Explicit world-knowledge and distributional semantic representations

ESSLLI 2017 Day 1: Foundations of representation

Asad Sayeed and Alessandra Zarcone

University of Gothenburg, Saarland University

Bird

I mean, the joke is. . .

. . . on a totally extensional concept of meaning
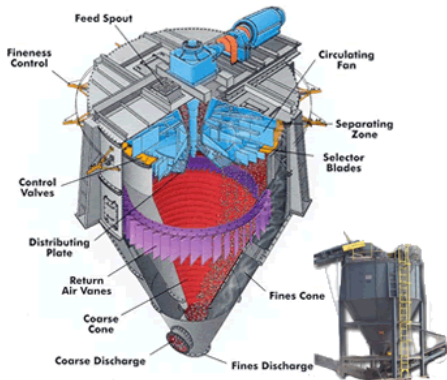

Obtaining an extensional definition of "bird":

1. Go through all physical (?) entities in the universe.
2. Classify them into bird and non-bird entities.
3. That classification is your definition of bird!

Except...

. . . how did you decide what to classify as a bird?
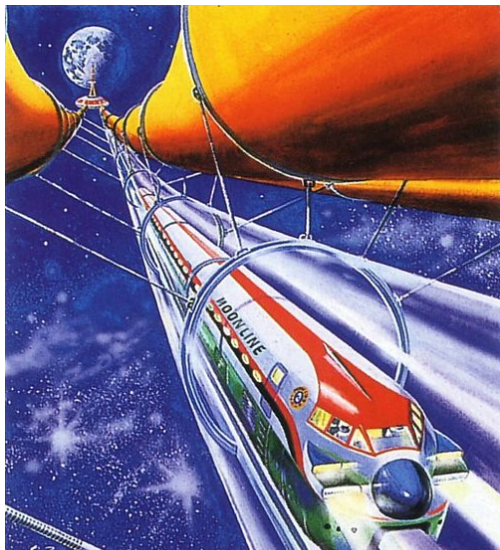
Easy! You train a classifier!

I mean, this is the modern world, right? We have MACHINE LEARNING.

But you do you train a classifier?

By presenting it with data (and some mathemagic and then a miracle happens)

But where did you get the data from?

By labelling some objects as birds and others not. . .

But seriously. . .

There are multiple problems here:

- We had to know what a bird was, before we could go around defining it. But how did we know that?
- Even "unsupervised" machine learning is still exposed to a potentially unrepresentative world.
  - I mean, how do you guarantee that it is getting the right training data? How do you even define what the right training data is?
- What about the human mind? Do we care that the machine picks up human intuitions?

This has lots of implications.

Semantics matters for everything:

- In linguistics and cognitive science:
  - The classic question: what is the exact role of experience in lexical meaning?
  - How much psycholinguistic data is explainable via some fancy transformation on frequency of experience?
- In natural language processing, particularly dialog systems:
  - Statistical modeling: how much does the inherent bias in our models prevent us from replicating human expectation?
  - How do we build systems that make linguistic inferences that people make?

So, to begin. . .
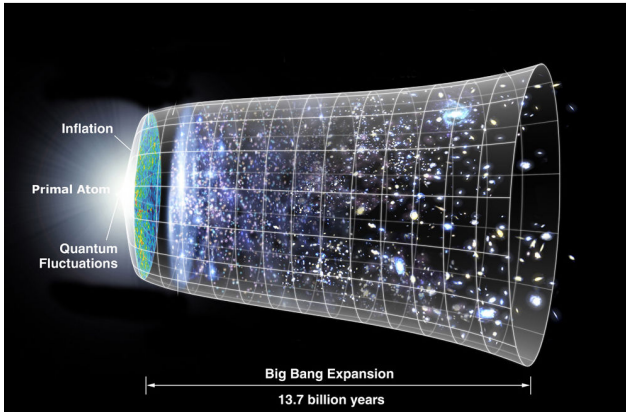
Part 1: introduction and overview

Extensional and intensional meaning

A traditional way to think of the relationship between meaning and the world.

Just a very cursory idea:

- Extensional meaning of a predicate – all the things in the world that make that predicate true.
- Intensional meaning of a predicate – a specification of the conditions under which the predicate is true.

Problem: the relationship between intensional and extensional meaning.

In a sense: between our knowledge and the world.

Where does lexical meaning come from?

If an extensional approach to lexical meaning is not enough, maybe we need intensional definitions.

- Conditions under which a word can be used – e.g. properties, relations.
- Mathematically/logically manipulable.
- Traditionally highly subjective.

But this doesn't solve the problem: where do we get the properties from?

What can we do with lexical meaning?

Attempting to "define bird" is an interesting exercise, but why do we want to define what a bird is?

- We want to be able to identify future instances of bird.
- We want to be able to perform "reasoning" with the lexical characteristics of birds.

This means that we need a lexical semantics that is compatible with compositional semantics.

Traditional viewpoint: formal representation

Meaning as reasoning:

- When we talk about semantic reasoning, we traditionally mean via logical representations/rules.
- Since logical reasoning takes place over symbols, a "meaning" is thus a symbol or a complex structure of symbols.
- Lexical entries are augmented by a set of axioms/higher-order functions that define symbol-manipulation processes.

Traditional viewpoint has traditional problem: meaning isn't discrete!
What is a symbolic representation of "bird"?

More recently: distributional representation

Meaning as quantified context:

- Part of "bird" meaning comes from where you see "bird"
  - You see birds flying in the world.
  - You see birds mentioned flying in text.
  - So "bird" is a "fly"-word.
- Many insects also fly, no beavers fly. So insects should be more similar to birds than beavers.

More recently: distributional representation

Meaning as quantified context:

- Similarity reasoning easily possible in distributional representations.
- What about other logical implications? Event structure? Discourse?
- For language: does linguistic co-occurrence really reflect real-world co-occurrence?

But centrally...

. . . how much knowledge do we have to "pre-encode" in order to capture "human" meaning?

Not clear where the dividing line is

Think about flying:

- Represents some interaction with the world – can you imagine a world in which we could not represent the concept of flying?
- Can the representation of flying be derived trivially from observing flying, or do we need foundational concepts to derive flying?
- From where do these "foundational concepts" come from?
  - How much of them are innate? Do some of them have to be taught?

How much knowledge that has to be represented without inference from co-occurrence?

Implications in multiple areas

Vast expansion of distributional modeling over large data sets.

- Sometimes advertised as panacea for all modeling and application problems.
- But if linguistic use is not representative of the capacity of human inference, a little more than just training data may be required. (?)
- How is multimodal data represented/used alongside linguistic data? Both in computational application and in psycholinguistic processing.

Who this course is aimed at

Intended to be a high level overview plus "argument" with elements of:

- Formal lexical semantics – without going too deeply into the mathematics of formalism.
- Psycholinguistics – without going too deeply into experimental technique or analysis.
- Natural language processing/distributional modeling – with only a touch of programming demo, and not going too far into the technical details

Goal: spur/expand interdisciplinary debate over these issues.

Who this course is aimed at

This is an "advanced" course in ESSLLI parlance...

- ... but because of the mixed group, some parts will be "introductory" for some people.
- **I really want you to argue with me.**
    - It's an unsettled area that touches on many things, your different expertise will help us understand it better.
- Note: some overlap with Generative Lexicon lecture just before lunch – but we aren't technically "introductory", while they are.

Our agenda for the week

- Today: introduction, history, some cognitive foundations
- Tomorrow: lexical semantics from a logical and cognitive perspective
- Wednesday: introduction to distributional semantics, partly technical
  - Formation of discussion groups on future challenges
- Thursday: evidence for explicit vs. distributional knowledge
  - Also: "pot-pourri" technical demonstration of distributional semantics for event analysis.
- Friday: future avenues in modeling and experimentation
  - Short group presentation from Wednesday groups.

Remaining items for today

- History of lexical semantics - structuralism, generativism, etc.
- Cognitivist approach to lexical knowledge - scripts, pragmatic knowledge
- Motivation: the problem of the dividing line between kinds of knowledge in lexical semantics.

Who are we?

- Asad Sayeed - Associate Senior Lecturer, University of Gothenburg
  - Background in natural language processing and "old school" theoretical syntax (PhD at University of Maryland).
  - Spent a few years at Saarland University before Gothenburg becoming a semanticist and psycholinguist.
  - At Gothenburg as of two months ago – part of CLASP ("Center for Linguistic Theory and Studies in Probability").
- Alessandra Zarcone - Research Scientist, Saarland University
  - Linguistics/psycholinguistics background with PhD from Stuttgart in modeling event knowledge.
  - Current work at Saarland University: script knowledge and information density in the DFG-funded project SFB 1102 "Information Density and Linguistic Encoding".

Where is Alessandra?

Alessandra Zarcone is my official co-lecturer, but I will be delivering all her lecture material.

The reason is adorable. (She's on maternity leave.)

Alessandra and I have different areas of expertise, so you must forgive me if I cannot do her work the justice it is due, but I will do my best!

But what about you?

Part 2: historical perspective on formal lexical semantics

Lexical semantics: an old effort

But not *very* old.

- Sure, ancient philosophers cared about meaning, quite a lot.
- Plato thought that there was a "natural correctness to names", according to Wikipedia.
- Also according to Wikipedia, Aristotle "thought that the meaning of a predicate was established through an abstraction of the similarities between various individual things." – not bad, eh?

But according to a paper by Geeraerts [2002] on the history of lexical semantics, only really appears as subdiscipline by about 1870.

Prestructuralist semantics

(We'll be following Geeraerts from now on.)

From about 1870-1930, can call semantics "prestructuralist":

- Diachronic focus (presumably to go with the tendencies in philology).
- "Semasiological" approach – "what does word X mean?"
    - (Um, isn't that the point of lexical semantics? – well, less focus on the word itself and more on the meaning)
- Psychological focus in diachronic context– meaning change considered in terms of change in thought.

Structuralist semantics

Started in the 20s and 30s, inspired (of course) by de Saussure.

- Move away from diachrony into a syncronic focus.
- Move away from individual meaning units into linguistically motivated structures (surprise, surprise).
- More subtly, emphasis on autonomy of linguistic structures from purely psychological explanation.
    - Treat linguistic signs as things-in-themselves, without reference to mental objects.

Structuralist semantics

Foci of structuralist semantics:

- Semantic similarity, once again.
- Introduction of lexical relationships like synonymy, antonymy, hypernymy, etc.
- Syntagmatic relationships, e.g., selectional preferences were considered part of the lexicon.

Neostructuralist semantics: occasional attempts to revive this approach. (Geeraerts mentions that they usually do not generate much new work under this label.)

Structuralist semantics

Structuralist semantics had a legacy in later approaches, however.

- In computational linguistics, e.g., WordNet.
- "Collocational" analysis – in a sense, doing distributional semantics informally.

In fact, much of what we'll be talking about in this course is a sort of structuralist revival with computers to help us fill in the missing bits.

Generativist semantics

Generativist semantics: from Katz and Fodor in the 60s and 70s, but mostly Katz.

- Built on top of structuralist semantics, but attempting to incorporate semantics directly into generative syntax.
- Rigid formalization of semantic structures for use in an algorithmic approach.
- Greater attempt to interconnect lexical semantics with sentential/compositional semantics.

Above all, a return to a "mentalistic"/cognitive/psychological notion of meaning.

Generativist semantics: digression on controversy

Generative semantics came up against a critique from generative syntacticians.

- In a nutshell, generative syntax favoured an "interpretive" semantics.
  - Syntactic "module" constructs representation of sentence stucture.
  - Representation "shipped" for interpretation to semantics.
- Generative semantics instead used transformational grammar "all the way down" to "deep structure".

Why might this be a problem?

Generativist semantics: digression on controversy

At least two challenges for generative semantics:

- Problem of syntactic redundancy: where does the "do" in English do-insertion work in a fully "syntactified" theory of lexical meaning?
- Problem of *ad hoc* compositionality: e.g. "switch on" vs. "climb on"
  - Why does "on" do one thing in one place, but another in another?

Some people accuse Chomsky's influence of also reducing the influence of generative semantics, but . . .

Neogenerativist and cognitive semantics

. . . attempts to "mentalistically" formalize lexical semantics continue to have traction.

- Pustejovsky's Generative Lexicon: focus on using the formalization to understand how new meaning can be produced.
- Cognitive semantics: attempt to approach meaning in terms of prototypes, metaphors, scripts, etc.
  $\Rightarrow$ incorporate mental representation of world-knowledge.

Part 3: cognitive foundations of abstract world representation

Situating lexical knowledge

Structuralist approaches: divorce from cognition.

- This provides analytical tools for identifying linguistic units – which is the goal of structuralism.
- But the problem we can't deny is:
  - We **do** things with language.
  - We **process** language.
- And that involves, somehow words.

Processing and doing

What is the common denominator in **processing** language and **doing things** with language.

EXPECTATION

To put a giant psycholinguistic story short:

- Learning, processing, experiencing with language all somehow seems to involve restricting what is possible later in the sequence.
- Language does not sit abstractly in sentences – there is a time factor.
- Interactions with language exist in a world of expectations.

Donald put the cake in the. . .

Expectations in the world: scripts

Script knowledge: knowledge of everyday activities [Schank and Abelson, 1977; *inter alia*]

- e.g. going to a restaurant.
- plays a key role in our interaction with the environment and in language comprehension.
- Script events, participants stored in memory in an order sensitive way.
- Chronological order of script events shown to determine if following events are cued. (e.g. cook, sit $\rightarrow$ dine).

Script and pragmatic knowledge

Script knowledge: knowledge of everyday activities

- In language, we expect plausible events to follow a sequence of script events.
  - We don't expect every single event to always be mentioned – we assume trivial events.
  - Weird to mention all of them all the time!

Presence of event in script does not predict mention frequency!

Script-text relationship

Crocker et al. [2016]:
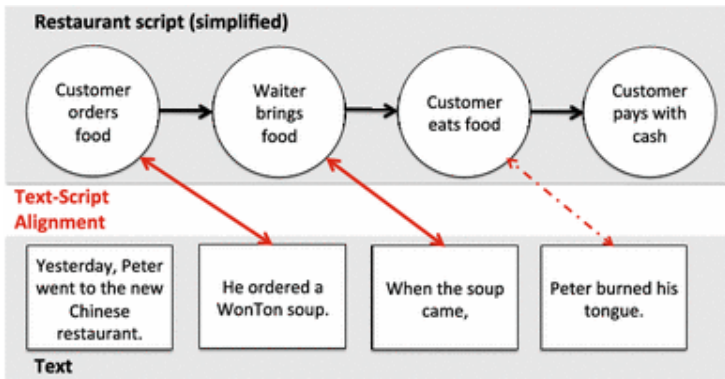
What can we get from this kind of characterisation?

Scripts and distributional evidence

Difficult to estimate event typicality in script context from corpus coocur-rences.

Padó, 2007

"Infrequent events may be perceived as more informative and interesting and therefore more worthy of being communicated, which may cause them to be discussed disproportionately more often than they are experienced . . . frequent events may be perceived as less newsworthy and therefore be mentioned less often than they occur."

Scripts and distributional knowledge

Imperfect parallelism between events in the world and events in language.

Bruni et al., 2012

It is difficult to model stereotypical color adjective information (e.g. bananas are yellow) on the basis of corpus-extracted information, arguably for the same reason that corpus-extracted information is not informative with regard to highly typical information.

In other words, people don't mention that bananas are yellow and don't mention that they have to sit in a car to drive it!

Part 4: motivation: finding the boundaries between distributional, formal, and pragmatic knowledge

A bird's eye view of the place of semantics
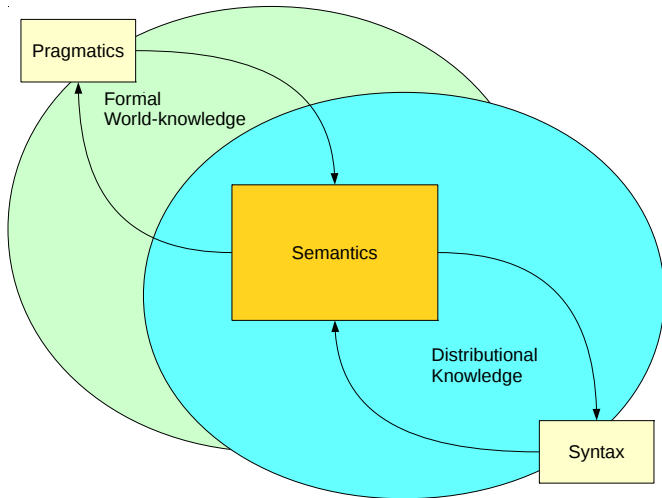
Types of knowledge

Dividing up the knowledge problem:

Implicit world-knowledge

Latent knowledge about the world that can be induced from indirect information sources (e.g. distributional characteristics of language).

Explicit world-knowledge

Knowledge about the world that is coded explicitly, deduced formally, innate, learned by being told, etc.

- Implicit world knowledge – somehow related to the "experiential" part of extensional meaning?
- Explicit world knowledge – somehow related to the "cognitive" part of intensional meaning?

Where are we now scientifically?

Convergence of modeling techniques and predictions.

- Traditional experimental practice in sentence processing:
  - Identify phenomenon, make hypothesis.
  - Collect subjects, expose to stimuli with a critical region.
  - Analysis: metrics of critical regions.

- How to obtain ecological validity: examine increasingly large quantities of behavioral data for hypothesis formulation and even in experiment!

Distributional approaches: creeping in the back door

Should make the assumptions explicit.

- For example: does cognitive workload track with semantic (mis)match? [Demberg, Kiagia, and Sayeed, 2013]
    - Pupillometric cognitive workload measurement using sentences (in German) like "Max is singing as a lawyer..."
    - Found signifciant effect of anomaly on workload via left-eye measurement.
- How can we develop a more fine-grained analysis of semantic mismatch effect?

An even bigger problem: affordances

How do we expect objects to interact with the world? Directly intersects with script knowledge effects.

- You **can** wipe your teeth with a towel, we just find brushes more convenient.
- That we can "misuse"/re-purpose objects and imagine talking about it makes it difficult to characterize meaning as context.
- We're not going to find that *directly* from collocations, but need some kind of model assumptions – what?

We know that there is an interaction. . .

. . . between collocational/distributional knowledge and our semantic "understanding" we don't know what that interaction is.

In the remainder of this week, we'll cover different aspects of this interaction.
The goal, and what we expect from you, is discussion and hopefully new thoughts on where next to go in this research topic.

See you tomorrow!