Machine Reading = Text Understanding
David Israel
What is it to Understand a Text?

understand a text — starting with a single sentence — is to:

• determine its truth (perhaps along with the evidence for its truth)

• determine its truth-conditions (roughly: what must the world be like for it to be true)

So one doesn’t have to check how the world actually is — that is whether the sentence is true

• calculate its entailments

• take — or at least determine — appropriate action in light of it (and some given preferences/goals/desires)

• translate it accurately into another language

• and what does accuracy come to?

• ground it in a cognitively plausible conceptual space

......

These are not necessarily competing alternatives!
or Contrast: What is it to Engage Intelligently in a Dialogue?

Respond appropriately to what your dialogue partner has just said (and done)

Typically, taking into account the overall purpose(s) of the interaction and the current state of that interaction (the dialogue state)

Compare and contrast dialogue between “equals” and between a human and a computer, trying to assist the human

- Siri is a very simple example of the latter
**A Little Ancient History of Dialogue Systems**

**ENT (Bobrow 1964)**

**Question-Answering: Word problems in simple algebra**

"If the number of customers Tom gets is twice the square of 20% of the number of advertisements he runs, and the number of advertisements he runs is 45, what is the number of customers Tom gets?"

**Overview of the method**

1. Map referential expressions to variables.

2. Use regular expression templates to identify and transform mathematical expressions.

3. Let the computer solve the resulting transformed set of statements using a kind of constraint satisfaction."
Eliza (Weizenbaum 1966)

1 User: You are like my father in some ways.

2 ELIZA: What resemblance do you see?

3 User: You are not very aggressive but I think you don’t want me to notice that.

4 ELIZA: What makes you think I am not aggressive?

5 User: You don’t argue with me.

6 ELIZA: Why do you think I don’t argue with you?

7 User: You are afraid of me.

8 ELIZA: Does it please you to believe I am afraid of you?

The user’s reply by (i) swapping 1st and 2nd person, (ii) interpolating stock phrases, and (iii) using hand-written (heuristically trained) scores to rank possible transformations.
A Little More Ancient History

rdlu (Winograd 1972)

RDLU is a program for understanding natural language, written by Terry Winograd at the M.I.T. Artificial Intelligence Laboratory in 1968-70. SHRDLU carried on a simple dialog (via teletype) with the user, about a small world of objects (the BLOCKS world) shown on an early display screen (DEC-340 attached to a PDP-6 computer).
Person: PICK UP A BIG RED BLOCK.

Computer: OK. (does it)

Person: GRASP THE PYRAMID.

Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.

Person: FIND A BLOCK WHICH IS TALLER THAN THE ONE YOU ARE HOLDING AND PUT IT INTO THE BOX.

Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.

Computer: OK. (does it)

Person: WHAT DOES THE BOX CONTAIN?

Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.

Person: WHAT IS THE PYRAMID SUPPORTED BY?

Computer: THE BOX.

Person: CAN THE TABLE PICK UP BLOCKS?

Computer: NO.

Person: CAN A PYRAMID BE SUPPORTED BY A BLOCK?

Computer: YES.

Person: CAN A PYRAMID SUPPORT A PYRAMID?

Computer: I DON'T KNOW.

Person: STACK UP TWO PYRAMIDS.

Computer: I CAN'T.

Person: WILL YOU PLEASE STACK UP BOTH OF THE RED BLOCKS AND EITHER A GREEN CUBE OR A PYRAMID?

Computer: OK. (does it)

Person: WHICH CUBE IS SITTING ON THE TABLE?

Computer: THE LARGE GREEN ONE WHICH SUPPORTS THE RED PYRAMID.

Person: IS THERE A LARGE BLOCK BEHIND A PYRAMID?

Computer: YES, THREE OF THEM: A LARGE RED ONE, A LARGE GREEN CUBE, AND THE BLUE ONE.

Person: PUT A SMALL ONE ONTO THE GREEN CUBE WHICH SUPPORTS A PYRAMID.

Computer: OK. (does it)

Person: PUT THE LITTLEST PYRAMID ON TOP OF IT.

Computer: OK. (does it)

Person: DOES THE SHORTEST THING THE TALLEST PYRAMID'S SUPPORT SUPPORTS SUPPORT ANYTHING GREEN?

Computer: YES, THE GREEN PYRAMID.

Person: WHAT COLOR IS IT?

Computer: BY "IT", I ASSUME YOU MEAN THE SHORTEST THING THE TALLEST PYRAMID'S SUPPORT SUPPORTS.

Computer: RED

Person: HOW MANY THINGS ARE ON TOP OF GREEN CUBES?

Computer: I'M NOT SURE WHAT YOU MEAN BY "ON TOP OF " IN THE PHRASE "ON TOP OF GREEN CUBES ". YOU MEAN:

- DIRECTLY ON THE SURFACE
- ANYWHERE ON TOP OF?

Person: 2

Computer: THREE OF THEM.

Person: HAD YOU TOUCHED ANY PYRAMID BEFORE YOU PUT THE GREEN ONE ON THE LITTLE CUBE?

Computer: YES, THE GREEN ONE.
Three Levels of Semantics

Lexical Semantics

The meanings of words

Sentential/compositional Semantics

How meanings of sentences are formed from meanings of words plus (syntactic) structure

Discourse-level semantics

How meanings of texts (multi-sentence texts) is formed from the meanings of constituent sentences plus … other facts about the context
Lexical Semantics

Words: senses and lexical semantic relations

WordNet and other lexical resources

Word similarity: thesaurus-based measures

Word similarity: distributional measures
Senses of Words

Unit of word meaning is a (word) sense

Usually represented explicitly in a “dictionary definition”

Most (non-rare) words have multiple senses/meanings

**Bank** is a financial intermediary that creates credit by lending money to a borrower, thereby creating a corresponding deposit on the bank’s balance sheet.

In geography, a **bank** generally refers to the land alongside a body of water.
Some Terminology

A **lexeme** is a pair of a word (**lemma** or **citation form**) and a sense or meaning

- *bank* is the lemma/citation form for *banks*
- *<bank, a financial institution>*

Specific surface forms (*bank, banks*) are **word-forms**
Homonymy

- Homonyms are lexemes that share a word-form
  - Orthographic, phonological or both
    - Homographs (*bank* and *bank*)
    - Homophones (*write* and *right*, *piece* and *peace*)

But have unrelated meanings
  - *<bank*, financial institution*> and *<bank*, riverside*>
Homonymy as Trouble

Speech Recognition (Speech-to-Text)

Text-to-Speech

Machine Translation

Information Retrieval (written or spoken queries)
Lexical Relation: Related Senses

bank was constructed in 1875 out of local limestone.

I drew the money from the bank.

Are these the same sense?

e 1: “The building belonging to a financial institution”

e 2: “A financial institution”

Consider this example:

Heavy snow caused the roof of the school to collapse.

School hired more teachers this year than ever before.
Polysemy: Systematically (?) Related Senses

Polysemy: multiple _related_ senses

, organizations and buildings housing (or otherwise associated with) them:

*school, university, hospital, church,…*

Authors (creators) and their works:

*Shakespeare died in 1616. I read Shakespeare every year.*

Animals and the meats/food derived from them:

*The chicken was domesticated in Asia. The chicken was overcooked.*

Trees and Fruit derived from them:

*Plums have beautiful blossoms. I ate some yummy plums yesterday.*
Distinguishing Multiple Senses

Do we know when a word has more than one sense?

Do flights serve breakfast?
Does America West serve Philadelphia?

“zeugma” test:

Does United serve breakfast and San Jose?

If this sounds weird, we say that these are two different senses of serve.
Lexical Relations

Synonymy

Antonymy

Hyponymy and Hypernymy

Meronymy and Holonymy

For all this and more, check out WordNet

http://wordnet.princeton.edu/
Intuitively: two words are synonymous when they have the same meaning.

- filbert/hazelnut
- automobile/car
- water/H$_2$O
- big/large

More formally: $w_1$ and $w_2$ are synonymous when they can be substituted, one for the other, in sentences without changing … the propositional meaning of the sentences.

- "car" is a one-syllable word
- "automobile" is a one-syllable word

Use vs mention

- But how about: Giovanni believes that H$_2$O boils at 100C. Giovanni believes that the boiling point of water is 100C. But what if Giovanni doesn’t know/believe that water is H$_2$O?

- what is our model of propositional meaning?

- truth-conditions?

- How are these represented?
Synonymy: As a Relation between Senses

big and large

Daniele was like a big brother to David.

Daniele was like a large brother to David.

big has a sense that mean being older or more grown up

large lacks this sense

To be synonyms must words share all their senses/meanings in common?

Surely not; for if so, there are probably no synonyms in any natural language.
synonymy as an equivalence relation

$s_1$ is synonymous with $s_1$ (Reflexivity)

If $s_1$ is synonymous with $s_2$, then $s_2$ is synonymous with $s_1$ (Symmetry)

If $s_1$ is synonymous with $s_2$ and $s_2$ with $s_3$, then $s_1$ is synonymous with $s_3$. (Transitivity)
Antonyms: words that — in at least one of their senses — have opposing contrary meanings

- loud/soft (pertaining to volume)
- hot/cold (pertaining to temperature)
- up/down (pertaining to position or direction on …the vertical axis)
- living/dead (binary opposition)
- Beware vampires and zombies!
- rise/fall

How about e.g., red/green or red/blue ??
Hyponyms and Hypernyms

- Hyponymy: \( \mathbf{X} \) is a hyponym of \( \mathbf{Y} \) if it denotes \( (?) \) a sub-class/sub-set of \( \mathbf{Y} \) or if it means/connotes \( (?) \) a concept that is included in the \( \mathbf{Y} \)-concept

  - asymmetric, transitive relation between senses

  - E.g. dog/mammal ; mango/fruit

- Hypernymy: the inverse relation

  - mammal/dog ; fruit/mango
Instances vs. Subclasses

Is Earth a hyponym of planet?

It is certainly an instance

And its singleton \{Earth\} is a sub-class of the class (set) of planets \{Mercury, Venus, Earth, ...\}

In my lingo: No! Instances are not sub-classes
Meronyms & Holonyms

Meronymy: X is a meronym of Y if it denotes a part of Y
asymmetric, transitive relation between senses

wheel/car

nose/face

Holonomy: its inverse

wheel/car

face/nose
Meronym subtypes

Part vs. Substance vs. Member meronyms

- Part: nose/face ; porch/house
- Substance: rubber/tire ; water/ocean
- Member: professor/faculty ; player/team
WordNet

http://wordnet.princeton.edu/

Via web interface:

http://wordnetweb.princeton.edu/perl/webwn

Hierarchically organized lexical database

machine-readable thesaurus (+ dictionary-like features)

Versions for other languages under development ....

Italian: Istituto di Linguistica Computazionale; C.N.R., Pisa

visit: http://www.illc.uva.nl/EuroWordNet/
Synsets

A synset is a set of synonyms representing a (single) sense

Each such sense — every member of the synset — shares the same gloss (the material in parens)

branch" (n)

branch, subdivision, arm (a division of some larger or more complex organization) "a branch of Congress"; "botany is a branch of biology"; "the Germanic branch of Indo-European languages"

branch, leg, ramification (a part of a forked or branching shape) 'broke off one of the branches"

branch (a stream or river connected to a larger one)
Word Similarity