Compositional and Distributional Semantics

ESSLLI 2008, Hamburg
Workshop Lexical Semantics: Bridging the gap between semantic theory and computational simulations
08. August, 2008

Peter Bosch

Institute of Cognitive Science
University of Osnabrück

The limited agenda

Distributional Semantics develops methods

- for the investigation of *human linguistic behaviour*
- and for applications in *linguistic engineering* and *information engineering* that may be, but are not necessarily related to human linguistic behaviour.

>> *This talk is only concerned with the former.*

The limited agenda

In particular I want to find out about the relation between

- ***Distributional Semantics* (broadly understood) and**
- ***Compositional Semantics* (a.k.a. Logical Semantics, Truth-Conditional Semantics, Fregean Semantics, Montague Semantics; again: broadly understood)**

First Impressions 1: Coverage

Distributional Semantics (DS) can apparently account for the meaning of lexical items from open word classes: nouns, verbs, adjectives

but has little to say about the semantics of lexical items from closed word classes: Determiners, quantifiers, conjunctions, auxiliaries, modals,…

(-> "stop words")

Compositional Semantics (CS) can account for the meaning of lexical items from closed word classes,

but has little to say about the semantics of lexical items from open word classes.

First Impressions 2: Goals

CS **aims at an account for *semantic productivity*:**

How can speakers of a natural language understand potentially infinitely many composite expressions (phrases, sentences), given finite lexical resources, i.e., a finite repertoire of simplex expressions?

DS **aims at the description of *distributional relations*:**

How similar are the distributional contexts in which any two expressions occur?

and puts up a surprising hypothesis:

That *distribution can be identified with meaning*, so that similarity of distribution is taken as a measure for semantic similarity.

The difference is in the conception of Meaning

For CS the meaning of a linguistic expression is the contribution it makes to the truth conditions of sentences

> Understanding a sentence means to know what is the case if it is true. (Thus one can understand it without knowing whether it is true.) One understands it if one understands its constituent parts.
> (Wittgenstein, Tractatus 4.024)

For DS the meaning of an expression is identified with the set of distributional contexts in which the expression occurs.

> You shall know a word by the company it keeps
> (Firth 1957:11)

Distributional Meaning

Even though the Firthian bit does not amount to much of theory, the notion of meaning as distribution intuitively makes sense.

Consider everyday parlance:

> *Clouds may mean rain. Smoke means fire. Fire often means smoke. A humid spring means lots of snails. Never doing your homework means that you'll fail.*

These "meaning" relations are just relations of significant co-occurrence that are interpreted as licensing inference:

> ***For event types A and B, if A is said to mean B, then A permits the inference that B.***

Distributional Meaning

It makes sense that linguistic expression types should acquire significance or meaning via co-occurrence relations.

> ***You learn that A "means" B, when you experience regular co-occurrence of A and B, or experience that A and B regularly occur in the same environments.***

There is plenty of evidence for such relations, e.g. in priming relations, linguistically as well as cross-modally – and not only for open class words

Disambiguation by determiners

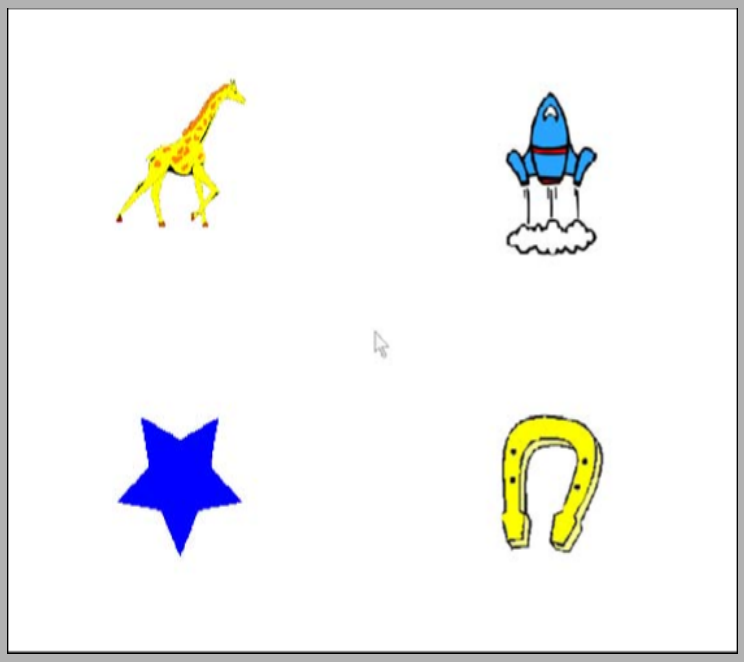
default nouns:

die$_{[fem]}$ *gelbe Giraffe*$_{[fem]}$
the yellow giraffe

die$_{[fem]}$ *blaue Rakete*$_{[fem]}$
the blue rocket

der$_{[masc]}$ *blaue Stern*$_{[masc]}$
the blue star

das$_{[neut]}$ *gelbe Hufeisen*$_{[neut]}$
the yellow horse shoe



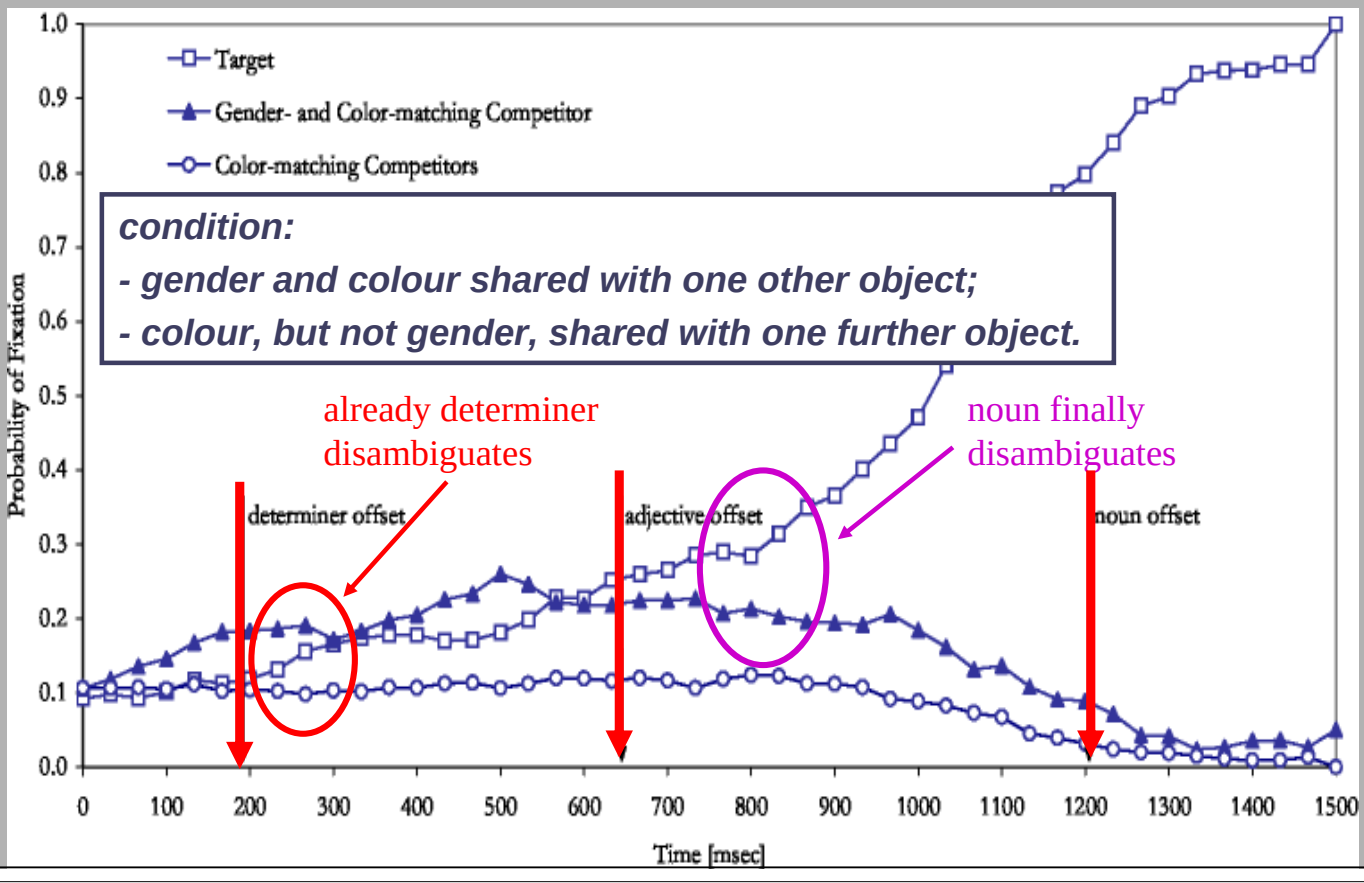
Klicken Sie auf die blaue Rakete.

click on **... [followed by a def. determiner, adjective, and noun]**

NadineHartmann 2005

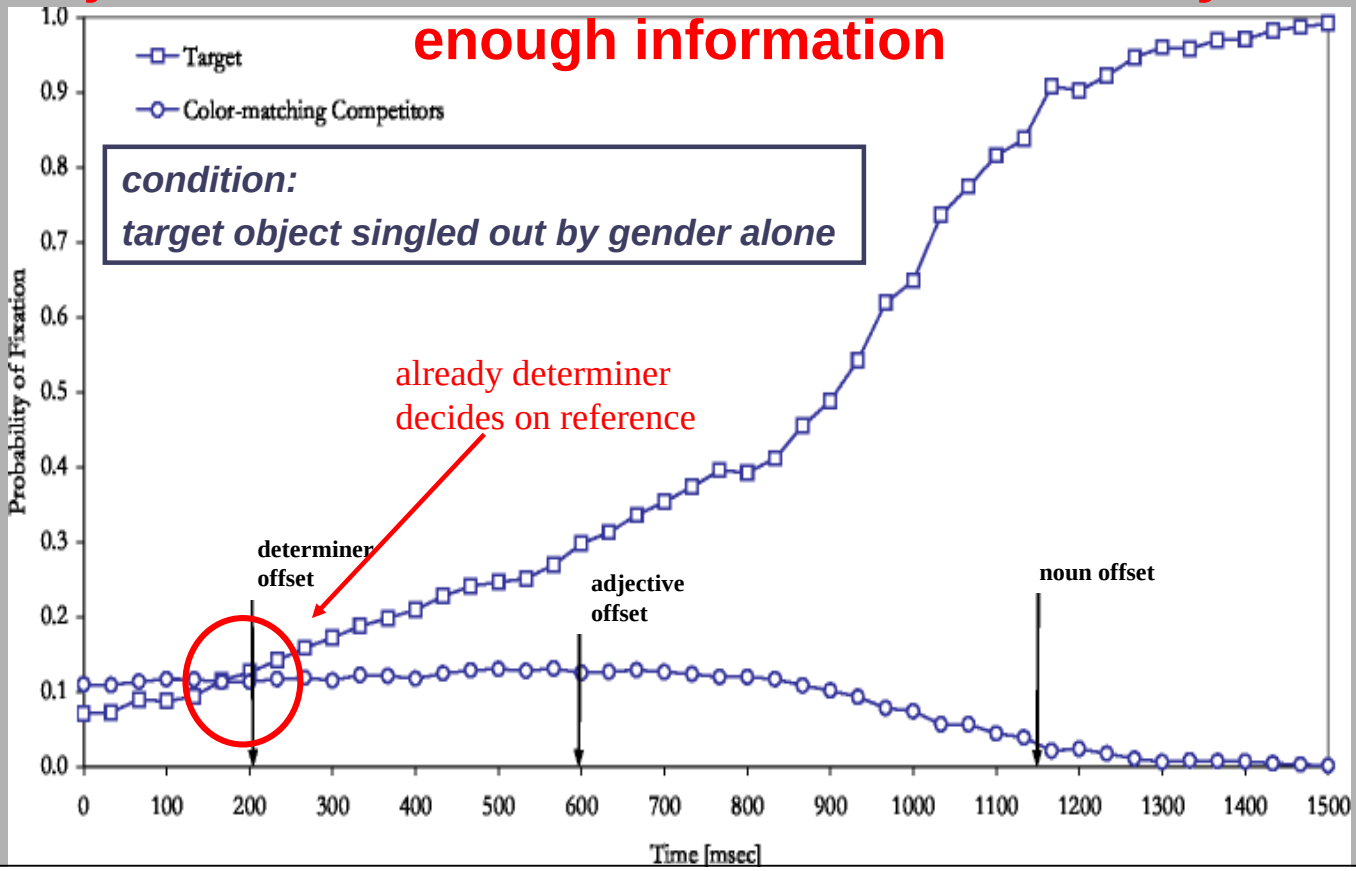
Klicken Sie auf die gelbe Giraffe

(die gelbe Giraffe/ die gelbe Banane/ das rote Hufeisen/der blaue Stern)



Klicken Sie auf die gelbe Giraffe
(die gelbe Giraffe/ das rote Auto/ der blaue Stern/ der grüne Baum)

subjects decide on reference as soon as they have enough information



Distributional Meaning is insufficient

Distributional Meaning will not get us any meaning for expression or utterance types that are not part of our linguistic experience:

> ***If a type has never been observed, and is still understood, it cannot have acquired its meaning via co-occurrence.***

This limits distributional meaning to expression types that re-occur in linguistic behaviour: non-composite, and in any case only lexically fixed expressions.

> ***Composite expressions (of which there are more than any of us will ever experience) must receive their meaning from their constituent parts.***

No account for linguistic productivity

From the point of view of linguistic theory, or any theory of linguistic behaviour, the limitation of DS to lexical expressions is serious:

> ***If, as many of us believe, linguistic productivity is a central property of human linguistic behaviour, then DS can at most be part of the story we need for semantics.***

> ***Association-based learning, for linguistic expressions as for other event types, can also be found in animal cognition and does not get at the core of the human linguistic faculty.***

Two notions of concepts: 1. Fregean concepts

Let ⟦*runs*⟧ be the denotation of the intransitive verb form *runs*.

CS models ⟦*runs*⟧ *as a Fregean concept,* or truth function:

> ***The concept ⟦runs⟧ is a function f_{runs} that is defined over a suitable domain D of arguments and assigns to each $a_i \in$ D either 1 of 0, depending on whether a_i runs or not.***

This is the simplest case; more complex denotations are modelled by more complex functions.

>> ***The concept f_{runs} thus models (part of) the knowledge a speaker must have in order to understand the verb form runs.***

Two notions of concepts: 2. Conceptual Distance

> ***In DS, ⟦runs⟧ can also be represented as a function, g_{runs} , that assigns to each lexical expression φ (φ ≠ runs) a numerical value for the frequency with which runs and φ co-occur in a corpus.***

The DS model of the concept ⟦runs⟧ thus may be interpreted as a measure for the semantic distance between the concepts ⟦runs⟧ and ⟦φ⟧.

>> ***The concept g_{runs} thus models a speaker's comprehension of the verb form runs as being closer to walks, presumably, than to rides, or drives, or drinks.***

Two notions of concepts & their relation

What is the relation between the two notions of concepts?

$$⟦runs⟧ = \lambda x \in D_{<denotation>}.\ f_{runs}(x)=1$$

$$⟦runs⟧ = \lambda \varphi \in D_{<expression>}.\ g_{runs}(\varphi)=n$$

The former is a function over entities, or more generally, ***denotations***: The argument in f_{runs} (x) must be a thing.

The latter is a function over ***linguistic expressions***: The argument in g_{runs} (φ) must be an expression.

And this is how it should be: f_{runs} is concerned with denotation, while g_{runs} is concerned with the distribution of expressions.

Two notions of concepts & their relation

$[\![\textbf{runs}]\!] = \lambda x \in D_{<\text{denotation}>}.\ f_{\text{runs}}(x)=1$

$[\![\textbf{runs}]\!] = \lambda \varphi \in D_{<\text{expression}>}.\ g_{\text{runs}}(\varphi)=n$

Now suppose the relation between expressions and concepts were 1:1 and we could speak of *the expression "runs"* and its denotation, *the concept* $[\![\textbf{runs}]\!]$, interchangeably.

The ***distributional distances of expressions*** would then be ***distances between CS concepts***, and we had a nice unification of DS and CS.

So why don't we do just that ?

Expressions are not denotations

Because it would simply be false:

"runs" does not stand in a 1:1 relation to any concept.

- "runs" is morphologically / syntactically ambiguous; it represents different grammatical words (syncretism)
- "runs" is polysemous: It denotes different concepts when applied to, e.g., people, cars, engines, watches, shows, water taps, etc.

and "runs" is in no way exceptional; most word forms in most languages are like this.

This affects both DS and CS concepts:

- ***Neither map 1:1 onto string types.***
- ***Neither distribution nor denotation are directly properties of string types.***

Contextual disambiguation

But suppose we had a method for "disambiguating" the expression – concept relation…

Suppose that the lexical expression already does this job, i.e., that

by knowing a lexical expression φ we already know that an occurrence of φ denotes a different concept in each context c:

$$\varphi(c) = [\![\varphi^c]\!]$$

"context" is here taken very broadly: Including syntactic, situational, discourse, and possibly further parameters.

I believe that the capacity for contextual disambiguation is part of the human cognitive abilities – Still: The modelling of this ability is one of the really big problems of NL semantics, and it is not specific to out current problem.

The different notions of "meaning"

Still, just supposing we already *had* a method for contextual disambiguation, then our two functions would be nicely complementary

$$[\![runs^c]\!] = \lambda x \in D_{<e>}.\boldsymbol{f}(x)=1$$

$$[\![runs^c]\!] = \lambda y \in D_{<e,t>}.\boldsymbol{g}(y)=n$$

- The former then gives us the contextual truth conditions under which $[\![runs^c]\!]$ applies, and
- the latter gives us the distance between $[\![runs^c]\!]$ and any other concepts in the domain of our distance function.

Different Notions of "Meaning"

The two functions

$[\![runs^c]\!] = \lambda x \in D_{<e>}.f(x)=1$

$[\![runs^c]\!] = \lambda y \in D_{<e,t>}.g(y)=n$

show us ***two sides of linguistically expressed context-specific concepts***, both of which are part of human linguistic capabilities:

- **judging a *concept's applicability to an argument*, and**
- **judging a *concept's closeness or similarity to other concepts***

Closeness is partly also covered by truth conditions (e.g. by semantic type relations) but non-truth-relevant experiential properties are not: This includes expectability properties, evaluative, affective, and also stylistic and register differences.

Conclusion

Cooperation between CS and TS seems to make sense
– but requires a lot more thinking.

Linguistic productivity will not go away – and there is not even a sketch of an account for it apart from what Compositional Semantics has to offer.

Distributional Semantics has the core of an account for learning lexical expressions from experience – which is unaccounted for in Compositional Semantics.

A common problem both need to tackle is context relativity. - A solution to contextual disambiguation probably also holds the key to a better understanding of the relation between denotations and distributions.