AI INSTITUTE #AIISC
UNIVERSITY OF SOUTH CAROLINA

UNIVERSITY OF SOUTH CAROLINA

CSCE 771: Computer Processing of Natural Language
Lecture 9: Semantics

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

15TH SEPTEMBER, 2022

Carolinian Creed: "I will practice personal and academic integrity."

Acknowledgement: Used materials by Jurafsky & Martin,

Organization of Lecture 9

- Opening Segment
  - Announcement
  - Review of Quiz 1
  - Project Report Format
- Main Lecture

Main Section
- Semantics
  - Shallow: similarity, relatedness; frames
  - Propbank
  - Deep: AMR
  - ConceptNet
- Review projects

- Concluding Segment
  - Reading material:
  - About Next Lecture – Lecture 10

7	Sep 8 (Th)	Statistical Parsing, QUIZ
8	Sep 13 (Tu)	Review Parsing, Quiz review, Review Project, Introduce Evaluation
9	Sep 15 (Th)	Semantics
10	Sep 20 (Tu)	Review: Machine Learning for NLP, Evaluation – Metrics
11	Sep 22 (Th)	Language Model – Vector embeddings, CNN/ RNN
12	Sep 27 (Tu)	Guest Lecture – Dr. Amitava Das: Glove, Word2Vec, Transformer Review: Reasoning for NLP
13	Sep 29 (Th)	Representation: Ontology, Knowledge Graph, QUIZ
14	Oct 4 (Tu)	Representation: Embeddings, Language Models
15	Oct 6 (Th)	Entity extraction
16	Oct 11 (Tu)	Guest Lecture – Dr. Amitava Das: Using lang models to solve NLP tasks

Announcements

GUEST LECTURES ON LANGUAGE MODELS BY DR. AMITAVA DAS

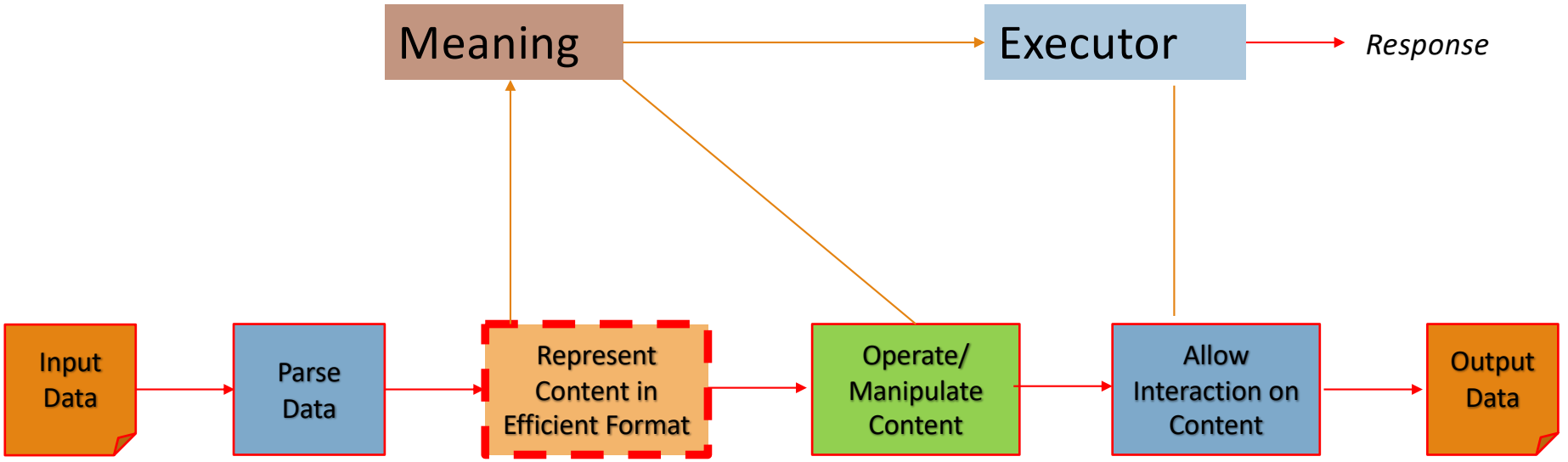
Project Report Format (A Starting Point*)

- Problem
- Related Work
- Data Sources
- Method / Solution Steps/ Algorithm
- Evaluation
- Significance of work / Discussion
- References

* Exact format depends on your project details. So, please change as appropriate.

Main Lecture

Semantics, Parsing and Representation



Semantics

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

Source: Jurafsky & Martin, Wikipedia (https://en.wikipedia.org/wiki/Semantics)

From Text to Meaning

- Shallow semantics
  - Input: text
  - Output: ***lexical semantics***
- Deep semantics
  - Input: text
  - Output: ***formal semantics***

LOGIC format:

$\exists$ w, b, g:
instance(w, want-01) $\wedge$ instance(g, go-01) $\wedge$
instance(b, boy) $\wedge$ arg0(w, b) $\wedge$
arg1(w, g) $\wedge$ arg0(g, b)

AMR format (based on PENMAN):

```
(w / want-01
  :arg0 (b / boy)
  :arg1 (g / go-01
          :arg0 b))
```

GRAPH format:

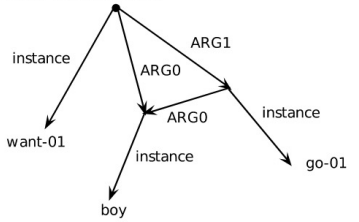


Figure 1: Equivalent formats for representating the meaning of "The boy wants to go".

Source: Abstract Meaning Representation for Sembanking, https://amr.isi.edu/a.pdf

Review: Common Definitions

- **Corpus** (plural corpora): a computer-readable corpora collection of text or speech.
- **Lemma**: A lemma is a set of lexical forms having the same stem, the same major part-of-speech, and the same word sense. Example: Cat and cats have same lemma.
- **Word form**: The word form is the full inflected or derived form of the word. Example: Cat and cats have different word forms.
- **Word type**: Types are the number of distinct words in a corpus. if the set of words is V, the number of types is the word token vocabulary size |V|.
- **Word tokens**: The total number N of running words in the sentence / document of interest.
- **Code switching**: use multiple languages in a code switching single communicative act – Example: Hindlish (Hindi English), Spanish (Spanish English)

"They picnicked by the pool, then lay back on the grass and looked at the stars."

- 16 tokens, 14 word types

Source: Jurafsky & Martin

Lexical Semantics

- Lemma
  - Sing, Mouse
- Word form
  - Sing, sang, sung
  - Mouse, mice
- Word sense
  - Mouse: a rodent
  - Mouse: an electronic pointing device

A lemma having many senses is called **Polysemous**

Synonymous and Similar Words

- **Synonym** - one word has a sense whose meaning is identical to a sense of another word
  - Two words are **synonymous** if they are substitutable one for the other in any sentence without changing the truth conditions of the sentence, the situations in which the sentence would be true
  - **Propositional meaning** – synonym words have the same propositional meaning (truth preserving)
- **Principle of contrast** – An assumption in linguistics is that difference in linguistic form (e.g., word form) is always associated with at least some difference in meaning
  - Water and H20 are truth preserving but used in different context
  - Synonym words are used for approximate synonymy. Then, how similar are the words?

Source: Jurafsky & Martin

Word Similarity - SimLex-999

- Captures similarity between word pairs, mining the opinions of 500 annotators via Amazon Mechanical Turk on a scale of 1 to 10

Note*: similarity,* rather than *relatedness* or *association*

- Contains
  - 666 Noun-Noun pairs,
  - 222 Verb-Verb pairs
  - 111 Adjective-Adjective pairs

vanish	disappear	9.8
behave	obey	7.3
belief	impression	5.95
muscle	bone	3.65
modest	flexible	0.98
hole	agreement	0.3

Source: Jurafsky & Martin

- **Usage**: Evaluation of learning based approaches for finding word similarity by correlation

SimLex-999: Evaluating Semantic Models with (Genuine) Similarity Estimation. 2014. Felix Hill, Roi Reichart and Anna Korhonen. *Computational Linguistics.* 2015
Website: https://fh295.github.io/simlex.html

Meaning (Semantics) versus Structure (Lexical)

Pair	Simlex-999 rating	WordSim-353 rating
coast - shore	9.00	9.10
clothes - closet	1.96	8.00

Example courtesy: https://fh295.github.io/simlex.html

Word Relatedness/ Association

- **Semantic Field:** related words from the same particular domain and bear structured relations with each other.
  - Example 1: cup, coffee
  - Example 2: scalpel, surgeon
  - Usually determined by experts in a field

Source: Jurafsky & Martin

- **Word Association Test/ Task:** how word meaning is stored in memory
  - Have people respond to word associations as a game; e.g., say the first word that comes to mind when one says "Doctor"
  - Applications
    - Used in marketing
    - Also evaluation of learning procedures discovering meaning (e.g., word embedding)

Sources:
- https://psychology.jrank.org/pages/656/Word-Association-Test.html,
- Establishing the Reliability of Word Association Data for Investigating Individual and Group Differences , Tess Fitzpatrick, David Playfoot, Alison Wray, Margaret J. Wright *Applied Linguistics*, Volume 36, Issue 1, February 2015, Pages 23–50, https://doi.org/10.1093/applin/amt020

Discovering Word Relatedness

- **Topic model**: a statistical notion of related words in a document. Hope is that meaningful topics will be from the same semantic field, but there is no guarantee
- Key idea
  - Topic: group of words
  - Counting words and grouping similar word patterns to infer topics within unstructured data.
  - Assumptions
    - Distributional hypothesis: similar topics make use of similar words
    - Statistical mixture hypothesis: documents talk about several topics
  - Perform unsupervised analysis/ clustering: given a corpus and number of topics (k), find k topics that are representative of key ideas in the corpus

References:

- Blog: https://monkeylearn.com/blog/introduction-to-topic-modeling/
- Tool: Gensim
- Paper: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Frames, Slots: Frame Semantics

- Examples
  - "John sold a car to Mary”
  - "Mary bought a car from John”
  - “Mary paid John a undisclosed amount to get his car”
- To understand a word, one needs to understand the knowledge related to the word
  - In example: sell, buy, pay
- Capture knowledge in structures called **semantic frames** which have placeholders called slots (variables)
  - During parsing of sentences, values are filled
- **Frame semantics** is a theory of linguistic meaning developed by Charles J. Fillmore; related notion is semantic parsing

PropBank FrameSet

- A repository of formalized predicates
  https://propbank.github.io/

```
<roleset id="care.01”
        name="having an opinion, feeling tenderly/strongly
for/about">
<roleset id="care.02"
        name="liking/desiring/wanting">
<roleset id="care.03"
        name="tending, taking care of">
<roleset id="care.04"
        name="being cautious, taking care to">
```

Example: Care
https://github.com/propbank/propbank-frames/blob/main/frames/care.xml

Hindi – भेजा - ***Beja***

Credits: https://verbs.colorado.edu/propbank/framesets-hindi/Beja-v.html

Example: Hindi Propbank

Roleset id: Beja.01 , *to send, transport, ship something*
- **Arg0**: *the one who sends something*
- **Arg2**: *the recipient to whom something is sent*
- **Arg1**: *the thing that is sent*

Roleset id: Beja.02 , *to send, transport, ship something*
- **Arg0**: *the one who sends something*
- **Arg2-gol**: *the place where something is sent*
- **Arg1**: *the thing that is sent*

Roleset id: Beja.03 , *to make someone send something to someone*
- **Argc**: *the causer- the one who makes someone send something*
- **Arga**: *the intermediate causer*
- **Arg0**: *the agent- the one who sends something*
- **Arg2**: *the one to whom something is sent*
- **Arg1**: *the thing that is sent*

Roleset id: Beja.04 , *to make someone send something to someplace*
- **Argc**: *the causer- the one who makes someone send something*
- **Arga**: *the intermediate causer*
- **Arg0**: *the agent- the one who sends something*
- **Arg2-gol**: *the place where something is sent*
- **Arg1**: *the thing that is sent*

Abstract Meaning Representation (AMR)

- Example: “The boy wants to go”
- AMR concepts are
  - English words (“boy”),
  - PropBank framesets (“want-01”), or
  - special key-words.
- Keywords include special entity types(“date-entity”, “world-region”, etc.), quantities(“monetary-quantity”, “distance-quantity”, etc.)
- logical conjunctions (“and”, etc).
- AMR uses approximately 100 relations

Source: Abstract Meaning Representation for Sembanking, https://amr.isi.edu/a.pdf

LOGIC format:

∃ w, b, g:
instance(w, want-01) ∧ instance(g, go-01) ∧
instance(b, boy) ∧ arg0(w, b) ∧
arg1(w, g) ∧ arg0(g, b)

AMR format (based on PENMAN):

```
(w / want-01
  :arg0 (b / boy)
  :arg1 (g / go-01
          :arg0 b))
```

GRAPH format:

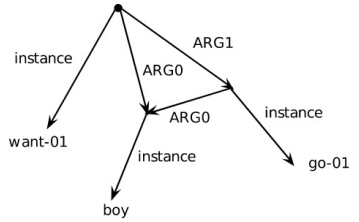
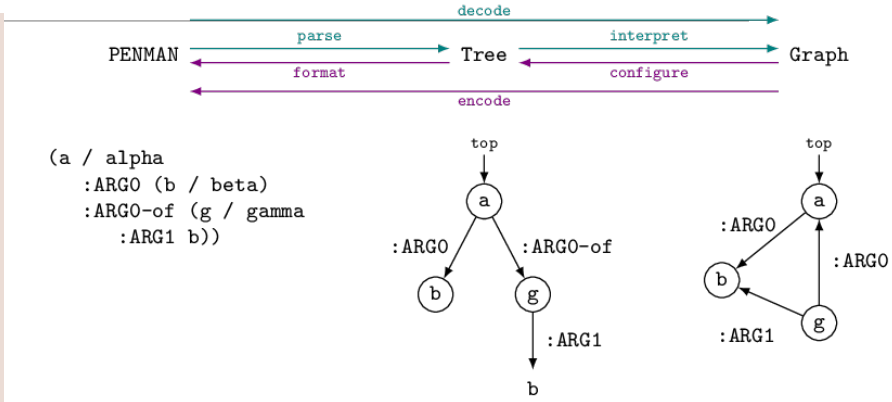


Figure 1: Equivalent formats for representating the meaning of “The boy wants to go”.

PENMAN Notation

```
;  ┌──────────────────────────── Variable (this one is the graph's top)
;  │   ┌──────────────────────── Instance relation
;  ┴ ____┴_____
   (d / drive-01
;    ┬ ──┬────
;    │   └──────────────────── Concept (node label)
;    └──────────────────────── Indicates the node's concept
;        ┌──────────────────── Edge relation
;    ______┴______
     :ARG0 (h / he)
;    ──┬──
;      └──────────────────────── Role (edge label)
     :manner (c / care-04
;               ┌──────── Attribute relation
;           ─────┴─────
            :polarity -))
;                  ┬
;                  └── Atom (or "constant")
```

Credit: https://penman.readthedocs.io/en/latest/notation.html



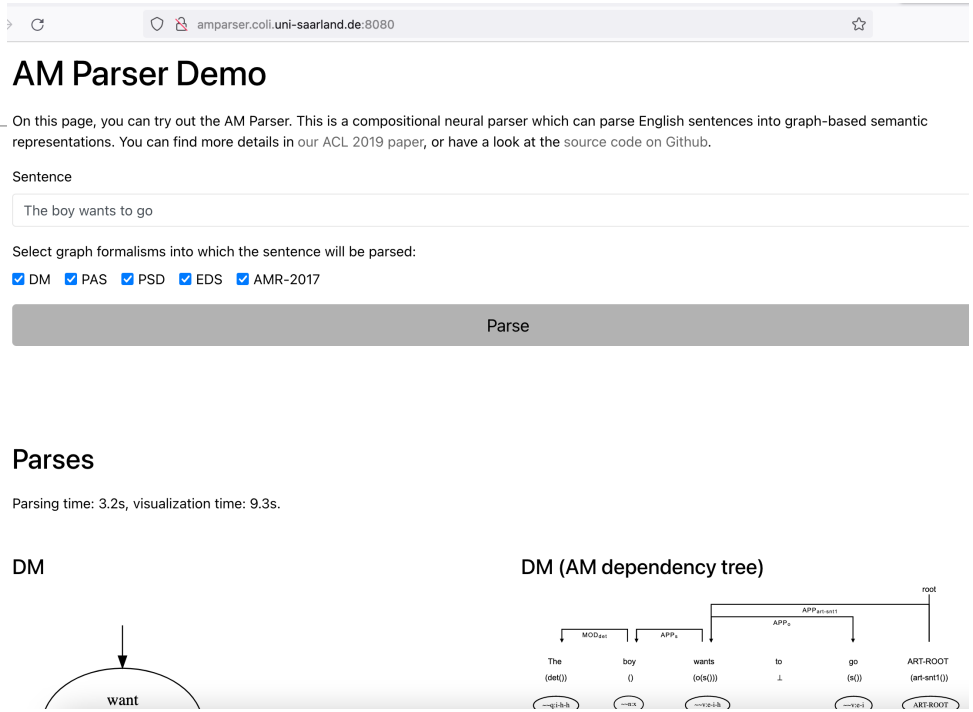
Credit: https://penman.readthedocs.io/en/latest/structures.html

Sample Code – PENMAN/ AMR

Sample code:
https://github.com/biplav-s/course-nl-f22/blob/main/sample-code/l9-semantics/PENMAN%20Notation%20-%20AMR.ipynb

AMR Demo

http://amparser.coli.uni-saarland.de:8080/



Exercise: 5 mins

- Try your sentences online
- Look at output in different formats

Semantic Parsing

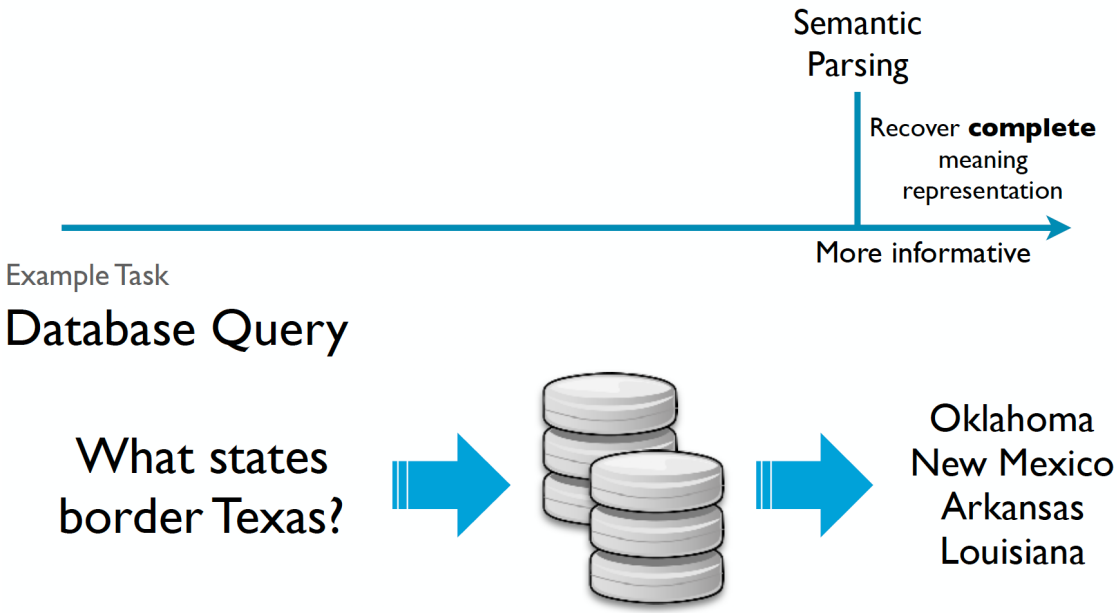
- Shallow semantic parsing
  - Also called: slot-filling or frame semantic parsing
  - "show me flights from Boston to Dallas"
- Deep semantic parsing
  - “show me flights from Boston to anywhere that has flights to Dallas”
  - Reference to quantifiers

Applications
- Paraphrasing
- Machine comprehension
- Question-answering
- Dialog

References:
- ACL 2020 Tutorial on Semantic Parsing
- https://en.wikipedia.org/wiki/Semantic_parsing

Semantic Parsing

Language to Meaning



Task-specific parsing

Source:
ACL 2020 Tutorial on Semantic Parsing

Resources: Semantic Parsing Libraries

- Open Sesame
  - Given English sentence, predicts FrameNet frames
  - https://github.com/swabhs/open-sesame

- AMRLib
  - Python library for AMR parsing, generation and visualization simple
  - https://github.com/bjascob/amrlib

Review: Lexical Meaning – Common Terms

- **Synonym**: same/ similar meaning
  - start-begin, finish-end, far-distant
- **Antonym**: opposite meaning
  - Far – near, clever - stupid, high - low, big – small
- **Homonym**: identical in spelling and pronunciation
  - bear, bank, ...
- **Homophones**: sounds identical but are written differently
  - site-sight, piece-peace.
- **Homograph**: written identically but sound differently
  - Potato, tomato, lead, wind, minute
- **Polysemy**: a word or phrase which has two (or more) different meanings (i.e., senses)
  - Duck, sharp

Source: Mausam

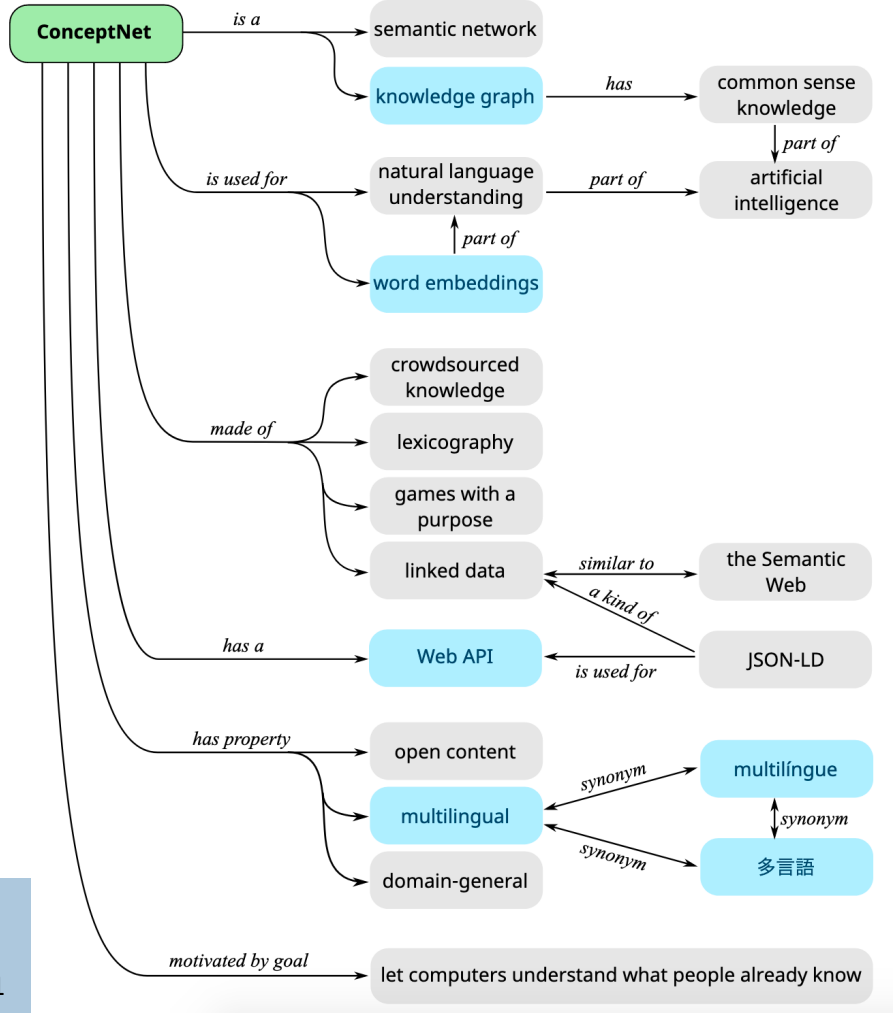
More Terms

- **Affective meanings** or **connotation**: word's meaning that are related to a writer or reader's emotions, sentiment, opinions, or evaluations
  - Positive evaluation: good, happy
  - Negative evaluation:
- **Sentiment**: Positive or negative evaluation expressed through language
  - Scherer's Typology of Affective States

Source: Jurafsky & Martin

ConceptNet

- NLP focused graph knowledge graph that connects words and phrases of natural language with labeled edges.
- Concepts collected from experts, crowd-sourcing, and games with a purpose
- Supports multiple languages
- Provides ”loose” semantics - relatedness



Details: http://conceptnet.io/, https://github.com/commonsense/conceptnet5/wiki,
Paper: https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/viewFile/14972/14051

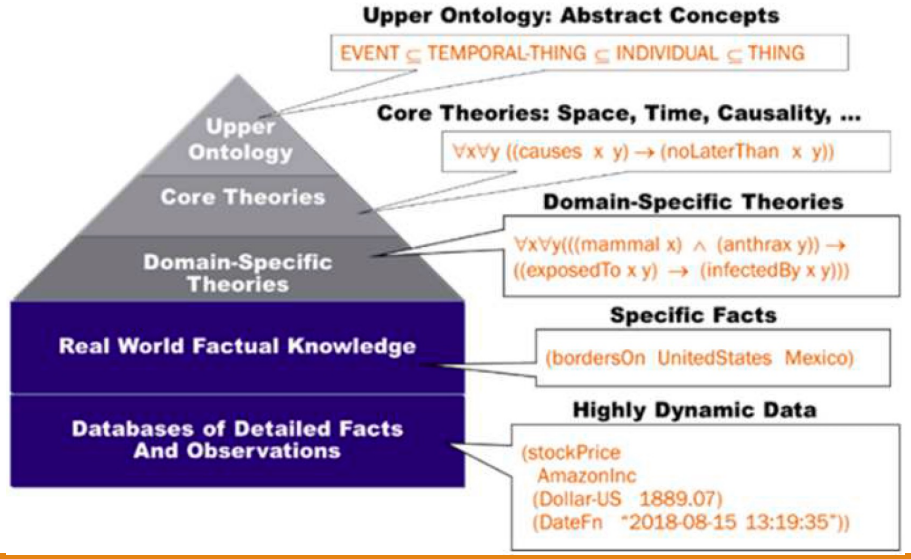
Demonstration - ConceptNet

Examples:

- Concepts:
  - Word: http://conceptnet.io/c/en/word ,
  - duck: http://conceptnet.io/c/en/duck
- Relationships:
  - http://conceptnet.io/s/resource/wordnet/rdf/3.1

Project CYC

- A large ontology to capture the world and human common sense
  - Doug Lenat lead team of computer scientists, computational linguists, philosophers, and logicians
  - Identify and formally axiomatize the tens of millions of rules about world
  - 35+ years effort by Cycorp
- Reasoners on the ontology to make decisions
  - 1000+ specialized reasoners



Details: https://www.cyc.com/

Source: Cyc White Paper

Cyc Details

- Ontology of about 1.5 million general concepts (e.g., taxonomically "placing" terms like eyes, sleep, night, person, unhappiness, hours, posture, being woken up, etc.);
- More than 25 million general rules and assertions involving those concepts
  - *"Most people sleep at night, for several hours at a time, lying down, with their eyes closed, they can be awakened by a loud noise but don't like that, "*
- Domain-specific extensions to the common sense ontology and knowledge base
  - healthcare, intelligence, defense, energy, transportation and financial services.
- Promoting synergistic use of ontology and learning based approaches (now)

Source: White Paper – Cyc Technology Overview

Lecture 9: Concluding Comments

- We reviewed how to give semantics to words and documents
- Can be human supervised or learning based or combined
- Can be generic or task-oriented

Concluding Segment

About Next Lecture – Lecture 10

Lecture 10 Outline

- Machine Learning for NLP
  - Supervised learning
  - Unsupervised learning
  - Neural methods