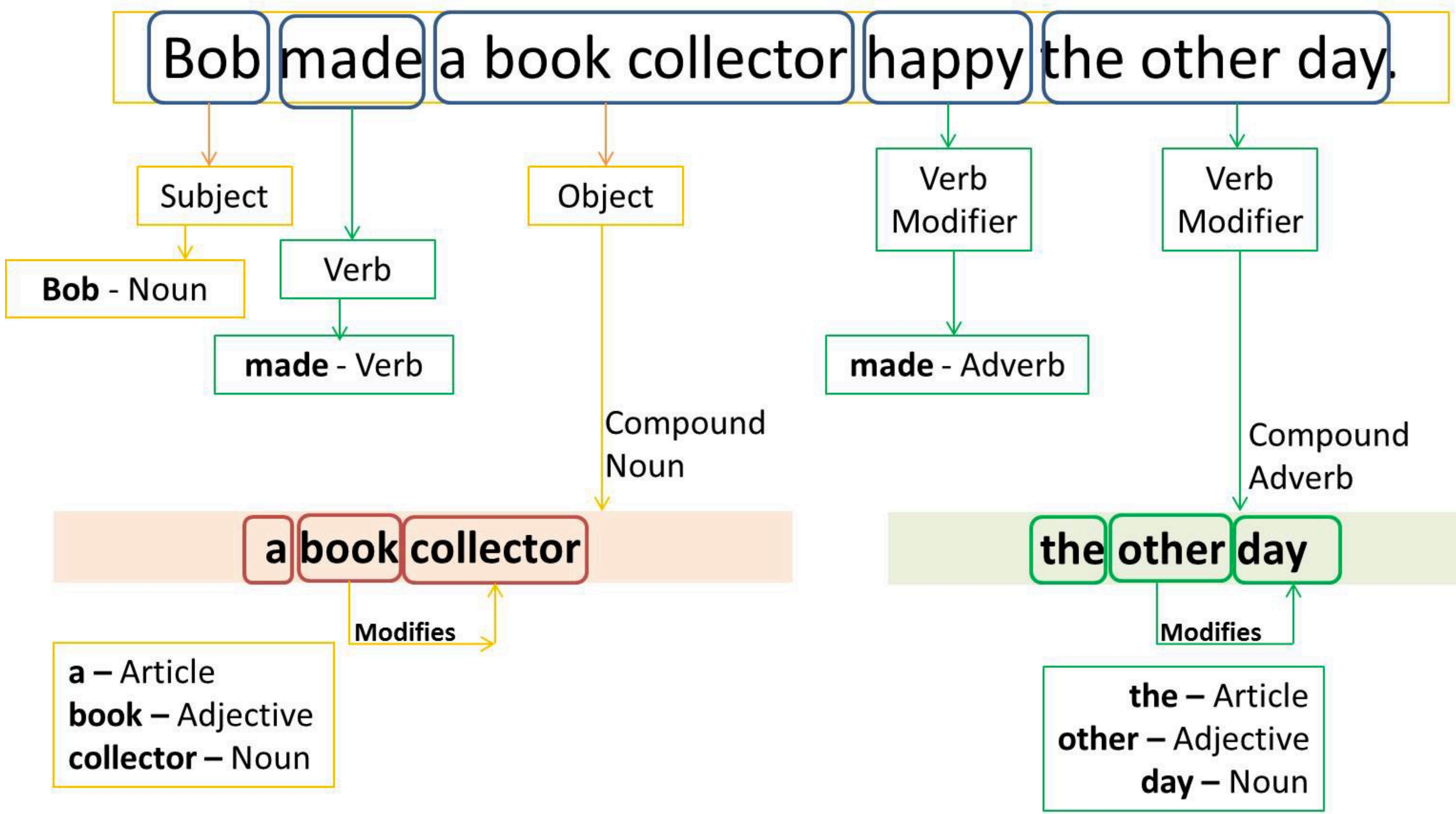
Lemmatization/Stemming

Basic foundations of language

Identification of a symbol as a Verb,
Noun, Preposition, Adjective, etc.

Part of Speech Tagging

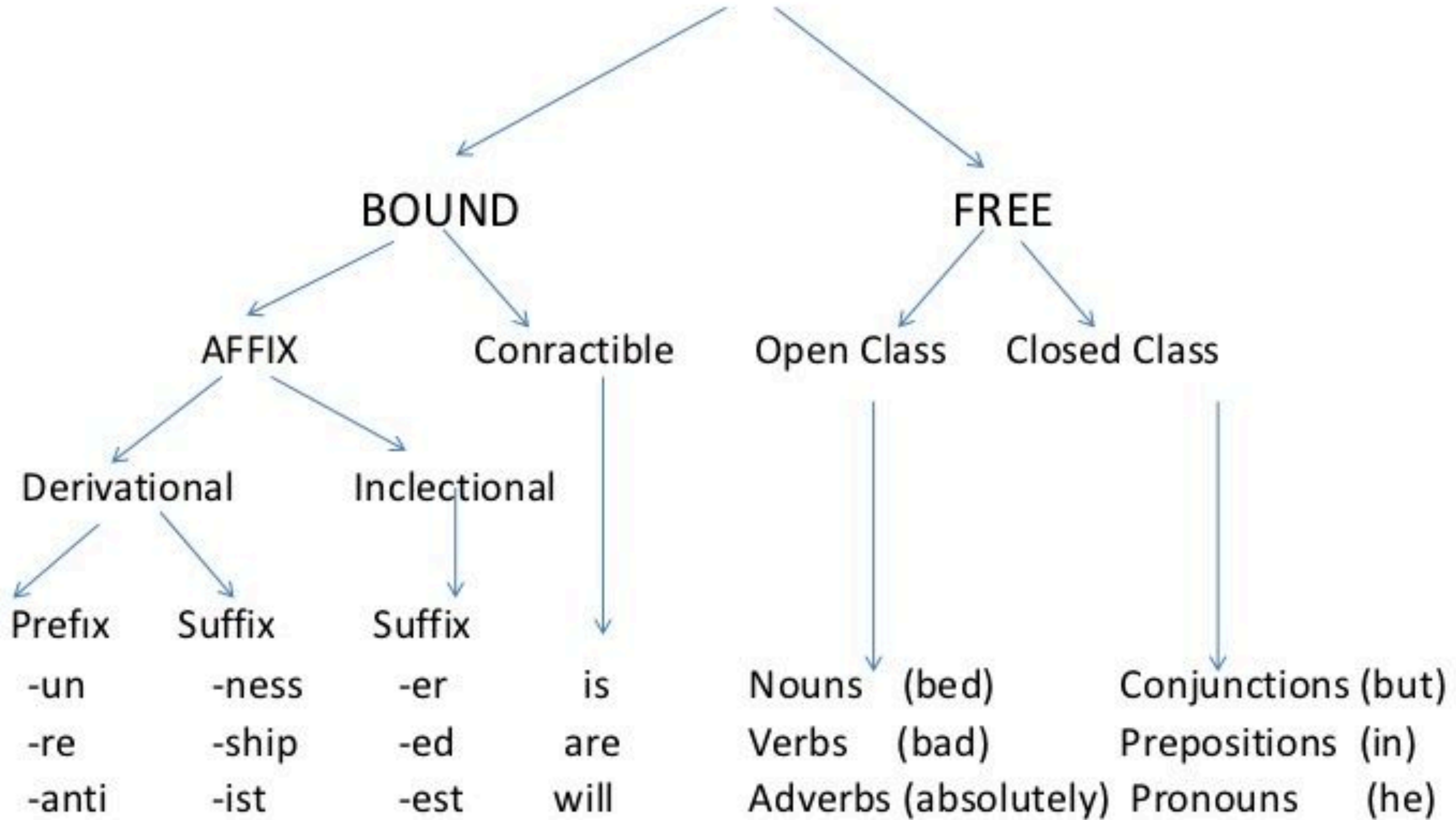
Basic foundations of language



Tagging of morphemes in symbols and sentences. A morpheme is the smallest unit of meaning in language and is not necessarily a word.

Morphology
Basic foundations of language

MORPHEMES



Semantics

Identification of higher-ordered meaning
of symbols in sentences.

Named Entity Recognition

Basic foundations of language

<Richard Paul Astley>₁ (born <6 February 1966>₁₇ <6>₂₇ February <1966>₂₃) is an <English>₅ <singer>₁, <songwriter>₃ and <radio personality>₄. He rose to <fame>₉ through his <association>₈ with the <production trio>₇ <Stock Aitken Waterman>₁₁; his <1987>₁₈ <1987>₂₈ <recording>₂ of their <song>₆ " <Never Gonna Give You Up>₁₃ " was a number <1>₂₁ <hit single>₂ in <25>₂₆ <countries>₁₅, winning the <1988>₁₉ <1988>₂₄ <Brit Award>₁₆ for <Best British Single>₁₄. By the time of his <retirement>₁₀ in <1993>₂₀ <1993>₂₂, <Astley>₁ had sold approximately <40 million>₂₅ <records>₁₂ worldwide.

1. Richard Paul Astley

[Wikipedia Article](#)

Salience: 0.72

PERSON

2. hit single

Salience: 0.06

OTHER

3. songwriter

Salience: 0.05

PERSON

4. radio personality

Salience: 0.04

OTHER

5. English

[Wikipedia Article](#)

Salience: 0.02

LOCATION

6. song

Salience: 0.02

WORK OF ART

Identification of higher-ordered meaning
of symbols in sentences.

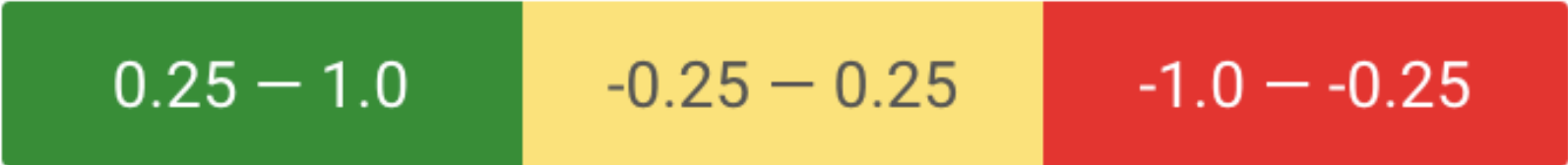
Sentiment Analysis

Basic foundations of language

Document and Sentence Level Sentiment

	Score	Magnitude
Entire Document	0.2	0.8
Richard Paul Astley (born 6 February 1966) is an English singer, songwriter and radio personality.	0.1	0.1
He rose to fame through his association with the production trio Stock Aitken Waterman; his 1987 recording of their song "Never Gonna Give You Up" was a number 1 hit single in 25 countries, winning the 1988 Brit Award for Best British Single.	0.4	0.4
By the time of his retirement in 1993, Astley had sold approximately 40 million records worldwide.	0.2	0.2

Score Range



Entity Level Sentiment

1. Richard Paul Astley Sentiment: Score 0 Magnitude 0.6	PERSON	2. hit single Sentiment: Score 0 Magnitude 0.1	OTHER
3. songwriter	PERSON	4. radio personality	OTHER

Understanding the multiple words a symbol
could be associated with following a
specific context.

Word Similarity Analysis

Basic foundations of language

Kobe Bryant wife Vanessa spoke 🥺🥲

an hour ago

Context

- name
- education
- sports
- profession
- politics

Similar Words

- james
- joshua
- cstand
- aurora
- irithel
- nadal
- ivan
- jinkee
- abueva
- slena

Entities

- name

Discourse?