CS565: Intelligent Systems and Interfaces

Vector Semantics

Semester: Jan – May 2019

Ashish Anand

Associate Professor, Dept of CSE

IIT Guwahati

Announcements

- Scribe
  - Ankit (184101005), Abhinav Reddy (150101039): 18th Mar, 2019

- 2nd Assignment is Posted

- Tentative Mid-Term Project Report Submission: 3rd April, Wednesday

- Mid-Term Project Discussion/Evaluation: 22nd March (2nd-half)/23rd March

Objective

- Discuss Vector Representation of Words
  - What we expect from a representation?
  - What is meaning of words?
  - Basis and intuition of vector representation

Vector Semantics

- Study of vector representation of words
  - model to represent meaning of words
- Question is what are the different aspects of meaning ?
  - Answer: Lexical semantics – linguistic study of word meaning

Desiderata

- Examples
  - Class : Teaching group / Economic Group / Rank
  - Right: Correct / A direction
  - Mouse: Animal / a specific computer peripheral
  - Multiple meaning: *word sense*

- *Relation 1:* Synonyms/Antonyms
  - Synonym: one word has a sense whose meaning is identical to a sense of another word, or nearly identical
  - Examples: couch/sofa vomit/throw up
  - Antonym: meanings are opposite
  - Examples: long/short big/little fast/slow rise/fall

Word Similarity: Relations beyond synonyms/antonyms

- Word Similarity
  - Words with similar meanings but not synonyms
  - *Cat / Dog*

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)

Adapted from Jurafsky and Martin's slide

Word Relatedness / Word Association: Relation beyond similarity

- Words are related by semantic frame or field
  - Cat, Dog : **similar**
  - Student, Teacher: **related** but not similar
  - **Relatedness:** co-participation in a shared event

- **Semantic Field**
  - Set of words covering a particular semantic domain, and
  - Have structured relations among them
  - Example
    - **University**
      - **Teacher, student, study, class, lecture, assignment, project**
    - **House**
      - **Room, door, furniture, bedroom**

Adapted from Jurafsky and Martin's slide

Semantic Frame and Field

- Topic Modeling: example of semantic field

- Semantic Frames and Roles
  - Set of words denoting perspectives or participants in a particular event type
  - Different agents playing distinct role in a single event
    - Teaching/Learning
    - Doctor/Patient
    - Buyer/Seller

Taxonomic Relations

One sense is a **subordinate/hyponym** 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 / hypernym**

- *vehicle* is a superordinate of *car*
- *fruit* is a subodinate of *mango*

Superordinate	vehicle	fruit	furniture
Subordinate	car	mango	chair

Adapted from Jurafsky and Martin's slide

Connotation: Affective Meaning

- Aspects of word’s meaning that are related to a writer’s or reader’s emotions, opinions, or evaluations.
  - Happy: Positive connotations vs Sad: Negative connotations
  - Great: Positive evaluation vs Terrible: negative evaluation

- Three important dimension of affective meaning
  - **Valence:** pleasantness of the stimulus (happy vs annoyed)
  - **Arousal:** intensity of emotion provoked by the stimulus (excited vs calm)
  - **Dominance:** degree of control exerted by the stimulus (controlling vs awed)

Connotation: Three dimensional vector representation

	Valence	Arousal	Dominance
courageous	8.05	5.5	7.38
music	7.67	5.57	6.5
heartbreak	2.45	5.65	3.58
cub	6.71	3.95	4.24
life	6.68	5.59	5.89

In Summary

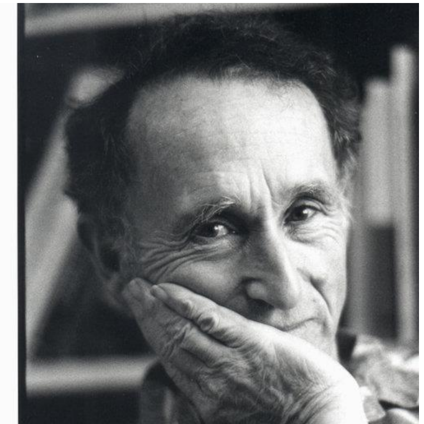
- Words
  - Have multiple senses, leading to complex relations between words
  - Synonymy / Antonymy
  - Similarity
  - Relatedness
  - Taxonomic Relations: Hypernym/Hyponym
  - Connotation

- Challenge is how we obtain an appropriate representation

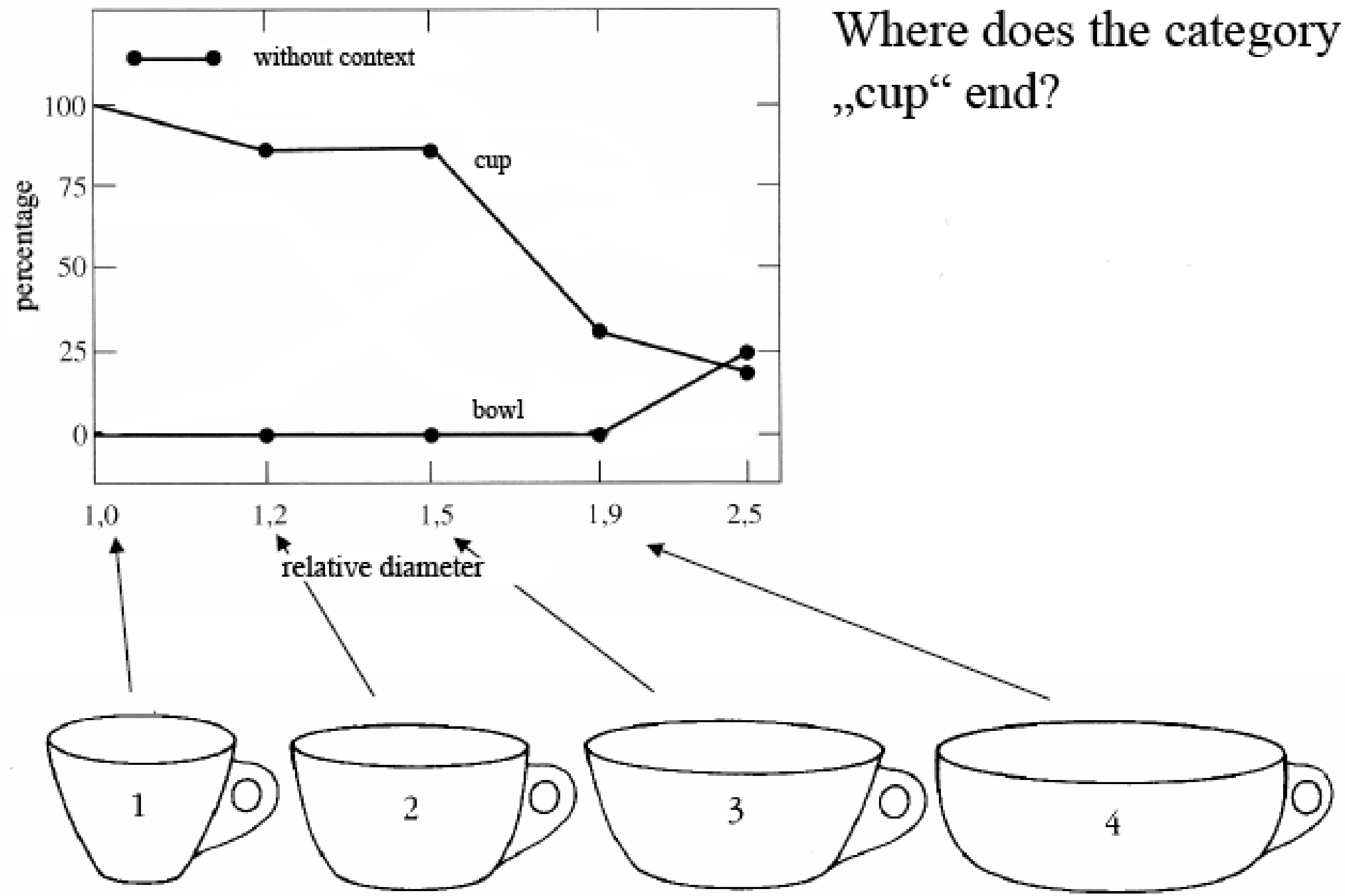
Earlier attempts to define meaning of word or concept

- William Labov. 1975

- What are these?
- Cup or bowl?

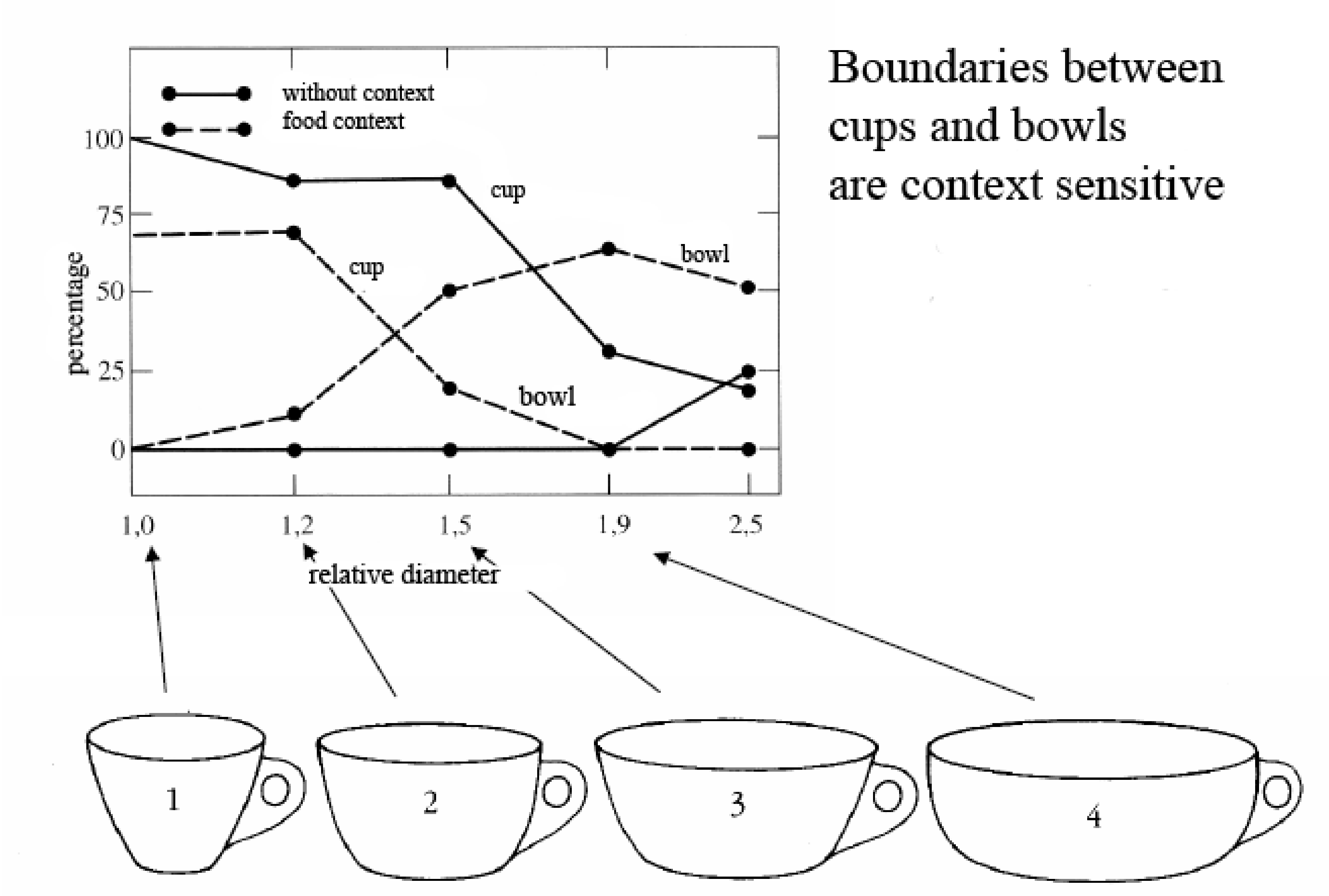


The category depends on complex features of the object (diameter, etc)



Where does the category „cup“ end?

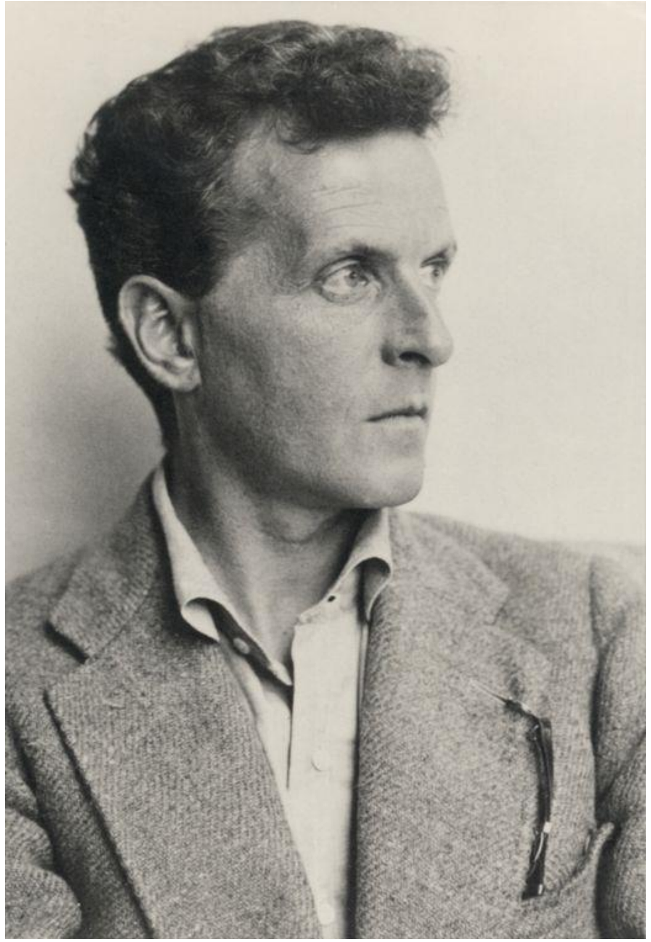
The category depends on the context! (If there is food in it, it's a bowl)



Boundaries between cups and bowls are context sensitive

Distributional hypothesis: radically different approach

- Ludwing Wittgenstein
  - Linguist / Philosopher of language
  - Meaning of a word is its use in language
- Joos, Harris and Firth
  - Define a word by the distribution it occurs in language use



Context determines meaning of words

- Harris (1954)
  - "*Oculist and eye-doctor … occur in almost the same environments*"
  - Generalize it: *"If A and B have almost identical environments … we say that they are synonyms"*

- Firth (1957)
  - "*You shall know a word by the company it keeps!*"

Context determines meaning of word

A bottle of *tesguino* is on the table.

Everybody likes *tesguino*.

Tesguino makes you drunk.

We make *tesguino* out of corn.

Broad categories of vector space models

- Long and Sparse vector representation
  - Co-occurrence matrix based methods (term-doc, term-term matrices based on MI, tf-idf etc.)
- Short and Dense vector representation
  - Dimensionality reduction techniques such as Singular value decomposition (Latent Semantic Analysis) on co-occurrence matrix
  - Neural language inspired models (skip-grams, CBOW)
- Other Methods
  - Clustering methods: Brown Clusters [Collins lecture]
  - Hybrid methods: GloVe

Co-occurrence Matrix

Building block of vector space models

Term Document Matrix: Document Vector

- Each cell: count of word *w* in a document *d*:
  - Each document is a $\boxed{\text{count vector}}$ in $\mathbb{N}^{V}$: a column below

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	1	8	15
soldier	2	2	12	36
fool	37	58	1	5
clown	6	117	0	0

Term-Document Matrix: Document Vector

- Initially defined as vector representation for documents.
- Each document is being represented in |V|- dimensional vector space.
- Notion: Similar documents tend to use similar words.
- Document vectors used in document clustering and several other Information Retrieval (IR) tasks.

Term-Document Matrix: Document vector

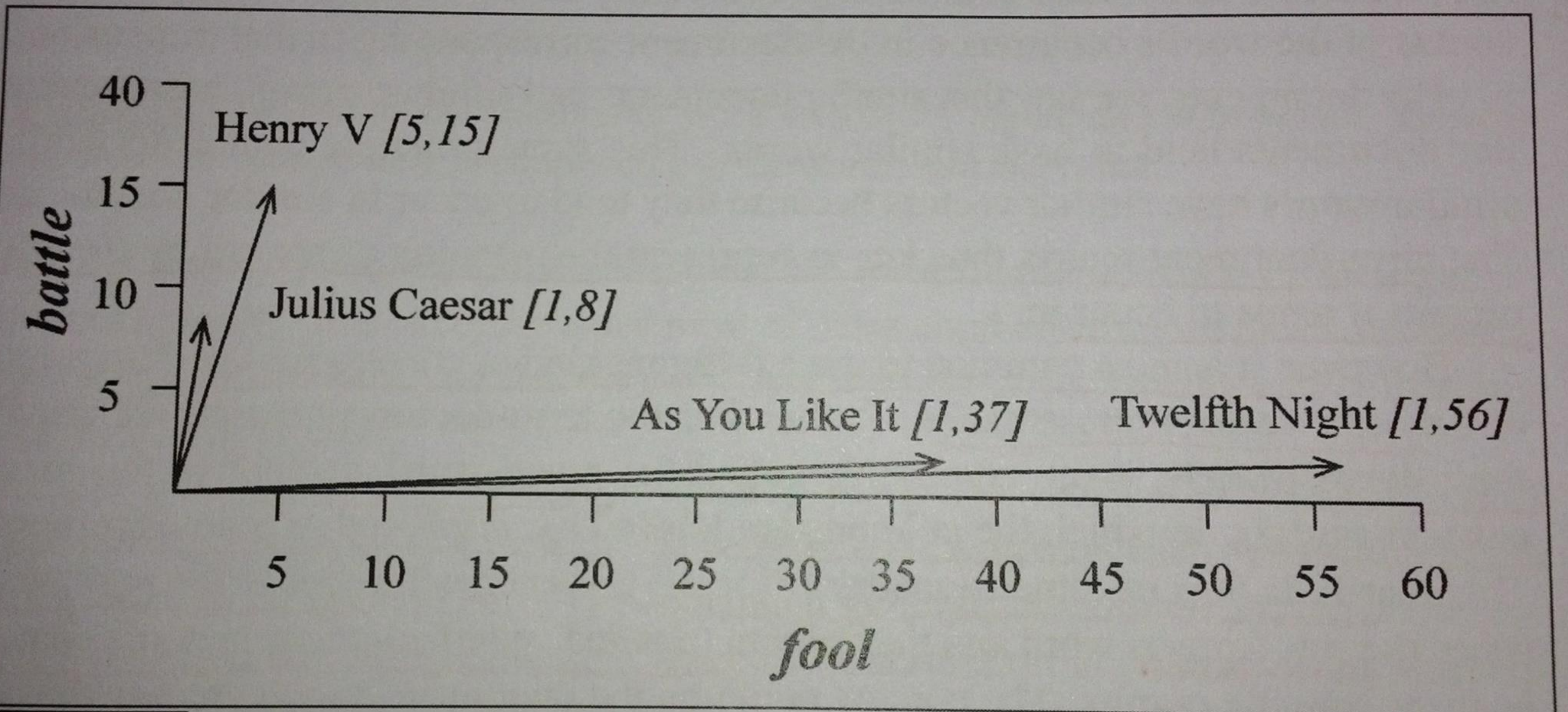


Figure 19.3 A spatial visualization of the document vectors for the four Shakespeare play documents, showing just two of the dimensions, corresponding to the words *battle* and *fool*. The comedies have high values for the *fool* dimension and low values for the *battle* dimension.

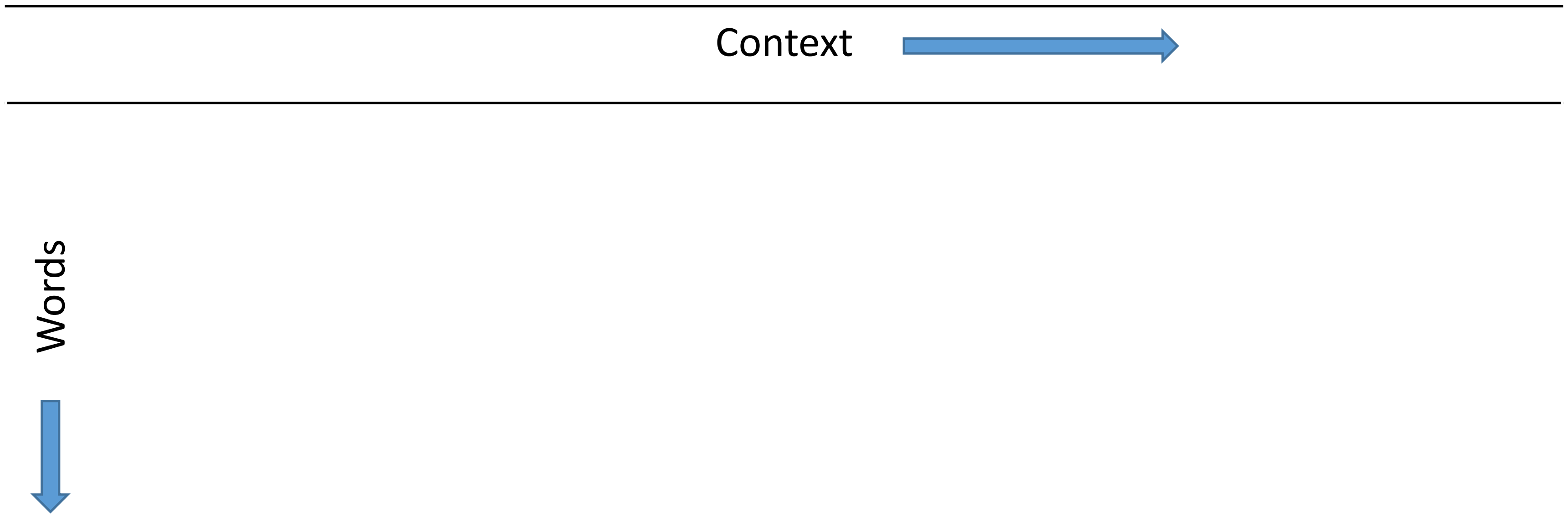
Term-Document Matrix: Word vector

- Row-vector can be used as vector representation of word
- Notion: meaning of a word can be inferred from the documents it tends to occur in.
- Two words are similar if their vectors are similar.

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	1	8	15
soldier	2	2	12	36
fool	37	58	1	5
clown	6	117	0	0

Term-Term Matrix: Word vector

- Alternate names
  - Word-word matrix
  - word-context matrix



Term-term Matrix

- Multiple ways to fill the |V| x |V| matrix
  - Each cell records number of times the *row (target)* words co-occur with the *column (context)* words.
  - ***context***: document, then "how many times the two words co-occur in the same document.
  - ***context***: window of *n* words around the word, then “number of times column words occur within *n* words either side of the row word”.

Co-occurrence takes into account two kinds of association

- Syntagmatic Association (First-order association)
  - They occur nearby each other.
  - *Drink* is first-order associate of *water*

- Paradigmatic Association (Second-order association)
  - They occur with similar words.
  - *Drink* is second-order associate of words like *sip*, *swallow*

Word-context matrix: An Example

sugar, a sliced lemon, a tablespoonful of	**apricot**	preserve or 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	

Word-context matrix: what determines size of context ?

- Objective at hands
  - Shorter window (1-3) , more **<u>syntactic</u>** representation
  - Longer window (4-10), more **<u>semantic</u>** representation

Word-context matrix: Issue with raw count

- Raw word count or frequency is not a good measure. [Why?]
  - May not be very informative
  - Example of very frequent and common words such as "*the*" and *"of"* not having discriminative power.

- Can you think of measure which can say whether a ***context word*** is informative about the ***target word*** ?
  - Answer lies in realizing similarity with a particular topic we discussed earlier in the course.

Word-context matrix: Alternative Measures

- Positive Pointwise Mutual Information [PPMI] [DIY]
  - Definition
  - Why positive adjective?
  - What happens with rare context words?
  - Do we need smoothing methods here?

- Tf-idf (Term Frequency – Inverse Document Frequency)

tf-idf: combine two factors

- **tf: term frequency**. frequency count (usually log-transformed):

$$\mathrm{tf}_{t,d} = \begin{cases} 1 + \log_{10} \mathrm{count}(t,d) & \text{if } \mathrm{count}(t,d) > 0 \\ 0 & \text{otherwise} \end{cases}$$

- **Idf: inverse document frequency: tf-**

Total # of docs in collection

$$\mathrm{idf}_i = \log\left(\frac{N}{\mathrm{df}_i}\right)$$

Words like "the" or "good" have very low idf

# of docs that have word i

tf-idf value for word t in document d:

$$w_{t,d} = \mathrm{tf}_{t,d} \times \mathrm{idf}_t$$

Summary: tf-idf

- Compare two words using tf-idf cosine to see if they are similar
- Compare two documents
  - Take the centroid of vectors of all the words in the document
  - Centroid document vector is:

$$d = \frac{w_1 + w_2 + \ldots + w_k}{k}$$

References

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

- Slides adapted from https://web.stanford.edu/~jurafsky/slp3/slides/vector1.pptx