Foundations of Language Science and Technology

Semantics 4

Manfred Pinkal
Saarland University

Overview

- Semantic Processing - Introduction
- Logic-based meaning representation and processing: Truth-conditional interpretation, entailment, deduction
- Word Meaning: Lexical-semantic resources, ontologies, similarity-based approaches
  - Informal overview
  - Semantic Relations, WordNet
  - Semantic similarity measures
  - Comparison
- Semantic Composition: Composing sentence and text meaning from word meaning
- Textual Entailment and Inference

WordNet vs. VSM Semantics

- WordNet
  - uses directed and truth-conditionally grounded semantic relations supporting both logical and similarity-based reasoning
  - decribes „paradigmatic" semantic relations only (i.e., concepts that are substitutable for each other, like *dolphin - whale - mammal*)
  - is in some sense more reliable/less noisy than VSM semantics, but varies in granularity of sense distinctions, classification steps
  - is an expensive, hand-crafted resource
- VSM models of lexical meaning
  - uses a symmetric concept of word similarity
  - supporting similarity-based approaches of information access, but no transparent relation to logic and truth-conditions
  - describes all kinds of similarity phenomena, including the „syntagmatic" phenomena of collocation, script- or scenario-based relatedness (e.g., dolphin - water - sea)
  - is obtained by unsupervised statistical methods, therefore inexpensive and easily to obtain for new languages and sub-languages.

WordNet plus VSM Semantics

- WordNet also contributes to similarity-based approaches of information access:
  - Query Expansion
  - WN similarity measures
- Note: WN similarity measures are symmetric (like corpus-based similarity measures), but they describe paradigmatic relations: (*lunch- dinner; starter - main dish - dessert*), not scenario-based ones (*dinner - waiter - menu - bill*). WN and distribution-based similarity measures complement each other.
- Corpus-based similarity measures can be used for resource extension, and thereby indirectly contribute to knowledge-based approaches: E.g., find nearest neighbour of a word (with respect to a sim relation) which is not in WordNet, and add it to the synset.

Neither WN nor VSM Semantics

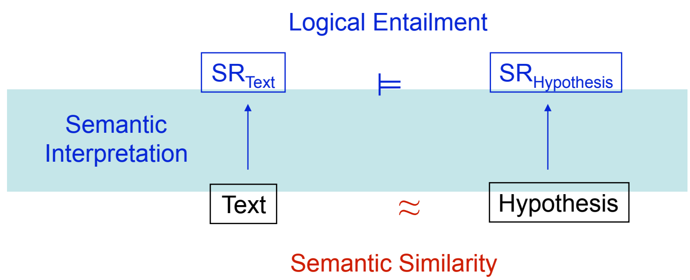
- … have a good answer to the problem of Word-Sense Disambiguation: How do we get from a word (token, occurrence of a word in a text or utterance) to its contextually appropriate sense?
- For offline computation of word similarity (e.g., through WordNet distance) and computation of document similarity often very simple heuristics is used:
  - Select the first WN sense (WN senses are ordered by frequency in WordNet)
  - Take all sentences to be equally probable, and compute the distance between words w1 and w2 as the arithmetic mean of the distances between all senses of w1 and w2, respectively.

WSD, Selectional Preferences

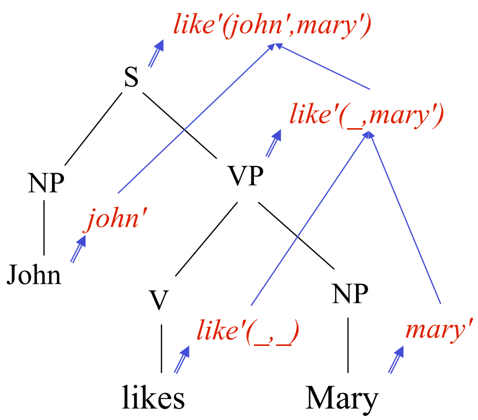
- For the online interpretation of individual NL expressions in context, we have to solve the WSD task, which is one of the hardest problems in NL interpretation.
- A partial answer comes with semantic composition, through so-called selectional constraints or selectional preferences, requirements on the semantic type of arguments. Example:
  - *They begin at ten* is ambiguous between two readings of the verb *begin*.
  - *The classes begin at ten* is disambiguated by the information that the subject denotes an event, not a person.

Overview

- Semantic Processing - Introduction
- Logic-based meaning representation and processing: Truth-conditional interpretation, entailment, deduction
- Word Meaning: Lexical-semantic resources, ontologies, similarity-based approaches
- Semantic Composition: Composing sentence and text meaning from word meaning
- Textual Entailment and Inference



Basic Semantic Composition



A Challenge for Semantic Composition

Every student presented a paper

$\forall d\ (student(d) \rightarrow \exists p\ (paper(p) \wedge present(d,p)))$

- In general, FOL semantic representations do not structurally correspond to the syntactic structure of NL sentences.
- How do we model the semantic composition process?
- We approach the problem via a detour: Looking into higher-order phenomena in NL semantics

FOL Limitations

John is a married piano player	piano-player(j) ∧ married(j)
John is a blond criminal	criminal(j) ∧ blond(j)
John is a poor piano player	piano-player(j) ∧ poor(j) ?
John is an alleged criminal	criminal(j) ∧ alleged(j) ???

FOL Limitations

Yesterday, it rained.
Probably, it is raining.
Unfortunately, it is raining.

Bill is blond. Blond is a hair colour.

$\not\models$ *Bill is a hair colour*

The Language of Type Theory

Types:

- The set of basic types is {e, t} :
  - e (for entity) is the type of individual terms
  - t (for truth value) is the type of formulas
- All pairs $\langle\sigma, \tau\rangle$ made up of (basic or complex) types σ, τ are types. $\langle\sigma, \tau\rangle$ is the type of functions which map arguments of type σ to values of type τ.
- In short: The set of types is the smallest set **T** such that e,t$\in$**T**, and if $\sigma,\tau \in$**T**, then also $\langle\sigma,\tau\rangle \in$**T**.

Some Useful Types for NL Semantics

- Proper name — bill: e
- Sentence — it_rains: t
- One-place predicate constant:
  work, student: <e,t>
- Two-place relation:
  like, larger_than: <e,<e,t>>
- Sentence adverbial:
  yesterday, unfortunately: <t,t>
- Attributive adjective:
  married, poor, alleged: <<e,t>,<e,t>>
- Degree modifier:
  very, relatively: <<<e,t>,<e,t>>,<<e,t>,<e,t>>>

Second-order predicates

Bill is blond. Blond is a hair colour.

bill: e   blond: <e,t>

blond(bill): t

Blond is a hair colour.

blond: <e,t>   hair_colour : <<e,t>,t>

hair_colour (blond): t

Bill is a hair colour

- Hair-colour is a second-order predicate. *hair_colour(bill)* is not even a well-formed expression.

Type-theoretic syntax

- Vocabulary:
  - A (possibly empty) set of constants: Con_τ, for every type τ
  - A set of variables: Var_τ, for every type τ
  - The usual FOL operators: connectives, quantifiers, equality

- The sets of well-formed expressions WE_τ for every type τ are given by:
  - $Con_\tau \cup Var_\tau \subseteq WE_\tau$ for every type τ
  - $\boxed{\text{If } \alpha \in WE_{<\sigma, \tau>}, \beta \in WE_\sigma \text{, then } \alpha(\beta) \in WE_\tau .}$
  - If A, B are in WE_t, then so are $\neg$ A, ($A \wedge B$), ($A \vee B$), ($A \rightarrow B$),($A \leftrightarrow B$)
  - If A is in WE_t, then so are $\forall v A$ and $\exists v A$, where *v* is a variable of arbitrary type.
  - If α, β are well-formed expressions of the same type, then $\alpha=\beta \in WE_t$.

Function Application

- The most important syntactic operation in type-theory is function application:

  If $\alpha \in WE_{<\sigma, \tau>}$, $\beta \in WE_\sigma$, then $\alpha(\beta) \in WE_\tau$.
- A functor of complex type combines with an appropriate argument to a (more complex) expression of less complex type.

Function Application: Examples

Bill drives fast

drive: <e,t>   fast: <<e,t>,<e,t>>

bill: e   fast(drive): <e,t>

fast(drive)(bill): t

Mary works in Saarbrücken

mary: e   work: <e,t>   in: <e,<t,t>>   sb: e

work(mary): t   in(sb): <t,t>

in(sb)(work(mary)): t

Using Higher-Order Variables

- *Bill has the same hair colour as John.*

  $\exists \mathcal{G}$ (hair_colour($\mathcal{G}$) $\wedge$ $\mathcal{G}$ (bill) $\wedge$ $\mathcal{G}$ (john))
- *Santa Claus has all the attributes of a sadist.*
- $\forall \mathcal{F} \forall a(\text{sadist}(a) \wedge \mathcal{F}(a) \rightarrow \mathcal{F}(b))$

Type-theoretic semantics [1]

- Let U be a non-empty set of entities. The domain of possible denotations D_τ for every type τ is given by:
  - $D_e = U$
  - $D_t = \{0,1\}$
  - $D_{<\sigma,\tau>}$ is the set of all functions from D_σ to D_τ
- A model structure for a type theoretic language:
  M = <U, V>, where
  - U (or U_M) is a non-empty domain of individuals
  - V (or V_M) is an interpretation function, which assigns to every member of Con_τ an element of D_τ.
- Variable assignment g assigns every variable of type τ a member of D_τ.

Type-theoretic semantics [2]

- Interpretation (with respect to model structure M and variable assignment g):

  $[[\alpha]]^{M,g} = V_M(\alpha)$, if α constant

  $[[\alpha]]^{M,g} = g(\alpha)$, if α variable

  $\boxed{[[\alpha(\beta)]]^{M,g} = [[\alpha]]^{M,g}([[\beta]]^{M,g})}$

  $[[\neg\varphi]]^{M,g} = 1$ iff $[[\varphi]]^{M,g} = 0$

  $[[\varphi \wedge \psi]]^{M,g} = 1$ iff $[[\varphi]]^{M,g} = 1$ and $[[\psi]]^{M,g} = 1$, etc.

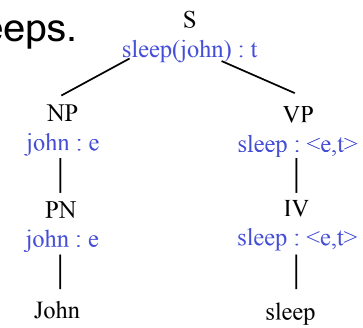
  If $v \in Var_\tau$, $[[\exists v\varphi]]^{M,g} = 1$ iff there is $a \in D_\tau$ such that $[[\varphi]]^{M,g[v/a]} = 1$

  If $v \in Var_\tau$, $[[\forall v\varphi]]^{M,g} = 1$ iff for all $a \in D_\tau$: $[[\varphi]]^{M,g[v/a]} = 1$

  $[[\alpha=\beta]]^{M,g} = 1$ iff $[[\alpha]]^{M,g} = [[\beta]]^{M,g}$

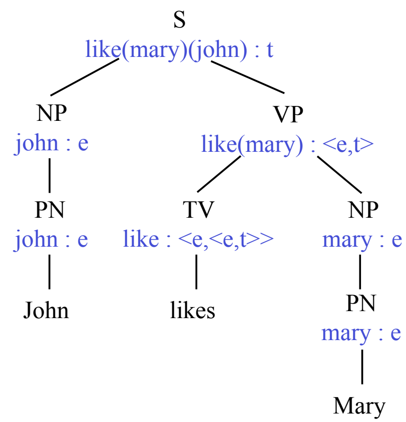
Semantics construction

- John sleeps.



Semantics construction

- John likes Mary.



The composition problem again

Every student presented a paper

∀d (student(d)→ ∃p (paper(p) ∧present(d,p)))

FLST 2008/2009 © Manfred Pinkal, Saarland University

The Semantics of Quantified NPs

John works.

john: e work: <e,t>

work(john): t

Every student works.

every-student: e work: <e,t>

every-student(work): t

This does not work !!!

FLST 2008/2009 © Manfred Pinkal, Saarland University

The Semantics of Quantified NPs

So we try it the other way round:

Every student works.

every-student: <<e,t>,t> work: <e,t>

every-student(work): t

'Every student' is a complex second-order predicate that is true of a first-order predicate, if all students are in the denotation of that predicate.

FLST 2008/2009 © Manfred Pinkal, Saarland University

What about Deduction?

∀x(truck(x)→motor_vehicle(x))
∃x(truck(x) ∧ sell(volvo, x)) ⊢ ∃x(motor_vehicle(x) ∧ sell(volvo, x))

Semantic Interpretation

Text

Hypothesis

- We have replaced the usual quantifier representation with higher-order non-logical constants (“every_student”), to facilitate semantic composition.
- This means that we cannot use FOL deduction anymore.
- Is there any way out of the dilemma?

FLST 2008/2009 © Manfred Pinkal, Saarland University

Another Detour

John drinks and drives

Drinking and driving is dangerous

Lambda-Abstraction

- $\lambda x[\mathrm{drive}(x) \wedge \mathrm{drink}(x)]$ is a composite predicate, whose meaning can be paraphrased with "an x such that x drinks and drives" or "to be somebody who drinks and drives"

drive: <e,t> x:e   drink: <e,t> x:e

drive(x): t   drink(x): t

$\mathrm{drive}(x) \wedge \mathrm{drink}(x)$: t

$\lambda x[\mathrm{drive}(x) \wedge \mathrm{drink}(x)]$: <e,t>

FLST 2008/2009 © Manfred Pinkal, Saarland University

Example

John drives and drinks.

drive: <e,t> x:e   drink: <e,t> x:e

drive(x): t   drink(x): t

$\mathrm{drive}(x) \wedge \mathrm{drink}(x)$: t

john : e $\lambda x[\mathrm{drive}(x) \wedge \mathrm{drink}(x)]$: <e,t>

$\boxed{(\lambda x[\mathrm{drive}(x) \wedge \mathrm{drink}(x)])(\mathrm{john}) : t}$

FLST 2008/2009 © Manfred Pinkal, Saarland University

β–Conversion

- β-conversion or β-reduction:
  $\lambda v\alpha(\beta) \Leftrightarrow \alpha^{[\beta/v]}$
- An application of a λ-expression $\lambda v\alpha$ to an argument β is equivalent to α, where all occurrences of the λ-variable v in α are replaced by β.

FLST 2008/2009 © Manfred Pinkal, Saarland University

Example

John drives and drinks.

$$\dfrac{\dfrac{\dfrac{\text{drive}: \langle e,t\rangle \quad x:e}{\text{drive}(x): t} \quad \dfrac{\text{drink}: \langle e,t\rangle \quad x:e}{\text{drink}(x): t}}{\text{drive}(x) \wedge \text{drink}(x): t}}{}$$

$$\dfrac{\text{john} : e \qquad \lambda x[\text{drive}(x) \wedge \text{drink}(x)]: \langle e,t\rangle}{(\lambda x[\text{drive}(x) \wedge \text{drink}(x)])(\text{john}) : t}$$

$\Rightarrow_\beta$ drive(john) $\wedge$ drink(john) : t

The composition problem again

Every student presented a paper

$\forall d\ (\text{student}(d) \rightarrow \exists p\ (\text{paper}(p) \wedge \text{present}(d,p)))$

Quantified NPs as λ–expressions

- The semantic interpretation of a universally quantified noun phrase can be straightforwardly encoded as a lambda term:

  $\lambda G\ \forall x(\text{student}(x) \rightarrow G(x))$

- Accordingly, the determiner *every* can be represented as:

  $\lambda F \lambda G \forall x(F(x) \rightarrow G(x))$

An example

- Every student works.

S
$(\lambda G \forall x(\text{student}(x) \rightarrow G(x)))(\text{work}) : t$
$\Rightarrow \forall x(\text{student}(x) \rightarrow \text{work}(x)) : t$

NP
$(\lambda F \lambda G \forall x(F(x) \rightarrow G(x)))(\text{student}) : \langle\langle e,t\rangle,t\rangle$
$\Rightarrow \lambda G \forall x(\text{student}(x) \rightarrow G(x)) : \langle\langle e,t\rangle,t\rangle$

VP
work : $\langle e,t\rangle$

Det
$\lambda F \lambda G \forall x(F(x) \rightarrow G(x)) : \langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle$
every

N
student : $\langle e,t\rangle$
student

VP
work : $\langle e,t\rangle$
works

Recommended Reading

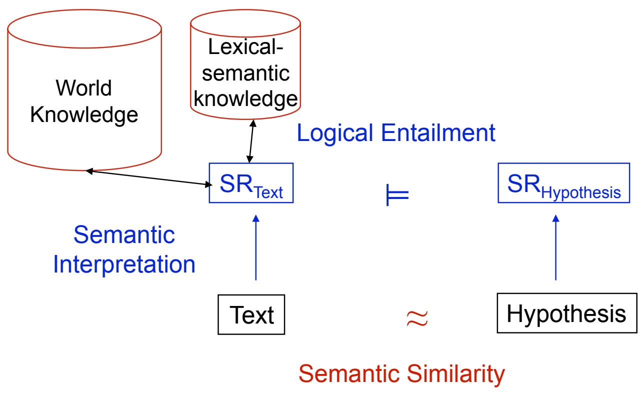
- Textbook: L.T.F. Gamut, Logic, Language, and Meaning. University of Chicago Press 1991

  Volume1: Introduction to Logic.

  Volume2: Intensional Logic and Logical Grammar.

Overview

- Semantic Processing - Introduction
- Logic-based meaning representation and processing: Truth-conditional interpretation, entailment, deduction
- Word Meaning: Lexical-semantic resources, ontologies, similarity-based approaches
- Semantic Composition: Composing sentence and text meaning from word meaning
- Textual Entailment and Inference



World-Knowledge and Inference

Text: *Security authorities have declared a state of maximum emergency in Guatemala, which is located directly in the path of the hurricane.*

Hypothesis: *There is a state of maximum emergency in Guatemala because of the hurricane.*

Default Inferences

Text: *As a real native Detroiter, I want to remind everyone that Madonna is from Bay City, Mich., a nice place in the thumb of the state's lower peninsula.*

Hypothesis: *Madonna was born in Bay City, Mich.*

Textual Entailment

„We say that *T entails H* if the meaning of *H* can be inferred from the meaning of *T*, as would typically be interpreted by people. This somewhat informal definition is based on (and assumes) *common human understanding of language* as well as *common background knowledge.”*

Monz, C. and de Rijke, M. (2001). Light-Weight Subsumption checking for computational semantics. ICOS 3

Modelling of Textual Entailment

- In principle, truth-based logical entailment nor distribution -based probabilistic similarity measures cannot give a full account of the intuitive concept of textual entailment:
- Logical entailment is always strict entailment, whereas the intuitive entailment concept of entailment is based on (degrees of) plausibility (take the hypothesis to be true, until you find counter-evidence).
- Similarity is symmetric: a≈b iff b≈a, where entailment is intuitively an assymmetric, directed relation.
- Both concepts can only count as rough approximations.

Modelling of Textual Entailment

- In principle, truth-based logical entailment nor distribution -based probabilistic similarity measures cannot give a full account of the intuitive concept of textual entailment:
- Logical entailment is always strict entailment, whereas the intuitive entailment concept of entailment is based on (degrees of) plausibility (take the hypothesis to be true, until you find counter-evidence).
- Similarity is symmetric: a≈b iff b≈a, where entailment is intuitively an asymmetric, directed relation.
- Both concepts can only count as rough approximations.
- In practice, the performance of alternative approaches has been (approximately) assessed in the RTE Shared Task.

The RTE Task

- RTE: Recognizing Textual Entailment
- Training corpus and test corpus
  - 800 T-H pairs each
  - 400 true, 400 false ones
  - formed on the basis of material taken from IR, IE, Q&A, Summarization tasks
  - no domain restriction
- Task: Build a system that matches the Y/N annotation of the corpus as close as possible
- Dagan, Glickmann, Magnini, *RTE 2004 Workshop Proceedings*

RTE: Examples

Text-Hypothesis Pairs, Example:

Text: *The Arabic-language television network Al-Jazeera reports it has received a statement and a videotape from militants.*

Hypothesis: *Al-Jazeera is an Arabic-language television network.*

Entailed: Yes

RTE: Examples

Text: *His wife Strida won a seat in parliament after forging an alliance with the main anti-Syrian coalition in the recent election.*

Hypothesis: *Strida elected to parliament.*

Entailed: Yes

RTE Examples

Text: *With $549 million in cash as of June 30, Google can easily afford to make amends.*

Hypothesis: *Some 30 million shares have been assigned to the company's workers.*

Entailed: No

RTE: Examples

Text: *Oscar-winning actor **Nicolas Cage**'s new son and Superman have sth. in common*

Hypothesis: *Nicolas Cage's new son was awarded an Oscar.*

Entailed: No

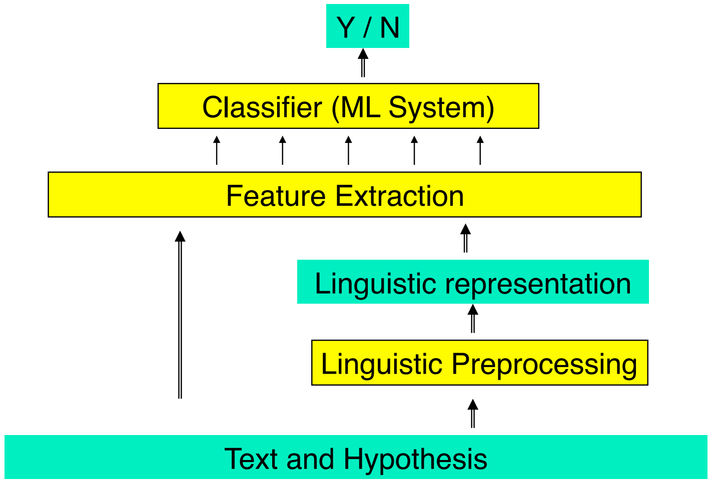
RTE: Examples

Text: *Wyniemko, now 54 and living in Rochester Hills, was arrested and tried in 1994 for a rape in Clinton Township.*

Hypothesis: *Wyniemko was accused of rape.*

Entailed: Yes

Architecture of RTE System



Information Used in different RTE Systems

- Word Overlap
- Semantic Similarity based on Vector-Space Models
- WordNet Information
- Syntactic Information
- World Knowledge
- Logical Inference

General Tendency of Results

- Systems relying on shallow information (word overlap, distributional similarity) perform better than naïve baseline of 50%, but only to some degree (60-65%).
- Systems relying on deep linguistic analysis and logical entailment perform drastically worse than naïve baseline. Reasons are, among other things:
  - Lack of robustness due to sequence of complex analysis steps requiring a large amount of precise input information.
  - Lack of reliable disambiguation techniques.
  - Lack of world-knowledge required for deduction
- But: Systems using deep processing techniques are significantly more precise on cases they can treat.
- The best results are obtained by combination of deep and shallow techniques.