From Syntax to Natural Logic

Lauri Karttunen, Annie Zaenen
Stanford

Natural logic for natural language

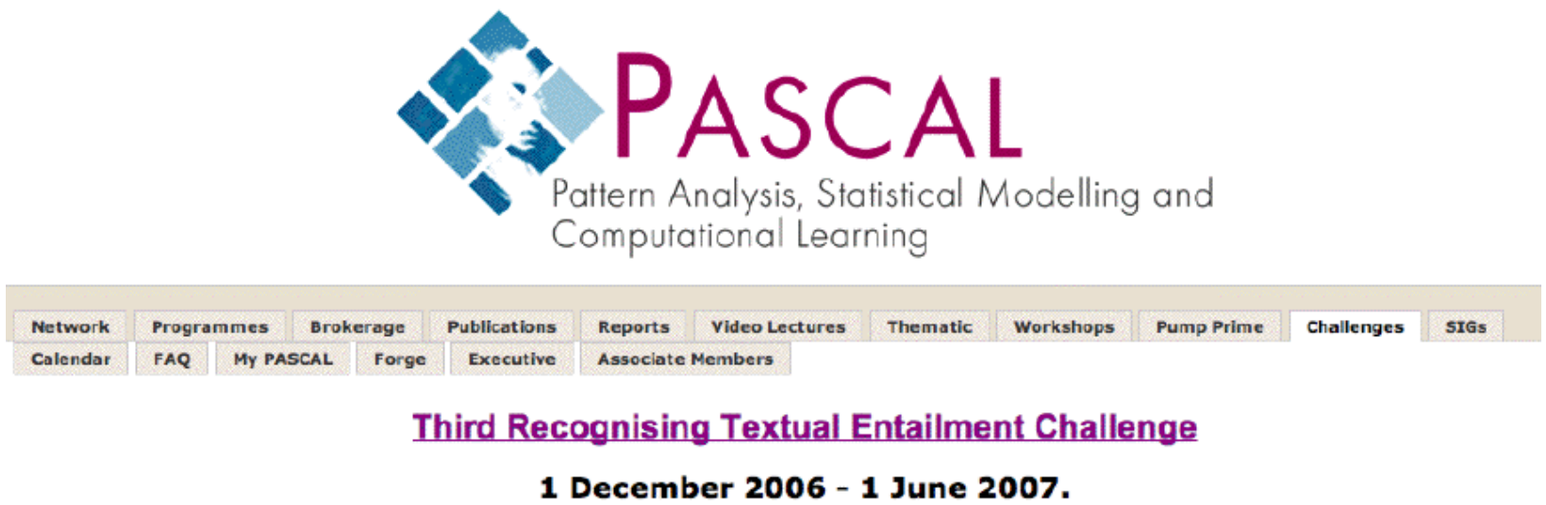
- Two dimensions
  - Proof theoretic derivations
  - Close to linguistic surface form

Natural Language Processing+ pressures

- NLP is moving away from concerns about search to Natural Language Understanding.
- Examples
  - PASCAL
  - Machine Reading

Pascal: textual entailment

- A measure of understanding a text is the ability to make inferences based on the information conveyed by it.



Ido Dagan, Oren Glickman, and Bernardo Magnini. 2006. The PASCAL Recognising Textual Entailment Challenge. In Quiñero-Candela et al., editors, MLCW 2005, Lecture Notes in Artificial Intelligence, Volume 3944. Springer-Verlag.

PASCAL task

PREMISE	Oswald killed Kennedy.
CONCLUSION	Kennedy died.
TRUE/FALSE	TRUE

Textual inference includes but goes beyond logical entailment and presupposition. It allows for background world knowledge (facts of geography, history, mathematics, physics), default reasoning (birds fly), conversational principles, statistical likelihood of situations and consequences of actions…

Participants are given a development set to help them prepare for the task typically 800 pairs, then a test set. Both are derived from naturally occurring data (newspapers and the like). The idea is that the test set and the development set are similar in nature but many factors are not controlled for.

The setup is geared to statistical learning methods but the knowledge needed includes clearly linguistic and lexical knowledge.

PASCAL notion of 'Entailment'

- Not a crisp notion:

  Romano Prodi will meet the US President George Bush in <u>his</u> capacity as the president of the European commission.

  George Bush is the president of the European commission.

  FALSE

  The real world knowledge that this relies on is that being US President excludes being the president of the European Commission. This knowledge excludes the anaphoric link between 'US President George Bush' and 'his'. But the grammar allows it.

- Can we make the notion crisper by sorting out 'real' entailment from more general inferencing?

PASCAL's notion of 'Entailment'

- There have been attempts to analyze the PASCAL data sets, to see what types of knowledge a system has to have to draw the right inferences.

- They are very varied in nature, from figuring out abbreviations to one step reasoning over syntactic equivalences. Here we are interested in the semantics/ reasoning side of things.

Fracas

- Dates from the 1990s, from before the overwhelming interest in search and from before the switch to statistical methods in NLP. It is more germane to our interests here but it has the disadvantage of being a made up set, no naturally occurring sentences.

Machine reading

- (D)ARPA programs
- sort of follow up on Acquaint
- Three big teams, five year program.
- “Machines will learn to read from a few examples and they will read to learn what they need in order to answer questions or perform some reasoning task. Three independent Reading Teams are building universal text engines which will capture knowledge from naturally occurring text and transform it into the formal representations used by Artificial Intelligence. An Evaluation Team is selecting and annotating text corpora with task domain concepts, creating model reasoning systems with which the reading systems will interact, and establishing question-answer sets and evaluation protocols to measure progress toward this goal.”
- ‘**Reading machine to snoop on Web**’

Question answering

- A long-standing goal of computational linguistics is to build a system for answering natural language questions.

  If the question is Did Shackleton reach the South Pole?, the sentence Shackleton failed to reach to the South Pole. contains the answer.

  None of the current search engines (Google, Yahoo!) is capable of delivering a simple NO answer in such cases. A successful question answering system has to recognize semantic relations within and between sentences.

Textual “Entailment ~ Machine Reading

- Textual entailment as define in PASCAL is very ‘local’: no long reasoning chains, just one or two steps.

- Machine Reading is more ambitious: a ‘real’ reasoning component is envisioned but At this point the project is mainly geared towards recognizing ‘events’ and relations between entitities.

What they have in common

- Not so much interested in the ‘meaning’ of each sentence than well in the inferential relation between (the meanings of) the textual fragments.

- Interest in partial success: we don’t need to get all the inferences.

  ‘The road meandered from Palo Alto to Menlo Park’ ➙ ‘The road goes from Palo Alto to Menlo Park.’

- At least part of the inferences can be traced back to the form of the sentences and the meaning of the words without needing to do much thinking about the world.

Mixing techniques

- Both RTE and MR envision using a variety of techniques. Many of them are machine learning techniques.
- We do not cover those in this course but note that they can have surprisingly good results: at first RTE employed bag-of-words approaches, e.g.
  - Several airlines polled saw costs grow more than expected
  - Some of the companies in the poll reported cost increases.
  - several ~ some; airlines ~ companies; polled ~ poll; cost ~ costs; grow ~ increases
- Whether one likes them or not they are here to stay. If one is interested in ‘applicable’ work one needs to take them into account.

Current formal semantics

- Most of natural language semantics aims to specify the meaning of linguistic expressions in model-theoretic terms
  - Model : set theory
  - Natural language expressions in their full glory need higher order logics
  - Interest in truth but not in consequences, if any reasoning is thought of it is supposed to take place on the set theoretical entities.
  - Interest in a couple of narrow classes of lexical items, e.g. quantifiers.

- Higher order logics do not come with efficient decision procedures, not suitable for applications

- Even first order logic is not decidable.

- Decidability: the existence of an effective method for determining membership in a set of formulas. Logical systems are decidable if membership in their set of logically valid formulas (or theorems) can be effectively determined.

- In computer science: attempts to get usuable systems that have lest power than first order logic: e.g. description logics

Some developments in logic

- Interest in “lower order” logics and proof theory
  - Johan van Benthem
  - Larry Moss
- Interest in logics that remain close to natural language expressions: classical logic reasoned with (regimented) natural language expressions

Natural Logic

- The term comes from the title of a paper by George Lakoff: Linguistics and Natural Logic, Synthese 22 (1970), unfortunately rather unreadable in the current context.

Two views on Natural Logic

Johan van Benthem:

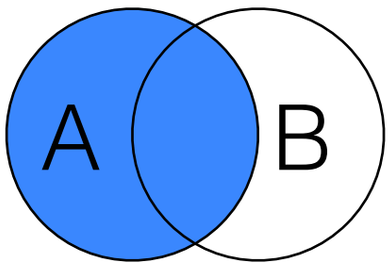
work on natural logic already in the 1980's and 90's.
The proposed ingredients of a logical system to satisfy his goals would consist of :

(a) Monotonicity Reasoning, i.e., Predicate Replacement,
(b) Conservativity, i.e., Predicate Restriction, and also
(c) Algebraic Laws for inferential features of specific lexical items.

Conservativity

All students work hard ↔ All students are students that work hard.

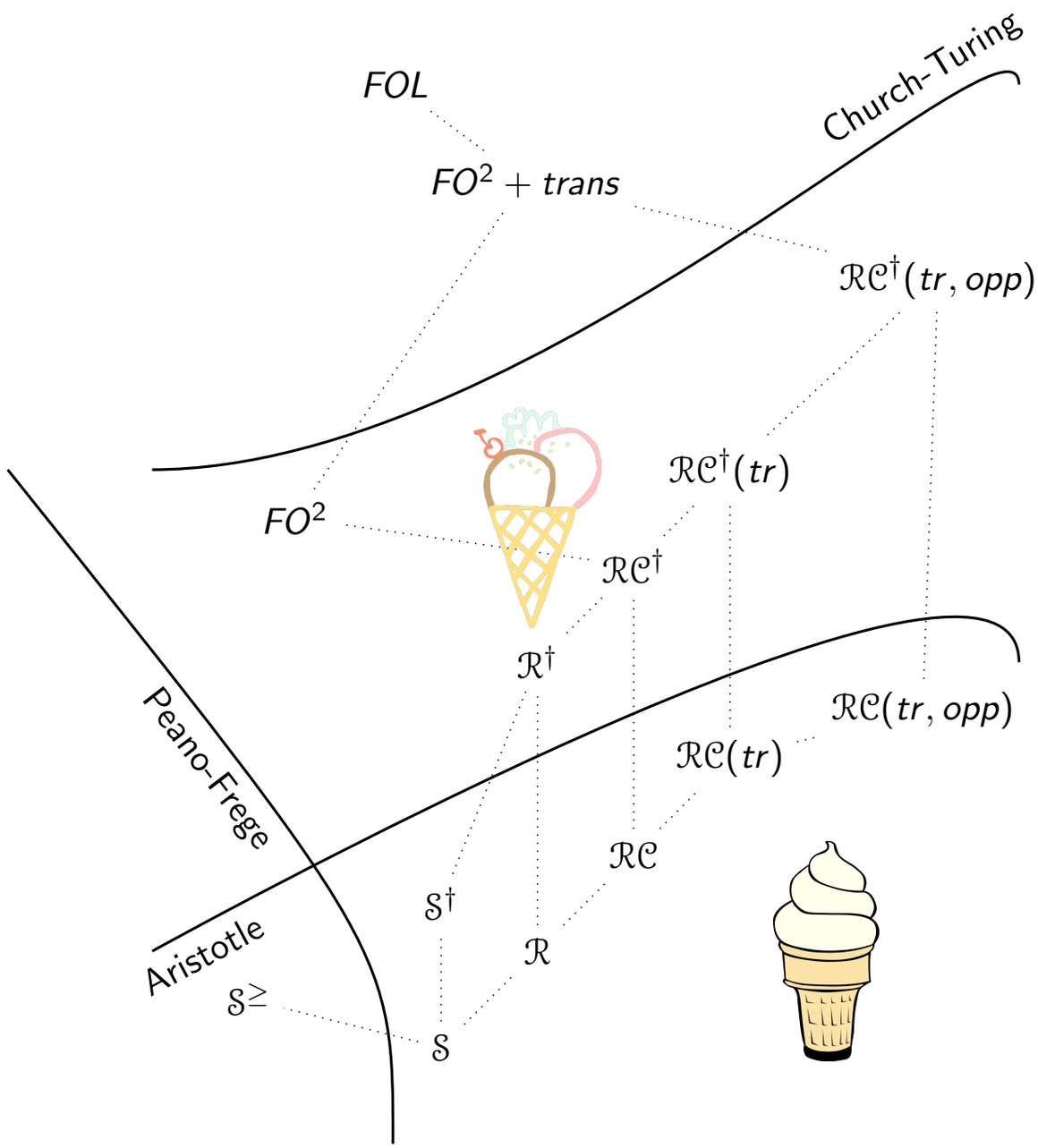
For all M and all A, B ⊆ M. $Q_M(A,B) \leftrightarrow Q_M(A, A \cap B)$



Larry Moss:

Re-think semantics based on computational linguistics.
Re-work the relation of logic and language, starting with inference.
First step: show that significant parts of natural language inference can be carried out in decidable logical systems whenever possible, to obtain complete axiomatizations, because the resulting logical systems are likely to be interesting.
To be completely mathematical and hence to work using all tools and to make connections to fields like complexity theory, (finite) model theory, proof theory, decidable fragments of first-order logic, and algebraic logic.
But these are all the first step, and they hardly touch upon the real goals.

Logics of the first two flavors



first-order logic

FO^2 + "R is trans"

2 variable FO logic

$\dagger$ adds full N-negation

$\mathcal{RC}(tr)$ + opposites
$\mathcal{RC}$ + (transitive) comparative adjs
$\mathcal{R}$ + relative clauses
$\mathcal{S}$ + full N-negation
$\mathcal{R}$ = relational syllogistic
$\mathcal{S}^{\geq}$ adds $|p| \geq |q|$
$\mathcal{S}$: all/some/no p are q

- We follow more the van Benthem line in this course but, please, go to Larry's talk tomorrow.

- Logicians like Larry Moss are interested in worst cases scenarios. The computational reality is that the worst case doesn't always happen.

- So decidablility results are not always a show stopper, although they are a reason to worry. Ultimately we need logicians to define an average case scenario, a 75% of the cases one, etc. as well as a worst case scenario.

Psychological argument

- It is also a truism that humans do not carry anything analogous to infinite sets of possible worlds or situations around in their heads, so the study of deduction -- inferential relations based on syntactic properties of some kind of “representations” of denotations -- are potentially of relevance to the psychology of language and to computational processing of meaning in a way that model-theoretic semantics alone is not.

Dowty: The Role of Negative Polarity and Concord Marking in Natural Language Reasoning

What the Moss and the Van Benthem approach both want

- a systematic relation between natural language syntax and logical form (best if they are the same)
- proof system based on that logical form (reasoning)

Proof theory for natural language semantics?

- Linear logic, versions of categorial grammar, Davidson
- We are not after complete semantic representations but after partial entailments from natural language sentences

Surface form?

- Not strings
- Phrase structure
- Categorial grammar
- Dependency structure
- ...

Examples of NL deductions

- A small bird was singing in the garden

  → a small bird was singing

  → a bird was singing in the garden

  → a bird was singing

- Not: a bird was singing →a small bird was singing

- No bird was singing
  - → No bird was singing in the garden
  - → No small bird was singing
  - → No small bird was singing in the garden
- NOT: no small bird was singing in the garden
  →no bird was singing

Temporal inferences

- Mary traveled to Athens in July 2001

  → Mary traveled to Athens in July

  → Mary traveled to Athens in 2001

- The deal lasted through August, until just before the government took over Freddie.

  → The deal lasted through August.

  → The deal lasted until just before the gov't took over Freddie.

  → The government took over after August.

Simple Davidsonian reasoning doesn't work

- Every boy bought a toy from Fred.

  → Every boy bought a toy.

- Last year, <u>in July</u>, he visited us every day.

  NOT: → Last year he visited us every day.

- Every boy bought a toy.

  NOT: → Every boy bought a toy <u>from Ed.</u>

- Last year he visited us every day.

  → Last year he visited us every day <u>in July</u>.

Spatial Inferences

- X is to the right of Y --> Y is to the left of X
- X is 20 miles from here, Y is 30 miles from here --> Y is farther from here than X
- Is this lexical knowledge or world knowledge?

Implicatives and Factives

- Shackleton failed to reach the South Pole. → Shackleton didn't reach the South Pole.
- John forgot to turn off the stove. → John didn't turn off the stove.
- John forgot that he turned off the stove. → John turned off the stove

Converses

- John is taller than Bill --> Bill is shorter than John
- Mary sold a book to Jane --> Jane bought a book from Mary

Outline of Course

today: Monotonicity

Tuesday, July 12
Syllogisms as instances of monotonicity reasoning and classical logic attempts to extend its scope.

Friday, July 15
Categorial Grammar (Lambek, van Benthem)

Tuesday, July 19
A categorial grammar treatment of monotonicity (Sánchez)

Friday, July 22
Beyond monotonicity (MacCartney)

Tuesday, July 26
Integrating natural logic into a working local inferencing system

Friday, July 29
Dependency grammar as a syntactic basis for a local inferencing system

Tuesday, August 2
Entailment and presupposition calculations in the Xerox Linguistic Environment (XLE)

- The validities of monadic predicate calculus with identity are decidable, however. This system is first-order logic restricted to signatures that have no function symbols and whose relation symbols other than equality never take more than one argument.

- More than half of the tenors have the flu. More than half of the tenors are sick.

Intuitively speaking, this inference is valid and obviously so. From a strictly logical point of view this is quite remarkable. On the one hand, it is a familiar fact that first-order logic cannot capture the meaning of the determiner ‘more than half’ (Barwise and Cooper 1981); a more powerful logic is needed for that. On the other hand, to the extent that it is decidable at all, first-order logic is known to computationally intractable, and therefore a more powerful logic can only be worse. So how do we humans manage to see that arguments like (the one above) are valid? The answer is fairly obvious: by trading completeness for efficiency. The inference strategies we employ cannot fail to be radically incomplete, but at least they are reasonably efficient.
One of the reasons why monotonicity inferences are important to us is that they are highly efficient.

... a system for producing monotonicity inferences can be very simple, because it requires only a shallow understanding of the representations it operates on. In the case of (above), for example, such a system would merely need to know that having the flu entails being sick and that ‘more than half’ is upward entailing; the exact meaning of ‘more than half’ is immaterial. That is why (it) is easy.

Bart Geurts and Frans van der Slik(2005)