Computational Semantics and Pragmatics

Autumn 2012

Raquel Fernández
Institute for Logic, Language & Computation
University of Amsterdam

Computational Semantics and Pragmatics

Lecturer: Raquel Fernández, <raquel.fernandez@uva.nl>

Timetable: Wed and Fri 11-13:00, room G3.13

Website: Slides, references, homework and other important information will be posted on the course's website:

http://www.illc.uva.nl/~raquel/teaching/cosp/cosp2012/

Prerequisites:

- some basic knowledge of semantics and pragmatics
- interest in computational methods of enquiry and evaluation
- there will be some programming, but programming skills are not required

⇒ *Please fill in the intake questionnaire if you have not yet done so.*

Evaluation

- Your grade will be based on:
  - regular homework exercises (min. 75% of overall grade)
  - readings and discussion of readings in class
  - occasional presentation

- Possible research project as a follow up to the course.

List of Topics as on Website

- Compositional semantics with functional programming
- Textual entailment
- Lexical semantics
  - psycholinguistic approaches to word meaning
  - computational representation and disambiguation of word senses
  - distributional semantics models
- Pragmatic inference and abductive reasoning
- Speech acts and dialogue modelling
- Generation of referring expressions

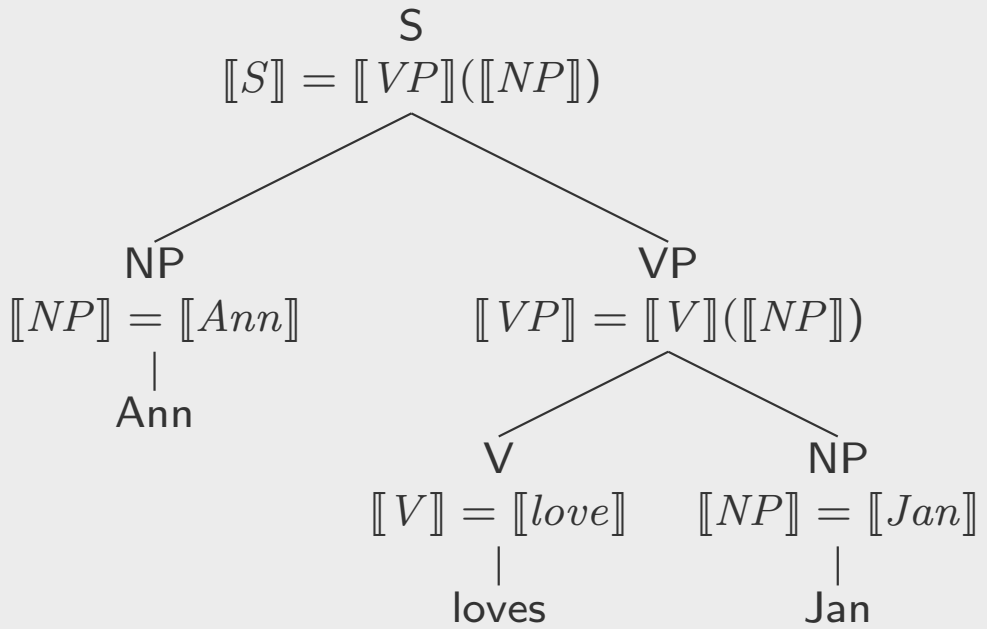
⇒ *The list and the order of the topics are tentative*

Compositional Semantics with FP
or
Computational Formal Semantics

Formal Semantics

- Contemporary formal semantics is based on the work of Montague
  - *English as a Formal Language* (1970)
  - *Universal Grammar* (1970)
  - *The Proper Treatment of Quantification in Ordinary English* (1974)
- Focus on compositional semantics $\approx$ the computation of propositional meaning at the sentence level.

$[\![Ann]\!] = a$
$[\![Jan]\!] = j$
$[\![love]\!] = \lambda xy.Love(x, y)$



- Precise and explicit (computational) interpretation algorithms.

Computational Formal Semantics

- Computational counterpart of formal semantics: automatic computation of semantic representations
- What do we gain from computational formal semantics?
  - possibility to reason automatically reasoning with the computed representations
  - from paper-and-pencil work to precise implementation that can rapidly compute the predictions of a theory
  - complement to formal semantics: implemented programs can give insights on how to refine and improve a theory
  - van Eijck and Unger: *"Implementing a rule system forces the linguist to be fully precise about the rules he or she proposes. You will find that once you are well-versed on functional programming, your programming efforts will give you immediate feedback on your linguistic theories."*
- Note that logic-based computational formal semantics is compatible with statistical approaches (probabilistic parsers).

Computational Formal Semantics

- Jan van Eijck and Christina Unger, *Computational Semantics with Functional Programming*.
- Two guest lectures by Jan van Eijck: 2 and 6 November.

$\Rightarrow$ read the first 3 chapters as preparation for Friday

Lexical Semantics

Compositional vs. Lexical Semantics

Formal compositional semantics employs a rather crude notion of lexical meaning:

$[\![dolphin]\!] = \{x \mid x \text{ is a dolphin}\}$	$f : D \rightarrow \{1, 0\}$	$\langle e, t \rangle$
$[\![envy]\!] = \{\langle x, y \rangle \mid x \text{ envies } y\}$	$f : D \rightarrow (D \rightarrow \{1, 0\})$	$\langle e, \langle e, t \rangle \rangle$

- Focus of formal semantics: how the truth-conditional meaning of sentences is compositionally built from the semantic value of basic expressions.
- Words are considered "basic expressions" associated with an entity, a property, or a relation in the world.

Compositional vs. Lexical Semantics

Dolphins **are** mammals, **not** fish. They **are** warm blooded like man, **and** give birth to **one** calf at **a** time. At birth **a** bottlenose dolphin calf **is** about 90-130 cms long **and** will grow to approx. 4 metres, living up to 40 years.

Function words (closed class)
- connectives and quantifiers
- copula, auxiliary and modal verbs
- temporal and modal adverbials
- pronouns, articles, degree modifiers...

Content words (open class)
- nouns
- adjectives
- verbs

$$\forall d(\text{dolphin}(d) \rightarrow \text{mammal}(d) \wedge \neg\text{fish}(d))$$
$$\forall d(\text{dolphin}(d) \rightarrow \forall xyt(\text{givebirth}(d,x,t) \wedge \text{givebirth}(d,y,t) \rightarrow x = y))$$

- Compositional semantics focuses on those function words that constitute the *glue* required for composition.
- But not a lot of emphasis is put on content words...

The Meaning of Words

Lexical semantics is about word meaning.

The relation between *word form* and *word meaning* is not one-to-one:

- Several words can have the same meaning → synonymy
  - *'buy'* / *'purchase'*
  - *'car'* / *'automobile'*
- One word can mean different things → homonymy/polysemy
  - *'bank'*$_1$: the slope of land adjoining a body of water
  - *'bank'*$_2$: a business establishment in which money is kept

The notion of word sense is used to refer to the concept expressed by a word form.

The Meaning of Words: Main Issues

1. What are word senses really? How can we represent them?
2. When there is lexical ambiguity (1 form, more than one sense) how do we disambiguate?

Issue 1:

- Psychological theories of concepts/categories and word meaning
  - classic definitional approach
  - prototype theory
  - exemplar-based theories
- Computational representations of lexical meaning
  - dictionary-like representation, e.g. WordNet
  - distributional semantic models

Distributional Semantic Models

Distributional Semantic or Vector Space Models:

- take a *usage-based* view of word meaning.
- Their basic underlying idea is that word meaning depends on the contexts in which words are used.
- An example by Stefan Evert: what's the meaning of *'bardiwac'*?

  * He handed her her glass of bardiwac.
  * Beef dishes are made to complement the bardiwacs.
  * Nigel staggered to his feet, face flushed from too much bardiwac.
  * Malbec, one of the lesser-known bardiwac grapes, responds well to Australia's sunshine.
  * I dined on bread and cheese and this excellent bardiwac.
  * The drinks were delicious: blood-red bardiwac as well as light, sweet Rhenish.

  $\Rightarrow$ *'bardiwac'* is a heavy red alcoholic beverage made from grapes

The Distributional Hypothesis

- DH: The degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear (Harris, 1954)
- DSMs make use of mathematical and computational techniques to turn the informal DH into empirically testable semantic models.
- Contextual semantic representations from data about language usage: an abstraction over the linguistic contexts in which a word is encountered.

	see	use	hear	...
boat	39	23	4	...
cat	58	4	4	...
dog	83	10	42	...

$\Rightarrow$ We will study the philosophical ideas behind these models and the computational techniques currently used to build them.

Issue 2: WSD

Word Sense Disambiguation (WSD) is the task of determining which sense of a word is being used in a particular context.

- supervised vs. unsupervised methods

Textual Entailment

Grasping Meaning: Inference

A necessary condition for natural language understanding is the ability to recognise entailment and contradiction.

- If you understand these sentences, you can recognise that (1) and (2) are contradictory ...

(1) No civilians were killed in the Najaf suicide bombing.
(2) Two civilians died in the Najaf suicide bombing.

- ... and that if (3) is true then (4) is true as well.

(3) Apple filed a lawsuit against Samsung for patent violation.
(4) Samsung has been sued by Apple.

Recognising whether entailment holds is a core aspect of our ability to grasp meaning.

Recognising Textual Entailment

Textual Entailment is a notion broader than logical entailment defined by the computational linguistics community as follows:

Textual entailment is a relation that holds between a pair $\langle T, H \rangle$ of natural language expressions (a *text* and a *hypothesis*), such that a human who reads (and trusts) T would infer that H is most likely true.

T	H	TE
Eyeing the huge market potential, currently led by Google, Yahoo took over search company Overture Services Inc last year.	Yahoo bought Overture.	✓
Since its formation in 1948, Israel fought many wars with neighboring Arab countries.	Israel was established in 1948.	✓
The National Institute for Psychobiology in Israel was established in May 1971 as the Israel Center for Psychobiology by Prof. Joel.	Israel was established in May 1971.	×
Arabic is used densely across North Africa and from the Eastern Mediterranean to the Philippines, as the key language of the Arab world and the primary vehicle of Islam.	Arabic is the primary language of the Philippines.	×

ACL RTE Portal:

http://aclweb.org/aclwiki/index.php?title=Textual_Entailment

Approaches to RTE

RTE can be seen as an abstract generic ability that captures inferential/semantic capabilities required by many tasks involving understanding.

$\Rightarrow$ How can we model this ability computationally?

Different types of approaches:

- Logic-based approaches
  - map expressions to logical representations and check logical entailment
  - automatic reasoning tools: theorem provers and model builders
- Shallower approaches
  - surface string features, e.g. string edit distance
  - syntactic similarity
  - semantic similarity

These approaches may be combined by using machine learning and treating RTE as a classification problem.

Pragmatic Inference or Implicature

Gricean Pragmatics

When we use language, we very often mean more than what we literally say:

(5) A: Are you going to Paul's party?
B: I have to work.
$\rightsquigarrow$ *I am not going.*

- B *implies* that she's not going to the party *without saying it.*
- This enrichment of the literal meaning is not a logical implication or entailment of B's utterance – it depends on features of the conversational context $\longrightarrow$ *conversational implicature*
- Grice proposes that conversational implicatures can be systematically accounted for by a set of general rationality principles for the efficient and effective use of language in conversation.

The CP and the Maxims

The Cooperative Principle: Make your contribution such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

- Maxim of Quality: be truthful
  - Do not say what you believe to be false.
  - Do not say that for which you lack adequate evidence.
- Maxim of Quantity:
  - Make your contribution as informative as is required (for the current purposes of the exchange).
  - Do not make your contribution more informative than is required.
- Maxim of Relation: be relevant
- Maxim of Manner: be perspicuous.
  - Avoid obscurity of expression / Avoid ambiguity.
  - Be brief / Be orderly.

Grice's point is not that we adhere to these maxims on a superficial level, rather that we interpret utterances assuming that the principles are being followed at some deeper level, often contrary to appearances.

Computational Exploration of CI?

Grice's proposals were brief and only suggestive of how work on the underlying ideas may proceed.

Work has indeed proceeded in several directions:

- Formal pragmatics: neo-gricean approaches, relevance theory, ...
- Experimental pragmatics: what do speakers/hearers actually do?
  - pragmatics and cognition
  - experimental methods to test pragmatic theories
- Computational pragmatics: can we account computationally for phenomena related to conversational implicature?

Generation of Referring Expression

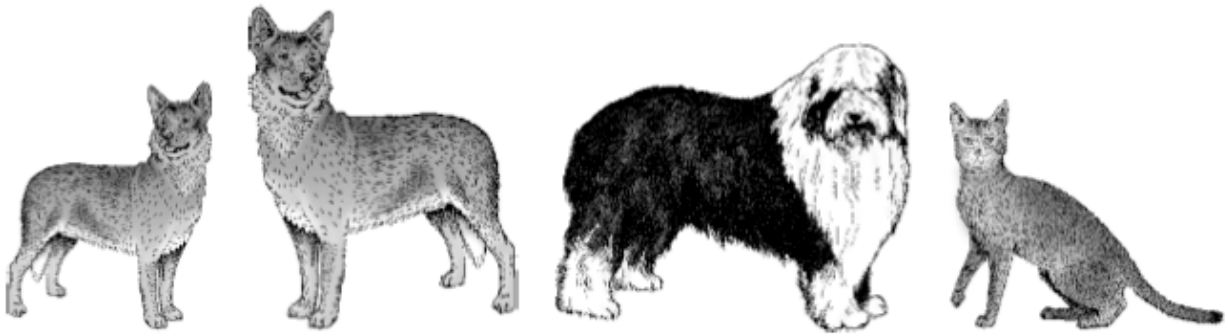
GRE is concerned with the production of linguistic expressions that enable the hearer to identify one or more entities in a given context.

Natural Language Generation is a subfield of CL/NLP. We can think of it as the reverse of the process of Natural Language Understanding (NLU):

- NLU: Mapping human language into non-linguistic representations.
- NLG: Mapping non-linguistic representations of information into human language.

GRE is an issue for NLG because the same entity may be referred to in many different ways.

- What constitutes "appropriate" language in a given communicative situation? How can the relevant pragmatic, semantic, syntactic, and psycholinguistic constraints be formalised?



$r = d_1$	Knowledge base representing the scene: d_1 : `type=dog, size=small, color=brown` d_2 : `type=dog, size=large, color=brown` d_3 : `type=dog, size=large, color=black+white` d_4 : `type=cat, size=small, color=brown`

Some examples of possible descriptions in this scenario:

content determination	possible realisation	distinguishing
$L = \{$`type=dog,size=small`$\}$	*'the small dog'*	✓
$L = \{$`type=dog,colour=brown`$\}$	*'the brown dog'*	×
$L = \{$`type=dog,size=small,colour=brown`$\}$	*'the small brown dog'*	✓

Speech acts and dialogue modelling

Conversation

Telephone conversation between two participants, Switchboard Corpus:

```
A.1:  Okay, {F um. } / How has it been this week for you? /
B.2:  Weather-wise, or otherwise? /
A.3:  Weather-wise. /
B.4:  Weather-wise. / Damp, cold, warm <laughter>. /
A.5:  <laughter> {F Oh, } no, / damp. /
B.6:  [ We have, + we have ] gone through, what might be called the four
        seasons, {F uh, } in the last week. /
A.7:  Uh-huh. /
B.8:  We have had highs of seventy-two, lows in the twenties. /
```

- Turns: stretches of speech by one speaker bounded by that speaker's silence – that is, bounded either by a pause in the dialogue or by speech by someone else.
- Utterances: units of speech delimited by prosodic boundaries (such as boundary tones or pauses) that form *intentional units* – that is, that can be analysed as an action performed with the intention of achieving something ($\longrightarrow$ *dialogue acts/speech acts*).

Dialogue Modelling

Intuitively, conversations are made up of sequences of actions (dialogue acts/speech acts) and turns.

- how can we derive the dialogue act performed by an utterance? computational models of the interpretation of speech acts
- content management: how can we account for the coherence of dialogue?
- interaction management: coordination between dialogue participants (feedback on the understanding process, turn-taking,...)

Relevant Local Seminars

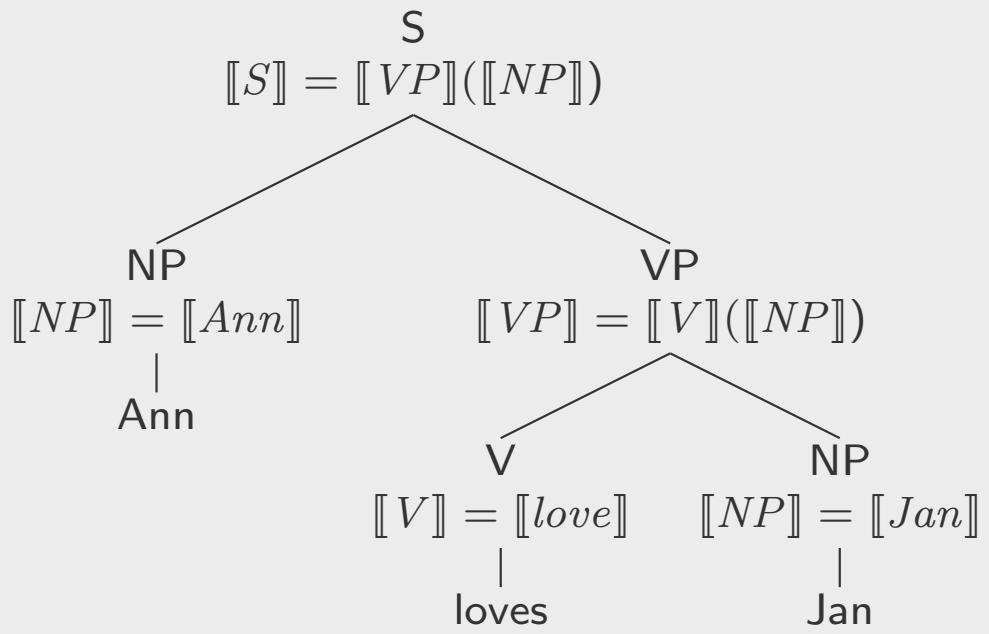
- Computational Linguistics Seminar (CLS)
  http://www.illc.uva.nl/LaCo/CLS/
- SMART lectures
  http://smartcognitivescience.wordpress.com/
- DIP (discourse processing) Colloquium
  http://sites.google.com/site/illcdip/

Learning Compositional Semantics

topic of today's CLS

Compositional semantics assumes a lexicon and a set of composition rules which tell us how to construct the meaning of complex expressions by systematically combining the meaning of words and phrases.

$[\![Ann]\!] = a$
$[\![Jan]\!] = j$
$[\![love]\!] = \lambda xy.Love(x, y)$



⇒ How can the semantic composition rules be learned?

Learning Compositional Semantics

Computational linguistics/ NLP:

- semantic interpretation is critical for many NLP applications
- corpora are often not annotated with semantic interpretations
- the right grammars are often not available
- can we use corpora to learn these rules?

Cognitive science:

- how do humans learn these rules?
- can we design a model that makes the same mistakes children make?

Example paper:

Piantadosi et al. (2008) A Bayesian Model of the Acquisition of Compositional Semantics. *Proc. of the 30th Annual Conference of the Cognitive Science Society*.