

# **CS 481**

## ***Artificial Intelligence Language Understanding***

**January 24, 2023**

# Announcements / Reminders

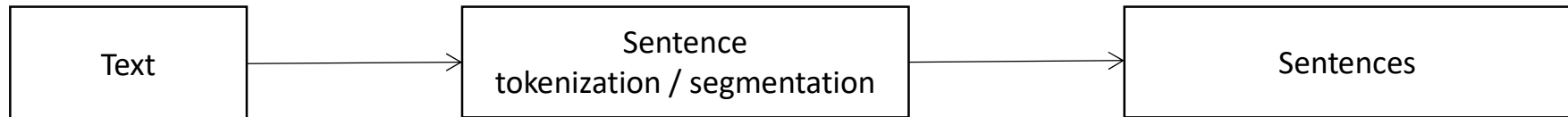
- Please follow the Week 02 To Do List instructions
- Quiz #02 due on Sunday (01/29/23) at 11:59 PM CST
- WA #01 due on Thursday (02/02/23) at 11:59 PM CST
- Thursday (01/26/23) at 12:45 PM: Dr. Dorr's lecture
  - <https://www.iit.edu/events/martha-evens-distinguished-lecture-series-presents-speaker-bonnie-j-dorr>
  - NO CS 481 lecture - attend Dr. Dorr's lecture instead
- Exam dates:
  - Midterm: 03/02/2023 during Thursday lecture time
  - Final: 04/27/2023 during Thursday lecture time

# Plan for Today

- Python libraries / packages for NLP
- Text corpora
- Words and their meanings

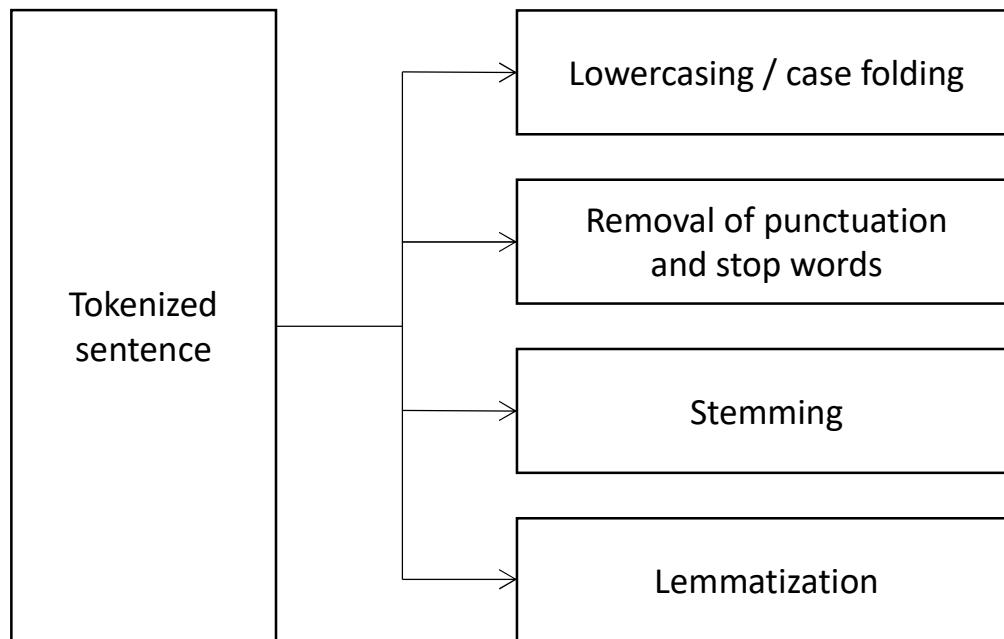
# Basic Pre-Processing: Normalization

Document(s) / text level:



---

Tokenized sentence level:



**Note:** depending on the nature of data, additional pre-processing steps may be required / important.

# Python NLP Libraries / Packages

- **Natural Language Toolkit (NLTK)** [more academic]
- TextBlob
- CoreNLP
- Gensim
- **spaCy** [industry / production]
- Polyglot
- scikit-learn (machine learning)
- pyTorch (machine learning)
- Pattern
- PyNLPI
  
- Hugging Face!

# Natural Language Toolkit (NLTK)

**“NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum.”**

**Link:** [\*\*https://www.nltk.org/\*\*](https://www.nltk.org/)

**Anaconda:** [\*\*https://anaconda.org/anaconda/nltk\*\*](https://anaconda.org/anaconda/nltk)

**Install:** [\*\*https://www.nltk.org/install.html\*\*](https://www.nltk.org/install.html)

# TextBlob

**“TextBlob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.”**

**Link:                <https://textblob.readthedocs.io/en/dev/>**

**Anaconda:        <https://anaconda.org/conda-forge/textblob>**

**Install:           <https://textblob.readthedocs.io/en/dev/install.html>**

# CoreNLP

**“CoreNLP is your one stop shop for natural language processing in Java! CoreNLP enables users to derive linguistic annotations for text, including token and sentence boundaries, parts of speech, named entities, numeric and time values, dependency and constituency parses, coreference, sentiment, quote attributions, and relations. CoreNLP currently supports 8 languages: Arabic, Chinese, English, French, German, Hungarian, Italian, and Spanish.”**

**Link:                <https://stanfordnlp.github.io/CoreNLP/>**

**Anaconda:        <https://anaconda.org/auto/corenlp>**



# Gensim

**“Gensim is a Python library for topic modelling, document indexing and similarity retrieval with large corpora. Target audience is the natural language processing (NLP) and information retrieval (IR) community.”**

**Link:                <https://github.com/RaRe-Technologies/gensim>**

**Anaconda:        <https://anaconda.org/anaconda/gensim>**

**Install:           <https://github.com/RaRe-Technologies/gensim>**

# spaCy

**“spaCy is a free, open-source library for advanced Natural Language Processing (NLP) in Python.”**

**Link:** [\*\*https://spacy.io/\*\*](https://spacy.io/)

**Anaconda:** [\*\*https://anaconda.org/conda-forge/spacy\*\*](https://anaconda.org/conda-forge/spacy)

**Install:** [\*\*https://spacy.io/usage\*\*](https://spacy.io/usage)

# Polyglot

**“Polyglot is a natural language pipeline that supports massive multilingual applications.”**

**Link: <https://polyglot.readthedocs.io/en/latest/index.html>**

**Anaconda: [https://anaconda.org/syllabs\\_admin/polyglot](https://anaconda.org/syllabs_admin/polyglot)**

**Install: <https://polyglot.readthedocs.io/en/latest/Installation.html>**

# scikit-learn

**“Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support-vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.”**

**Link:                <https://scikit-learn.org/stable/index.html>**

**Anaconda:        <https://anaconda.org/anaconda/scikit-learn>**

**Install:           <https://scikit-learn.org/stable/install.html>**

# Pattern

**“Web mining module for Python, with tools for scraping, natural language processing, machine learning, network analysis and visualization.”**

**Link: <https://github.com/clips/pattern>**

**Anaconda: <https://anaconda.org/conda-forge/pattern>**

**Install: <https://github.com/clips/pattern>**

# PyNLPI

“PyNLPI (**P**ython **N**atural **L**anguage **P**rocessing **l**ibrary), pronounced as 'pineapple', is a Python library for Natural Language Processing. It contains various modules useful for common, and less common, NLP tasks.”

Link: <https://github.com/proycon/pynlpl>

Anaconda: N/A?

Install: <https://github.com/proycon/pynlpl>

# Text Corpora

In linguistics, a **corpus** (Latin for “body” | plural: **corpora**) or **text corpus** is a language resource consisting of a large and structured set of texts (nowadays usually electronically stored and processed), written or transcribed.

Text corpora are:

- usually **purposefully collected**
- usually **structured**
- usually **annotated** (part of speech tags, etc.)

# Text Corpora

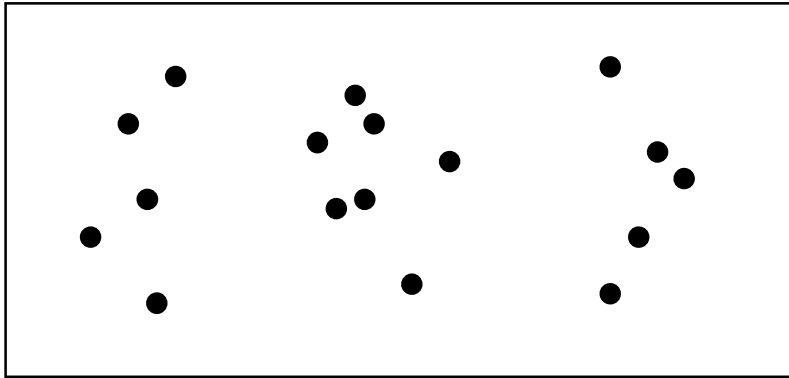
**Words / documents are produced within a context.**

**A text is generated by:**

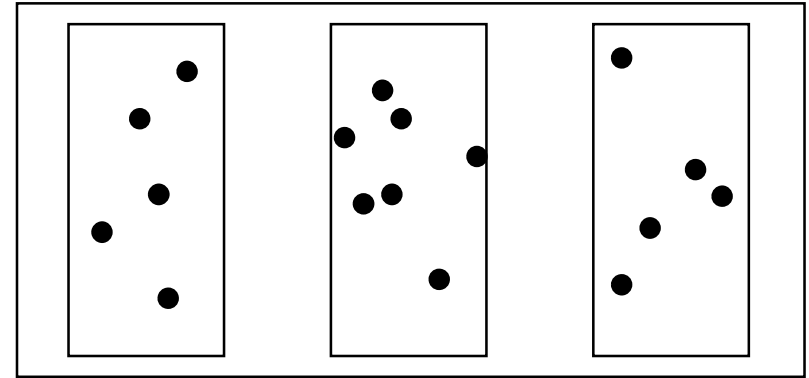
- **a specific writer(s),**
- **at a specific time,**
- **in a specific variety,**
- **of a specific language,**
- **for a specific function.**



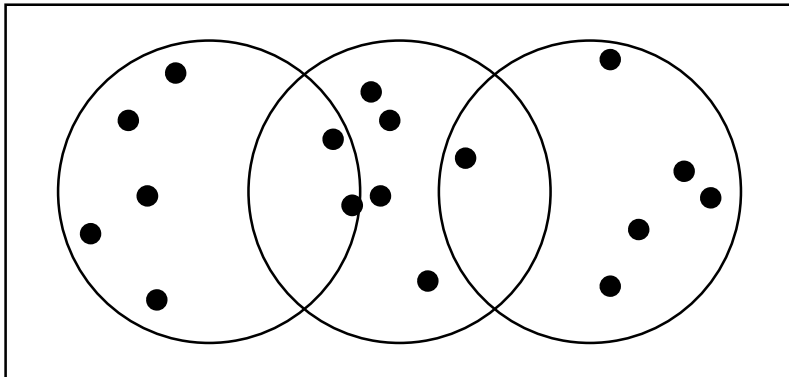
# Text Corpora Structures



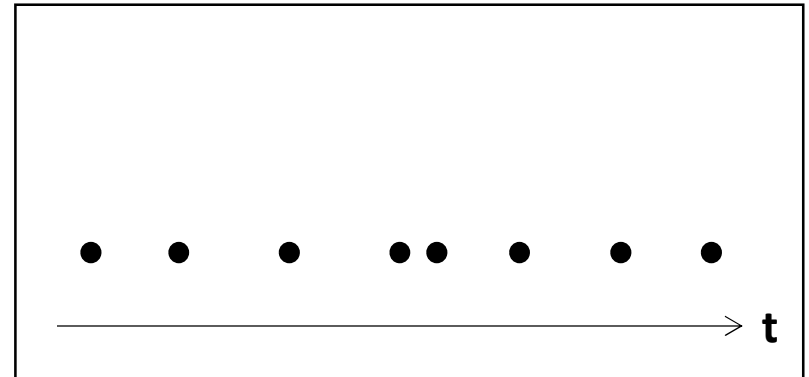
Isolated (e.g. Gutenberg)



Categorized (e.g. Brown)



Overlapping (e.g. Reuters)



Temporal (e.g. Inaugural Address)

# Text Corpora Variation

Text corpora can contain a lot of variation:

- **Language:** 7097 languages in the world
- **Variety,** like African American Language varieties.
- **Twitter posts:** might include forms like "*iont*" (*I don't*)
- **Code switching (e.g., Spanish/English, Hindi/English):**
  - *S/E: Por primera vez veo a @username actually being hateful! It was beautiful:)*
  - *[For the first time I get to see @username actually being hateful! it was beautiful:)]*
  - *H/E: dost tha or ra- hega ... dont worry ... but dherya rakhe*
  - *[“he was and will remain a friend ... don’t worry ... but have faith”]*
- **Genre:** newswire, fiction, scientific articles, Wikipedia
- **Author(s) demographics:** writer's age, gender, ethnicity

# Text Corpora Sizes

**V** = vocabulary = set of types

**|V|** = size (cardinality) of vocabulary

**N** = number of tokens (instances of types)

Heaps Law / Herdan's Law:  $|V| = kN^\beta$

where (often):  $0.67 < \beta < 0.75$  (i.e., vocabulary size grows with  $>$  square root of the number of word tokens)

Corpus	Tokens = <b>N</b>	Types = <b> V </b>
Switchboard phone conversations	2.4 million	20 thousand
Shakespeare	884,000	31 thousand
COCA (Corpus of Contemporary American English)	440 million	2 million
Google N-grams	1 trillion	13+ million

# Text Corpora Datasheet

Text corpora should be described by:





- **Motivation:** Why was the corpus collected, by whom, and who funded it?
- **Situation:** When and in what situation was the text written/spoken?
- **Language variety:** What language (including dialect/region) was the corpus in?
- **Speaker demographics:** What was, e.g., age or gender of the authors of the text?
- **Collection process:** How big is the data? If it is a subsample how was it sampled? Was the data collected with consent? How was the data pre-processed, and what metadata is available?
- **Annotation process:** What are the annotations, what are the demographics of the annotators, how were they trained, how was the data annotated?
- **Distribution:** Are there copyright or other intellectual property restrictions?

# English Corpora: Online Tour



The most widely used online corpora: [guided tour](#), [overview](#), [search types](#), [variation](#), [virtual corpora \(quick overview\)](#) [BYU](#).

The links below are for the online interface. But you can also  download the corpora for use on your own computer.

Corpus (online access)	Download	# words	Dialect	Time period	Genre(s)
<a href="#">News on the Web (NOW)</a>		14.3 billion+	20 countries	2010-yesterday	Web: News
<a href="#">iWeb: The Intelligent Web-based Corpus</a>		14 billion	6 countries	2017	Web
<a href="#">Global Web-Based English (GloWbE)</a>		1.9 billion	20 countries	2012-13	Web (incl blogs)
<a href="#">Wikipedia Corpus</a>		1.9 billion	(Various)	2014	Wikipedia
<a href="#">Coronavirus Corpus</a>		1.3 billion+	20 countries	Jan 2020-yesterday	Web: News
<a href="#">Corpus of Contemporary American English (COCA)</a>		1.0 billion	American	1990-2019	Balanced
<a href="#">Corpus of Historical American English (COHA)</a>		475 million	American	1820-2019	Balanced
<a href="#">The TV Corpus</a>		325 million	6 countries	1950-2018	TV shows
<a href="#">The Movie Corpus</a>		200 million	6 countries	1930-2018	Movies

Source: <https://www.english-corpora.org/>

# NLTK Corpora

## NLTK Corpora

NLTK has built-in support for dozens of corpora and trained models, as listed below. To use these within NLTK we recommend that you use the NLTK corpus downloader, `>>> nltk.download()`

Please consult the README file included with each corpus for further information.

1. *Australian Broadcasting Commission 2006* [[download](#) | [source](#)]  
id: `abc`; size: 1487851; author: Australian Broadcasting Commission; copyright: ; license: ;
2. *Alpino Dutch Treebank* [[download](#) | [source](#)]  
id: `alpino`; size: 2797255; author: ; copyright: ; license: Distributed with permission of Gertjan van Noord;
3. *Averaged Perceptron Tagger* [[download](#) | [source](#)]  
id: `averaged_perceptron_tagger`; size: 2526731; author: ; copyright: ; license: ;
4. *Averaged Perceptron Tagger (Russian)* [[download](#) | [source](#)]  
id: `averaged_perceptron_tagger_ru`; size: 8628828; author: ; copyright: ; license: ;
5. *Grammars for Basque* [[download](#) | [source](#)]  
id: `basque_grammars`; size: 4704; author: Kepa Sarasola; copyright: ; license: ;
6. *BioCreAtIvE (Critical Assessment of Information Extraction Systems in Biology)* [[download](#) | [source](#)]  
id: `biocreative_ppi`; size: 223566; author: ; copyright: Public Domain (not copyrighted); license: Public Domain;
7. *BLLIP Parser: WSJ Model* [[download](#) | [source](#)]  
id: `bllip_wsj_no_aux`; size: 24516205; author: ; copyright: ; license: ;
8. *Grammars from NLTK Book* [[download](#) | [source](#)]  
id: `book_grammars`; size: 9103; author: Ewan Klein; copyright: ; license: ;
9. *Brown Corpus* [[download](#) | [source](#)]  
id: `brown`; size: 3314357; author: W. N. Francis and H. Kucera; copyright: ; license: May be used for non-commercial purposes.;

Source: [https://www.nltk.org/nltk\\_data/](https://www.nltk.org/nltk_data/)



# NLTK: Brown Corpus

The **Brown University Standard Corpus of Present-Day American English** (or just **Brown Corpus**) is an electronic collection of text samples of **American English**, the first major structured corpus of varied genres.

# NLTK: Reuters Corpus

The Reuters Corpus (overlapping corpus):

- 10,788 news documents,
- 1.3 million words,
- documents have been classified into 90 topics



# NLTK: Gutenberg Corpus

NLTK includes a small **selection of texts from the Project Gutenberg electronic text archive**, which contains electronic books (hosted at <http://www.gutenberg.org/>)

# (Statistical) Language Model

- A (statistical) **language model** is a probability distribution over words or word sequences.
- In practice, a language model gives the probability of a certain word sequence being “valid”.
- Validity in this context does not need to mean grammatical validity at all.

Use lexical resources (corpora) to build LM.

# Word Prediction

Words do not randomly appear in text.

The **probability of a word appearing in a text** is to a large degree related to the words that have appeared before it.

- e. g. *I'd like to make a collect. . .*
- *call* is the most likely next word, but other words such as *telephone, international. . .* are also possible.
- other (very common) words are unlikely (e. g. *dog, house*).

# Word Prediction

- Word prediction is very useful for applications such as:
  - speech recognition: it is possible to select between words that are hard for a speech recognizer to distinguish
  - augmentative communication for the disabled: speech generation systems can become more effective
  - spelling error detection:
    - *They are leaving in about 15 minuets.*
    - *He is trying to fine out.*
- Word prediction is also related to the problem of computing the **probability of a sentence**

# Counting Words In Corpora

Text corpora sizes:

- Brown corpus: 1 M words of text.
- Switchboard corpus: 3 M words of speech

Word counting in corpora:

- Punctuation may count as words or not.
- Are “don’t”, “O’Reilly”, “non-carbonated” one or two words.
- Are “They” and “they” different or the same word.
- Many of these choices depend on the application.

# Words: Frequency and Rank

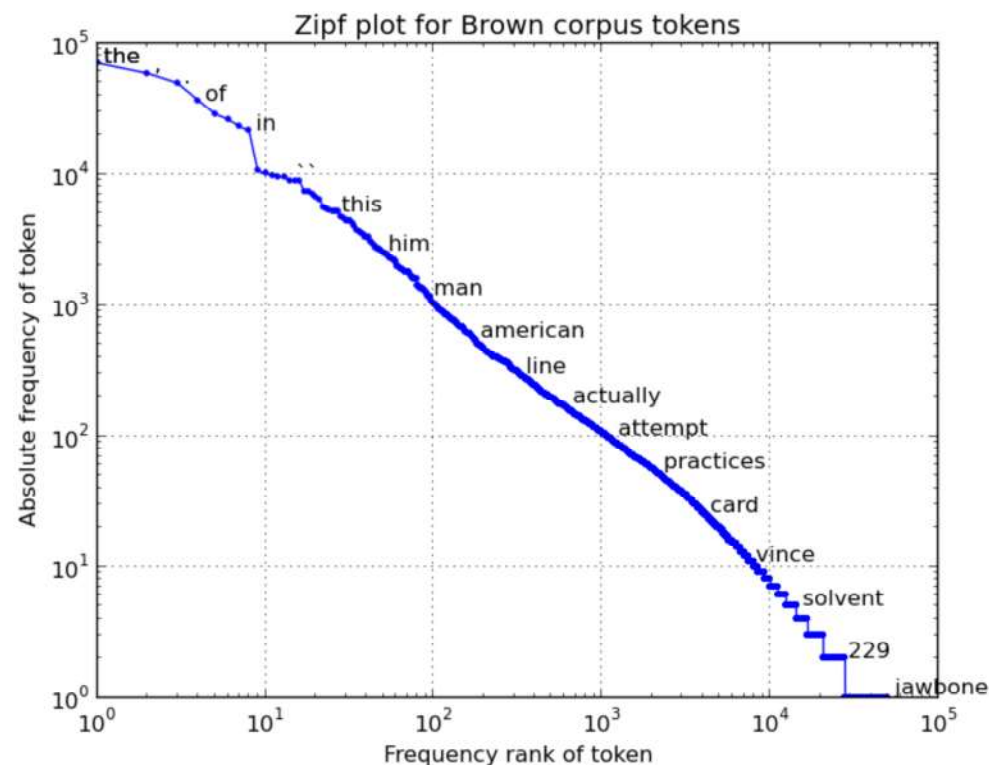
- **Frequency**: a the number of occurrences of a word in the given document or corpus.
- **Rank**: position occupied by a word within a given document or a corpus. A word with the highest frequency will have the highest rank.

# Words in Corpora

- **Tokens**: Total numbers of running words in corpus (possibly including punctuation)
- **Types**: number of distinct words in corpus.
- **Wordform**: the inflected word as it appears in the corpus (*cat* versus *cats*)
- **Lemma**: set of word forms having the same stem and the same word sense. e. g. *cat* and *cats* are not distinguished.

# Words: Frequency and Rank

According to Zipf's law the frequency of word tokens in a large corpus of natural language is inversely proportional to the rank.



Source: <https://finnaarupnielsen.wordpress.com/2013/10/22/zipf-plot-for-word-counts-in-brown-corpus/>



# Simple Language Models

- Probabilistic models of word sequence
- Simplest model:
  - every word may follow any other word
  - **all words have equal probability**
- More complex:
  - the probability of appearance of each word **depends on its frequency in the corpus**:
    - *the* appears 69 971 times in Brown corpus (7%)
    - *rabbit* appears 12 times (0.001%)
- But suppose we have the sentence:
  - Here comes the white. . .

# Collocations

- A **collocation** is any turn of phrase or accepted usage where somehow the **whole is perceived to have existence beyond the sum of its parts.**
- A **collocation** can be defined as a **sequence of words which co-occur more than would be expected by chance.**
- **Collocations** are phrases or expressions made of multiple words, that are very likely to occur together.
- Collocations: “phrases that act like single words”.

# Collocations: Examples and Use

## Examples:

- **compounds:** *disk drive*
- **phrasal verbs:** *make up*
- **stock phrases:** *bacon and eggs*

## Use:

- **keyword extraction**
- **N-gram concatenation:** *social media* → *social\_media*
  - can improve text analysis and processing

# Collocations

**Collocations can:**

- **have specialized meaning:** *CT scan*
- **be idiomatic:** *miss the boat*

**Collocations can be:**

- **several words long:** *international best practice*
- **discontinuous:** *make [something] up*

# Concordance: Key Word In Context

Displaying 25 of 27 matches:

only reiterated the United States' profound attachment to the alliance , `` co  
e not only religious dogma , but a profound and unshakable Weltanschauung . U  
ress to you , once again , his own profound determination to go to the Mainlan  
' ' . It is both great writing and profound religion . N.C. has said something  
say that the idea of death is more profound in Irenaeus than the idea of sin i  
of the American people is clear , profound and far-reaching . To try to oppos  
rable . The idea seems to threaten profound , barely conscious assumptions . A  
to the cause of revolution with a profound respect for legality . John Adams  
er and playing with orchids . More profound and more disturbing , however , is  
Shakespeare's wit and wisdom , his profound insight into human nature , have s  
electorate would have caused us a profound moral shock . About this man we ha  
al structure we find an additional profound difference between the third and t  
y Shelley and Wordsworth . Hegel's profound admiration for the insights of the  
t Woodruff had done , and it had a profound effect on him . Once he learned a  
aches me to wonder ' ' . This was a profound statement . In the face of the unf  
seriously wrestling with the most profound questions of meaning and value . A  
merge . However , there is a more profound consideration to this proposed mer  
of the hypothalamic balance has a profound influence on the clinical behavior  
ugh the methods of deciding may be profound and diverse , the possible conclus  
perience . That which I found most profound and most disturbing appeared to ev  
f Emma Hardy in 1912 , which had a profound emotional effect on Hardy for whic  
bring me closer to a knowledge of profound sorrow than the breakup of camp ,  
d it , they thought it exceedingly profound , though none of them understood i  
mall Polish nobleman with a really profound distaste for his pupils ; ; there  
and its beauties , reflecting its profound religious impulses . He was a prop

# Corpora and Other Lexical Resources

- **Corpora**
- **Lexicons: list of words**
- **Dictionary: list of words with meanings**
- **Word lists: A word list is a list of a language's lexicon within some given text corpus, serving the purpose of vocabulary acquisition**

# What do Words Mean?

In methods (N-grams, text classification, etc.) we've seen:

- words are just strings
- **meaning** is not considered

**Meaning** in logic:

- The meaning of "dog" is DOG (predicates and symbols)

$$\forall x \text{ DOG}(x) \rightarrow \text{MAMMAL}(x)$$

Old linguistics joke by Barbara Partee in 1967:

- Q: What's the **meaning** of life?
- A: LIFE

That is not very helpful.

# Words: Lemmas and Senses

**lemma**

**mouse** (Noun)

1. any of numerous small rodents...
2. a hand-operated device that controls a cursor...

**senses**

from the online thesaurus WordNet



# Words: Lemmas and Senses

**lemma**

**mouse** (Noun)

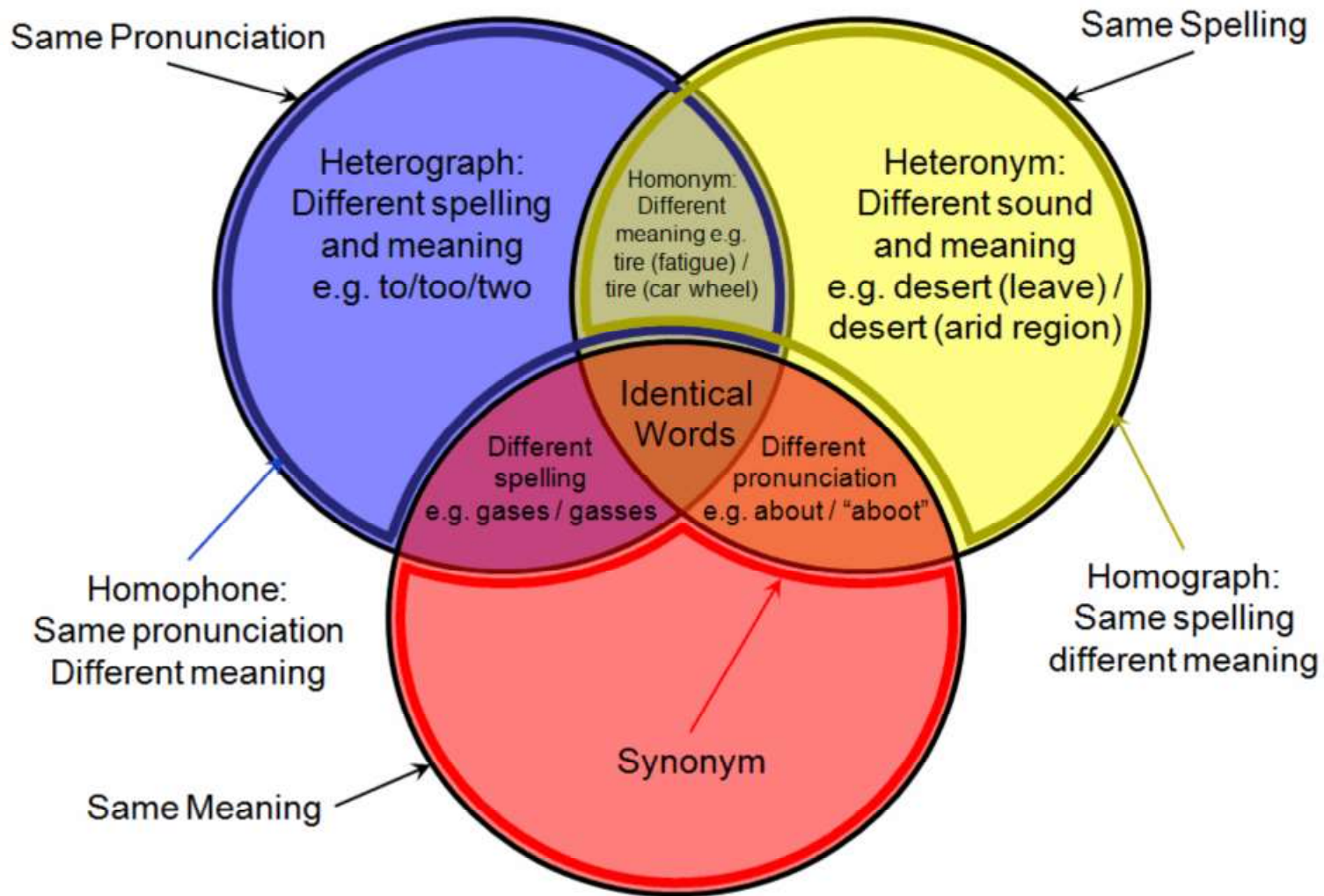
1. any of numerous small rodents...
2. a hand-operated device that controls a cursor...

**senses**

from the online thesaurus WordNet

- A **sense** or “**concept**” is the meaning component of a word
- **Lemmas** can be **polysemous** (have multiple **senses**)

# Relationships Between Words



Words Different In Pronunciation, Spelling, and Meaning

Source: <https://owlcation.com/humanities/Lexical-Relations-Describing-Similarities-In-The-English-Language>

# Lexical Relationships

Lexical relationships are the connections established between one word and another:

- **Synonymy** is the idea that some words have the same meaning as others
  - *quick* is similar to *fast*
- **Antonymy** is precisely the opposite of synonymy
  - *good* is the opposite *bad*
- **Hyponymy** is similar to the notion of embeddedness
  - *Human* ← *Female* (*Female* is a more specific concept than *Human*)
- **Holonomy** and **Meronymy** describe relationships between an object and its parts:
  - *tree* is a holonym of *bark* (*tree* has *bark*)
  - *bark* is a meronym of *tree* (*bark* is a part of *tree*)

# Lexical Semantics: Definition

## Lexical semantics:

*the branch of linguistics and logic concerned with **meaning**.*

*There are a number of branches and subbranches of semantics, including:*

- **formal semantics**, which studies the logical aspects of **meaning**, such as **sense**, **reference**, **implication**, and **logical form**,
- **lexical semantics**, which studies **word meanings** and **word relations**, and **conceptual semantics**, which studies the **cognitive structure of meaning**.

from Oxford Dictionary

# Sense Relationships: Synonymy

- Synonyms have the **same meaning** in **some or all contexts**:
  - *filbert / hazelnut*
  - *couch / sofa*
  - *big / large*
  - *automobile / car*
  - *vomit / throw up*
  - *water / H<sub>2</sub>O*

# Sense Relationships: Synonymy

- There are probably **no examples of perfect synonymy**:
  - **many** aspects of meaning maybe identical, but **not necessarily all** aspects
  - words may differ based on:
    - politeness
    - slang
    - register,
    - genre, etc.

# Sense Relationships: Synonymy?

- Some examples:

- *water / H<sub>2</sub>O*

- Would "*H<sub>2</sub>O*" be used in a surfing guide?

- *car / automobile*

- *big / large*

- *my big sister* is **NOT** always going to be synonymous **with** *my large sister*

# The Linguistic Principle of Contrast

- Substitutions between some pairs of words like *car* / *automobile* or *water* /  $H_2O$  are **truth preserving**, the words are still not **identical in meaning**
- The Linguistic Principle of Contrast  
difference in form → difference in meaning



# Sense Relationships: Similarity

- Words with similar meanings.
- Not synonyms, but sharing some element of meaning
- Some examples:
  - *cow / horse*
  - *car / bicycle*

# Sense Relationships: Similarity

- Knowing **how similar two words** are can:
  - help in computing **how similar the meaning of two phrases or sentences** are
  - assist in **higher level tasks**:
    - question answering
    - paraphrasing
    - summarization

# Sense Relationships: Similarity

## Human-evaluated word similarity:

Word 1	Word 2	Similarity [0-10]
<i>vanish</i>	<i>disappear</i>	9.8
<i>behave</i>	<i>obey</i>	7.3
<i>belief</i>	<i>impression</i>	5.95
<i>muscle</i>	<i>bone</i>	3.65
<i>modest</i>	<i>flexible</i>	0.98
<i>hole</i>	<i>agreement</i>	0.3

Source: SimLex-999 dataset (Hill et al., 2015) | <https://fh295.github.io/simlex.html>

# Sense Relationships: Relatedness

- Also called "**word association**"
- Words can be **related** in any way, for example via a semantic frame or field
- Some examples:
  - *coffee, tea*: **similar**
  - *coffee, cup*: **related**, not **similar**

# Semantic Frame: Definition

## Semantic Frame:

*a semantic frame is defined as a **coherent structure of concepts that are related** such that **without knowledge of all of them, one does not have complete knowledge of one of the either***

**from** *[https://cogling.fandom.com/wiki/Semantic\\_frame](https://cogling.fandom.com/wiki/Semantic_frame)*

# Semantic Field: Definition

## Semantic Field:

*a **lexical set** of semantically **related items***

from Oxford Dictionary

# Semantic Field

Words that

- cover a particular semantic domain
- bear structured relations with each other.

**hospitals**

*surgeon, scalpel, nurse, anaesthetic, hospital*

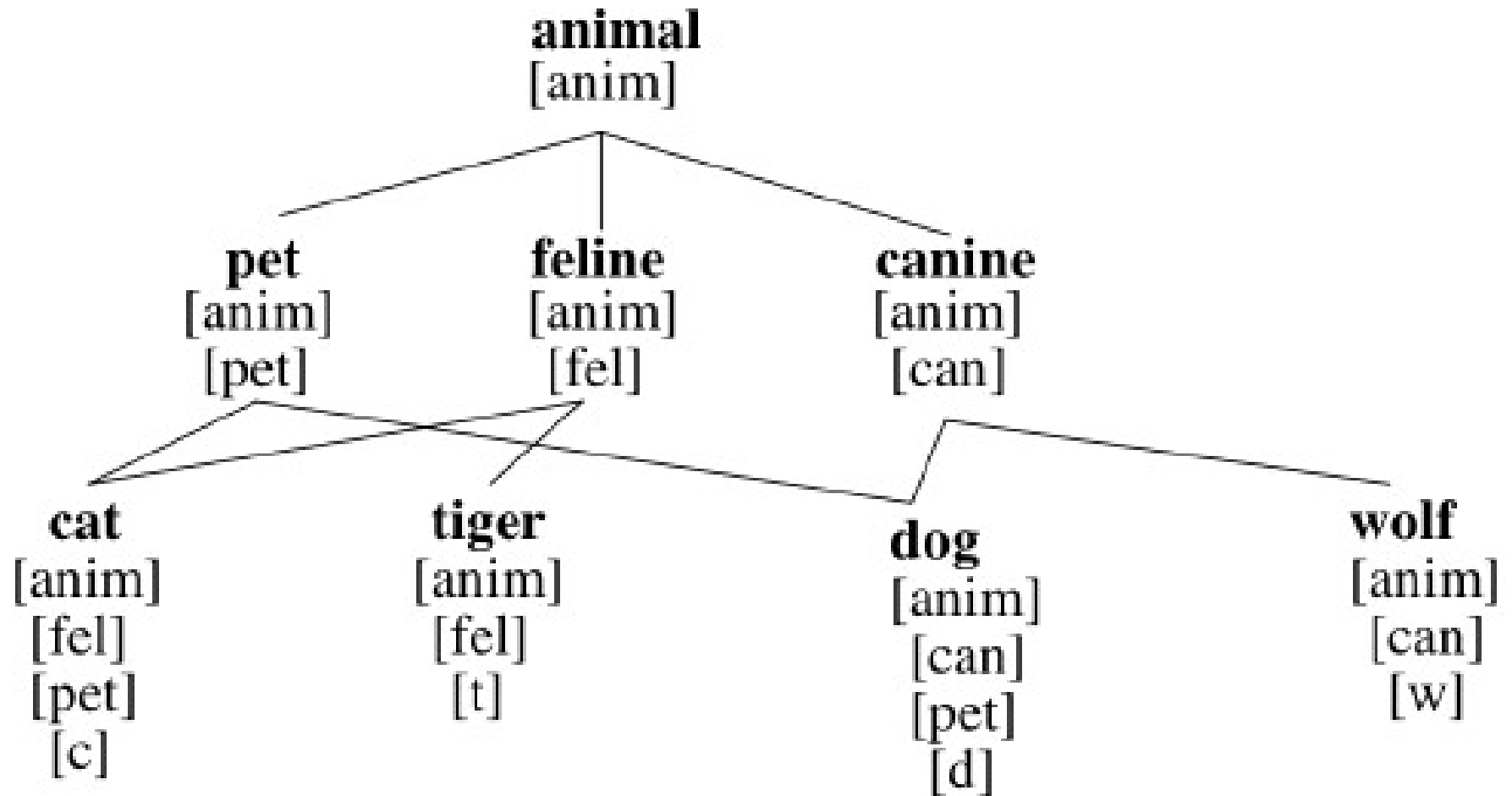
**restaurants**

*waiter, menu, plate, food, menu, chef*

**houses**

*door, roof, kitchen, family, bed*

# Semantic Field



Source: Helge Dyvik - "Translations as a semantic knowledge source"



# Sense Relationships: Antonymy

- Senses that are **opposites with respect to only one feature of meaning**
- Otherwise, they are very **similar (sharing some element of meaning)**:
  - *dark/light    short/long    fast/slow*
  - *hot/cold    up/down    in/out*

# Sense Relationships: Antonymy

- More formally, **antonyms** can
  - define a **binary opposition** or be at **opposite ends of a scale**:
    - *long / short, fast / slow*
  - be **reversives**:
    - *rise/fall, up/down*

# Words and Meaning: Connotations

- Words have **affective** meanings
  - positive connotations (*happy*)
  - negative connotations (*sad*)
- Connotations can be **subtle**:
  - positive connotation: *copy, replica, reproduction*
  - negative connotation: *fake, knockoff, forgery*
- Evaluation (sentiment):
  - positive evaluation (*great, love*)
  - negative evaluation (*terrible, hate*)

# Words and Meaning: Connotations

- Words seem to vary along **three affective dimensions**:
  - valence**: the pleasantness of the stimulus
  - arousal**: the intensity of emotion provoked by the stimulus
  - dominance**: the degree of control exerted by the stimulus

	Word	Score		Word	Score
valence	love	1.000		toxic	0.008
	happy	1.000		nightmare	0.005
arousal	elated	0.960		mellow	0.069
	frenzy	0.965		napping	0.046
dominance	powerful	0.991		weak	0.045
	leadership	0.983		empty	0.081

Source: NRC VAD Lexicon (<https://saifmohammad.com/WebPages/nrc-vad.html>)

# Words and Meaning: Summary

- Concepts or word **senses**
  - have a complex many-to-many association with words (homonymy, multiple **senses**)
- Have **relations** with each other
  - synonymy
  - antonymy
  - similarity
  - relatedness
  - connotation

# Concordance View

**A concordance view shows us every occurrence of a given word, together with some context.**

# WordNet

**WordNet®** is a large lexical database of English.

Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept.

Synsets are interlinked by means of **conceptual-semantic and lexical relations**.

Link: <https://wordnet.princeton.edu/>