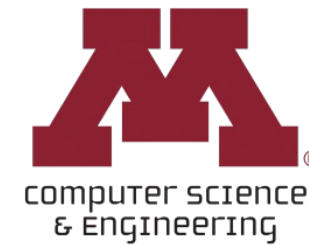


CSCI 5541: Natural Language Processing

Lecture 4: Lexical Semantics

Dongyeop Kang (DK), University of Minnesota

dongyeop@umn.edu | twitter.com/dongyeopkang | dykang.github.io



UNIVERSITY OF MINNESOTA

Driven to Discover®

Recap: Text Classification



$$f(x) = y$$



Text Classification

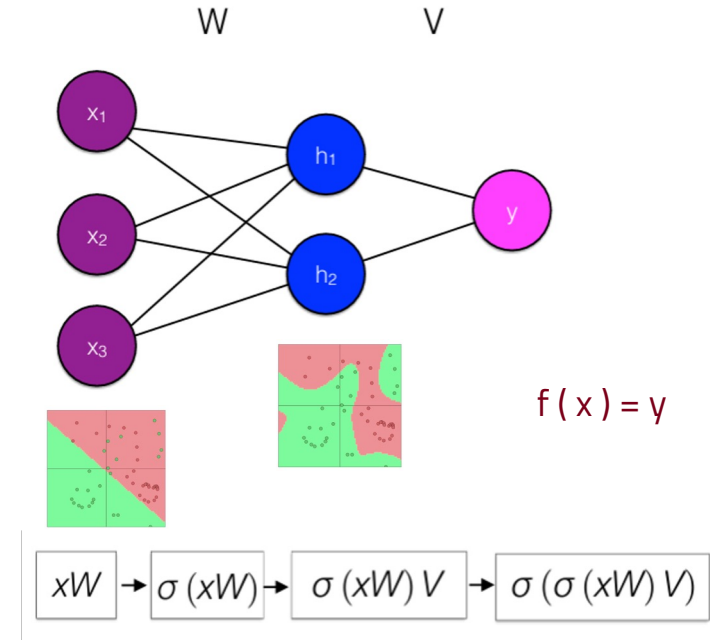


$$f \left(\begin{array}{l} \text{it} \quad 6 \\ \text{I} \quad 5 \\ \text{the} \quad 4 \\ \text{to} \quad 3 \\ \text{and} \quad 3 \\ \text{seen} \quad 2 \\ \text{vet} \quad 1 \end{array} \right) = y$$

$$f \left(\begin{array}{l} \text{I want} \quad 1 \\ \text{have you} \quad 2 \\ \text{how many} \quad 2 \\ \text{isn't it} \quad 1 \\ \dots \end{array} \right) = y$$

happy	1
love	2
recommend	2

Representation
of data (x)



Learning method to learn how
x and y mapped



Text Classification



it 6
I 5
the 4
to 3
and 3
seen 2
vet 1

$$f \left(\begin{array}{l} \text{I want} \quad 1 \\ \text{have you} \quad 2 \\ \text{how many} \quad 2 \\ \text{isn't it} \quad 1 \\ \dots \end{array} \right) = y$$

happy	1
love	2
recommend	2

Does this representation really capture **semantics** of your text (x) ?



Outline

- ❑ What is semantics? How do you define meaning?
- ❑ Three ways of looking at word meaning
 - Decompositional
 - Ontological
 - Distributional
- ❑ Different kinds of encoding “context”
 - Count-based
 - Distributed prediction-based (type) embeddings
 - Distributed contextual (token) embeddings from language models
 - Many more variants



What is semantics?
How do you define meaning?



Define "cup"





You

What is "cup"?



ChatGPT

The term "cup" can have multiple meanings depending on the context. Here are a few possible interpretations:

1. **Physical Object:** A cup is a small container typically used for drinking liquids. It usually has a handle and may be made of materials such as glass, ceramic, metal, or plastic.
2. **Unit of Measurement:** In cooking and baking, "cup" is often used as a unit of measurement for both dry and liquid ingredients. It is a standardized volume measurement, and one cup is equivalent to 8 fluid ounces in the United States or approximately 240 milliliters.
3. **Sports:** In sports, a "cup" can refer to a trophy or championship awarded in a competition, such as the FIFA World Cup in soccer or the Stanley Cup in ice hockey.
4. **Anatomical Reference:** Informally, people might use the term "cup" to refer to the concave shape formed by the hands when placed together, often used to catch or hold something.



Three ways of looking at word meaning

❑ Decompositional

- What **characteristics/components** of what the word represents

❑ Ontological

- How the meaning of the word **relates** to the meanings of other words

❑ Distributional

- What **contexts** the word is found in, relative to other words

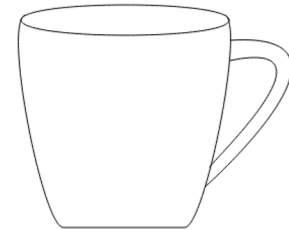


Decompositional semantics



Color: blue, black, etc

Shape:



Texture: ceramic, wood, glass, clay, etc

Limitations of decompositional semantics

- ❑ Where do the features come from?
 - How to divide semantic space into features?
 - How to settle on a “final” list?
- ❑ How do you assign features to concepts and map these features to real world objects in a principled way?

Computationally less useful due to its
subjectivity and lack of scalability

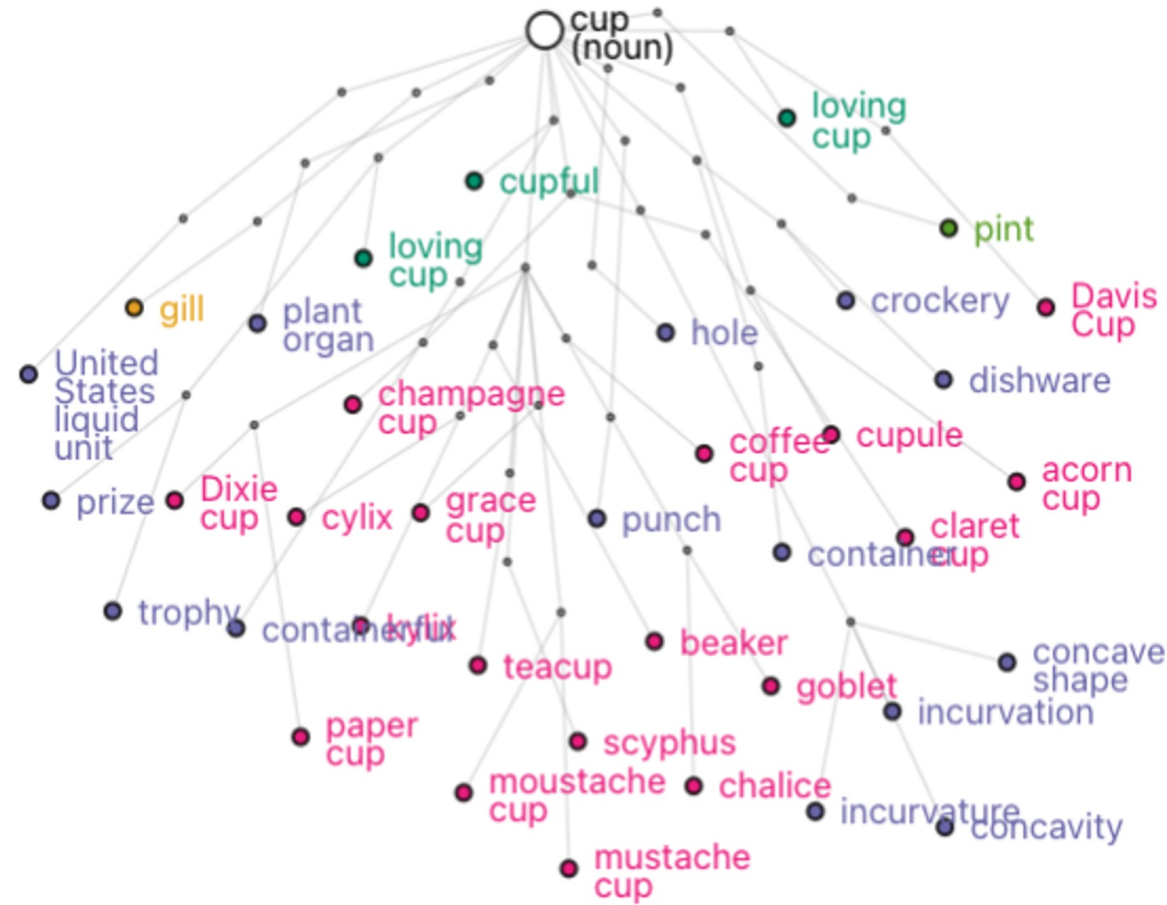


Ontological semantics

- synonym
- hyponym
- attribute

- antonym
- holonym
- entailment

- hypernym
- meronym
- cause

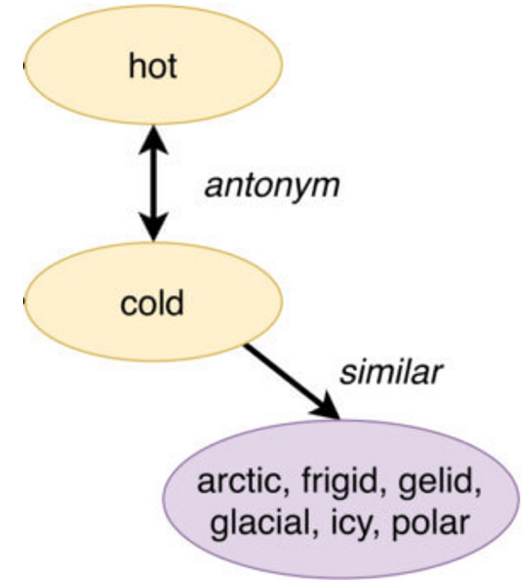


<https://lexical-graph.herokuapp.com/>



Semantic relations

- **Synonymy** — equivalence
 - <small, little>
- **Antonymy** — opposition
 - <small, large>



Semantic relations

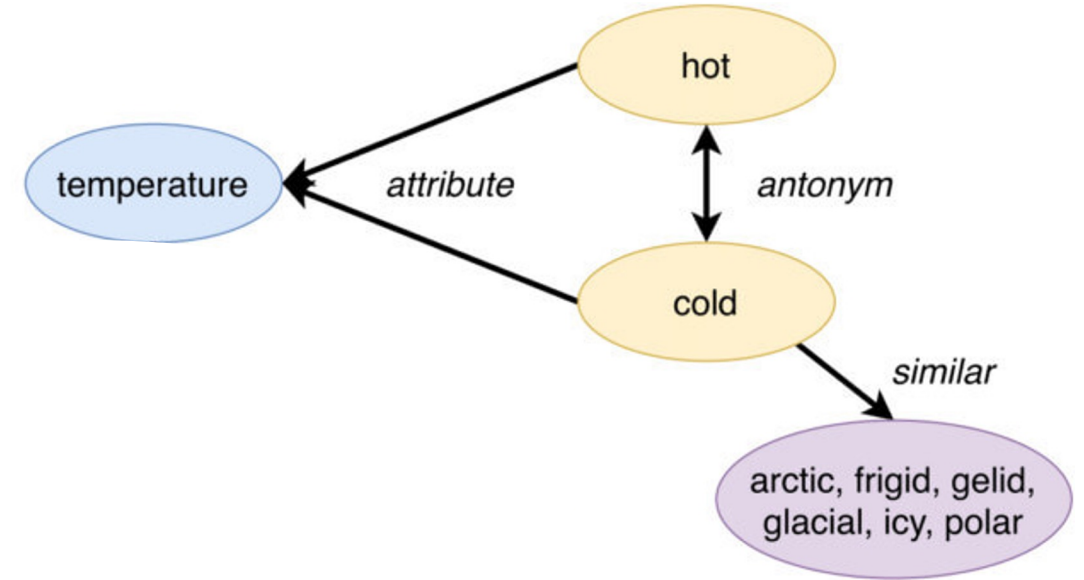
☐ **Synonymy** — equivalence

○ <small, little>

☐ **Antonymy** — opposition

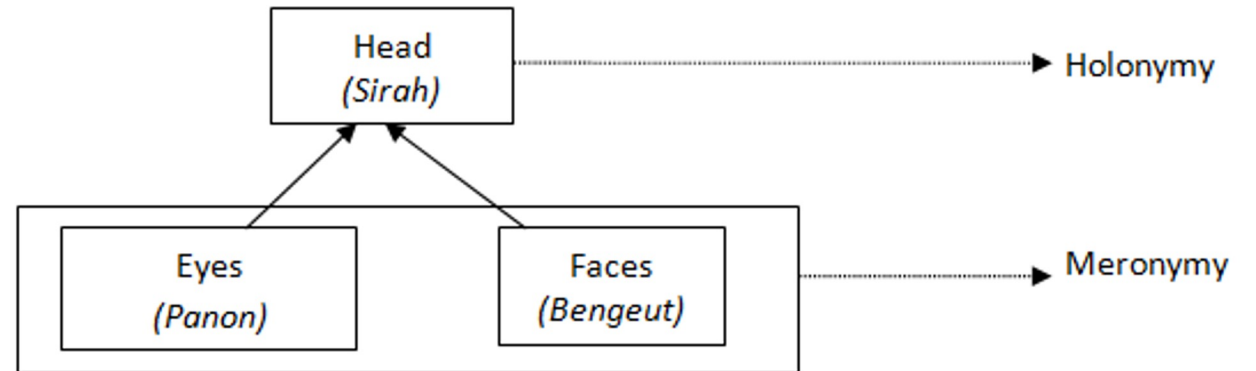
○ <small, large>

☐ **Attribute**



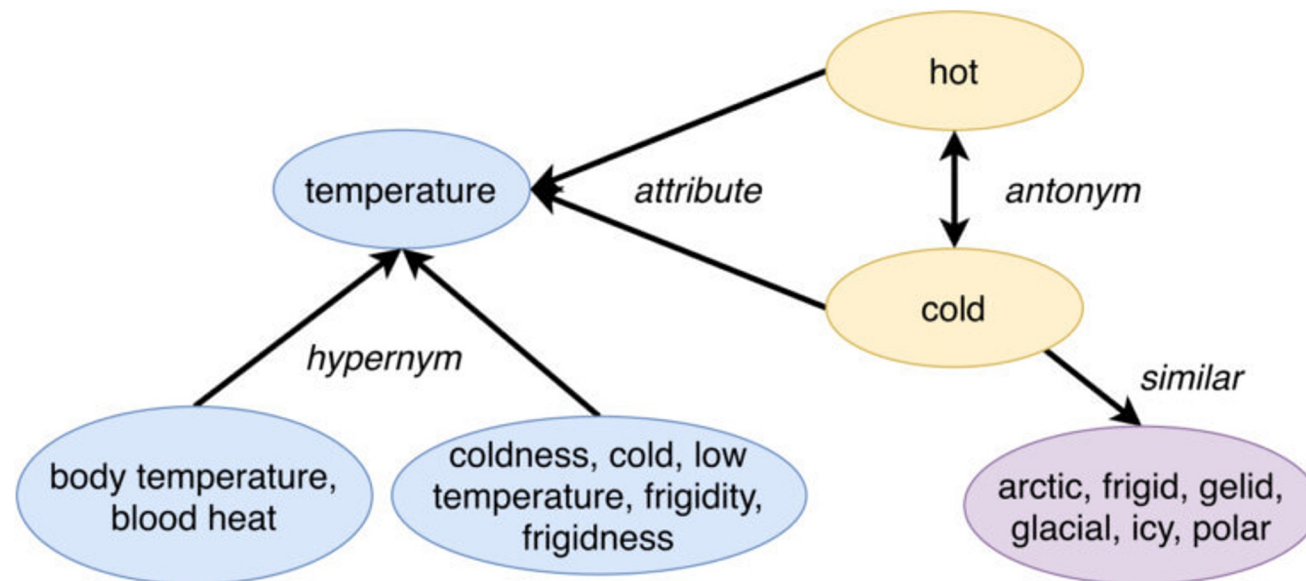
Semantic relations

- Synonymy — equivalence
 - <small, little>
- Antonymy — opposition
 - <small, large>
- **Meronymy** — part-of relation
 - <liver, body>
- **Holonymy** — has-a relation
 - <body, liver>



Semantic relations

- Synonymy — equivalence
 - <small, little>
- Antonymy — opposition
 - <small, large>
- Meronymy — part-of relation
 - <liver, body>
- Holonymy — has-a relation
- **Hyponymy** — subset; is-a relation
 - <dog, mammal>
- **Hypernymy** — superset
 - <mammal, dog>



Tools for WordNet

- ❑ Original English WordNet project:
 - <https://wordnet.princeton.edu/related-projects>
- ❑ WordNets in the World (many languages available):
 - <http://globalwordnet.org/resources/wordnets-in-the-world/>

WordNet Search - 3.1
 - [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
 Display options for sense: (gloss) "an example sentence"

Noun

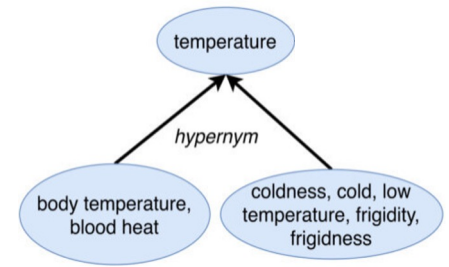
- [S:](#) (n) [jump](#), [leap](#) (a sudden and decisive increase) *"a jump in attendance"*
- [S:](#) (n) [leap](#), [jump](#), [saltation](#) (an abrupt transition) *"a successful leap from college to the major leagues"*
- [S:](#) (n) [jump](#) ((film) an abrupt transition from one scene to another)
- [S:](#) (n) [startle](#), [jump](#), [start](#) (a sudden involuntary movement) *"he awoke with a start"*
- [S:](#) (n) [jump](#), [parachuting](#) (descent with a parachute) *"he had done a lot of parachuting in the army"*
- [S:](#) (n) [jump](#), [jumping](#) (the act of jumping; propelling yourself off the ground) *"he advanced in a series of jumps"; "the jumping was unexpected"*

Taxonomy	Synsets		Lemmas		New words	
	Nouns	Verbs	Nouns	Verbs	Nouns	Verbs
<i>WordNet 1.6</i>	66 025	12 127	94 474	10 319	-	-
<i>WordNet 1.7</i>	75 804	13 214	109 195	11 088	11 551	401
<i>WordNet 2.0</i>	79 689	13 508	114 648	11 306	4 036	182
<i>WordNet 2.1</i>	81 426	13 650	117 097	11 488	2 023	158
<i>WordNet 3.0</i>	82 115	13 767	117 798	11 529	678	33

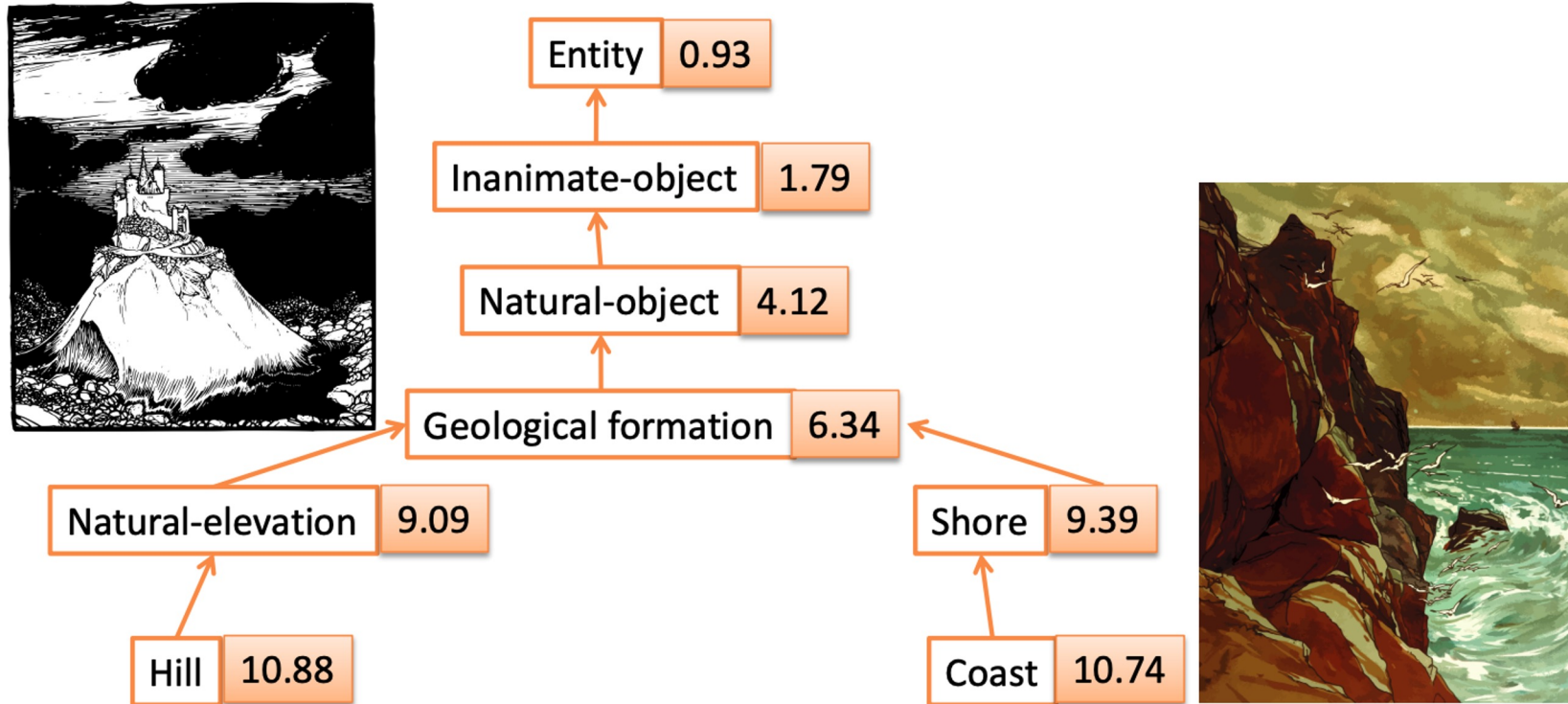
Taxonomy Enrichment with Text and Graph Vector Representations



Quantifying Information Content (IC)



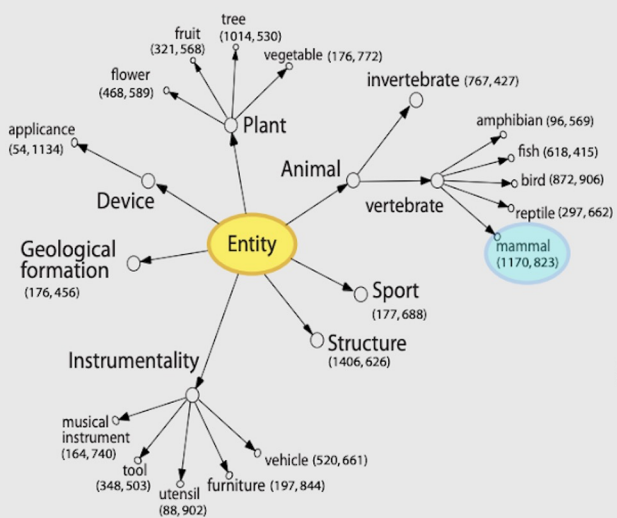
$$IC(w) = -\log \frac{\# \text{ words that are equivalent to or are hyponyms of } w}{\# \text{ words in corpus}}$$



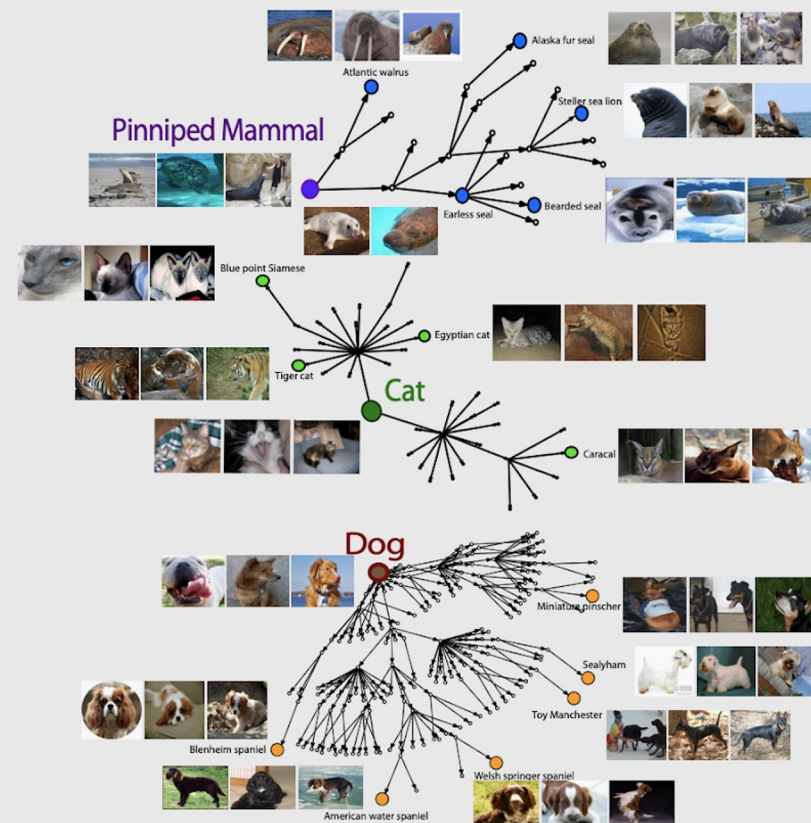
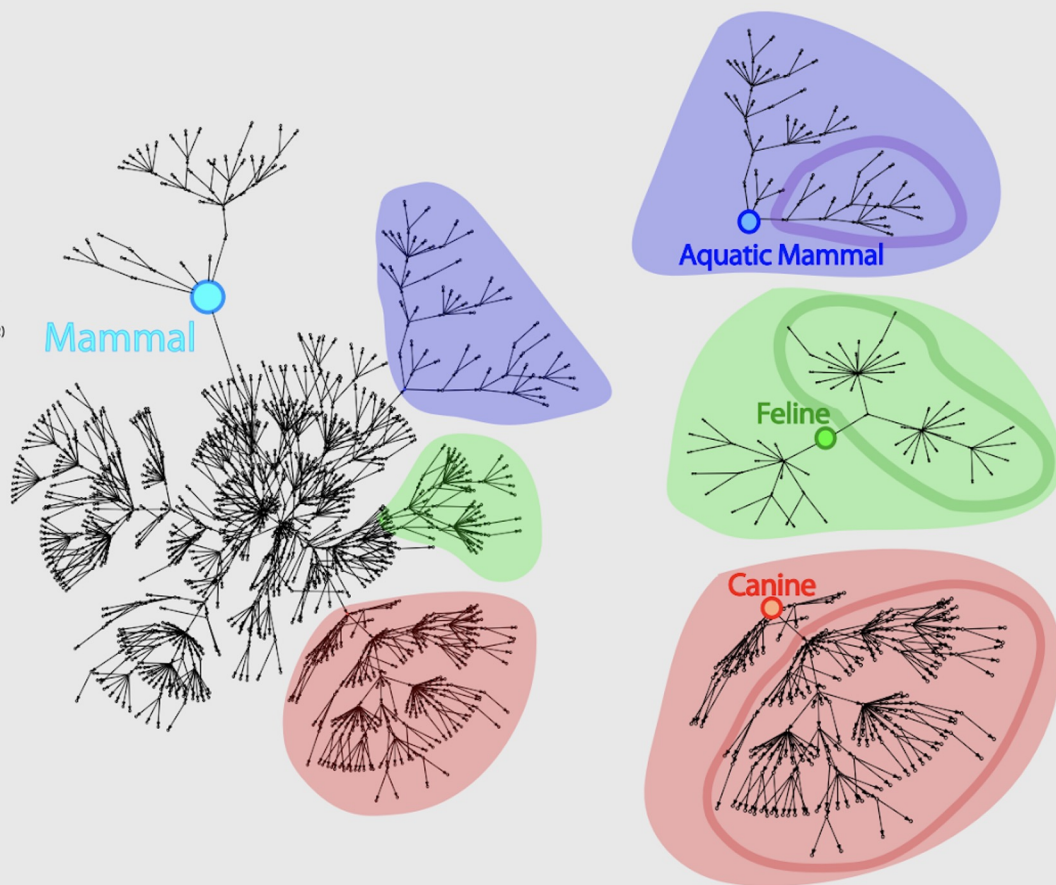
Adapted from Lin. 1998. An information Theoretic Definition of Similarity. ICML.

ImageNet based on WordNet

ImageNet Trees



* synset name (# of synsets in the sub tree, average # of images per synset)



L. Fei-Fei and J. Deng. *ImageNet: Where have we been? Where are we going?*, *CVPR Beyond ImageNet Large Scale Visual Recognition Challenge workshop, 2017*, [pdf](#)



Dictionary: A Database for Lexical Semantics




Lemma

Sense


Definition

Usage

cup *noun*


  /kʌp/  /kʌp/


Idioms

1 ★  [countable] a small container that is like a bowl in shape, usually with a handle, used for drinking tea, coffee, etc.

- He filled the cup with water.
- a coffee cup
- a cup and saucer
- a plastic/paper cup

SEE ALSO **Dixie Cup™**, loving cup, sippy cup




2 ★  [countable] the contents of a cup


- She drank the whole cup.
- **cup of something** *Would you like a cup of tea?*

+ Extra Examples


+ Oxford Collocations Dictionary

3 ★  [countable] a gold or silver cup on a stem, often with two handles, that is given as a prize in a competition

- She's won several cups for skating.
- He lifted the cup (= won) for the fifth time this year.



+ Extra Examples

TOPICS **Sports: ball and racket sports** 

+ Oxford Collocations Dictionary

Extra Examples

- Customers don't like drinking out of plastic cups.
- She raised her cup to her lips.
- She was so thirsty that she drained her cup.

<https://www.oxfordlearnersdictionaries.com/>



No exact match found for “borked” in English

Did you mean:

- barked
- bonked
- booked
- bored
- corked

Nearest results from our other dictionaries and grammar usage guide:

American English

- barked
- bonked
- booked
- bored
- corked

Academic English



Urban Dictionary

borked



To have **totally fucked** something up. Usually by **doing something** stupid. Specifically used to describe **technology** that is broken.

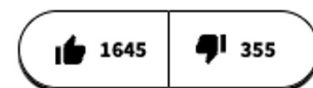
Example 1....

*Admin: I totally borked my machine **installing Win XP SP2**.*

Example 2....

"I can't come over at the moment...my car is borked".

by **Senhor Salvador Guido** September 7, 2004



<https://www.urbandictionary.com/>



What NLP tasks are there in
lexical semantics?



Ask humans how similar 2 words are

Word 1	Word 2	similarity
vanish	disappear	?

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



Ask humans how similar 2 words are

Word 1	Word 2	similarity
vanish	disappear	9.8

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



Ask humans how similar 2 words are

Word 1	Word 2	similarity
vanish	disappear	9.8
behave	obey	

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



Ask humans how similar 2 words are

Word 1	Word 2	similarity
vanish	disappear	9.8
behave	obey	7.3

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



Ask humans how similar 2 words are

Word 1	Word 2	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)



Polysemy

	Sense 1	Sense 2
	Of poor quality or little worth.	slang (orig. U.S.). Formidable, good.

A word sense is a representation of one aspect of a word's meaning.



Polysemy

	Sense 1	Sense 2
bad	Of poor quality or little worth.	slang (orig. U.S.). Formidable, good.

A word sense is a representation of one aspect of a word's meaning.



Polysemy

	Sense 1	Sense 2
bad	Of poor quality or little worth.	slang (orig. U.S.). Formidable, good.
sick	Suffering from illness of any kind; ill, unwell, ailing.	slang (now esp. Skateboarding and Surfing). Excellent, impressive; risky.

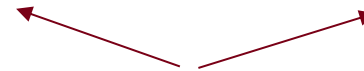
A word sense is a representation of one aspect of a word's meaning.



Polysemy

	Sense 1	Sense 2
bad	Of poor quality or little worth.	slang (orig. U.S.). Formidable, good.
sick	Suffering from illness of any kind; ill, unwell, ailing.	slang (now esp. Skateboarding and Surfing). Excellent, impressive; risky.

Word Sense
Disambiguation (WSD)



"thats **sick** man"

<https://www.urbandictionary.com/define.php?term=sick>

A word sense is a representation of one aspect
of a word's meaning.



Semcor

- ❑ 200K+ words from Brown corpus tagged with Wordnet senses.
- ❑ Benchmark for word sense disambiguation (WSD)

original

It urged that the city take steps to remedy this problem



Semcor

- ❑ 200K+ words from Brown corpus tagged with Wordnet senses.
- ❑ Benchmark for word sense disambiguation (WSD)

original	It urged that the city take steps to remedy this problem
lemma sense	It urge¹ that the city² take¹ step¹ to remedy¹ this problem²



Semcor

- ❑ 200K+ words from Brown corpus tagged with Wordnet senses.
- ❑ Benchmark for word sense disambiguation (WSD)

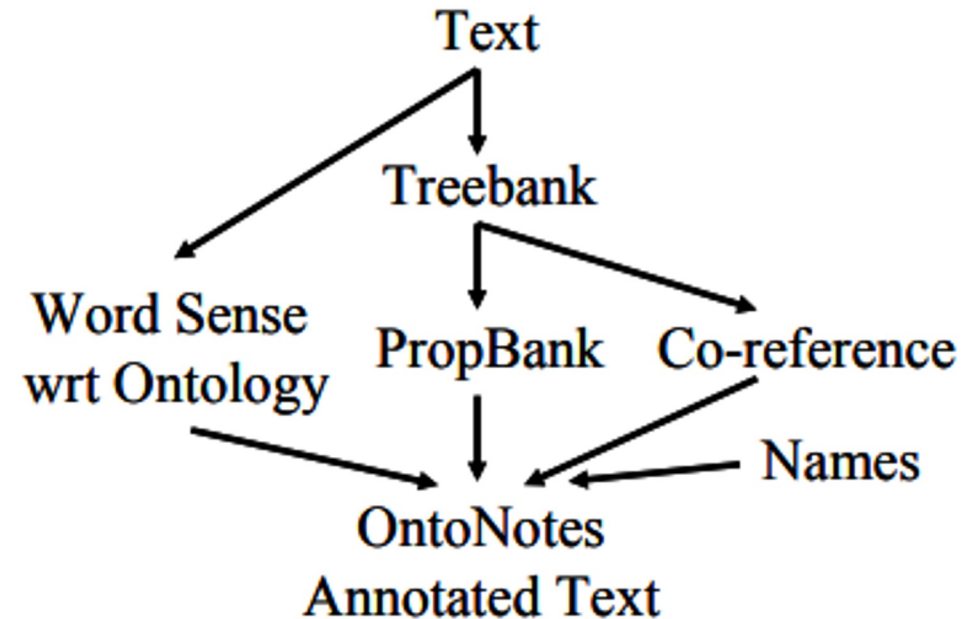
original	It urged that the city take steps to remedy this problem
lemma sense	It urge¹ that the city² take¹ step¹ to remedy¹ this problem²
synset number	It urge^{2:32:00} that the city^{1:15:01} take^{2:41:04} step^{1:04:02} to remedy^{2:30:00} this problem^{1:10:00}



OntoNotes

- Annotate a large corpus comprising **various genres of text** (news, conversational telephone speech, weblogs, usenet newsgroups, broadcast, talk shows) in **three languages** (English, Chinese, and Arabic) with **structural information** (syntax and predicate argument structure) and **shallow semantics** (**word sense** linked to an ontology and coreference).

- Another benchmark for WSD



Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., & Weischedel, R. (2006, June) [OntoNotes: The 90% Solution](#) NAACL, Companion Volume: Short Papers (pp. 57-60).



The founder of Pakistan's nuclear program, Abdul Qadeer Khan, has admitted that he transferred nuclear technology to Iran, Libya, and North Korea”.



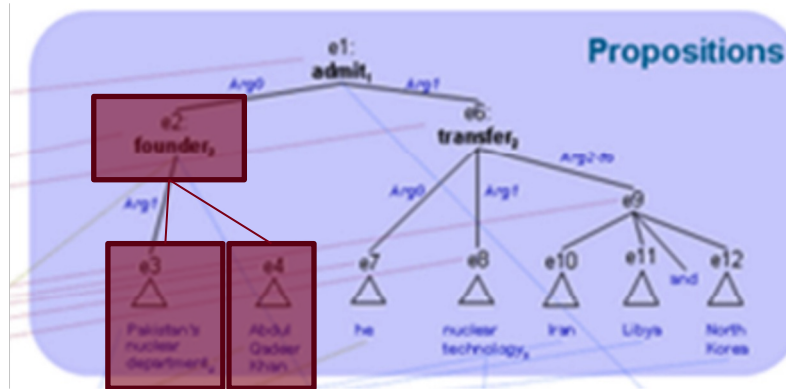
The founder of Pakistan's nuclear program, Abdul Qadeer Khan, has admitted that he transferred nuclear technology to Iran, Libya, and North Korea".

WordNet Ontology



The founder of Pakistan's nuclear program, Abdul Qadeer Khan, has admitted that he transferred nuclear technology to Iran, Libya, and North Korea".

Propositional connection



the semantic kernel of a sentence that determines its truth conditions, independent of its syntactic form and lexical realization.

The founder of Pakistan's nuclear program, Abdul Qadeer Khan, has **admitted** that he **transferred** nuclear technology to Iran, Libya, and North Korea”.

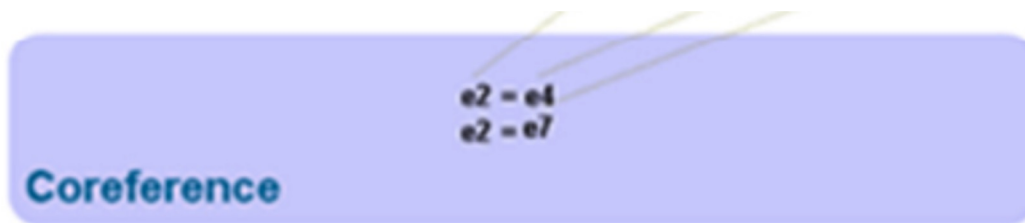
The screenshot shows the VerbNet interface. On the left, a search results list includes 'admit-64.3-1' which is highlighted with a blue box. A red arrow points from this box to the word 'admitted' in the text above. The main content area shows the details for 'ADMIT-64.3-1', including parent classes, member verb lemmas (ADMIT, INCLUDE, PERMIT, RECEIVE, WELCOME), and an example sentence: 'She admitted us here.'

VerbNet
(https://uvi.colorado.edu/uvi_search)

FrameNet
(<https://framenet.icsi.berkeley.edu/>)

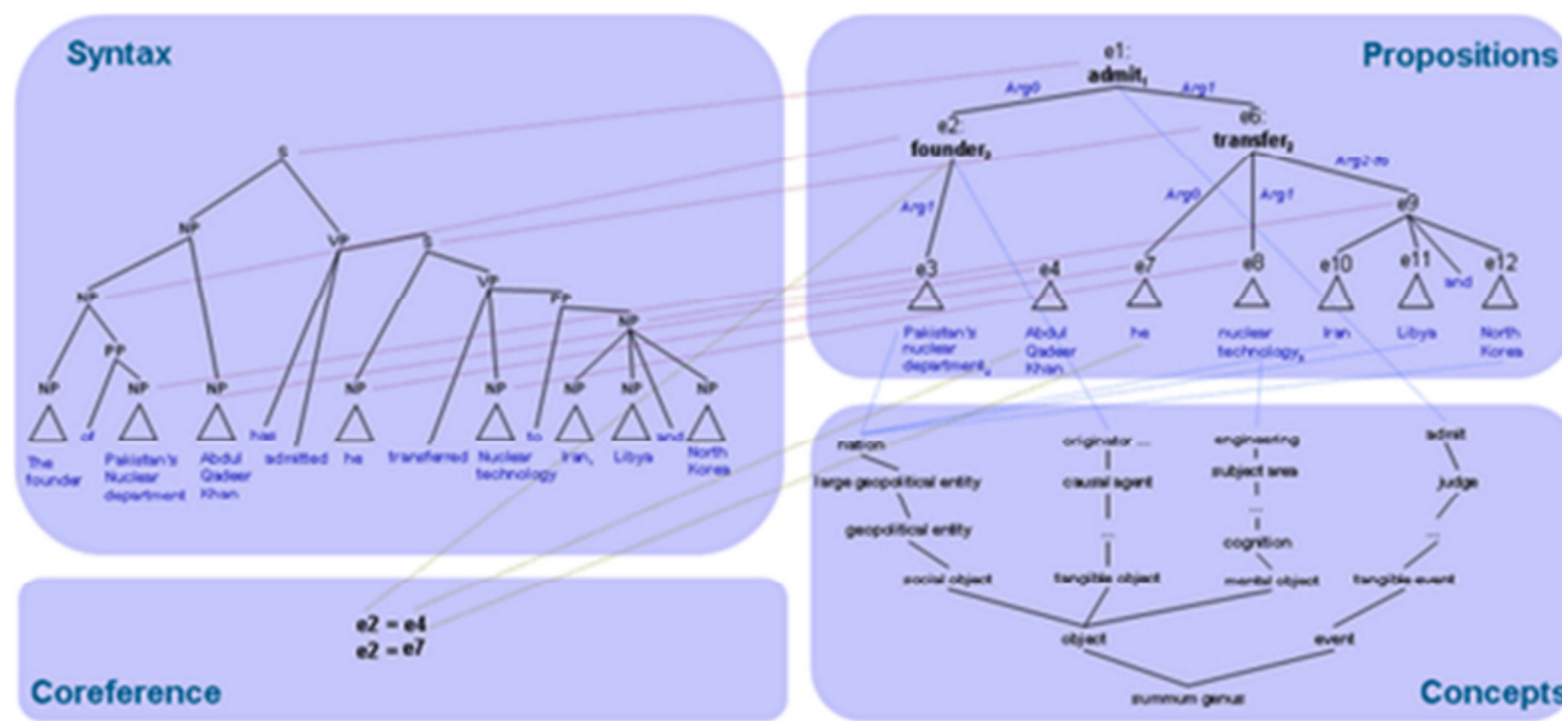


The founder of Pakistan's nuclear program, Abdul Qadeer Khan, has admitted that he transferred nuclear technology to Iran, Libya, and North Korea".



OntoNotes

The founder of Pakistan's nuclear program, Abdul Qadeer Khan, has admitted that he transferred nuclear technology to Iran, Libya, and North Korea”.



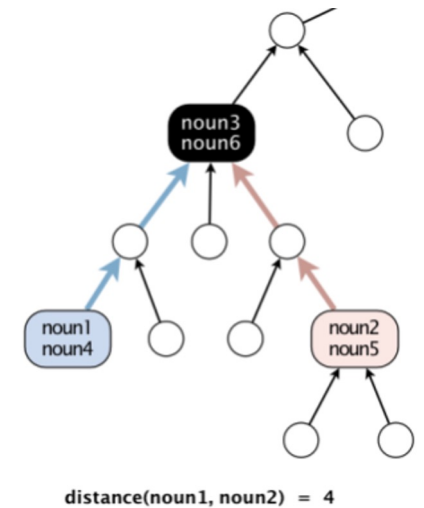
Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., & Weischedel, R. (2006, June) [OntoNotes: The 90% Solution](#) NAACL, Companion Volume: Short Papers (pp. 57-60).



Limitations of WordNet and ontological semantics

- ❑ WordNet is a useful resource, but there are intrinsic limits
 - It requires many years of manual effort by experts like skilled lexicographers
 - Some lexicographers are not skilled, and this has led to inconsistencies (<https://en.wikipedia.org/wiki/WordNet>)
 - Ontology is only as good as ontologists; not driven by data
- ❑ Sparse
 - wicked, sick, badass, ninja..
- ❑ Hard to compute word relationships

Let's look into lexical semantics driven by data
and not rely on lexicographers



Three ways of looking at word meaning

❑ Decompositional

- What **characteristics/components** of what the word represents

❑ Ontological

- How the meaning of the word **relates** to the meanings of other words

❑ **Distributional**

- What **contexts** the word is found in, relative to other words



Assumptions in distributional semantics

“The meaning of word is its **use** in the language”

Wittgenstein PI 43

“You shall know a word by the **company** it keeps”

Firth, J. R. 1957:11

“If A and B have almost identical **environments**
we say that they are synonyms.”

Harris 1954



What does “beef” mean?



Beef

001. □ t but of different schools. Cook had discovered a BEEF in his possession a few days earlier and, whe
002. □ ity to available canned pork products. Tests with BEEF have been largely unsuccessful because of the
003. □ ermaster Corps program is to find the reasons for BEEF's low palatability and means of overcoming it
004. □ rads can extend the shelf life (at 35 F) of fresh BEEF from 5 days to 5 or 6 weeks. However, the pro
005. □ radiation blanching process discolors the treated BEEF and liquid accumulates in prepackaged cuts. C
006. □ nd liquid accumulates in prepackaged cuts. Cooked BEEF irradiated in the absence of oxygen assumes a
007. □ the improvised counter of boards laid across two BEEF barrels. There was, of course, no real need t
008. □ ey of the hut across from him was surmounted by a BEEF barrel with ends knocked out. In this heavy a
009. □ secret employers their money's worth. A good many BEEF-hungry settlers were accepting the death of W
010. □ ogrammes and cost-cutting measures are planned to BEEF up performance. Analysts at Paribas are looki
011. □ rsion into animal feed, produce 600,000 tonnes of BEEF, which alone is worth £1,000 million at whol
012. □ pay on the French wards. We served them up corned BEEF, cheese, pickles and hot cocoa, and they thre
013. □ they threw it back at us. "Good gracious, corned BEEF, cheese and bread and butter, they were lucky
014. □ Greece) Ingredients 1 small packet feta cheese 2 BEEF tomatoes ½ cucumber 1 small jar stoned
015. □ for the past twenty minutes!" Waiter! Waiter! The BEEF surprise was lovely, but what's the surprise?
016. □ as hands who rode for a Mr. Wolgast who supplied BEEF to the reservation up at San Carlos. I would
017. □ "Every week or so I'd see them come in for their BEEF ration. And they're allowed to hunt. They can
018. □ ll specialize in steaks, chops, chicken and prime BEEF as well as Tom's favorite dish, stuffed shrim
019. □ close, she said. She had raised a calf, grown it BEEF-fat. She had, with her own work-weary hands,
020. □ and get ready to bear. She was ready to kill the BEEF, dress it out, and with vegetables from her g
021. □ 't know what to say. He did say she could get her BEEF and vegetables in cans this summer. He did sa
022. □ 12. _HAMBURGER PATTIES WITH NUTS_ 1 pound ground BEEF 2 teaspoons grated onion Dash of pepper 1/2 t
023. □ a highly competitive business more profitable for BEEF, dairy, and sheep men. The target chart quick
024. □ cle assumes that the rations you are feeding your BEEF, dairy cattle, and sheep are adequately balan
025. □ -infective properties of this drug. _HOW TO FEED: BEEF CATTLE (FINISHING RATION)_ - To increase rate
026. □ in the prevention of liver abscesses in feed-lot BEEF cattle. Prevention of bacterial pneumonia, sh
027. □ founder, and in controlling scours. _HOW TO FEED: BEEF AND DAIRY CALVES_ - 0.2 gram Dynafac per head
028. □ gain and improves feed efficiency. _HOW TO FEED: BEEF CATTLE_ - 10 milligrams of diethylstilbestrol
029. □ oves growth rate and feed efficiency of fattening BEEF animals. _HOW TO FEED:_ At the rate of 2-1/2



Beef

- Suppose you see these sentences:
 - "there was the smell of **roast beef**"
 - "I smell **seared beef**, and my stomach twists from hunger."

- Suppose you've also seen these:
 - ..a piece of **meat roasted** or for roasting..
 - ..So long as the **meat** was **seared**, they gulped down several times their own weight every day, ...



Beef is a **meat roasted** or **seared** with smell.

Beef is the culinary name for **meat** from cattle.

Wikipedia, "beef"

Beef

001. □ t but of different schools. Cook had discovered a BEEF in his possession a few days earlier and, whe
002. □ ity to available canned pork products. Tests with BEEF have been largely unsuccessful because of the
003. □ ermaster Corps program is to find the reasons for BEEF's low palatability and means of overcoming it
004. □ rads can extend the shelf life (at 35 F) of fresh BEEF from 5 days to 5 or 6 weeks. However, the pro
005. □ radiation blanching process discolors the treated BEEF and liquid accumulates in prepackaged cuts. C
006. □ nd liquid accumulates in prepackaged cuts. Cooked BEEF irradiated in the absence of oxygen assumes a
007. □ the improvised counter of boards laid across two BEEF barrels. There was, of course, no real need t
008. □ ey of the hut across from him was surmounted by a BEEF barrel with ends knocked out. In this heavy a
009. □ secret employers their money's worth. A good many BEEF-hungry settlers were accepting the death of W
010. □ ogrammes and cost-cutting measures are planned to BEEF up performance. Analysts at Paribas are looki
011. □ rsion into animal feed, produce 600,000 tonnes of BEEF, which alone is worth £1,000 million at whol
012. □ pay on the French wards. We served them up corned BEEF, cheese, pickles and hot cocoa, and they thre
013. □ they threw it back at us. "Good gracious, corned BEEF, cheese and bread and butter, they were lucky
014. □ Greece) Ingredients 1 small packet feta cheese 2 BEEF tomatoes ½ cucumber 1 small jar stoned
015. □ for the past twenty minutes!" Waiter! Waiter! The BEEF surprise was lovely, but what's the surprise?
016. □ as hands who rode for a Mr. Wolgast who supplied BEEF to the reservation up at San Carlos. I would
017. □ "Every week or so I'd see them come in for their BEEF ration. And they're allowed to hunt. They can
018. □ ll specialize in steaks, chops, chicken and prime BEEF as well as Tom's favorite dish, stuffed shrim
019. □ close, she said. She had raised a calf, grown it BEEF-fat. She had, with her own work-weary hands,
020. □ and get ready to bear. She was ready to kill the BEEF, dress it out, and with vegetables from her g
021. □ 't know what to say. He did say she could get her BEEF and vegetables in cans this summer. He did sa
022. □ 12. _HAMBURGER PATTIES WITH NUTS_ 1 pound ground BEEF 2 teaspoons grated onion Dash of pepper 1/2 t
023. □ a highly competitive business more profitable for BEEF, dairy, and sheep men. The target chart quick
024. □ cle assumes that the rations you are feeding your BEEF, dairy cattle, and sheep are adequately balan
025. □ -infective properties of this drug. _HOW TO FEED: BEEF CATTLE (FINISHING RATION)_ - To increase rate
026. □ in the prevention of liver abscesses in feed-lot BEEF cattle. Prevention of bacterial pneumonia, sh
027. □ founder, and in controlling scours. _HOW TO FEED: BEEF AND DAIRY CALVES_ - 0.2 gram Dynafac per head
028. □ gain and improves feed efficiency. _HOW TO FEED: BEEF CATTLE_ - 10 milligrams of diethylstilbestrol
029. □ oves growth rate and feed efficiency of fattening BEEF animals. _HOW TO FEED:_ At the rate of 2-1/2



001. t but of different schools. Cook had discovered a BEEF in his possession a few days earlier and, whe
002. ity to available canned pork products. Tests with BEEF have been largely unsuccessful because of the
003. ermaster Corps program is to find the reasons for BEEF's low palatability and means of overcoming it
004. rads can extend the shelf life (at 35 F) of fresh BEEF from 5 days to 5 or 6 weeks. However, the pro
005. radiation blanching process discolors the treated BEEF and liquid accumulates in prepackaged cuts. C
006. nd liquid accumulates in prepackaged cuts. Cooked BEEF irradiated in the absence of oxygen assumes a
007. the improvised counter of boards laid across two BEEF barrels. There was, of course, no real need t
008. ey of the hut across from him was surmounted by a BEEF barrel with ends knocked out. In this heavy a
009. secret employers their money's worth. A good many BEEF-hungry settlers were accepting the death of W
010. ogrammes and cost-cutting measures are planned to BEEF up performance. Analysts at Paribas are looki
011. rsion into animal feed, produce 600,000 tonnes of BEEF, which alone is worth £1,000 million at whol
012. pay on the French wards. We served them up corned BEEF, cheese, pickles and hot cocoa, and they thre
013. they threw it back at us. "Good gracious, corned BEEF, cheese and bread and butter, they were lucky
014. Greece) Ingredients 1 small packet feta cheese 2 BEEF tomatoes ¼ cucumber 1 small jar stoned
015. for the past twenty minutes!" Waiter! Waiter! The BEEF surprise was lovely, but what's the surprise?
016. as hands who rode for a Mr. Wolgast who supplied BEEF to the reservation up at San Carlos. I would
017. "Every week or so I'd see them come in for their BEEF ration. And they're allowed to hunt. They can
018. ll specialize in steaks, chops, chicken and prime BEEF as well as Tom's favorite dish, stuffed shrim
019. close, she said. She had raised a calf, grown it BEEF-fat. She had, with her own work-weary hands,
020. and get ready to bear. She was ready to kill the BEEF, dress it out, and with vegetables from her g
021. 't know what to say. He did say she could get her BEEF and vegetables in cans this summer. He did sa
022. 12. _HAMBURGER PATTIES WITH NUTS_ 1 pound ground BEEF 2 teaspoons grated onion Dash of pepper 1/2 t
023. a highly competitive business more profitable for BEEF, dairy, and sheep men. The target chart quick
024. cle assumes that the rations you are feeding your BEEF, dairy cattle, and sheep are adequately balan
025. -infective properties of this drug. _HOW TO FEED: BEEF CATTLE (FINISHING RATION)_ - To increase rate
026. in the prevention of liver abscesses in feed-lot BEEF cattle. Prevention of bacterial pneumonia, sh
027. founder, and in controlling scours. _HOW TO FEED: BEEF AND DAIRY CALVES_ - 0.2 gram Dynafac per head
028. gain and improves feed efficiency. _HOW TO FEED: BEEF CATTLE_ - 10 milligrams of diethylstilbestrol
029. oves growth rate and feed efficiency of fattening BEEF animals. _HOW TO FEED:_ At the rate of 2-1/2



Beef

0.7
1.3
-4.5

e.g., shape
e.g., color
e.g., smell

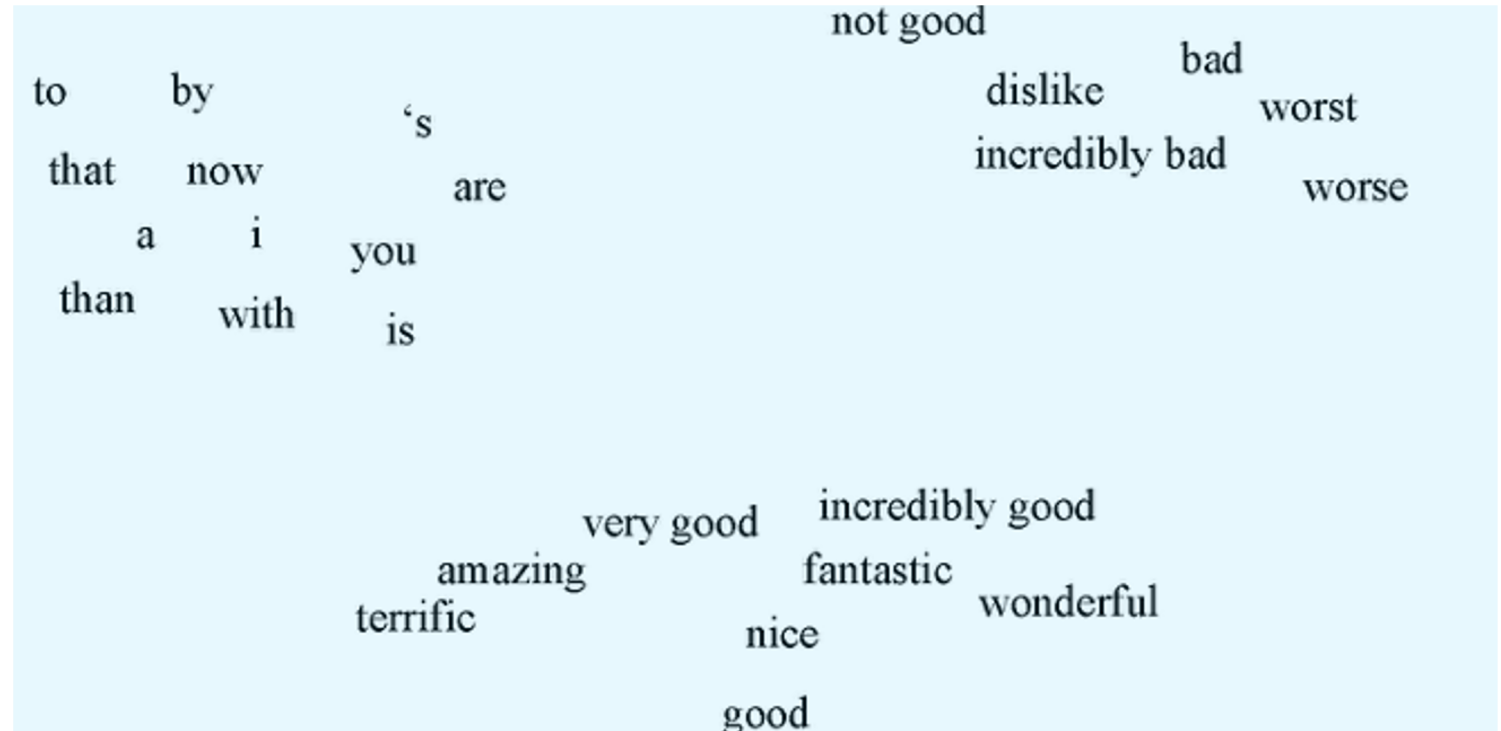
Learned latent feature/component:

Learning a vector representation of "Beef" from context

Model of meaning focusing on similarity

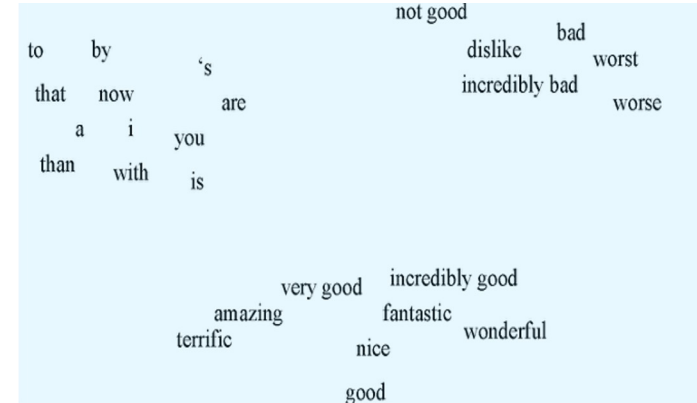
- Each word = a vector
- Not just “word”; Similar words are nearby in space
- The standard way to represent meaning in “NLP” these days

0.7
1.3
-4.5



Distributed representation

- Vector representation that encodes information about the **distribution** of contexts a word appears in
- Words that appear in similar contexts have similar representations (and similar meanings, by the **distributional hypothesis**).
- We have several different ways we can encode the notion of "**context**."



001. □ t but of different schools. Cook had discovered a BEEF in his possession a few days earlier and, who
002. □ ity to available canned pork products. Tests with BEEF have been largely unsuccessful because of the
003. □ ermaster Corps program is to find the reasons for BEEF's low palatability and means of overcoming it
004. □ rads can extend the shelf life (at 35 F) of fresh BEEF from 5 days to 5 or 6 weeks. However, the pro
005. □ radiation blanching process discolors the treated BEEF and liquid accumulates in prepackaged cuts. C
006. □ nd liquid accumulates in prepackaged cuts. Cooked BEEF irradiated in the absence of oxygen assumes a
007. □ the improvised counter of boards laid across two BEEF barrels. There was, of course, no real need t
008. □ ey of the hut across from him was surmounted by a BEEF barrel with ends knocked out. In this heavy t
009. □ secret employers their money's worth. A good many BEEF-hungry settlers were accepting the death of W
010. □ ogrammes and cost-cutting measures are planned to BEEF up performance. Analysts at Faribas are looki
011. □ rison into animal feed, produce 600,000 tonnes of BEEF, which alone is worth £1,000 million at whol
012. □ pay on the French wards. We served them up corned BEEF, cheese, pickles and hot cocoa, and they thro
013. □ they threw it back at us. "Good gracious, corned BEEF, cheese and bread and butter, they were lucky
014. □ Greece) Ingredients 1 small packet feta cheese 2 BEEF tomatoes ¼ cucumber 1 small jar stoned
015. □ for the past twenty minutes!" Walter! Walter! The BEEF surprise was lovely, but what's the surprise?
016. □ as hands who rode for a Mr. Wolgast who supplied BEEF to the reservation up at San Carlos. I would
017. □ "every week or so I'd see them come in for their BEEF ration. And they're allowed to hunt. They can
018. □ ll specialize in steaks, chops, chicken and prime BEEF as well as Tom's favorite dish, stuffed shrim
019. □ close, she said. She had raised a calf, grown it BEEF-fat. She had, with her own work-wear hands,
020. □ and get ready to bear. She was ready to kill the BEEF, dress it out, and with vegetables from her g
021. □ "t know what to say. He did say she could get her BEEF and vegetables in cans this summer. He did sa
022. □ 12. HAMBURGER PATTIES WITH NUTS 1 pound ground BEEF 2 teaspoons grated onion Dash of pepper 1/2 t
023. □ a highly competitive business more profitable for BEEF, dairy, and sheep men. The target chart quick
024. □ cle assumes that the rations you are feeding your BEEF, dairy cattle, and sheep are adequately balen
025. □ -infective properties of this drug. HOW TO FEED: BEEF CATTLE (FINISHING RATION) - - To increase rate
026. □ in the prevention of liver abscesses in feed-lot BEEF cattle. Prevention of bacterial pneumonia, sh
027. □ founder, and in controlling scours. HOW TO FEED: BEEF AND DAIRY CALVES. - 0.2 gram Dynafac per head
028. □ gain and improves feed efficiency. HOW TO FEED: BEEF CATTLE - 10 milligrams of diethylstilbestrol
029. □ oves growth rate and feed efficiency of fattening BEEF animals. HOW TO FEED: At the rate of 2-1/2



Another example of polysemy

have a beef with (someone or something)

To have an outstanding or unsettled dispute or disagreement with someone or something

Why does she have a beef with me? I'm always nice to her!

Dad has a real beef with the phone company because they keep raising their rates.

Beef

0.7
1.3
-4.5
0.1
0.4
...

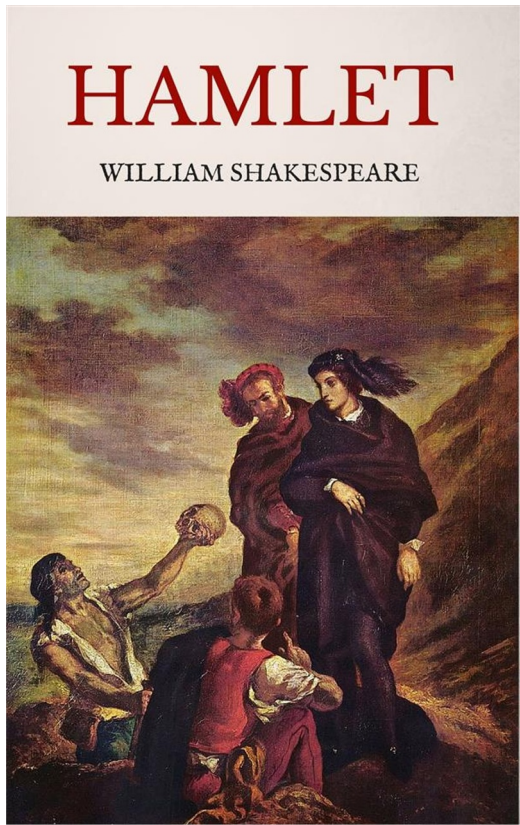
Polysemy should be learned to some extent, assuming that the word vector is trained using ***diverse textual contexts*** and the ***dimensions of the vector*** is large enough



Different kinds of encoding “context”

- ❑ **Count-based**
 - PMI, TF-IDF
- ❑ **Distributed prediction-based (type) embeddings**
 - Word2vec, GloVe, Fasttext
- ❑ **Distributed contextual (token) embeddings from language models**
 - ELMo, BERT, GPT
- ❑ **Many more variants**
 - Multilingual / multi-sense / syntactic embeddings, etc





MICHAEL FASSBENDER MARION COTILLARD
MACBETH

A FILM BY JUSTIN KURZEL
STORY BY ANDREW COOPER AND JUSTIN KURZEL. BASED UPON THE PLAY BY WILLIAM SHAKESPEARE. CASTING BY JAMES HAMILTON. COSTUME DESIGNER ANDREW COOPER. HAIR AND MAKEUP BY JAMES HAMILTON. PRODUCTION DESIGNER ANDREW COOPER. EXECUTIVE PRODUCERS ANDREW COOPER AND JUSTIN KURZEL. PRODUCED BY ANDREW COOPER AND JUSTIN KURZEL. WRITTEN BY ANDREW COOPER AND JUSTIN KURZEL. DIRECTED BY JUSTIN KURZEL. MICHAEL FASSBENDER MARION COTILLARD ANDREW COOPER AND JUSTIN KURZEL. MUSIC BY JAMES NEWTON HOWARD. EDITOR ANDREW COOPER. EXECUTIVE PRODUCERS ANDREW COOPER AND JUSTIN KURZEL. PRODUCED BY ANDREW COOPER AND JUSTIN KURZEL. WRITTEN BY ANDREW COOPER AND JUSTIN KURZEL. DIRECTED BY JUSTIN KURZEL.



Term-document matrix

	Hamlet	Macbeth	Romeo & Juliet	Richard III	Julius Caesar	Tempest
knife	1	1	4	2		2
dog				6	12	2
sword	2	2	7	5		5
love	64		135	63		12
like	75	38	34	36	34	41
...						

Context = appearing in the same document.



Word vectors

knife	1	1	4	2		2
sword	2	2	7	5		5

Vector representation of the **term**;
vector size = **number of documents**



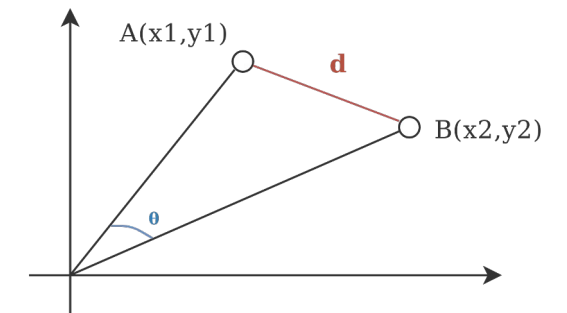
Cosine Similarity

- Calculate the cosine similarity between the two word vectors, to judge the degree of their similarity [Salton 1971]

$$\cos(x, y) = \frac{\sum_{i=1}^F x_i y_i}{\sqrt{\sum_{i=1}^F x_i^2} \sqrt{\sum_{i=1}^F y_i^2}}$$

Note:

- Euclidean distance measures the **magnitude** of distance between two points
- Cosine similarity measures their **orientation**



<https://cmry.github.io/notes/euclidean-v-cosine>



	Hamlet	Macbeth	Romeo & Juliet	Richard III	Julius Caesar	Tempest
knife	1	1	4	2		2
dog				6	12	2
sword	2	2	7	5		5
love	64		135	63		12
like	75	38	34	36	34	41
...						

$\cos(\text{knife}, \text{knife})$ 1.0
 $\cos(\text{knife}, \text{dog})$ 0.11
 $\cos(\text{knife}, \text{sword})$ 0.99
 $\cos(\text{knife}, \text{love})$ 0.65
 $\cos(\text{knife}, \text{like})$ 0.61

Not all dimensions are equally informative.
 Let's weight dimensions!



TF-IDF

- Term frequency ($TF_{t,d}$) = the number of times terms t occurs in document d
 - Several variants: e.g., passing through log function
- Inverse document frequency (IDF_d) = inverse function of number of documents containing (D_t) among total number of documents N .

$$tfidf(t, d) = tf_{t,d} \times \log \frac{N}{D_t}$$



	Hamlet	Macbeth	Romeo & Juliet	Richard III	Julius Caesar	Tempest
knife	1	1	4	2		2
dog				6	12	2
sword	2	2	7	5		5
love	64		135	63		12
like	75	38	34	36	34	41
...						

IDF
0.07
0.30
0.07
0.20
0.00

$$tfidf(t, d) = tf_{t,d} \times \log \frac{N}{D_t}$$

IDF indicates the **informativeness** of the terms when comparing documents.



knife	0.07	0.07	0.28	0.14	0	0.14
dog	0	0	0	1.8	3.6	0.6

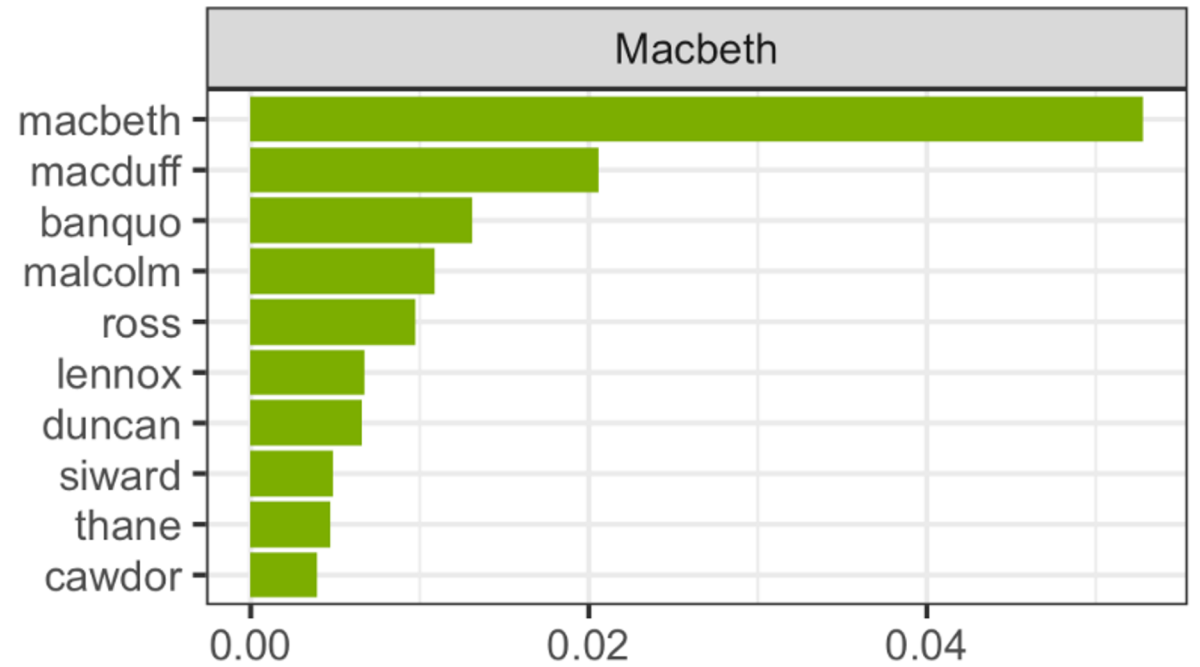
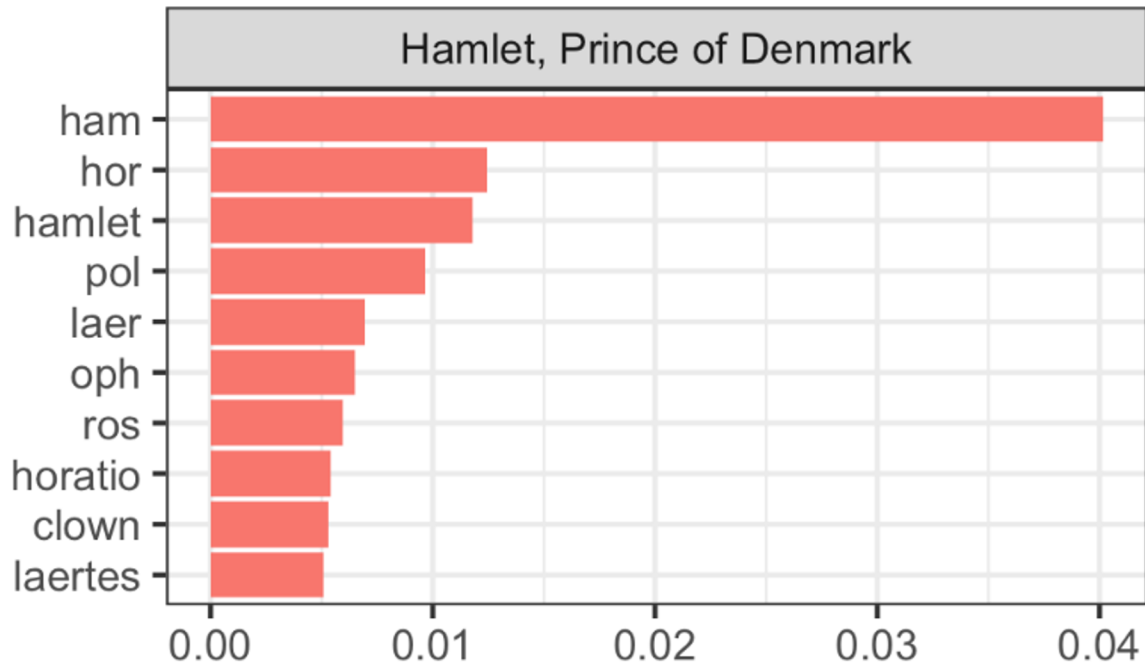
	Hamlet	Macbeth	Romeo & Juliet	Richard III	Julius Caesar	Tempest
knife	1	1	4	2		2
dog				6	12	2
sword	2	2	7	5		5
love	64		135	63		12
like	75	38	34	36	34	41
...						

IDF
0.07
0.30
0.07
0.20
0.00

$$tfidf(t, d) = tf_{t,d} \times \log \frac{N}{D_t}$$

IDF indicates the **informativeness** of the terms when comparing documents.





<https://datavizs21.classes.andrewheiss.com/example/13-example/>



#	Name	Formula
1.	Joint probability	$P(xy)$
2.	Conditional probability	$P(y x)$
3.	Reverse conditional probability	$P(x y)$
4.	Pointwise mutual information	$\log \frac{P(xy)}{P(x)P(y)}$
5.	Mutual dependency (MD)	$\log \frac{P(xy)^2}{P(x)P(y)}$
6.	Log frequency biased MD	$\log \frac{P(xy)^2}{P(x)P(y)} + \log P(xy)$
7.	Normalized expectation	$\frac{2f(xy)}{f(x)+f(y)}$
8.	Mutual expectation	$\frac{2f(xy)}{f(x)+f(y)} \cdot P(xy)$
9.	Salience	$\log \frac{P(xy)}{P(x)P(y)} \cdot \log f(xy)$
10.	Pearson's χ^2 test	$\sum_{i,j} \frac{(f_{ij} - f_{i.}f_{.j})^2}{f_{i.}f_{.j}}$
11.	Fisher's exact test	$\frac{f(x+y)!f(y)!f(x+y)!f(x+y)!}{N!f(x)!f(y)!f(x)!f(y)!}$
12.	t test	$\frac{f(xy) - f(x)f(y)}{\sqrt{f(x)(1-f(x)/N)}}$
13.	z score	$\frac{f(xy) - f(x)f(y)}{\sqrt{f(x)(1-f(x)/N)}}$
14.	Poisson significance measure	$\frac{f(xy) - f(x)f(y) \log f(xy) + \log f(x)f(y)}{\log N}$
15.	Log likelihood ratio	$-2 \sum_{i,j} f_{ij} \log f_{ij} / f_{ij}$
16.	Squared log likelihood ratio	$-2 \sum_{i,j} \log f_{ij}^2 / f_{ij}$
17.	Russel-Rao	$\frac{a+b+c+d}{a+d}$
18.	Sokal-Michiner	$\frac{a+d}{a+b+c+d}$
19.	Rogers-Tanimoto	$\frac{a+d}{a+2b+c+d}$
20.	Hamann	$\frac{(a+d)-(b+c)}{a+b+c+d}$
21.	Third Sokal-Sneath	$\frac{a+d}{a+b+c}$
22.	Jaccard	$\frac{a}{a+b+c}$
23.	First Kulczynski	$\frac{a}{a+b+c}$
24.	Second Sokal-Sneath	$\frac{a+2(b+c)}{a+2(b+c)}$
25.	Second Kulczynski	$\frac{1}{4} \left(\frac{a}{a+b} + \frac{a}{a+c} \right)$
26.	Fourth Sokal-Sneath	$\frac{1}{4} \left(\frac{a}{a+b} + \frac{a}{a+c} + \frac{d}{d+b} + \frac{d}{d+c} \right)$
27.	Odds ratio	$\frac{ad}{bc}$
28.	Yulle's ω	$\frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$
29.	Yulle's Q	$\frac{ad-bc}{ad+bc}$
30.	Driver-Kroeber	$\frac{a}{\sqrt{(a+b)(a+c)}}$
31.	Fifth Sokal-Sneath	$\frac{ad}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$
32.	Pearson	$\frac{ad-bc}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$
33.	Baroni-Urbani	$\frac{a+\sqrt{ad}}{a+b+c+\sqrt{ad}}$
34.	Braun-Blanquet	$\frac{a}{\max(a+b, a+c)}$
35.	Simpson	$\frac{a}{\min(a+b, a+c)}$
36.	Michael	$\frac{a}{(a+d)(b+c)}$
37.	Mountford	$\frac{ab}{2bc+a+b+c}$
38.	Fager	$\frac{a}{\sqrt{(a+b)(a+c)}} = \frac{1}{2} \max(b, c)$
39.	Unigram subtuples	$\log \frac{ad}{bc} = 3.29 \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$
40.	U cost	$\log \left(1 + \frac{\min(b, c) + a}{\max(b, c) + a} \right)$
41.	S cost	$\log \left(1 + \frac{\min(b, c)}{a+1} \right)^{-\frac{1}{2}}$
42.	R cost	$\log \left(1 + \frac{a}{a+b} \right) \cdot \log \left(1 + \frac{a}{a+c} \right)$
43.	T combined cost	$\sqrt{U \times S \times R}$
44.	Phi	$\frac{P(xy) - P(x)P(y)}{\sqrt{P(x)P(y)(1-P(x))(1-P(y))}}$
45.	Kappa	$\frac{P(xy) + P(x)P(y) - P(x)P(y) - P(x)P(y)}{1 - P(x)P(y) - P(x)P(y)}$

#	Name	Formula
46.	J measure	$\max\{P(xy) \log \frac{P(xy)}{P(x)P(y)} + P(x) \log \frac{P(xy)}{P(x)P(y)}, P(xy) \log \frac{P(xy)}{P(x)P(y)} + P(y) \log \frac{P(xy)}{P(x)P(y)}\}$
47.	Gini index	$\max\{P(x)(P(y x)^2 + P(y x)^2) - P(x+y)^2, P(x)(P(y x)^2 + P(y x)^2) - P(x+y)^2, P(y)(P(x y)^2 + P(x y)^2) - P(x+y)^2, P(y)(P(x y)^2 + P(x y)^2) - P(x+y)^2\}$
48.	Confidence	$\max\{P(y x), P(x y)\}$
49.	Laplace	$\max\left\{\frac{NP(x)+1}{NP(x)+2}, \frac{NP(y)+1}{NP(y)+2}\right\}$
50.	Conviction	$\max\left\{\frac{P(xy)P(x)}{P(x)}, \frac{P(xy)P(y)}{P(y)}\right\}$
51.	Piatersky-Shapiro	$P(xy) - P(x)P(y)$
52.	Certainty factor	$\max\left\{\frac{P(xy) - P(x)P(y)}{1 - P(x)}, \frac{P(xy) - P(x)P(y)}{1 - P(y)}\right\}$
53.	Added value (AV)	$\max\{P(y x) - P(y), P(x y) - P(x)\}$
54.	Collective strength	$\frac{P(xy) + P(x)P(y)}{P(x)P(y) + P(x)P(y)}, \frac{1 - P(x)P(y) - P(x)P(y)}{1 - P(x) - P(y)}$
55.	Klosgen	$\sqrt{P(xy)} \cdot AV$
56.	Context entropy	$-\sum_w P(w C_x) \log P(w C_x)$
57.	Left context entropy	$-\sum_w P(w C_x^l) \log P(w C_x^l)$
58.	Right context entropy	$-\sum_w P(w C_x^r) \log P(w C_x^r)$
59.	Left context divergence	$P(x) \log P(x) - \sum_w P(w C_x^l) \log P(w C_x^l)$
60.	Right context divergence	$P(y) \log P(y) - \sum_w P(w C_x^r) \log P(w C_x^r)$
61.	Cross entropy	$-\sum_w P(w C_x) \log P(w C_x)$
62.	Reverse cross entropy	$-\sum_w P(w C_x) \log P(w C_x)$
63.	Intersection measure	$\frac{2 C_x \cap C_y }{ C_x + C_y }$
64.	Euclidean norm	$\sqrt{\sum_w (P(w C_x) - P(w C_y))^2}$
65.	Cosine norm	$\frac{\sum_w P(w C_x)P(w C_y)}{\sqrt{\sum_w P(w C_x)^2} \sqrt{\sum_w P(w C_y)^2}}$
66.	L1 norm	$\sum_w P(w C_x) - P(w C_y) $
67.	Confusion probability	$\sum_w \frac{P(w C_x)P(w C_y)}{P(w)}$
68.	Reverse confusion probability	$\sum_w \frac{P(w C_x)P(w C_y)}{P(w)}$
69.	Jensen-Shannon divergence	$\frac{1}{2} [D(p(w C_x) \frac{1}{2}(p(w C_x) + p(w C_y))) + D(p(w C_y) \frac{1}{2}(p(w C_x) + p(w C_y)))]$
70.	Cosine of pointwise MI	$\frac{\sum_w MI(w, x)MI(w, y)}{\sqrt{\sum_w MI(w, x)^2} \sqrt{\sum_w MI(w, y)^2}}$
71.	KL divergence	$\sum_w P(w C_x) \log \frac{P(w C_x)}{P(w C_y)}$
72.	Reverse KL divergence	$\sum_w P(w C_y) \log \frac{P(w C_y)}{P(w C_x)}$
73.	Skew divergence	$D(p(w C_x) \exp(w C_x) + (1-\alpha)p(w C_x))$
74.	Reverse skew divergence	$D(p(w C_y) \exp(w C_y) + (1-\alpha)p(w C_y))$
75.	Phrase word cooccurrence	$\frac{1}{2} \left(\frac{f(x C_x) + f(y C_x)}{f(x)} \right)$
76.	Word association	$\frac{1}{2} \left(\frac{f(x C_x) - f(x)}{f(x)} + \frac{f(y C_x) - f(y)}{f(y)} \right)$
Cosine context similarity:		$\frac{1}{2} (\cos(\mathbf{e}_x, \mathbf{e}_y) + \cos(\mathbf{e}_x, \mathbf{e}_y))$ $\mathbf{e}_x = (z_i); \cos(\mathbf{e}_x, \mathbf{e}_y) = \frac{\sum z_i x_i}{\sqrt{\sum z_i^2} \sqrt{\sum x_i^2}}$
77.	in boolean vector space	$z_i = \delta(f(w_i C_x))$
78.	in tf vector space	$z_i = f(w_i C_x)$
79.	in $tf \cdot idf$ vector space	$z_i = f(w_i C_x) \cdot \frac{N}{df(w_i)}; df(w_i) = \{x: w_i \in C_x\} $
Dice context similarity:		$\frac{1}{2} (\text{dico}(\mathbf{e}_x, \mathbf{e}_y) + \text{dico}(\mathbf{e}_y, \mathbf{e}_x))$ $\mathbf{e}_x = (z_i); \text{dico}(\mathbf{e}_x, \mathbf{e}_y) = \frac{2 \sum z_i x_i}{\sum z_i^2 + \sum x_i^2}$
80.	in boolean vector space	$z_i = \delta(f(w_i C_x))$
81.	in tf vector space	$z_i = f(w_i C_x)$
82.	in $tf \cdot idf$ vector space	$z_i = f(w_i C_x) \cdot \frac{N}{df(w_i)}; df(w_i) = \{x: w_i \in C_x\} $

Table 1: Inventory of lexical association measures for collocation extraction.



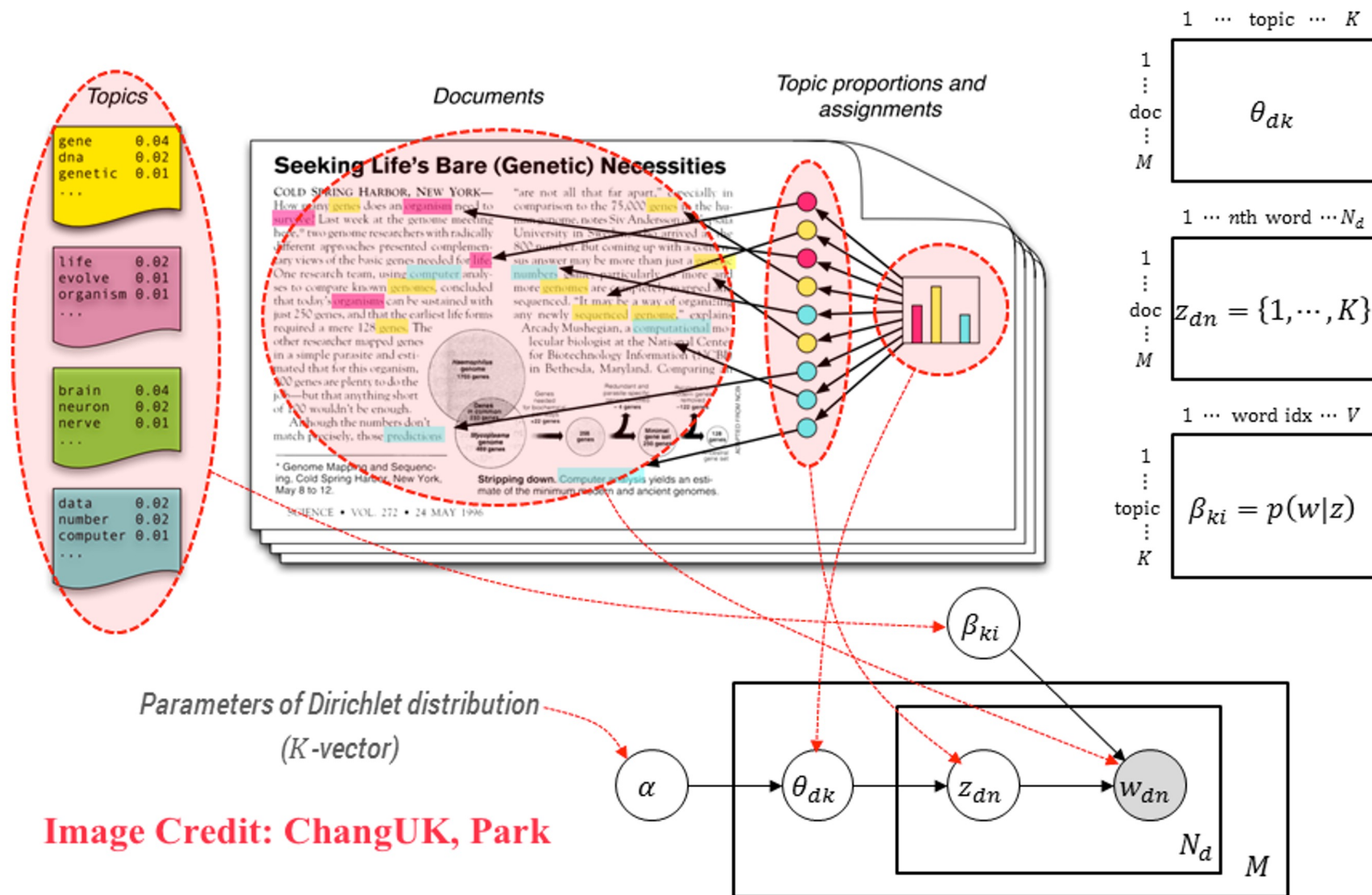
(Optional) Topic Modeling

- ❑ A probabilistic model for discovering hidden “topics” (groups of **terms** that tend to occur together) in documents.
- ❑ Unsupervised (find *interesting structure* in the data)
- ❑ Clustering **tokens** into **topics**

{album, band, music}	{government, party, election}	{game, team, player}
album	government	game
band	party	team
music	election	player
song	state	win
release	political	play
{god, call, give}	{company, market, business}	{math, number, function}
god	company	math
call	market	number
give	business	function
man	year	code
time	product	set



Topic Modeling



Different kinds of encoding “context”

~~Count-based~~

- PMI, TF-IDF

Distributed prediction-based (type) embeddings

- Word2vec, GloVe, Fasttext

Distributed contextual (token) embeddings from language models

- ELMo, BERT, GPT

Many more variants

- Multilingual / multi-sense / syntactic embeddings, etc



