An **Inferential** Approach

for Natural Language Semantics

Daniel Kayser
L.I.P.N. - Institut Galilée
Université Paris-Nord

Motivation

- Natural Language is important for Computer Science
  - For theoretical reasons (Artificial Intelligence)
  - For practical reasons (availability through the Web of huge amounts of texts)
- Despite over 2,500 years of thinking, its nature is still poorly understood

PLAN

- Motivation
- Critical presentation of traditional (**reference** based) semantics
- Is an **inference** based semantics really different?
- Some propositions : operators and operands
- Conclusion

Levels of analysis [1]

- Morphology
  - house - houses; class - classes; enemy - enemies ; goose - geese; …
  - love - loves; do - does; be - is;
  - not a problem for computers (with the possible exception of neologisms, foreign words, acronyms, misspellings, …)
- Lexicon
  - String matching is not a problem for computers (with the exception of compounds, neologisms, acronyms, misspellings, …)
  - Correspondence string-word not immediate…
    - syntactic ambiguity: *free* = adj, adv, verb, noun)
    - semantic ambiguity: *bank*

    … but needs not be solved at the lexical level

Levels of analysis [2]

- "**tripartition**" of Semiotics:
  - Syntax: relation between symbols
  - Semantics: relation of the symbols to the world
  - Pragmatics : relation of the symbols to the enunciative situation
- Syntax
  - only partially definable (the criterion of acceptability is not always well defined)
  - good computational models, not absolutely adequate
  - the "solvable" part reasonably solved:
    - tagging,
    - detection of:
      - noun phrases,
      - verb phrases,
      - many dependencies

Semantics

- Intuitively, **semantics = study of meaning**
  - meaning of words
  - of sentences
  - of texts
- Technically, relation of the symbols to the world
  - words ⇔ "elements" of the world
  - sentences, texts ⇔ "states of affairs" in the world

Levels of analysis [3]

- Semantics
  - Only a small part reasonably solved: the most sophisticated textbooks on Semantics treat
    - either very simple sentences (*a cat eats a mouse*)
    - or complex useless ones *(every man whose father is a doctor loves a woman*)
    - but fail to give a useful analysis of normal sentences taken e.g. in the newspaper
  - All the unsolved problems are said to resort to "pragmatics"
- Pragmatics: Juxtaposition of partial issues, e.g.
    - Speech acts
    - Conversational conventions
    - Contextual disambiguation, anaphoras
    - Non-literal meaning, metaphors
    - Argumentation, text analysis

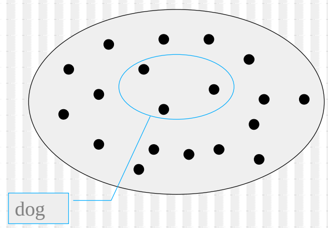
  With, most of the time, informal "solutions" (e.g. relevance theory) very difficult to give a computational account of.

Traditional Semantics

- The "world" (real or fictitious: *universe of discourse*) is known beforehand
- "elements" of the world are objects (or classes of objects, situations, states, processes, … involving objects) of that universe
- cf. Bible : God presents the animals to Adam, for him to give them a name.

Common Noun = unary predicate

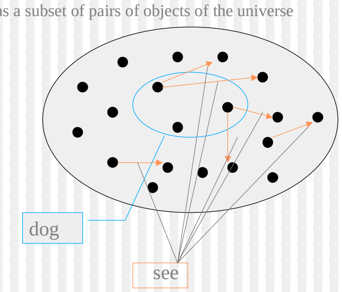
Gets interpreted as a subset of the universe



→ Language based on abstraction over **perception**
(a common noun designates those objects that we **perceive** as belonging to the same category)

Transitive Verb = binary predicate

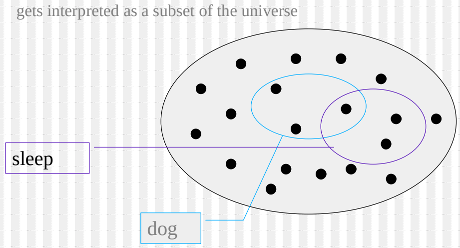
gets interpreted as a subset of pairs of objects of the universe



A dog sees an object ≡ (∃ x,y) (dog(x) ∧ see(x,y))

Intransitive Verb = unary predicate

gets interpreted as a subset of the universe



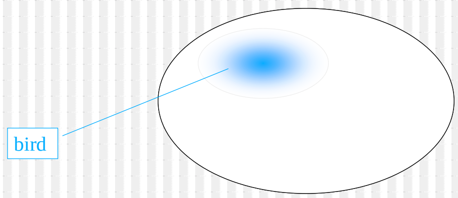
A dog sleeps ≡ (∃ x) (dog(x) ∧ sleep(x))

most common Objections

- Typicality
- Polysemy
- Modalities
- Properties of relations
- Collective nouns
- …

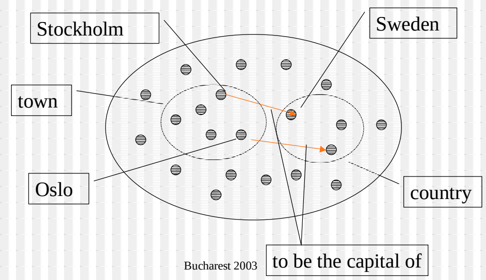
Typicality

A common noun gets interpreted
- as a subset of the universe = mapping U $\rightarrow$ {0,1}
- as a mapping U $\rightarrow$ [0,1]

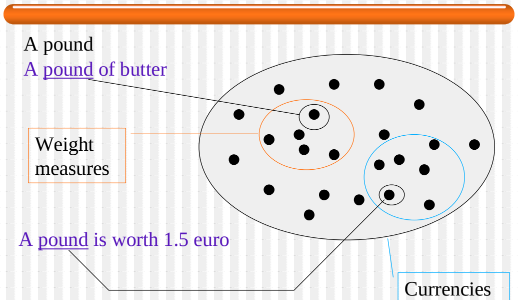


Modalities [1]

To believe Oslo to be Sweden's capital
(relationship between an agent and a proposition)

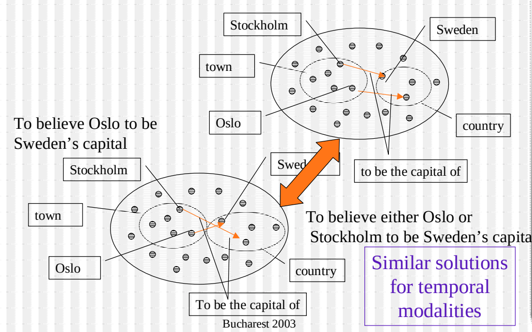


Polysemy



Common noun $\rightarrow$ n unary predicates
Frequent discrepancy about n …

Modalities [2]



Similar solutions for temporal modalities

Properties of relations

- To express that a given relation R is symmetrical, transitive, …
  - $(\forall x,y)\ (R(x,y) \Rightarrow R(y,x))$
  - $(\forall x,y,z)\ (R(x,y) \wedge R(y,z) \Rightarrow R(x,z))$
  - …
- To express what is symmetry, transitivity
  - Higher Order Logic
    - $(\forall R)\ (\mathrm{symmetrical}(R) \Leftrightarrow (\forall x,y)(R(x,y) \Rightarrow R(y,x)))$, …
  - or « false second order »
    - $(\forall R,x,y)\ (\mathrm{symmetrical}(R) \Leftrightarrow (\forall x,y)(\mathrm{true}(R,x,y) \Rightarrow \mathrm{true}(R,y,x)))$
    - …

Collective [2]

- So far, reference based semantics can solve the problems (in a more or less *ad hoc* way), but how could it find a referent for *the French* in:

For 2 centuries, The French have hesitated between an authoritarian and a democratic system of government

- Pure distribution: impossible
- Distribution at each time point on a "typical element"?
  At any time, the "typical" frenchman has no definite preference
- Collective interpretation at each time point?
  At any time, few frenchmen hesitate; but no majority emerges
- Distribution on temporal intervals of a collective interpretation?
  At any time, a majority exists, but it fluctuates across time

Co-presence of several of these interpretations?

Collective [1]

- Three men …
  $(\exists\ x,y)\ (=(\mathrm{card}(x),3) \wedge (\forall z)(\mathrm{member\text{-}of}(z,x) \Rightarrow \mathrm{man}(z) \wedge$
  - … wear a tie

  $(\exists\ t)(\mathrm{member\text{-}of}(t,y) \wedge \mathrm{tie}(t) \wedge \mathrm{wear}(z,t)\ \ldots)))$

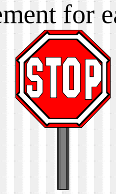
  - … carry a piano

  $\mathrm{piano}(y) \wedge \mathrm{carry}(z,y)))$
- The Paris métro carries 3.6 billion passengers per year

$(\exists\ x)\ (=(\mathrm{card}(x),\ 3.6\ \mathrm{billion}) \wedge (\forall z)\ (\mathrm{member\text{-}of}(z,x) \Rightarrow \mathrm{passenger}(z)))$ is dead wrong!

Reference-based semantics

adds a new ± « ad hoc » implement for each new difficulty.



Alternative : in the *perception-action* loop,
why should it be preferable to base language on abstraction of perception,
rather than on abstraction of **action**?

Hypothesis to examine

- Instead of considering a word as factoring out <u>perceptive</u> properties, e.g.
  - A *table* is made of a horizontal top, and of leg(s) in sufficient quantity for it to be stable, …
- Consider it as factoring out <u>active</u> properties, e.g.
  - One can sit at a *table* for eating, writing, …
- Would that really change much?

the referential « safeguard »

- avoids the risk of building an uncontrolled collection of inference rules
  - (the existence of models is a guarantee of consistency)
  - How to replace this guarantee?
- At every step,

a model of the set of conclusions inferred should exist
models such that every conclusion inferred holds true in at least one of them can co-exist

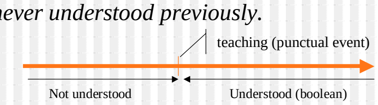
- Those models are not necessarily the same from one step to another

Approach based on models vs. Approach based on proofs

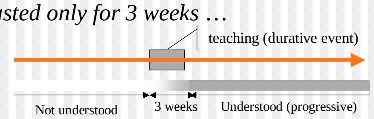
- In a formal system that is *correct*,
  - Every provable fact holds true in all models
- In a formal system that is *complete*,
  - Whatever is true in all models is provable
- First-order logic is both correct and complete …
- But correction and completeness make sense only if interpretation takes place in a fixed universe!

Evolution of models

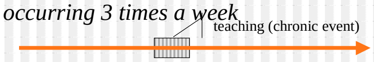
This teaching made me understand a theory that I never understood previously.



It lasted only for 3 weeks …

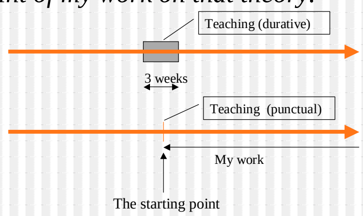


… occurring 3 times a week



Co-existence of models

- *This 3 week teaching has been the starting point of my work on that theory.*



Example : color [1]

1. ~~*This flower is red*~~
2. *red is a color*
3. ~~∴ *This flower is a color*~~

There is an ellipsis in 1. :

1' *The color of this flower is red*

But now, there is a problem with types:

$= \underbrace{(red, color(fl\#1))}_{1'} \wedge \underbrace{color(red)}_{2}$

color is a function, color is a predicate.

(alternative solution: red(fl#1) ∧ color(red) yields red predicate and red, constant = 0-ary function)

How ?

- Basic idea : the same word can have simultaneously or successively several interpretations,
- Let us consider it as the **operand** of a number of potential **operators**
  - E.g. a word designating an event is an argument for the operator PUNCTUAL and for the operator DURATIVE

Color [2]

- This is a very general problem!!
  - John is Sophie's father; he is a good father!
    - father : function, father : predicate
  - London is an important market place; London market place is very busy
    - market place : predicate, market place : function
  - HPLIPN is a printer; it is one of the printers of the B300 network
    - printer : unary predicate; printer: binary predicate
  - sin(x) is a function over reals; sin(x) is a real number
    - sin(x) : name of a function ; sin(x) : value of the function
  - [Montague] the temperature is 90°F ; the temperature is rising
    - temperature :value of function ; temperature : name of function

Operators and operands

- *color* has no interpretation
- There exists an operator PREDICATE such that when it takes *color* as argument, yields a unary predicate
  - Red is a color → PREDICATE(*color*) (red)
- There exists an operator FUNCTION such that when it takes *color* as argument, yields a unary function
  - This flower is red → =(red, FUNCTION(*color*) (fl#1))
- The other examples are solved in a similar way

Iteration of operators

- At first sight, the word *price* behaves the same way as the word *father*:
  - *15 Euros is a price* [predicate]
  - *15 Euros is the price of this book* [function]
- However, its semantics is far more complex
  - Price of an object having multiple copies
  - Price of a collection of objects
  - Price of substances
  - The price may depend on time and location

Relation between the results
of operators applied to the same operand

- Since FUNCTION(*father*) (x) computes x's father, we have:

(∀x) (PREDICATE(*father*) (FUNCTION(*father*) (x)))

- Similarly :

(∀x) (PREDICATE(*color*) (FUNCTION(*color*) (x)))

But if x stands for the French flag?

One should rather see FUNCTION as an operator yielding a **collection** and write:

(∀ *x*,y,z)(member-of(z,FUNCTION(*x*)(y)) ⇒ PREDICATE(*x*) (z))

15 Euros is the price of that book

Probably, co-presence of two entities within the same model
• *that book*: physical object, which I am currently pointing at
• *that book*: class of objects having the "same" characteristics.
Starting from an x, satisfying PREDICATE(*book*) (x), we build CLASS(x)

For every "relevant" *y* (∀x,z) (PREDICATE(*y*) (x) ∧ member-of(z,CLASS(x)) ⇒ PREDICATE(*y*) (z))

We get =(15 Euros, FUNCTION(*price*) (CLASS(book#1)))

A default rule says that:
• if 15 Euros is the price of the class,
• it is the price of an element of the class,
• *exceptions* (dog-eared book, book on display, …)

Price of a collection

- Do not confuse the price of a *class* of (presumably identical) objects with the price of a *collection* of objects
- Here too, there is a default rule :

$$\text{Collection}(x) \Rightarrow =(\text{FUNCTION}(\textit{price})\ (x), \sum_{y \in x} \text{FUNCTION}(\textit{price})(y))$$

Exceptions : price of a pack

This rule holds for all additive magnitudes

Adding parameters [1]

- From:
  - *The price of that book is 15 Euros* and:
  - *one pound is worth 1.5 Euro*, infer:
  - *The price of that book is 10 pounds,*
- So rather than:

  =(15 Euros, FUNCTION(*price*) (CLASS(book#1)))

  one should write:

  =(15, ADD-PARAM(*unit*) ((FUNCTION(*price*)(CLASS(book#1))),*Euro*))

with :

=(y1,ADD-PARAM(*unit*) (x,d1)) ∧ =(d1/2,conversion(d1,d2)) ⇒ =(*(y1,d1/2), ADD-PARAM(*unit*) (x,d2))

Price of a substance

- The price of gasoline does exist; now, it is not a price!
- For PREDICATE(*price*) to become applicable, a volumetric unit must be specified.
- This is a special case of a general phenomenon: the operator **add-parameters**

(∀x,y) (substance(x) ⇒ ¬PREDICATE(*price*)(FUNCTION(*price*)(x)) ∧ volumetric-unit(y) ⇒ PREDICATE(*price*) [ADD-PARAM(*unit*) ((FUNCTION(*price*)(x)),y)])

Adding parameters [2]

- Price depends not only on currency:
  - *The price of that book is 15 Euros but at Barnes & Noble, they sell it 5% cheaper*
  - *The price of that book is 15 Euros but last month, it costed 1 Euro less*
  - *The price of that book is 15 Euros but for good customers, they are allowing a discount*
  - …
- For each of these sentences, a new kind of parameter appears, which can combine with the previous ones:

¬=(y,Barnes & Noble) ∧ =(p, ADD-PARAM(*seller*) (FUNCTION(*price*) (book#1),y)) ⇒ =(*(p,0,95), ADD-PARAM(*seller*) (FUNCTION(*price*) (book#1), Barnes & Noble))

=(15, ADD-PARAM(*unit*) (ADD-PARAM(*time*) ((FUNCTION(*price*) (book#1)), m), *Euro*)) ∧ =(14, ADD-PARAM(*unit*) (ADD-PARAM(*time*) ((FUNCTION(*price*) (book#1)), -(m,1)), *Euro*))

the fall of the prices of fresh products (4,8%) [1]

- First attempt:

=(p, ADD-PARAM(*time*) ((FUNCTION(*price*) (x)), t_1)) ∧
=(*(p,1,048), ADD-PARAM(*time*) ((FUNCTION(*price*) (x)), t_0))

<u>Missed</u> : there is no such *object* x « fresh products », the price of which has changed from one month to the next.

Notice : the product operates directly on p, not on its value expressed in a currency.

- x is a *collection*; however
  - Its elements are variable
  - There is no additivity but a weighted sum

the fall of the prices of fresh products (4,8%) [2]

- Instead of:

Collection(x) ⇒ =(FUNCTION(*price*) (x), $\sum_{y \in x}$ FUNCTION(*price*)(y))

We must write: =(z,Weighted-collection(x,P)) ⇒

=(FUNCTION(*price*) (z), $\sum_{y \in x}$ *(FUNCTION(*price*) (y), P(y, x)))

With object P such that : $\sum_{y \in x} P(y, x) = 1$

- Second attempt: = (p, ADD-PARAM(*time*) ((FUNCTION(*price*) (<u>Weighted-collection(x,P)</u>)),t_1)) ∧
=(*(p,1,048), ADD-PARAM(*time*) ((FUNCTION(*price*) (<u>Weighted-collection(x,P)</u>)), t_0))

the fall of the prices of fresh products (4,8%) [3]

- Once more **<u>missed</u>**!
  - The fresh products are not the same in one season or the other
  - Their weight in the index varies
- Third attempt: there exists a stable list x of products that at least at some periods of the year, belong to the index of fresh products. At every time, there is a weight function P
  - The weight of y in the list x is zero when y ∉ x
  - Therefore P is a time-varying object, so ADD-PARAM(*time*) (P,t) allows to « project » this object at a particular time-point t
  - $\forall t : \sum_{y \in x} \text{ADD-PARAM}(time)(P, t)(y, x) = 1$
- Finally, the sentence gets the following translation:

=(p, ADD-PARAM(*time*) ((FUNCTION(*price*) (Weighted-collection(x, <u>ADD-PARAM(*time*) (P,t_1)</u>))), t_1)) ∧
=(*(p,1,048), ADD-PARAM(*time*) ((FUNCTION(*price*) (Weighted-collection(x, <u>ADD-PARAM(*time*) (P,t_0)</u>))), t_0))

Conclusion from the example

- *Price* is not a symbol (function, predicate, …) interpretable as an object, subset, mapping, … in a pre-given universe of discourse
- It is the operand of a (sequence of) operators, built as the comprehension of the text proceeds
- The universe of interpretation is <u>modified according to the operators</u> used
- Some conclusions should be « protected » when the universe of interpretation changes
  - Example : if p is the price of x, and I have q>p, then
    - I can buy x
    - If I do so, I will possess x and (q-p)

Conclusion [1]

- The protection of "protected" inferences has priority over the construction of models
- They result from *actions* to be carried out, and not from *perceptions*
- Once comprehension has taken place, there exists a correspondence of:
  - references in a universe with words,
  - states of affairs with sentences,
  - sequences of interpretations with texts,

  ...but this is clearly a posteriori
- The main task of comprehension consists in constructing the universe: it is not an initial data!

Conclusion [2]

- Constructing the universe is itself an inferential process
- Words act as « inference triggers » for that process
- It is better to consider words as **factoring out**
  - similar inferences, rather than
  - similar objects.