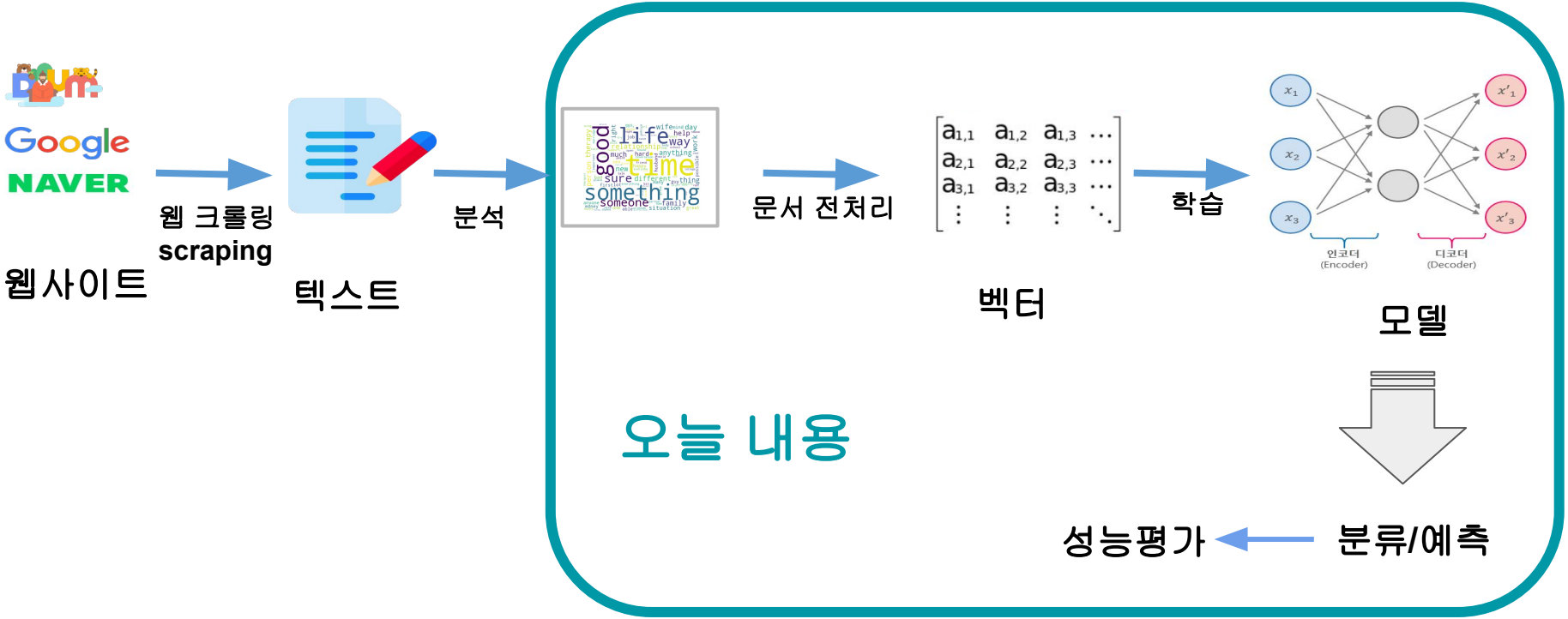
# Vector Semantics

SLP 6. Vector Semantics

https://web.stanford.edu/~jurafsky/slp3/6.pdf

# 한눈에 보는 자연어 처리 과정



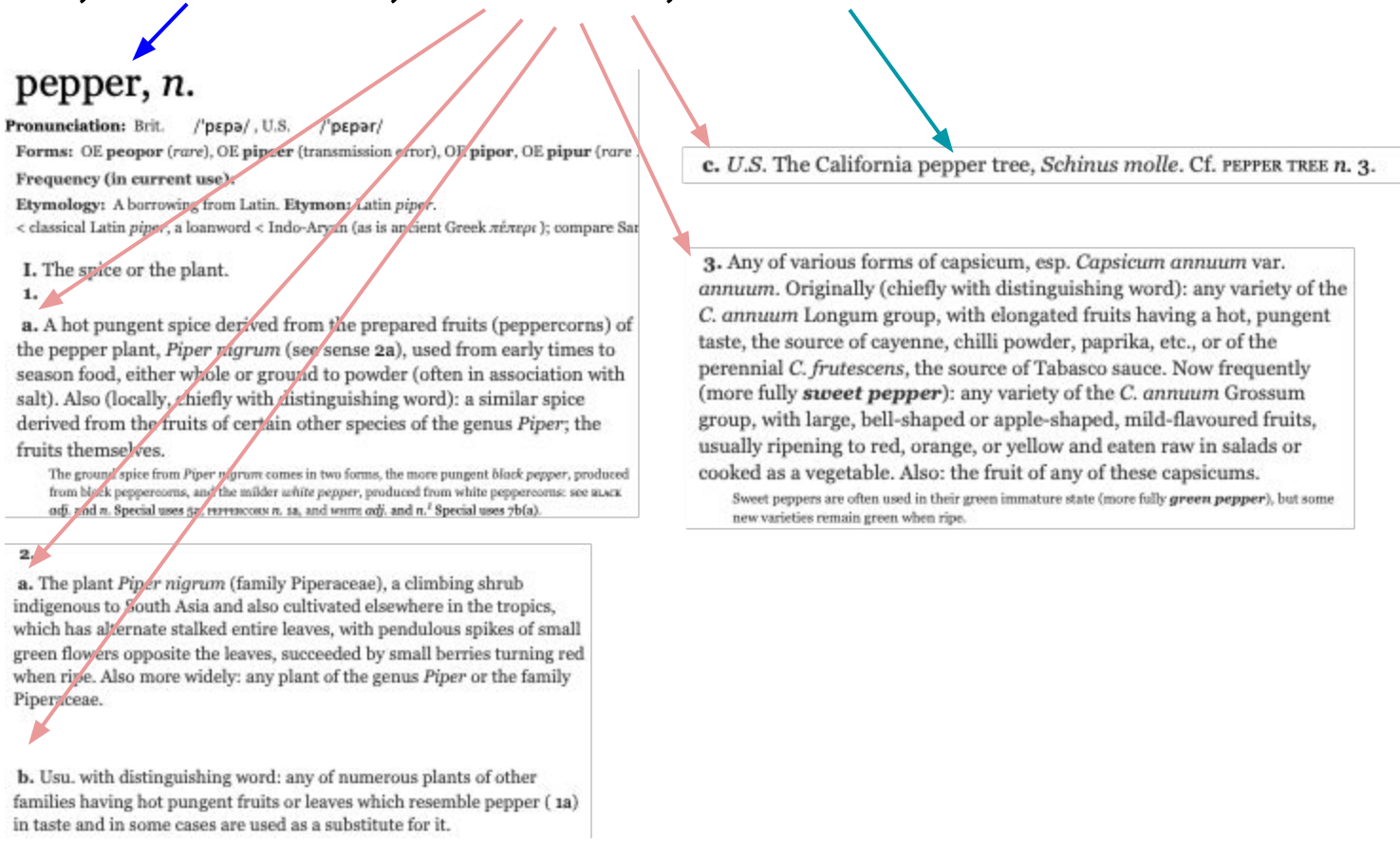
# Outline

- Word Vectors
- Cosine similarity
- Tf-idf
- Classification with word vectors

# What words mean?

look in a dictionary: http://www.oed.com/

# Words, Lemmas, Senses, Definitions



**pepper, *n.***

**Pronunciation:** Brit. /ˈpɛpə/ , U.S. /ˈpɛpər/

**Forms:** OE **peopor** (*rare*), OE **pipcer** (transmission error), OE **pipor**, OE **pipur** (*rare*

**Frequency (in current use):**

**Etymology:** A borrowing from Latin. **Etymon:** Latin *piper*.

< classical Latin *piper*, a loanword < Indo-Aryan (as is ancient Greek *πέπερι*); compare San

**I.** The spice or the plant.

**1.**

**a.** A hot pungent spice derived from the prepared fruits (peppercorns) of the pepper plant, *Piper nigrum* (see sense **2a**), used from early times to season food, either whole or ground to powder (often in association with salt). Also (locally, chiefly with distinguishing word): a similar spice derived from the fruits of certain other species of the genus *Piper*; the fruits themselves.

The ground spice from *Piper nigrum* comes in two forms, the more pungent *black pepper*, produced from black peppercorns, and the milder *white pepper*, produced from white peppercorns: see BLACK *adj.* and *n.* Special uses 5a; PEPPERCORN *n.* 1a, and WHITE *adj.* and *n.*[1] Special uses 7b(a).

**2.**

**a.** The plant *Piper nigrum* (family Piperaceae), a climbing shrub indigenous to South Asia and also cultivated elsewhere in the tropics, which has alternate stalked entire leaves, with pendulous spikes of small green flowers opposite the leaves, succeeded by small berries turning red when ripe. Also more widely: any plant of the genus *Piper* or the family Piperaceae.

**b.** Usu. with distinguishing word: any of numerous plants of other families having hot pungent fruits or leaves which resemble pepper (**1a**) in taste and in some cases are used as a substitute for it.

**c.** *U.S.* The California pepper tree, *Schinus molle*. Cf. PEPPER TREE *n.* 3.

**3.** Any of various forms of capsicum, esp. *Capsicum annuum* var. *annuum*. Originally (chiefly with distinguishing word): any variety of the *C. annuum* Longum group, with elongated fruits having a hot, pungent taste, the source of cayenne, chilli powder, paprika, etc., or of the perennial *C. frutescens*, the source of Tabasco sauce. Now frequently (more fully ***sweet pepper***): any variety of the *C. annuum* Grossum group, with large, bell-shaped or apple-shaped, mild-flavoured fruits, usually ripening to red, orange, or yellow and eaten raw in salads or cooked as a vegetable. Also: the fruit of any of these capsicums.

Sweet peppers are often used in their green immature state (more fully ***green pepper***), but some new varieties remain green when ripe.

# A sense or “concept” is the meaning component of a word

## Lemma pepper

- Sense 1: spice from pepper plant
- Sense 2: the pepper plant itself
- Sense 3: another similar plant (Jamaican pepper)
- Sense 4: another plant with peppercorns (California pepper)
- Sense 5: capsicum (i.e. chili, paprika, bell pepper, etc)

# There are relations between senses

- Synonym
- Antonym
- Similarity
- Relatedness
- Superordinate/ subordinate
- Connotation

# Relation: Synonymity

- Synonyms have the same meaning in some or all contexts
  - filbert / hazelnut
  - couch / sofa
  - big / large
  - automobile / car
  - vomit / throw up
  - Water / H2O
- Note that there are probably no examples of perfect synonymy
- The Linguistic Principle of Contrast:
  - Difference in form -> difference in meaning

# Relation: Antonymy

- Senses that are opposites with respect to one feature of meaning
- Otherwise, they are very similar!
  - dark/light short/long fast/slow rise/fall
  - hot/cold up/down in/out
- Antonyms can define a binary opposition or be at opposite ends of a scale
  - long/short, fast/slow
- Be reversives:
  - rise/fall, up/down

# Relation: Similarity

- Words with similar meanings.  Not synonyms, but sharing some element of meaning
  - car, bicycle
  - cow, horse
- Ask humans how similar 2 words are

| word1 | word2 | similarity |
|---|---|---|
| vanish | disappear | 9.8 |
| behave | obey | 7.3 |
| belief | impression | 5.95 |
| muscle | bone | 3.65 |
| modest | flexible | 0.98 |
| hole | agreement | 0.3 |

SimLex-999 dataset (Hill et al., 2015)

# Relation: Word relatedness (or association)

- Words be related in any way, perhaps via a semantic frame or field
  - car, bicycle:  similar
  - car, gasoline:  related, not similar
- Semantic field is the words that cover a particular semantic domain
  - Hospitals:  surgeon, scalpel, nurse, anaesthetic, hospital
  - Restaurants: waiter, menu, plate, food, menu, chef),
  - Houses: door, roof, kitchen, family, bed

# Relation: Superordinate/ subordinate

- One sense is a **subordinate** of another if the first sense is more specific, denoting a subclass of the other
  - car is a subordinate of vehicle
  - mango is a subordinate of fruit
- Conversely **superordinate**
  - vehicle is a superordinate of car
  - fruit is a superordinate of mango

| **Superordinate** | vehicle | fruit | furniture |
|---|---|---|---|
| **Subordinate** | car | mango | chair |

# Relation: Semantic Frames and Roles

- A semantic frame is a set of words that denote perspectives or participants in a particular type of event
  - Sam bought the book from Ling.
  - Ling sold the book to Sam
  - Sam has the role of the buyer in the frame and Ling the seller
- Important for question answering, and can help in shifting perspective for machine translation

# Relation: Connotation

- An idea or feeling that a word invokes in addition to its literal or primary meaning
  - E.g., “discipline” has unhappy connotations of punishment and repression
- Cultural or emotional association that some word or phrase carries, in addition to its literal meaning, which is its *denotation*
- Words have affective meanings
  - positive connotations (happy)
  - negative connotations (sad)
- positive evaluation (great, love)
- negative evaluation (terrible, hate)

# Words and Vectors

# Classical (“Aristotelian”) Theory of Concepts

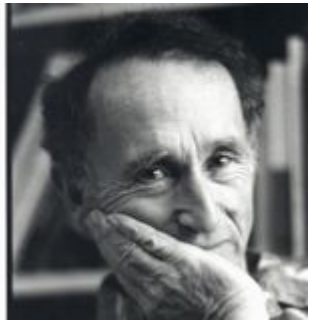
- meaning of a word: a concept defined by necessary and sufficient conditions
- A **necessary condition** for being an X is a condition C that X must satisfy in order for it to be an X. If not C, then not X
  - ”Having four sides” is necessary to be a square.
- A **sufficient condition** for being an X is condition such that if something satisfies condition C, then it must be an X. If and only if C, then X
- The following necessary conditions, jointly, are sufficient to be a square
  - x has (exactly) four sides
  - each of x's sides is straight
  - x is a closed figure
  - x lies in a plane
  - each of x's sides is equal in length to each of the others
  - each of x's interior angles is equal to the others (right angles)
  - the sides of x are joined at their ends

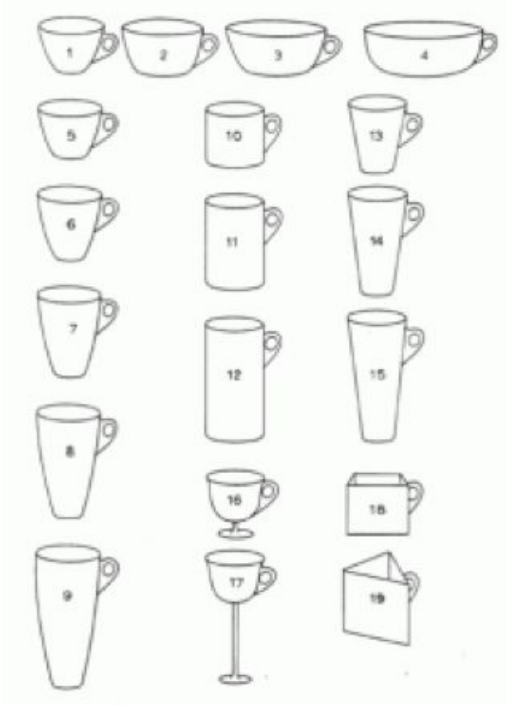
Example from Norman Swartz, SFU

# Features are complex and may be context-dependent

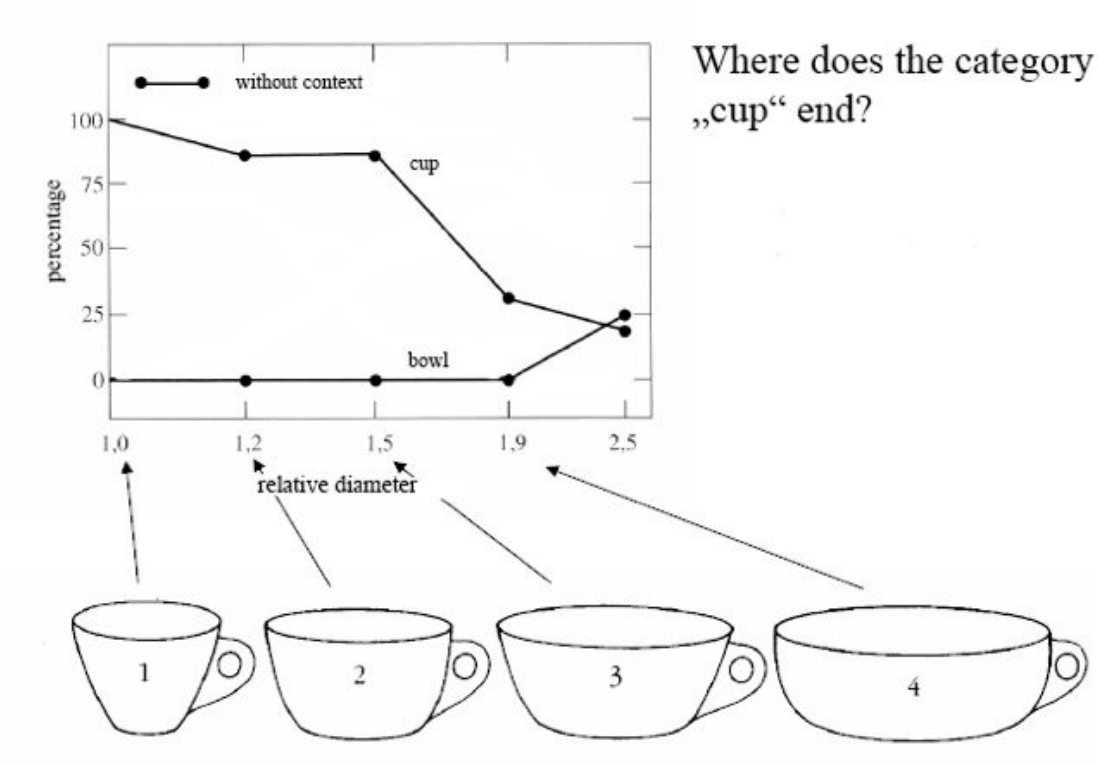
William Labov. 1975

What are these? Cup or bowl?



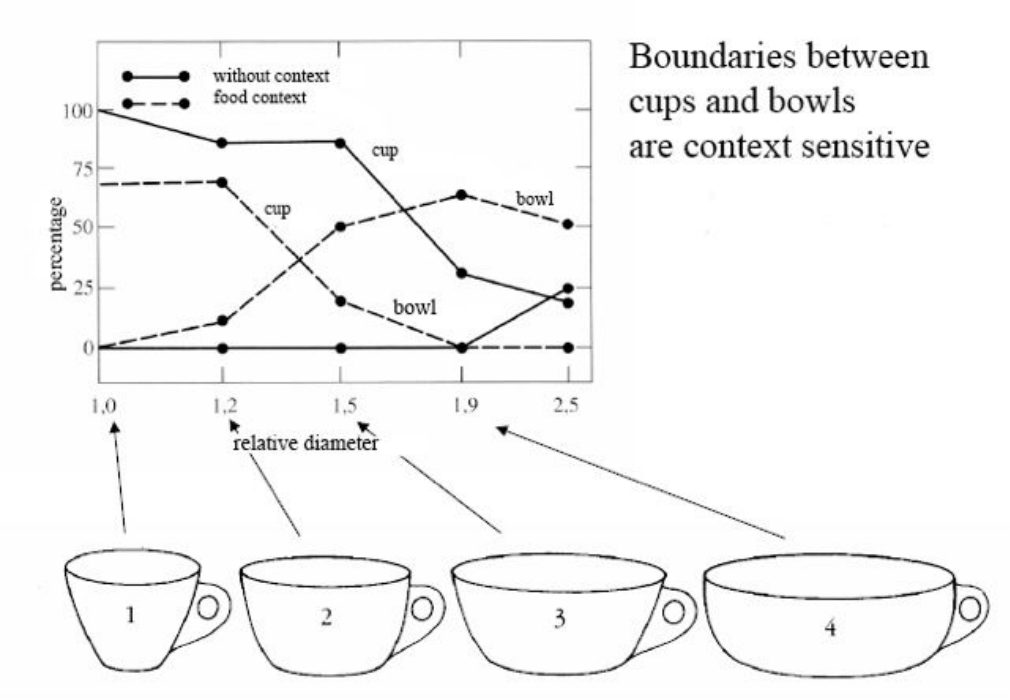


The category depends on complex features of the object (diameter, etc)




Where does the category „cup“ end?

The category depends on the context! (If there is food in it, it's a bowl)




Boundaries between cups and bowls are context sensitive

# Labov's definition of cup

The term *cup* is used to denote round containers with a ratio of depth to width of $1 \pm r$ where $r \leq r_b$, and $r_b = \alpha_1 + \alpha_2 + \ldots \alpha_v$ and $\alpha_1$ is a positive quality when the feature i is present and 0 otherwise.

feature 1 = with one handle
2 = made of opaque vitreous material
3 = used for consumption of food
4 = used for the consumption of liquid food
5 = used for consumption of hot liquid food
6 = with a saucer
7 = tapering
8 = circular in cross-section

*Cup* is used variably to denote such containers with ratios width to depth $1 \pm r$ where $r_b \leq r \leq r_1$ with a probability of $r_1 - r/r_t - r_b$. The quantity $1 \pm r_b$ expresses the distance from the modal value of width to height.

# Ludwig Wittgenstein (1889-1951)

- skeptical of building a formal theory of meaning definitions for each word
- “**The meaning of a word is its use in the language**” (Wittgenstein, 1953, PI 43)
- Define words by some representation of how the word was used by actual people in speaking and understanding

# Let's define words by their usages

- In particular, words are defined by their environments (the words around them)
- Zellig Harris (1954): **If A and B have almost identical environments we say that they are synonyms**

# What does ongchoi mean?

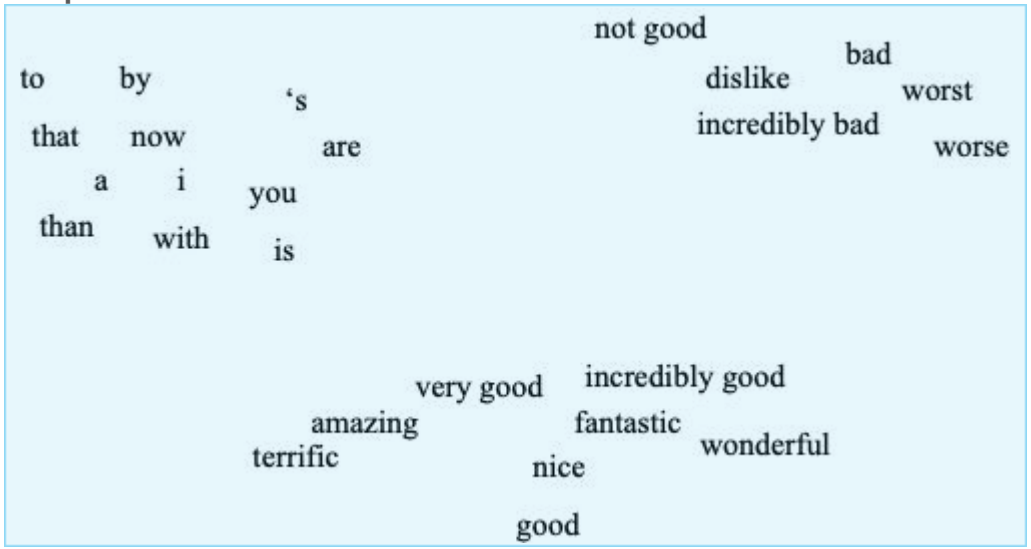
- Suppose you see these sentences:
  - Ong choi is **delicious sautéed** with **garlic**.
  - Ong choi is **superb** over rice
  - Ong choi **leaves** with **salty** sauces
- And you've also seen these:
  - …spinach **sautéed** with **garlic** over rice
  - Chard stems and **leaves** are **delicious**
  - Collard greens and other **salty leafy** greens
- Conclusion:
  - Ongchoi is a leafy green like spinach, chard, or collard greens

# Ong choi: Ipomoea aquatica  "Water Spinach"

Yamaguchi, Wikimedia Commons, public domain

# Build a new model of meaning focusing on similarity

- Word as a **vector (or embedding)**, a list of numbers where the numbers are based in **counts of neighboring words**
- Similar words are "nearby in space"




Embeddings learned for sentiment analysis

# Define a word as a vector

- Called an "**embedding**" because it's embedded into a space
- Standard way to represent meaning
- Fine-grained model of meaning for similarity
  - With words, requires same word to be in training and test
  - With embeddings: ok if similar words occurred!!!
- Practical because they can be learned automatically from text without labeling

# 2 kinds of embeddings

- Tf-idf
  - A common baseline model
  - Sparse vectors
  - Words are represented by a simple function of the counts of nearby words
- Word2vec
  - Dense vectors
  - Representation is created by training a classifier to distinguish nearby and far-away words

# Term-document matrix (Salton, 1971)

Initially defined to find similar documents for information retrieval

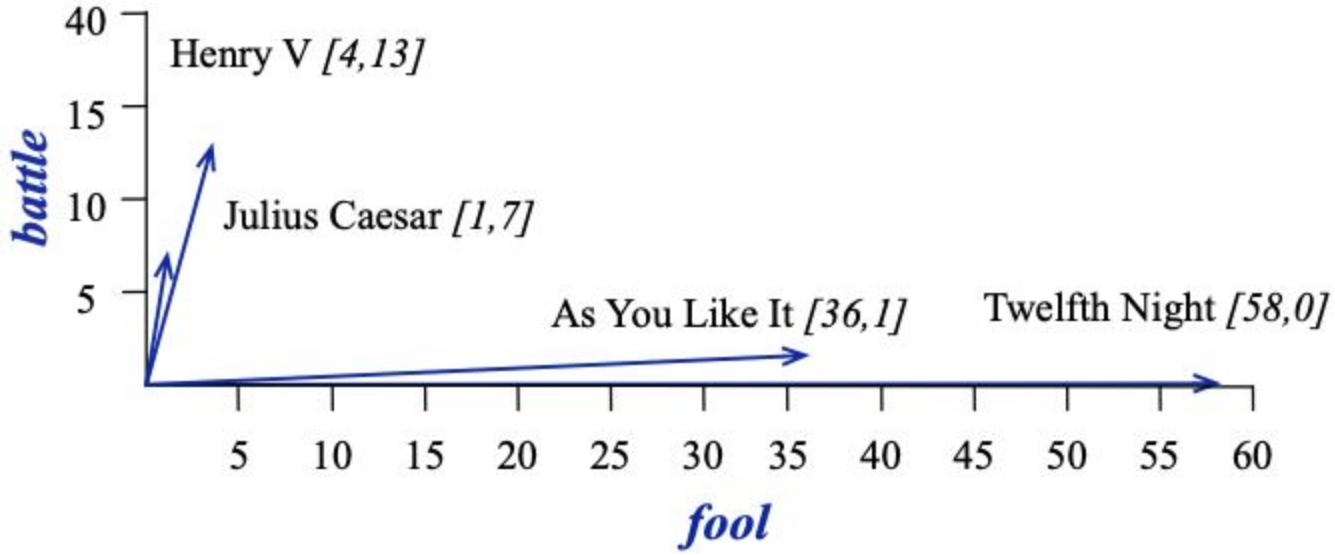
Each document is represented by a vector of words

- Each row represents a word
- Each column represents a document

| | As You Like It | Twelfth Night | Julius Caesar | Henry V |
|---|---|---|---|---|
| battle | 1 | 0 | 7 | 13 |
| good | 114 | 80 | 62 | 89 |
| fool | 36 | 58 | 1 | 4 |
| wit | 20 | 15 | 2 | 3 |

As you Like it: [1, 114, 36, 20]
Julius Caesar: [7,62,1,2]

| | As You Like It | Twelfth Night | Julius Caesar | Henry V |
|---|---|---|---|---|
| battle | 1 | 0 | 7 | 13 |
| good | 114 | 80 | 62 | 89 |
| fool | 36 | 58 | 1 | 4 |
| wit | 20 | 15 | 2 | 3 |




Vectors are similar for the two comedies different than the history

Comedies have more fools and wit and fewer battles.

# Words can be vectors too

| | As You Like It | Twelfth Night | Julius Caesar | Henry V |
|---|---|---|---|---|
| **battle** | 1 | 0 | 7 | 13 |
| **good** | 114 | 80 | 62 | 89 |
| **fool** | 36 | 58 | 1 | 4 |
| **wit** | 20 | 15 | 2 | 3 |

- *battle* is the kind of word that occurs in Julius Caesar and Henry V
- *fool* is the kind of word that occurs in comedies, especially Twelfth Night

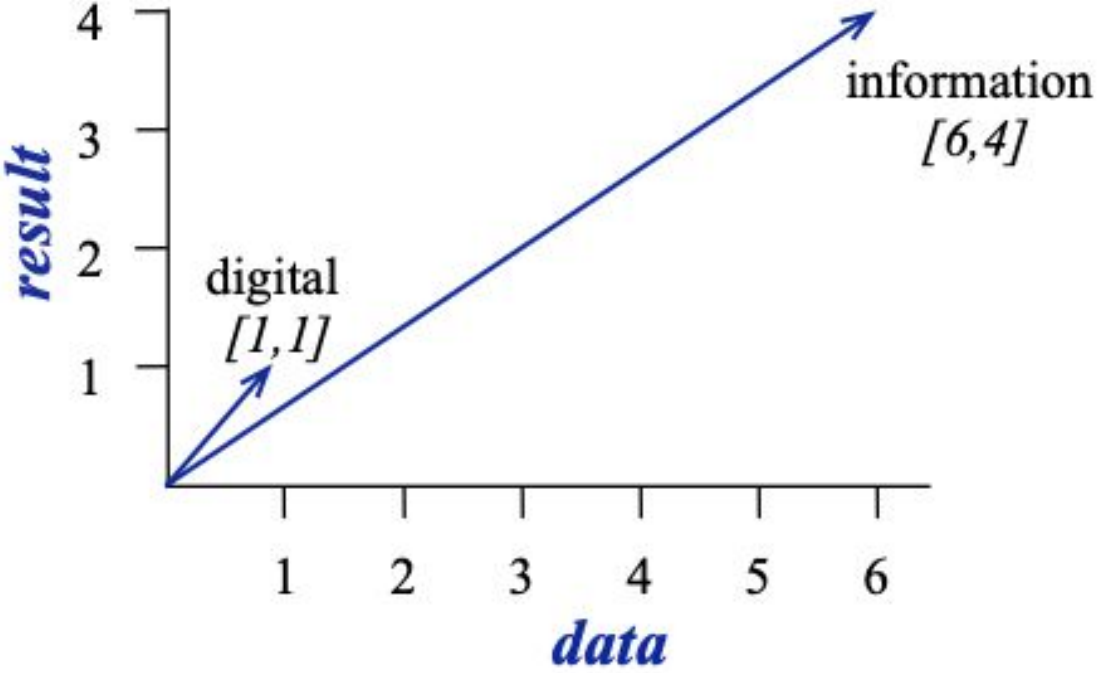
# More common: word-word matrix (or "term-context matrix")

- Columns are labeled by words
- Two words are **similar in meaning if their context vectors are similar**
- +-4 word window around the row word are usually used

*e.g., 7-word window example from the Brown corpus*

sugar, a sliced lemon, a tablespoonful of **apricot** jam, a pinch each of,
their enjoyment. Cautiously she sampled her first **pineapple** and another fruit whose taste she likened
well suited to programming on the digital **computer**. In finding the optimal R-stage policy from
for the purpose of gathering data and **information** necessary for the study authorized in the

| | aardvark | computer | data | pinch | result | sugar | ... |
|---|---|---|---|---|---|---|---|
| apricot | 0 | 0 | 0 | 1 | 0 | 1 | |
| pineapple | 0 | 0 | 0 | 1 | 0 | 1 | |
| digital | 0 | 2 | 1 | 0 | 1 | 0 | |
| information | 0 | 1 | 6 | 0 | 4 | 0 | |


4
3
2
1
result
information
[6,4]
digital
[1,1]
1
2
3
4
5
6
data

# Cosine

- Standard way to use embeddings to compute functions like **semantic similarity** between two words, two sentences, or two documents
- Cosine of the angle between the vectors as a measure of vector similarity
- Important tool in practical applications like question answering, summarization, or automatic essay grading

# Dot product

$$\text{dot-product}(\vec{v},\vec{w}) = \vec{v}\cdot\vec{w} = \sum_{i=1}^{N} v_i w_i = v_1 w_1 + v_2 w_2 + \ldots + v_N w_N$$

- **Dot products** are higher if vectors are longer, with higher values in each dimension. Frequent words have longer vectors and have higher co-occurrence values
- Measures how similar two words are **regardless of their frequency**
- **Solution: Normalized dot product**

# Cosine for computing similarity

$$\frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|} = \cos\theta$$

$$\text{cosine}(\vec{v},\vec{w}) = \frac{\vec{v}\cdot\vec{w}}{|\vec{v}||\vec{w}|} = \frac{\sum_{i=1}^{N} v_i w_i}{\sqrt{\sum_{i=1}^{N} v_i^2}\sqrt{\sum_{i=1}^{N} w_i^2}}$$

$v_i$ is the count for word $v$ in context $i$

$w_i$ is the count for word $w$ in context $i$.

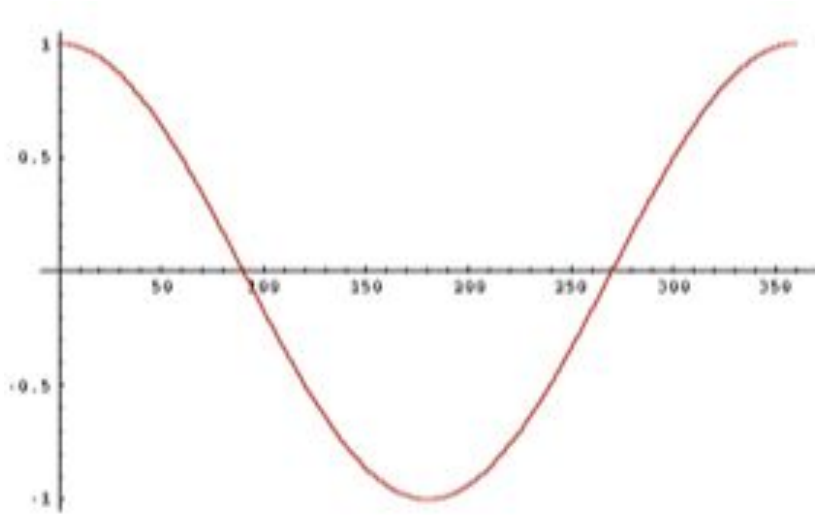
Cosine($v$,$w$) is the cosine similarity of $v$ and $w$

# Cosine as a similarity metric

-1: vectors point in opposite directions

**+1:  vectors point in same directions**

0: vectors are orthogonal




Frequency is non-negative, so  cosine range 0-1

$$\cos(\vec{v},\vec{w}) = \frac{\vec{v}\bullet\vec{w}}{|\vec{v}||\vec{w}|} = \frac{\vec{v}}{|\vec{v}|}\bullet\frac{\vec{w}}{|\vec{w}|} = \frac{\sum_{i=1}^{N} v_i w_i}{\sqrt{\sum_{i=1}^{N} v_i^2}\sqrt{\sum_{i=1}^{N} w_i^2}}$$

| | large | data | computer |
|---|---|---|---|
| apricot | 1 | 0 | 0 |
| digital | 0 | 1 | 2 |
| information | 1 | 6 | 1 |

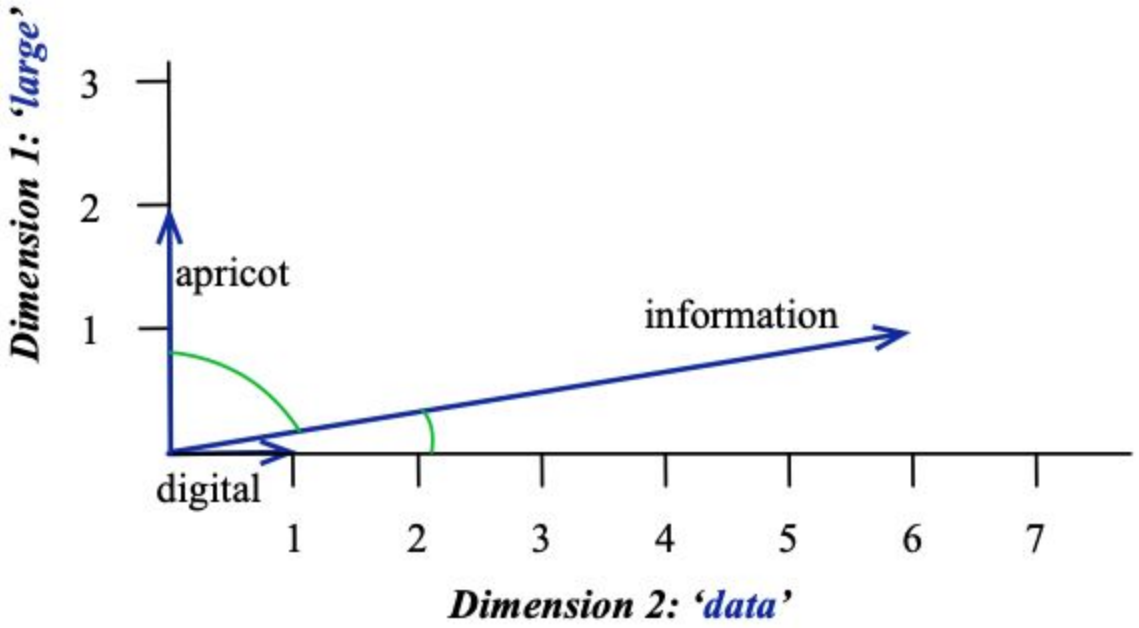
Which pair of words is more similar?

cosine(apricot,information) = $\frac{1+0+0}{\sqrt{1+0+0}\ \sqrt{1+36+1}} = \frac{1}{\sqrt{38}} = .16$

cosine(digital,information) = $\frac{0+6+2}{\sqrt{0+1+4}\ \sqrt{1+36+1}} = \frac{8}{\sqrt{38}\sqrt{5}} = .58$

cosine(apricot,digital) = $\frac{0+0+0}{\sqrt{1+0+0}\ \sqrt{0+1+4}} = 0$

# Visualizing cosines (well, angles)



# But raw frequency is a bad representation

- Frequency is useful
  - e.g., if sugar appears a lot near apricot, that's useful information
- But overly frequent words like *the*, *it*, or *they* are not very informative about the context
- Need a function that resolves this frequency paradox!

# tf-idf