Perceptual Semantics and Coordination in Dialogue

Staffan Larsson

Department of Philosophy, Linguistics and Theory of Science
Centre of Language Technology
University of Gothenburg

PReLiM Workshop 2014-07-10

Outline

Introduction

- ► Main questions
  - ► How is linguistic meaning related to perception?
  - ► How do we learn and agree on shared meanings of words and expressions?
- ► Goal
  - ► Provide a formal semantic/pragmatic account addressing these questions

Outline

- ► We will present a dynamic semantic approach to perceptual aspects of meaning
- ► This shows how perceptual aspects of meaning can be incorporated with formal semantics
- ► Furthermore, we show how subsymbolic aspects of meaning can be updated as a result of observing language use in interaction, thereby enabling fine-grained semantic plasticity and semantic coordination.
- ► If time permits, one or more of: compositionality, vagueness, relation to possible worlds semantics

Outline

Semantic coordination

► Research on alignment shows that agents negotiate domain-specific microlanguages for the purposes of discussing the particular domain at hand (Clark and Wilkes-Gibbs, 1986; Garrod and Anderson, 1987; Pickering and Garrod, 2004; Brennan and Clark, 1996; Healey, 1997; Larsson, 2007)

► Two agents do not need to share exactly the same linguistic resources (grammar, lexicon etc.) in order to be able to communicate

► An agent's linguistic resources can change during the course of a dialogue when she is confronted with a (for her) innovative use

► *Semantic coordination*: the process of interactively coordinating the meanings of linguistic expressions.

Private perception and public meaning

▶ Perception, it seems, is inherently agent-specific – there is no notion of objective perception – so if we are to include perception in semantics we need to somehow explain how individual perception relates to public meaning.

▶ The key to this is semantic coordination
  ▶ By interacting with each other, agents reciprocally learn from each other and thereby come to have more or less coordinated (shared) takes on the world and on language (Fernández *et al.*, 2011)

▶ Interactive coordination and reciprocal learning require semantic plasticity, i.e. the ability to modify meanings.

▶ A requirement on our semantics is therefore that it enables the kinds of modifications needed to account for semantic coordination of perceptual meanings.

Mechanisms for semantic coordination in dialogue

Some mechanisms for semantic coordination in dialogue:

▶ Corrective feedback (a.k.a. embedded correction), where one speaker implicitly corrects the way an expression is used by another speaker (Clark, 2007; Clark and Wong, 2002; Saxton, 1997; Saxton, 2000; Larsson and Cooper, 2009)

▶ Explicit definitions and negotiations of meanings (Linell and Noren, 2005; Ludlow, 2014)

▶ **"Silent" coordination, by speakers observing the language use of others and adapting to it** (Carey and Bartlett, 1978; Larsson, 2010)

The left-or-right game

- ▶ A and B are facing a framed surface on a wall, and A has a bag of objects which can be attached to the framed surface.
- ▶ A round of the game is played as follows:
    1. A places an object in the frame
    2. B orients to the new object
    3. A says either "left" or "right"
    4. B interprets A's utterance based on B's take on the situation.
    5. If (B's interpretation of) A's utterance is consistent with B's take on the situation, B assumes A is right, says "aha", and learns from this exchange; otherwise, B says "okay"

- ▶ The left-or-right game can be regarded as a considerably pared-down version of the "guessing game" in Steels and Belpaeme (2005), where perceptually grounded colour terms are learnt from interaction.

- ▶ The kinds of meanings learnt in the left-or-right game may be considered trivial.

- ▶ However, at the moment we are mainly interested in the basic principles of combining formal semantics with learning of perceptual meaning from dialogue

- ▶ The hope is that these can be formulated in a general way which can later be used in more interesting settings.

Outline

The Symbol Grounding Problem

▶ If a speaker of English is unable to distinguish gloves from mittens, most people would probably agree that something is missing in this person's knowledge of the meaning of "glove".

▶ Similarly, if we tell A to find some nice pictures of dogs chasing cats, and A comes back happily with an assortment of pictures displaying lions chasing zebras, we would question whether A really knows the full meaning of the words "dog" and "cat"

Perception and meaning

- ▶ Part of learning a language is learning to identify individuals and situations that are in the extension of the phrases and sentences of the language

- ▶ For many concrete expressions, this identification relies crucially on the ability to
  - ▶ perceive the world
  - ▶ use perceptual information to classify individuals and situations as falling under a given linguistic description or not

- ▶ This view was put forward by (Harnad, 1990) as a way of addressing the "symbol grounding problem" in artificial intelligence:

  How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?"

  (Harnad, 1990)

How to solve the symbol grounding problem

- ▶ Harnad's own sketch of a solution to the symbol grounding problem:
  - ▶ A **hybrid** system encompassing both symbolic and non-symbolic representations, the latter such that they "can pick out the objects to which they refer, via connectionist networks that extract the invariant features of their analog sensory projections"
  - ▶ **Learning** non-symbolic representations from interaction; "a connectionist network that learns to identify icons correctly from the sample of confusable alternatives it has encountered by dynamically adjusting the weights of the features"
  - ▶ **Compositionality**, where complex constructions "will all inherit the intrinsic grounding of [the grounded set of elementary symbols]"
- ▶ All these components are needed for a solution to the symbol grounding problem
- ▶ We follow these ideas, specify them further and formalize them

Statistical classifiers

► Harnad proposed using connectionist networks to ground symbols

► Connectionist networks are one kind of *(statistical) classifier*, a computational device determining what class an item belongs to, based on various properties of the item.

► Crucially, these properties need not be encoded in some high-level representation language (such as logic or natural language)

► Instead, it may consist entirely of numeric data encoding more or less "low-level" information about the item in question, for example perceptual data.

Classifiers, intensions and extensions

▶ Classifiers can be defined formally as functions (or programs)

▶ Typically, the domain of a classifier function is numerical (e.g. real-valued, integer or binary) vectors and the range is a set of categories

▶ When making use of classifiers in formal semantics we will regard them as (parts of) representations of (agents' takes on) intensions of linguistic expressions.

▶ Classifiers (as intensions) produce judgements whether some perceived thing or situation falls within the extension of a linguistic expression

Perceptual meaning

▶ Perceptual meaning is an important aspect of the meaning of linguistic expressions referring to physical objects (such as concrete nouns or noun phrases).

▶ Knowing the perceptual meaning of an expression allows an agent to identify perceived objects and situations falling under the meaning of the expression.

▶ For example, knowing the perceptual meaning of "blue" would allow an agent to correctly identify blue objects.

▶ Similarly, an agent which is able to compute the perceptual meaning of "a boy hugs a dog" will be able to correctly classify situations where a boy hugs a dog.

Using classifiers to represent perceptual meanings

▶ Steels & Belpaeme (2005): Robots coordinating on colour terms through a simple language game of pointing and guessing; meanings of colour terms are captured in (weight vectors describing) neural networks; utterances describe single objects

▶ This can be seen as a further specification implementation of Harnad's ideas, adding interaction to the mix

▶ We follow Steels & Belpaeme in representing (takes on) meanings using classifiers, and training these classifiers based on dialogue interaction

▶ We add a connection to formal semantics as well as an account of compositionality

Outline

Formal semantics for perceptual meanings

- ▶ We want to integrate perceptual meanings and low-level perceptual data into formal semantics
- ▶ This means mixing low-level (perceptual) and high-level (logical-inferential) meaning in a single framework
  - ▶ A hybrid system, as proposed by Harnad
- ▶ To enable learning and coordination, we need a framework where intensions are (1) represented independently of extensions, and are (2) structured objects which can be modified (updated)

Formal semantics for perceptual meaning

► We want to use a framework which also encompasses accounts of many problems traditionally studied in formal semantics[1]

► We will be using Type Theory with Records, or TTR.

► As many other type theories, TTR is based on the notion of *judgements* of entities being of certain types – for example, a judgement that a certain situation is of a certain type.

► TTR starts from the idea that information and meaning is founded on our ability to perceive and classify the world, i.e., to perceive objects and situations as being of types.

---

[1]Semantic phenomena which have been described using TTR include modelling of intensionality and mental attitudes (Cooper, 2005), dynamic generalised quantifiers (Cooper, 2004), co-predication and dot types in lexical innovation, frame semantics for temporal reasoning, reasoning in hypothetical contexts (Cooper, 2011), enthymematic reasoning (Breitholtz and Cooper, 2011), clarification requests (Cooper, 2010), negation (Cooper and Ginzburg, 2011), and information states in dialogue (Cooper, 1998; Ginzburg, 2012).

Why TTR?

TTR is well suited for dealing with the problems we are interested in.

- ▶ Types are first-class objects, which allows perceptual classifier functions to be formalised
- ▶ TTR integrates logical techniques such as binding and the lambda-calculus into feature-structure like objects called record types.
- ▶ More structure than in a traditional formal semantics and more logic than is available in traditional unification-based systems.
- ▶ Feature structure like properties are important for the straightforward definition of meaning modifications involving refinement and generalization.
- ▶ Logical aspects are important for relating our semantics to the model and proof theoretic tradition associated with compositional semantics.

TTR: An extremely brief introduction

We can here only give a brief and partial introduction to TTR; see also Cooper (2005) and Cooper (2012).

- $a : T$ means that a is of type T
- One *basic type* in TTR is Ind, the type of an individual
- Another basic type is Real, the type of real numbers.
- Given that T_1 and T_2 are types, $T_1 \rightarrow T_2$ is a *function type* whose domain is objects of type T_1 and whose range is objects of type T_2.

Records and record types

- If $a_1 : T_1, a_2 : T_2(a_1), \ldots, a_n : T_n(a_1, a_2, \ldots, a_{n-1})$
- where $T(a_1, \ldots, a_n)$ represents a type T which depends on the objects $a_1, \ldots, a_n$
- the record to the left is of the record type to the right:

$$
\begin{bmatrix}
\ell_1 & = & a_1 \\
\ell_2 & = & a_2 \\
\ldots & & \\
\ell_n & = & a_n \\
\ldots & &
\end{bmatrix}
:
\begin{bmatrix}
\ell_1 & : & T_1 \\
\ell_2 & : & T_2(l_1) \\
\ldots & & \\
\ell_n & : & T_n(\ell_1, l_2, \ldots, l_{n-1})
\end{bmatrix}
$$

- $\ell_1, \ldots \ell_n$ are *labels* which can be used elsewhere to refer to the values associated with them.

Dependent types, ptypes and proofs

A sample record and record type:

$$
\left[\begin{array}{lll} \text{ref} & = & \text{obj}_{123} \\ c_{\text{man}} & = & \text{prf}_{456} \\ c_{\text{run}} & = & \text{prf}_{678} \end{array} \right] : \left[\begin{array}{lll} \text{ref} & : & \text{Ind} \\ c_{\text{man}} & : & \text{man}(\text{ref}) \\ c_{\text{run}} & : & \text{run}(\text{ref}) \end{array} \right]
$$

▶ Types constructed with predicates may be *dependent*.

▶ Above, the type of c_{man} is dependent on ref (as is c_{run}).

▶ Types can be constructed from predicates, e.g., "run" or "man"

▶ Such types are called *ptypes* and correspond roughly to propositions in first order logic.

▶ "propositions are types of proofs": something of a ptype $P(a_1, \ldots, a_n)$ is whatever it is that counts as a proof of $P(a_1, \ldots, a_n)$.

Paths and nesting

- If r is a record and ℓ is a label in r, we can use a *path* $r.\ell$ to refer to the value of ℓ in r
- Similarly, if T is a record type and ℓ is a label in T, $T.\ell$ refers to the type of ℓ in T.
- Records (and record types) can be nested, so that the value of a label is itself a record (or record type).

Manifest fields

Some of our types will contain *manifest fields* (Coquand *et al.*, 2004) like the c_{man}-field:

$$\left[\begin{array}{lll} \text{ref} & : & \text{Ind} \\ c_{man}{=}prf_{23} & : & man(ref) \end{array}\right]$$

- $\left[\begin{array}{lll} c_{man}{=}prf_{23} & : & man(ref) \end{array}\right]$ is a convenient notation for $\left[\begin{array}{lll} c_{man} & : & man(ref)_{prf_{23}} \end{array}\right]$ where $man(ref)_{prf_{23}}$ is a *singleton type*
- If $a : T$, then T_a is a singleton type and $b : T_a$ iff $b = a$
- Manifest fields allow us to progressively specify what values are required for the fields in a type.

Outline

The Perceptron

- ▶ As a simple example of how perceptual classifiers can be integrated in formal semantics, we will use the perceptron (Rosenblatt, 1958)
- ▶ Classification of perceptual input can be regarded as a mapping of sensor readings (corresponding to situations) to types
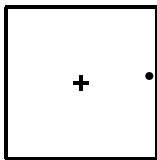- ▶ The perceptron is a very simple neuron-like object with several inputs and one output.

$$o(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w} \cdot \mathbf{x} > t \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbf{w} \cdot \mathbf{x} = \sum_{i=1}^{n} w_i x_i = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n$

- ▶ Limited to learning problems which are linearly separable; the distinction between left and right is one such problem.

Classifying objects as being to the left or to the right

- ▶ Suppose we have a square surface, and object are placed on the surface
- ▶ To classify objects as being to the right or not:
  - ▶ Direct a sensor (e.g. a camera) towards the surface
  - ▶ Get a sensor reading (a picture from the camera)
  - ▶ Apply an algorithm which returns a vector of the coordinates of the object on the surface (assuming there is only one); this is a slightly higher-level rendering of our initial sensor reading
  - ▶

Classifying objects as being to the left or to the right

- ▶ Suppose we have a square surface, and object are placed on the surface
- ▶ To classify objects as being to the right or not:
  - ▶ Direct a sensor (e.g. a camera) towards the surface
  - ▶ Get a sensor reading (a picture from the camera)
  - ▶ Apply an algorithm which returns a vector of the coordinates of the object on the surface (assuming there is only one); this is a slightly higher-level rendering of our initial sensor reading
  - ▶ Apply a perceptron classifier to the coordinate vector and returns 1 or 0

$$\Rightarrow 1$$

Staffan Larsson (GU)   Perceptual Semantics
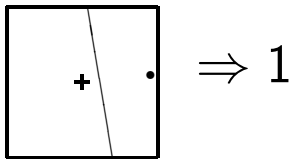
The TTR perceptron cont'd

A TTR perceptron classifier can be represented as a record:

$$\kappa = \left[\begin{array}{rcl} \mathsf{w} & = & \begin{bmatrix} 0.800 & 0.010 \end{bmatrix} \\ \mathsf{t} & = & 0.090 \\ \mathsf{f} & = & \lambda v : \mathsf{RealVector}. \; (\left\{ \begin{array}{ll} 1 & \text{if } v \cdot \mathsf{w} > \mathsf{t} \\ 0 & \text{otherwise} \end{array} \right) \end{array} \right]$$

Where $\kappa.\mathsf{f}$ will evaluate to

$$\lambda v : \mathsf{RealVector}.(\left\{ \begin{array}{ll} 1 & \text{if } v \cdot \begin{bmatrix} 0.800 & 0.010 \end{bmatrix} > 0.090 \\ 0 & \text{otherwise} \end{array} \right)$$

This representation allows training the classifier by modifying $\kappa.\mathsf{w}$ and $\kappa.\mathsf{t}$

The TTR perceptron

- ▶ The basic perceptron returns a real-valued number (1 or 0) but when we use a perceptron as a classifier of situations we want it to instead return a type.
- ▶ Typically, such types will be built from a predicate and some number of arguments; a type of proof, or a "proposition".
- ▶ A TTR classifier perceptron for a type P can be represented as a record:

$$
\kappa = \left[
\begin{array}{lll}
\mathsf{w} & = & \begin{bmatrix} 0.800 & 0.010 \end{bmatrix} \\
\mathsf{t} & = & 0.090 \\
\mathsf{f} & = & \lambda v : \mathsf{RealVector}.\left(\left\{ \begin{array}{ll} P & \text{if } v \cdot \mathsf{w} > \mathsf{t} \\ \neg P & \text{otherwise} \end{array} \right.\right)
\end{array}
\right]
$$

Staffan Larsson (GU)   Perceptual Semantics
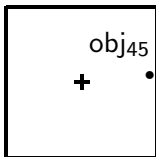
The meaning of "(that is to the) right" in TTR

Using a TTR classifier perceptron to represent a agent's take on perceptual meaning:

$[\text{right}]^{Agt} =$

$$
\begin{bmatrix}
w = \begin{bmatrix} 0.800 & 0.010 \end{bmatrix} \\
t = 0.090 \\
bg = \begin{bmatrix} sr_{pos} & : & \text{RealVector} \\ foo & : & \text{Ind} \\ spkr & : & \text{Ind} \end{bmatrix} \\
f = \lambda r : bg. \left(\begin{bmatrix} c_{right} = \begin{bmatrix} foo = r.foo \\ sr_{pos} = r.sr_{pos} \end{bmatrix} \end{bmatrix} : \left\{ \begin{array}{ll} \text{right}(r.foo) & \text{if } r.sr_{pos} \cdot w > t \\ \neg\, \text{right}(r.foo) & \text{otherwise} \end{array} \right. \right)
\end{bmatrix}
$$

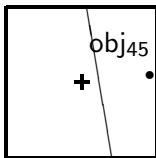# Classifying objects as being to the right or not, TTR style

- ▶ Representation of current situation s
  - ▶ Coordinates of object in focus of attention
  - ▶ Label for object (obj_{45})

$$s = \left[\begin{array}{lcl} sr_{pos} & = & \begin{bmatrix} 0.900 & 0.100 \end{bmatrix} \\ foo & = & obj_{45} \\ spkr & = & A \end{array} \right]$$

- ▶ Apply [right] to s:

$$\left[c_{right} = \begin{bmatrix} foo & = obj_{45} \\ sr_{pos} & = \begin{bmatrix} 0.900 & 0.100 \end{bmatrix} \end{bmatrix} : right(obj_{45}) \right]$$

Outline

Updating perceptual meaning

Perceptrons are updated using the *perceptron training rule*:

$$w_i \leftarrow w_i + \Delta w_i$$

where

$$\Delta w_i = \eta(o_t - o)x_i$$

where o_t is the target output, o is the actual output, and w_i is associated with input x_i.

▶ Note that if $o_t = o$, there is no learning.

▶ This rule can be formulated as a TTR update function (see Larsson, 2013)

▶ In the LoR-game, training results in moving the line dividing "(to the) right" from "not (to the) right"
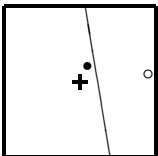
Agent B's initial take on the meaning of "right":

$[\text{right}]^{Agt} =$

$$
\begin{bmatrix}
\text{w} = \begin{bmatrix} 0.800 & 0.010 \end{bmatrix} \\
\text{t} = 0.090 \\
\text{bg} = \begin{bmatrix} \text{sr}_{\text{pos}} & : & \text{RealVector} \\ \text{foo} & : & \text{Ind} \\ \text{spkr} & : & \text{Ind} \end{bmatrix} \\
\text{f} = \lambda r : \text{bg.} \left(\begin{bmatrix} \text{c}_{\text{right}} = \begin{bmatrix} \text{foo} = r.\text{foo} \\ \text{sr}_{\text{pos}} = r.\text{sr}_{\text{pos}} \end{bmatrix} \end{bmatrix} : \left\{ \begin{array}{ll} \text{right}(r.\text{foo}) & \text{if } r.\text{sr}_{\text{pos}} \cdot \text{w} > \text{t} \\ \neg\, \text{right}(r.\text{foo}) & \text{otherwise} \end{array} \right. \right)
\end{bmatrix}
$$
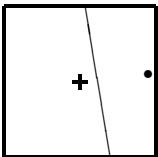
A: "right"
B: "okay"

A: "right"

- B's classifier applied to this situation yields that the object is not to the right
- B applies the perceptron training rule to adjust the classifier

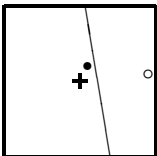Agent B's revised take on the meaning of "right":
$[\text{right}]^{Agt} =$

$$
\begin{bmatrix}
\text{w} = \begin{bmatrix} \mathbf{0.808} & \mathbf{0.200} \end{bmatrix} \\
\text{t} = 0.090 \\
\text{bg} = \begin{bmatrix} \text{sr}_{\text{pos}} & : & \text{RealVector} \\ \text{foo} & : & \text{Ind} \\ \text{spkr} & : & \text{Ind} \end{bmatrix} \\
\text{f} = \lambda r :\text{bg}.\left(\begin{bmatrix} \text{c}_{\text{right}} = \begin{bmatrix} \text{foo} = r.\text{foo} \\ \text{sr}_{\text{pos}} = r.\text{sr}_{\text{pos}} \end{bmatrix} \end{bmatrix} : \left\{ \begin{array}{ll} \text{right}(r.\text{foo}) & \text{if } r.\text{sr}_{\text{pos}} \cdot \text{w} > \text{t} \\ \neg\, \text{right}(r.\text{foo}) & \text{otherwise} \end{array} \right. \right)
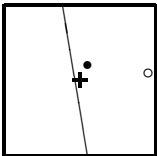\end{bmatrix}
$$
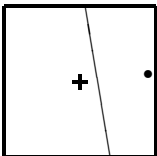
A: "right"

B: "okay"
A: "right"

B: "aha"

From learning to coordination

- ▶ In the left-or-right game, as described above, there is an asymmetry in that agent A is assumed to be fully competent at judging whether objects are to the right or not, whereas agent B is to learn this.
- ▶ By contrast, when humans interact they *mutually* adapt to each others' language use on multiple levels
  - ▶ alignment (Pickering and Garrod, 2004), entrainment (Brennan, 1996), negotiation (Mills and Healey, 2008) or coordination (Garrod and Anderson, 1987; Healey, 1997; Larsson, 2007)
- ▶ The LoR game could quite easily be altered to illustrate coordination directly
  - ▶ Let A and B switch roles after each round
  - ▶ In this symmetric LoR game, the agents would converge on a meaning of "right" that neither of them may subscribe to initially.

Social meanings and individual representations

▶ We take it that a central task of semantic theory is to model semantic plasticity and semantic coordination

▶ By modelling how individuals (1) represent meanings, (2) use meanings to form judgements and (3) coordinate on meanings and judgements, we indirectly model the emergence, perpetuation and variation of meaning in a linguistic community.

▶ Although perception and mental representations concern individual agents, meaning itself is inherently social and dependent on learning and adaptation through interaction.

▶ This view implies that a central task of semantic theory is to model semantic plasticity and semantic coordination, i.e. how meanings change as a result of language use in interaction.

Outline

Compositionality

- ▶ A crucial step in demonstrating the usefulness of the proposed approach is to show how the principle of compositionality can be applied also to perceptual meaning
- ▶ Exploring compositionality in something like the left-or-right game requires extending it.
  - ▶ add more words (e.g. "upper" and "lower") and some simple grammar ("upper left", "lower right" etc).

Compositionality: Basic Example

▶ Proof of concept of compositionality: show how to compute the meaning of "upper right" from the meanings of "upper" and "right".

$[\text{upper}]^B =$

$$
\begin{bmatrix}
\text{w}_{\text{upper}} = \ldots \\
\text{t}_{\text{upper}} = \ldots \\
\text{bg} = \begin{bmatrix} \text{sr}_{\text{pos}} & : & \text{RealVector} \\ \text{foo} & : & \text{Ind} \\ \text{spkr} & : & \text{Ind} \end{bmatrix} \\
\text{f} = \lambda r : \text{bg}(\begin{bmatrix} \text{c}_{\text{upper}} = \begin{bmatrix} \text{sr}_{\text{pos}} = r.\text{sr}_{\text{pos}} \\ \text{foo} = r.\text{foo} \end{bmatrix} : \pi_{\text{upper}}(\text{w}_{\text{upper}}, \text{t}_{\text{upper}})(r) \end{bmatrix})
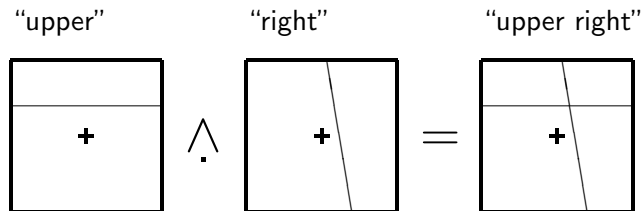\end{bmatrix}
$$

Compositionality: Basic Example

Compositional meaning of "upper right" obtained by merging of meanings of "upper" and "right"

$[\text{upper right}]_B = [\text{upper}]_B \wedge [\text{right}]_B =$

$$
\begin{bmatrix}
w_{upper} = \ldots \\
t_{upper} = \ldots \\
w_{right} = \ldots \\
t_{right} = \ldots \\
bg = \begin{bmatrix} sr_{pos} & : & \text{RealVector} \\ foo & : & \text{Ind} \\ spkr & : & \text{Ind} \end{bmatrix} \\
f = \lambda r{:}bg(\begin{bmatrix} c_{upper} = \begin{bmatrix} sr_{pos} = r.sr_{pos} \\ foo = r.foo \end{bmatrix}{:}\pi_{upper}(w_{upper}, t_{upper})(r) \\ c_{right} = \begin{bmatrix} sr_{pos} = r.sr_{pos} \\ foo = r.foo \end{bmatrix}{:}\pi_{right}(w_{right}, t_{right})(r) \end{bmatrix})
\end{bmatrix}
$$

Compositionality: Basic Example

"upper" "right" "upper right"

Compositionality: Degree modifiers

- What are the compositional semantics for degree modifiers, e.g. "far" in "far right"?
- Proposal: "far" takes parameters of the "right" classifier and yields modified classifier for "far rightness" (increased threshold)
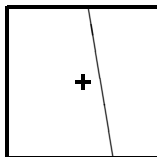
$$[far] = \begin{bmatrix} \alpha = 1.4 \\ f = \lambda m : \begin{bmatrix} t & : & \text{Real} \end{bmatrix} \\ \quad (m \; \boxed{\wedge} \; \begin{bmatrix} t & = & \alpha * m.t \end{bmatrix}) \end{bmatrix}$$

$$[far\ right] = [far].f([right]) =$$

$$\begin{bmatrix} t = 0.090 \\ bg = \dots \\ f = \dots \end{bmatrix} \boxed{\wedge} \begin{bmatrix} t & = & 1.4*0.090 \end{bmatrix} = \begin{bmatrix} t = 0.126 \\ bg = \dots \\ f = \dots \end{bmatrix}$$

Compositionality: Degree modifiers

"right":

"far right":
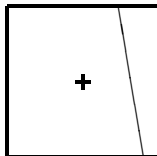
Outline

Vagueness

- ▶ (Based on Fernández and Larsson, 2014)
- ▶ In the perceptron account, perceptual meanings are categorical
- ▶ However, most (if not all) perceptual meanings are more or less vague
- ▶ Many are also context-dependent (e.g. '*tall*')

- ▶ Case study: scalar predicates
    - ▶ e.g. '*tall*', '*long*' and '*expensive*'
    - ▶ Interpreted with respect to a scale, i.e., a dimension such as height, length, or cost along which entities for which the relevant dimension is applicable can be ordered.
    - ▶ Have a relatively simple semantics (they are often uni-dimensional) and thus constitute a perfect case-study for investigating the properties and effects of vagueness on language use.

- ▶ (However, our account should also work for n-dimensional concepts, e.g. colours, shapes)

Modeling vagueness using a noisy threshold

▶ There are several ways in which one can account for vagueness—amongst others, by introducing perceptual uncertainty (possibly inaccurate sensor readings).

▶ Here, in line with Lassiter (2011), we opt for substituting the precise threshold with a noisy, probabilistic threshold.

▶ We consider the threshold to be a normal random variable, which can be represented by the parameters of its Gaussian distribution, the mean μ and the standard deviation σ (the noise width).

  ▶ Which noise function may be the most appropriate is an empirical question we do not tackle here.

  ▶ Our choice of Gaussian noise follows Schmidt *et al.* (2009).

The meaning of '*Tall*'

$$\textbf{tall} = \begin{bmatrix} \mathsf{T}_{ctxt} = \begin{bmatrix} \mathsf{c} & : & \mathsf{Type} \\ \mathsf{x} & : & \mathsf{c} \\ \mathsf{h} & : & \mathbb{R}^{+} \end{bmatrix} \\ \mu & = \mu_{tall} \\ \sigma & = \sigma_{tall} \\ \mathsf{f} & = \lambda r : \mathsf{T}_{ctxt}. \begin{bmatrix} \mathsf{sit} = r \\ \mathsf{sit\text{-}type} = \begin{bmatrix} \mathsf{c}_{tall} : \mathsf{tall}(r.\mathsf{x}) \end{bmatrix} \\ \mathsf{prob} = \kappa_{tall}(\sigma, \mu, r) \end{bmatrix} \end{bmatrix}$$

- $\mathsf{T}_{ctxt}.\mathsf{c}$ is the comparison class (allowing us to model context sensitivity)
- $\mathsf{T}_{ctxt}.\mathsf{x}$ is an individual of type $\mathsf{T}_{ctxt}.\mathsf{c}$
- The output of the function **tall**.f is now a *probabilistic* Austinian proposition (Cooper *et al.*, 2014).

Staffan Larsson (GU)   Perceptual Semantics

A classifier for tallness

- ▶ We define a tallness classifier κ_{tall} that takes as parameters μ_{tall} and σ_{tall}, both of them dependent on a comparison class and hence of type $\text{Type} \rightarrow \mathbb{R}^+$.
  - ▶ The comparison class here specifies a type, e.g. *Human*, *Child* or *BasketballPlayer*
- ▶ The output of the classifier is a probability

$$\kappa_{tall}(\mu, \sigma, r) = \frac{1}{2} \left[1 + \text{erf}\left(\frac{r.\mathsf{h} - \mu(r.\mathsf{c})}{\sigma(r.\mathsf{c})\sqrt{2}} \right) \right]$$

$$\kappa_{tall} : (\text{Type} \rightarrow \mathbb{R}^+, \text{Type} \rightarrow \mathbb{R}^+, T_{ctxt}) \rightarrow [0, 1]$$

Staffan Larsson (GU)　　　　　　Perceptual Semantics

▶ Here erf is the error function, defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{t=0}^{x} e^{-t^2} \mathrm{d}t$$

▶ The error function defines a sigmoid shape, in line with the upward monotonicity of '*tall*'.

▶ The output of $\kappa_{tall}(\mu, \sigma, r)$ corresponds to the probability that h will exceed the normal random threshold with mean μ and deviation σ.

Staffan Larsson (GU)   Perceptual Semantics
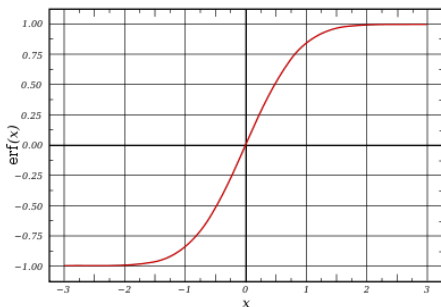
Example

- Assume that we have $\mu_{tall}(\text{Human})=1.87$ and $\sigma_{tall}(\text{Human})=0.05$.

- Let's also assume $ctxt = \begin{bmatrix} c & = & \text{Human} \\ x & = & \text{john\_smith} \\ h & = & 1.88 \end{bmatrix}$

- In this case, **tall**.f($ctxt$) will compute as follows:

$$\lambda r : T_{ctxt}. \begin{bmatrix} \text{sit} = r \\ \text{sit-type} = \begin{bmatrix} c_{tall} : \text{tall}(r.x) \end{bmatrix} \\ \text{prob} = \kappa_{tall}(\mu_{tall}, \sigma_{tall}, r) \end{bmatrix} (\begin{bmatrix} c & = & \text{Human} \\ x & = & \text{john\_smith} \\ h & = & 1.88 \end{bmatrix}) =$$

$$\begin{bmatrix} \text{sit} = \begin{bmatrix} c & = & \text{Human} \\ x & = & \text{john\_smith} \\ h & = & 1.88 \end{bmatrix} \\ \text{sit-type} = \begin{bmatrix} c_{tall} : \text{tall}(\text{john\_smith}) \end{bmatrix} \\ \text{prob} = 0.579 \end{bmatrix}$$

since $\kappa_{tall}(\mu_{tall}, \sigma_{tall}, \begin{bmatrix} \text{c=Human} \\ \text{x=john\_smith} \\ \text{h=1.88} \end{bmatrix}) = \frac{1}{2} \left[1 + \text{erf}\left(\frac{1.88-1.87}{0.05\sqrt{2}} \right) \right] = 0.579$

- This probability can now be used in further probabilistic reasoning, to decide whether to refer to an individual x as tall, or to evaluate someone else's utterance describing x is tall.

- For example, an agent may map different probabilities to different adjective qualifiers of tallness to yield compositional phrases such as '*sort of tall*', '*quite tall*', '*very tall*', '*extremely tall*', etc.

- The meanings of these composed adjectival phrases could specify probability ranges trained independently.

- Compositionality for vague perceptual meanings, and the interaction between compositionality and learning, is an interesting area for future research.

Computing the Noisy Threshold

▶ for a vague scalar predicate like '*tall*', we assume that an agent will have at its disposal a set of *observations* Ω_{tall}^T consisting of entities of a particular type T (a comparison class such as *Human*) that have been judged to be tall, together with their observed heights.

▶ Different functions can be used to compute μ_{tall} and σ_{tall} from Ω_{tall}^T.

▶ What constitutes an appropriate function for a certain predicate is an empirical matter; Schmidt *et al.* (2009) collect judgements of people asked to indicate which items are tall given distributions of items of different heights.

Computing the Noisy Threshold, cont'd

▶ The best performing threshold model in their study is the *relative height by range* model, where (in our notation):

$$\mu_{tall}(T) = \max(\Omega_{tall}^T) - k \cdot (\max(\Omega_{tall}^T) - \min(\Omega_{tall}^T))$$

▶ $\max(\Omega_{tall}^T)$ and $\min(\Omega_{tall}^T)$ stand for the maximum and the minimum height, respectively

▶ The model includes two parameters, k and a noise-width parameter that in our approach corresponds to σ_{tall}.

  ▶ Any item within the top $k\%$ of the range of heights that have been judged to be tall counts as tall.
  ▶ Schmidt *et. al.* report that the best fit of their data was obtained with $k = 29\%$ and $\sigma_{tall} = 0.05$.

Updating Vague Meanings

▶ How is the vague meaning of '*tall*' updated as an agent is exposed to new judgements via language use?

▶ If a new entity $x : T$ with height h is referred to as tall, the agent adds h to its set of observations Ω_{tall}^{T} and recomputes $\mu_{tall}(\text{Human})$, for instance using RH-R

▶ This in turn will trigger an update to the probability outputted by κ_{tall}.

Connection to probabilistic TTR

▶ Generally, we want classifiers for vague perceptual terms which take real-valued input (derived from sensor input) and give probabilistic judgements as output

▶ These judgements can be used as input to probabilistic reasoning

▶ For example, we can imagine an agent having vague and context-sensitive classifiers for shape and colour, taking real-valued vector input derived from digitized pictures

▶ The output of these classifiers can be used as input to a classifier of objects, e.g. fruits, in a Bayes net

▶ The fruit classifier would be used to specify the perceptual meanings of words denoting fruits ('apple', 'pear', 'orange' etc.)

▶ All classifiers are continually updated as interaction proceeds (semantic plasticity)

Learning in probabilistic TTR

The fruit classifier would be trained from interaction using the learning theory of probabilistic TTR (Cooper *et al.*, 2014; below is a modified version)

κ: $\mathsf{Sit} \to \mathsf{Set}(\left[\begin{smallmatrix} \text{sit} & : & \text{Sit} \\ \text{sit-type} & : & \text{Type} \\ \text{prob} & : & [0,1] \end{smallmatrix}\right])$ such that if $s{:}\mathsf{Sit}$ then

$$\kappa(s) = \left\{ \left[\begin{smallmatrix} \text{sit} & = & s \\ \text{sit-type} & = & T \\ \text{prob} & = & p_{A,\mathfrak{J}}(s : T \mid s : T_{e_1}, \ldots, s : T_{e_n}) \\ & & p_{A,\mathfrak{J}}(s : T_{e_1}) \ldots p_{A,\mathfrak{J}}(s : T_{e_n}) \end{smallmatrix}\right] \middle| \; T \in \langle T_{c_1}, \ldots, T_{c_m} \rangle \right\}$$

where

- $p_{A,\mathfrak{J}}(r : T_c \mid r : T_{e_1}, \ldots, r : T_{e_n}) = \mathsf{prior}_{\mathfrak{J}}(T_c) \frac{p_{A,\mathfrak{J}}(s : T_{e_1} \mid s : T_c) \ldots p_{A,\mathfrak{J}}(s : T_{e_n} \mid s : T_c)}{\mathsf{prior}_{\mathfrak{J}}(T_{e_1}) + \ldots + \mathsf{prior}_{\mathfrak{J}}(T_{e_n})}$

- $p_{A,\mathfrak{J}}(s : T_1 \mid s : T_2) = \frac{||T_1 \wedge T_2||_{\mathfrak{J}}}{||T_2||_{\mathfrak{J}}}$, if $|| T_2 ||_{\mathfrak{J}} \neq 0$, and 0 otherwise.

- $\mathsf{prior}_{\mathfrak{J}}(T) = \frac{||T||_{\mathfrak{J}}}{\mathcal{P}(\mathfrak{J})} = \frac{\sum_{j \in \mathfrak{J}_T} j.\mathsf{prob}}{\sum_{j \in \mathfrak{J}} j.\mathsf{prob}}$ if $\mathcal{P}(\mathfrak{J}) > 0$, and 0 otherwise.

- An agent, A, makes judgements based on a finite string of probabilistic Austinian propositions, $\mathfrak{J}$

- For a type, T, $\mathfrak{J}_T = \{j \mid j \in \mathfrak{J}$ and $j.\mathsf{sit\text{-}type} \sqsubseteq T\}$

- ▶ Putting this together with other semantic phenomena modeled in TTR, we can model, for example, how an agent who has learnt the meanings of "tall", "pear", "far" and "right" could take a sentence like "Most of the tall pears are to the far right" and correctly classify visual scenes as falling under this description or not
- ▶ The idea is that this approach can be gradually extended to larger fragments of the language used to refer to perceived situations

Outline

Classifiers in possible worlds semantics?

▶ The work presented here builds on Larsson (2011), where the idea of including perceptual meaning (modeled as classifiers of low-level perceptual input) into a compositional formal semantics is introduced.

▶ Recently, researchers in computational semantics have begun exploring the idea of connecting perceptual classifiers to compositional formal semantic representations (Matuszek *et al.*, 2012; Krishnamurthy and Kollar, 2013).

▶ Although motivated by the practical problem of allowing robots to learn to connect language and the world, this work connects to semantic theory by using meaning representations similar to those used in possible worlds semantics.

▶ This leads to the issue regarding if and how classifiers could be included into possible worlds semantics.

Barker (2002)

▶ Barker (2002), in his accounting of learning the meaning of "tall" seems to provide a method which could be used for implementing classifiers extensionally in (an extension of) possible worlds semantics.

▶ Barker first encodes the information that some person f has height h (in a world w) as propositions tall(e,f) for all $e \leq h$.

▶ Crucially, e is a "degree of tallness", i.e. a degree to which e is tall, and anyone who is tall to degree h is also tall to degree e for all $e \leq h$.

▶ Then, given a threshold t for tallness, one can check that f is tall (to degree t) by checking that f's height h is greater or equal to the threshold ($h \geq t$).

▶ This, in turn, is done by checking that tall(t, f) is true (in w).

▶ A delineation function d associated with each world supplies the threshold (or "standard") for each predicate in each world.

Barker (2002), cont'd

▶ We may regard this as a method for implementing simple threshold classifiers, which works by requiring explicitly listing (for each possible world) all the degrees of x-ness (where x is some degree predicate) of an individual which yield positive output from the classifier

▶ Here, tall(d, f) for all the degrees d to which f is tall.

▶ Alternatively, a more straightforward way of encoding classifier functions in an extensional framework would be to simply provide the characteristic function of the classifier

▶ It thus seems to be in principle possible to model perceptual meanings in a possible worlds framework (with some additions).

Barker (2002), cont'd

- ▶ Can this approach be generalized to classifiers taking more than one input and to complex classifiers such as, e.g., those made of up neural networks with several interconnected layers of artificial neurons?

- ▶ Potential problem: accounting for learning in a framework allowing only extensional representations of classifiers.

- ▶ The parameters which are typically modified in learning (such as the weights of the perceptron) are not represented as such.

- ▶ Other potential problem: compositionality beyond the simple conjunctive compositionality covered by Krishnamurthy and Kollar (2013) and exemplified above with "upper right".

- ▶ Our account of compositionality for degree modifiers ("far right") again relies on parameters being explicitly represented and available for modification

Local extensionality?

- ▶ Barker can be regarded as attempting to account for basic perceptual classification while sticking to what we may call "universal extensionality", i.e. the modeling of the complete extension of each predicate in the form of a characteristic function in each possible world.

- ▶ Alternatively, and as implied by (Matuszek *et al.*, 2012, Krishnamurthy and Kollar, 2013), one could give up on universal extensionality and model only temporary and local extensions which apply only to the presently observable situation.

- ▶ This would take us closer to the TTR account presented above

Give up on extensionality to model learning?

- ▶ An interesting question is how to account for continuity of meaning is such an approach, where the extension of a word may shift from one situation to another (for instance depending on which dogs happen to be present in the situation).
- ▶ One answer is to include representation of intensions such as classifiers
- ▶ If these classifiers are to be taken as part of the semantic theory proper, the assumption of extensionality in possible-worlds semantics may be problematic
- ▶ It seems likely that an intensional and situation-based model-theoretic semantics would closer to a type-theoretic approach such as TTR.

Keeping extensionality, leaving out learning?

▶ Alternatively, one may leave the intensional classifiers out of the semantic theory proper, and argue that the latter is an abstract level of representation not concerned with lower-level issues of practical implementation in the form of algorithms such as those used to specify classifiers intensionally.

▶ This, however, seems to mean that one is left with a semantic theory which is not able on its own to account for some aspects of meaning that we would argue are quite central, namely learning and coordinating on perceptual meanings.

Desiderata on formal theories of learning etc.

▶ Accounting for learning in individual agents requires relativising interpretation processes to individual agents.

▶ Intensions of linguistic expressions (modeled as classifiers) are best treated as first-class objects

▶ This allows semantic learning to be formulated in terms of modifying structured representations of meaning, including functions and their parameters.

▶ We believe that in the end, a type-theoretical approach with structured representations such as TTR will have significant advantages over a purely extensional approach when it comes including perceptual classifiers in formal semantics.

Outline

Conclusions

We have presented a formal semantics for coordination and learning of perceptual meaning, combining the following ideas:

► Perceptual meanings as classifiers of perceptual information

► Updating perceptual meanings based on language use in interaction

► Compositionality of perceptual meanings

► Vagueness of perceptual meanings modeled using statistical classifiers outputting probabilistic judgements

Future work

Much work remains to be done, including

- ▶ Connecting low-level transformations on perceptual data and low-level classifiers
- ▶ Extending probabilistic learning theory to full Bayesian nets
- ▶ Learning the structure of probabilistic dependencies in Bayesian nets
- ▶ Putting together various kinds of classifiers (neural nets, noisy thresholds, knn, etc.) and associated learning methods
- ▶ Learning which kind of classifier (if any) to connect to a word
- ▶ Exploring the interaction between compositionality and learning
- ▶ Providing richer models of the interaction patterns involved in semantic coordination, and their relation to semantic learning
- ▶ Implementing and verifying models in terms of their behaviour in interaction
- ▶ . . .

📄 Barker, C. 2002.
The Dynamics of Vagueness.
Linguistics and Philosophy 25(1):1–36.

📄 Breitholtz, Ellen and Cooper, Robin 2011.
Enthymemes as rhetorical resources.
In *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2011)*, Los Angeles (USA). 149–157.

📄 Brennan, S. E. and Clark, H. H. 1996.
Conceptual pacts and lexical choice in conversation.
Journal of Experimental Psychology: Learning, Memory and Cognition 22:482–493.

📄 Brennan, Susan E. 1996.
Lexical Entrainment in Spontaneous Dialog.
In *International Symposium on Spoken Dialog*. 41–44.

📄 Carey, Samuel and Bartlett, E. 1978.
Acquiring a single new word.
Papers and Reports on Child Language Development 15:17–29.

📄 Clark, H. H. and Wilkes-Gibbs, D. 1986.
Refering as a collaborative process.
Cognition 22:1–39.

📄 Clark, Eve V. and Wong, Andrew D. W. 2002.
Pragmatic directions about language use: Offers of words and
relations.
Language in Society 31:181–212.

📄 Clark, E. V. 2007.
Young children's uptake of new words in conversation.
Language in Society 36:157–82.

📄 Cooper, Robin and Ginzburg, Jonathan 2011.
Negation in dialogue.

In *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2011)*, Los Angeles (USA).

📄 Cooper, Robin; Dobnik, Simon; Lappin, Shalom; and Larsson, Staffan 2014.
A probabilistic rich type theory for semantic interpretation.
In *Proceedings of the EACL Workshop on Type Theory and Natural Language Semantics (TTNLS)*.

📄 Cooper, Robin 1998.
Information states, attitudes and dependent record types.
In *ITALLC98*.
85–106.

📄 Cooper, Robin 2004.
Dynamic generalised quantifiers and hypothetical contexts.
In *Ursus Philosophicus, a festschrift for Björn Haglund*. Department of Philosophy, University of Gothenburg.

📄 Cooper, Robin 2005.

Austinian truth, attitudes and type theory.
Research on Language and Computation 3:333–362.

📄 Cooper, Robin 2010.
Generalized quantifiers and clarification content.
In Łupkowski, Paweł and Purver, Matthew, editors 2010, *Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue*, Poznań. Polish Society for Cognitive Science.

📄 Cooper, Robin 2011.
Copredication, quantification and frames.
In Pogodalla, Sylvain and Prost, Jean-Philippe, editors 2011, *LACL*, volume 6736 of *Lecture Notes in Computer Science*. Springer. 64–79.

📄 Cooper, Robin 2012.
Type theory and semantics in flux.

In Kempson, Ruth; Asher, Nicholas; and Fernando, Tim, editors 2012, *Handbook of the Philosophy of Science*, volume 14: Philosophy of Linguistics. Elsevier BV.
General editors: Dov M. Gabbay, Paul Thagard and John Woods.

📄 Coquand, Thierry; Pollack, Randy; and Takeyama, Makoto 2004.
A logical framework with dependently typed records.
Fundamenta Informaticae XX:1–22.

📄 Fernández, Raquel and Larsson, Staffan 2014.
Vagueness and learning: A type-theoretic approach.
In *Proceedings of the 3rd Joint Conference on Lexical and Computational Semantics (∗SEM 2014)*.

📄 Fernández, Raquel; Larsson, Staffan; Cooper, Robin; Ginzburg, Jonathan; and Schlangen, David 2011.
Reciprocal learning via dialogue interaction: Challenges and prospects.
In *Proceedings of the IJCAI 2011 Workshop on Agents Learning Interactively from Human Teachers (ALIHT)*, Barcelona, Catalonia, Spain.

📄 Garrod, Simon C. and Anderson, Anthony 1987.
Saying what you mean in dialogue: a study in conceptual and
semantic co-ordination.
Cognition 27:181–218.

📄 Ginzburg, Jonathan 2012.
The Interactive Stance.
Oxford University Press, New York.

📄 Harnad, Stevan 1990.
The symbol grounding problem.
Physica D: Nonlinear Phenomena 42(1990):335–346.

📄 Healey, P.G.T. 1997.
Expertise or expertese?: The emergence of task-oriented
sub-languages.
In Shafto, M.G. and Langley, P., editors 1997, *Proceedings of the 19th
Annual Conference of the Cognitive Science Society*.
301–306.

📄 Krishnamurthy, Jayant and Kollar, Thomas 2013.
Jointly Learning to Parse and Perceive : Connecting Natural Language to the Physical World.
Transactions of the Association for Computational Linguistics 1:193–206.

📄 Larsson, Staffan and Cooper, Robin 2009.
Towards a formal view of corrective feedback.
In Alishahi, Afra; Poibeau, Thierry; and Villavicencio, Aline, editors 2009, *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition, EACL.*
1–9.

📄 Larsson, Staffan 2007.
Coordinating on ad-hoc semantic systems in dialogue.
In *Proceedings of the 10th workshop on the semantics and pragmatics of dialogue.*

📄 Larsson, Staffan 2010.
Accommodating innovative meaning in dialogue.

In Łupkowski, Paweł and Purver, Matthew, editors 2010, *Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue*, Poznań. Polish Society for Cognitive Science.
83–90.

📄 Larsson, Staffan 2011.
The ttr perceptron: Dynamic perceptual meanings and semantic coordination.
In *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2011)*, Los Angeles (USA).

📄 Larsson, Staffan 2013.
Formal semantics for perceptual classification.
Journal of Logic and Computation.

📄 Lassiter, Dan 2011.
Vagueness as probabilistic linguistic knowledge.
In Nowen, R.; Rooij, R.van; Sauerland, U.; and Schmitz, H. C., editors 2011, *Vagueness in Communication*. Springer.

📄 Linell, Per and Noren, Kerstin 2005.
Ny och ny - en dialogisk analys av meningspotentialen hos lexemet ny i svenskan.
In Melander, Björn et. al., editor 2005, *Språk i tid*, volume 67.
Uppsala: Institutionen för nordiska språk.

📄 Ludlow, Peter 2014.
Living Words: Meaning Underdetermination and the Dynamic Lexicon.

Oxford University Press.

📄 Matuszek, Cynthia; FitzGerald, N; and Zettlemoyer, Luke 2012.
A joint model of language and perception for grounded attribute learning.
In *Proceedings of the 29th International Conference on Machine Learning*.

📄 Mills, Gregory J. and Healey, Patrick G. T. 2008.
Semantic negotiation in dialogue: The mechanisms of alignment.

In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, SIGdial '08, Stroudsburg, PA, USA. Association for Computational Linguistics.
46–53.

📄 Pickering, Martin J. and Garrod, Simon 2004.
Toward a mechanistic psychology of dialogue.
Behavioral and Brain Sciences 27(02):169–226.

📄 Rosenblatt, F 1958.
The perceptron: a probabilistic model for information storage and organization in the brain.
Psychological review 65(6):386–408.

📄 Saxton, Matthew 1997.
The contrast theory of negative input.
Journal of Child Language 24:139–161.

📄 Saxton, Matthew 2000.

Negative evidence and negative feedback: immediate effects on the grammaticality of child speech.
First Language 20(3):221–252.

📄 Schmidt, L.A.; Goodman, N.D.; Barner, D.; and Tenenbaum, J.B. 2009.
How tall is tall? compositionality, statistics, and gradable adjectives.
In *Proceedings of the 31st annual conference of the cognitive science society.*

📄 Steels, Luc and Belpaeme, Tony 2005.
Coordinating perceptually grounded categories through language: A case study for colour.
Behavioral and Brain Sciences 28(4):469–89.
Target Paper, discussion 489-529.