Autonomous construction of ecologically and socially relevant semantics.

Takáč M. (2008)

Grounded Cognition

Prof. Igor Farkas

WS2014
MEi:CogSci

Moderated by
Rahil Hosseini
University of Vienna
Constructing computational models of cognitive phenomena: **Understanding by Building**

**Focus of the current paper:** a synthetic modeling approach to study the processes of concept formation and language acquisition, and their mutual relations.
Constructivist approach to overcome “symbol grounding problem”

- In each communication participant constructs its own individual meanings by interacting with the environment and other participants.
- The communicative goal is to uniquely identify a chosen static object by discriminating it from all other objects currently present in the communicative situation.

However

- Language goes beyond a present situation.
- Meanings are more than object categories; they include properties, relations, dynamic changes, situations and events.

As a solution this paper

- Proposes a similarity-based semantic representation of various types of meanings rather than focusing on differences.

And also it

- Proposes individual and social mechanisms of autonomous construction of such semantic representations.
- Implements the hypothesized mechanisms in computational models and analyze the results of simulations.
Meanings

- Are not objectively existing “out there”, they are mental entities or internal representations.
- The internal representations get their meaning via structural coupling with the world.
- Are embodied conceptual structures.
- Are grounded in perception and action.
- Are physically corresponding to activations of neural structures.
- Are correlated with perceiving, performing, imagining or talking about the content they represent.

The coupling of the representations with the world

- Physical symbol grounding
- Social (or external) symbol grounding

Meanings are individually created subjective constructs, which need to be attuned to each other collectively.
Agents attribute meanings to parts of their **environment** by recognizing, via their **sensors and actuators**, information useful for achieving their **goals**.

**Meanings are preverbal**: The embodied knowledge coming from perceiving and acting in the environment can be found in living organisms well before the appearance of a language, both in phylogeny and ontogeny.

**Cued representations**
- Triggered by something present in the current situation.
- Cued representations observable as non-volitional behavioral reactions are innate and have evolved pylologenetically.

**Detached representations**
Stands for objects and events neither present nor triggered by the current situation.
Language evolved in order to make cooperation about future goals possible.

Paradox: Cooperation requires socially shared meanings, but meanings are constructed individually.

Possible explanations:
• Similar learning mechanisms in shared and similar environment lead to sufficiently similar meanings
• Meanings are not transferred, they are inferred by the hearer.
• Coherent language as a result of local interactions of language users (self-organizing process).
Gavagai Paradox: Is understanding fundamentally impossible?

It is particularly important to establish shared meanings of referential expressions

Children’s strategy:

• They assume that novel words refer to whole objects
• The mutual exclusivity constraint: A novel word cannot name an object that already has a name.
• The principle of contrast: Any difference in form marks a difference in the meaning
• They disambiguate meanings by occurrences of their referents in multiple situations

In computational models, the hearer’s inference of the meaning is constrained by the assumption that the scene only contains one referent of the speaker’s utterance.
**Discrimination**: A relative judgment between things that are present simultaneously.

**Identification (categorization)**: An absolute judgment of a thing alone answering the question whether or not a given input is a member of a particular category

**Basic level of categorization**
- The most general level, at which a common perceptual image and a common motor program can be created for members of a category
- With the highest intra-cluster similarity and inter-cluster distinctiveness
- Supports inductive inferences
- Understood and acquired by children first (used by adults to communicate with children)
Discussion

Introduction
Meanings and their origin
• Individual Meanings
• Social Meanings
• The Inference of Meanings
• Discrimination vs identification

Computational Modelling of Grounded Perception
• Perception
• Representation of meaning

Construction of environmentally and ecologically relevant meanings
• Model
• Results

Construction of socially relevant meanings
• Model
• Results

Evolution of meanings within iterated intergenerational Transmission
• Model
• Experiment 1
• Experiment 2

Discussion
Introduction
Meanings and their origin
• Individual Meanings
• Social Meanings
• The Inference of Meanings
• Discrimination vs identification

Computational Modelling of Grounded
• Perception
• Representation of meaning

Construction of environmentally and ecologically relevant meanings
• Model
• Results

Construction of socially relevant meanings
• Model
• Results

Evolution of meanings within iterated intergenerational Transmission
• Model
• Experiment 1
• Experiment 2

Discussion
Perceptual Input at time $t$

- A set of frames of the attribute
- Value pairs with one frame for each object in the environment

**Frames**

- All perceptual inputs relevant to the agent (even the agent proprioceptive input)
- A computational level description of structures resulting from a low-level preprocessing of the perceptual input (not YET a representation)
Input:
- A perceptual frame of one object
- Perceptual frames of several objects
- ...

Output:
The degree of the input’s membership in the category.

How?

**Locally tuned detectors**: extracting common statistical properties of examples of categories encountered during their lifetime (0-1).
Goal: construction of preverbal meanings by sensorimotor interactions with the environment

**Agent**
- Start: Empty categorical system
- Active Exploration
- Repertoire of actions: lift, put down
- Performance Parameters: force, arm angle

**Environment**
- 2-dimensional lattice
- Frames of toys, fruits and furniture

- The agent randomly chooses an action and performs it with different parameters
- The environment simulates the action effect
- After performing an action:
  - Agent Observes the resulting change
  - Represents the knowledge of causal relations between action, object and the change
    - If an action led to the same change on several objects, they would all fall in the same object category
    - A significantly different outcome of the action triggers creation of new categories
**Associated change criteria**: the difference in position of the involved object \((\Delta)\)

**Attribute values**: mean \(\pm\) the standard deviation

---

<table>
<thead>
<tr>
<th>Category</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(putDown(5 \pm 3))</td>
</tr>
<tr>
<td>C1</td>
<td>No change</td>
</tr>
<tr>
<td>C2</td>
<td>No change</td>
</tr>
<tr>
<td>C3</td>
<td>(\Delta = {posZ: -6 \pm 1})</td>
</tr>
<tr>
<td>C4</td>
<td>(\Delta = {posZ: -4 \pm 2})</td>
</tr>
</tbody>
</table>

---

**C1**: objects that cannot be put down, because they are already on the ground

**C2**: objects too heavy to be lifted

**C3**: Toys

**C4**: Fruits

---

**Table 2**

Number of objects of each type for a category they are most similar to

<table>
<thead>
<tr>
<th>Object type</th>
<th>Category</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fruit</td>
<td></td>
<td></td>
<td>8</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Toy</td>
<td></td>
<td>1</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Furniture</td>
<td></td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Number of time steps: 5000

Measurement: The resulting activity of the predicted change criterion applied to the actually perceived change (Prediction)

Fig. 4. Results of the experiment simulating construction of categories by sensorimotor interactions. The number of criteria saturates and the prediction value converges to that of the average intra-category distance (i.e. predictions are correct).
Discussion

Introduction
Meanings and their origin
• Individual Meanings
• Social Meanings
• The Inference of Meanings
• Discrimination vs identification

Computational Modelling of Grounded Perception
• Representation of meaning

Construction of environmentally and ecologically relevant meanings
• Model
• Results

Construction of socially relevant meanings
• Model
• Results

Evolution of meanings within iterated intergenerational Transmission
• Model
• Experiment 1
• Experiment 2

Discussion
Goal: study the influence of verbal instruction (naming) on category formation process

Two agents:
- Teacher: describes various aspects of the present situation
- Learner: Induces meanings of the teacher’s words by noticing cross-situation similarities between their referents

Induction assumptions:
- No true synonymy: Different words have different meanings, even if they share a referent
- No true homonymy: All referents of a single word across different situations were considered instances of the same category denoted by the word

Environment:
- 2D geometrical shapes characterized by five attributes
  - Vertices
  - Coordinates of the centroid of the shape (PosX, PosY)
  - Size of the bounding rectangle (SizeX, SizeY)
- Dynamic
In each time step, the environment undergoes some changes
- The learner describes the scene
- The teacher comments on the environment in each time step using its predefined lexicon
- Learner’s description and teacher’s description of each scene are compared in each time step
- After 500 learning epochs, 200 guessing games are played

Table 3
A predefined ontology and lexicon of the teacher

<table>
<thead>
<tr>
<th>Word</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Square</td>
<td>vertices = 4 ∧ sizeX = sizeY</td>
</tr>
<tr>
<td>Triangle</td>
<td>vertices = 3</td>
</tr>
<tr>
<td>Big</td>
<td>sizeX &gt; 15 ∧ sizeY &gt; 15</td>
</tr>
<tr>
<td>Slim</td>
<td>sizeX &lt; 0.2sizeY</td>
</tr>
<tr>
<td>Small</td>
<td>sizeX &lt; 10 ∧ sizeY &lt; 10</td>
</tr>
<tr>
<td>Grow</td>
<td>sizeX(t) &gt; sizeX(t-1) ∧ sizeY(t) &gt; sizeY(t-1)</td>
</tr>
<tr>
<td>Shrink</td>
<td>sizeX(t) &lt; sizeX(t-1) ∧ sizeY(t) &lt; sizeY(t-1)</td>
</tr>
</tbody>
</table>
Introduction
Meanings and their origin
• Individual Meanings
• Social Meanings
• The Inference of Meanings
• Discrimination vs identification
Computational Modelling of Grounded
• Perception
• Representation of meaning
Construction of environmentally and ecologically relevant meanings
• Model
• Results
Construction of socially relevant meanings
• Model
• Results
Evolution of meanings within iterated intergenerational Transmission
• Model
• Experiment 1
• Experiment 2
Discussion

\[
\text{Correctness} = \frac{1 - \text{number of wrong words in the learner's description}}{\text{the total number of words in the learner's description}}
\]

\[
\text{Completeness} = \frac{1 - \text{number of teacher's words missing in the learner's description}}{\text{total number of words in the teacher's description}}
\]

\[
\text{Description Similarity} = \text{Average} (\text{Correctness}, \text{Completeness})
\]

Fig. 5. The quality of the lexicon acquired by cross-situational learning within one generation. The measure “usage200” is an average usage of 200 guessing games played after learning. Each measure in the graph has been averaged over the time window of 30 last steps. The results of the experiment were averaged over 10 simulation runs with different random seeds.
**Goal:** To see whether meanings constructed by social interactions remain stable if the acquisition process iterate intergenerationally

**Framework:** Iterated learning model (ILM)
- A learner builds up its own internal language representation by observing external language input from its teacher
- The learner becomes a teacher and produces utterances, which are the input for the next generation learner

**First generation:**
- In each time step, the environment undergoes some changes
- The learner describes the scene
- The teacher comments on the environment in each time step using its predefined lexicon
- After a certain number of epochs, the teacher with predefined ontology was removed and the learner became a teacher for a new agent with an empty ontology and lexicon

And it continues...
**Condition:** the agent could neither modify nor add any new meanings, once it became a teacher (it only used the meanings acquired from it’s own teacher).

**Results:**

- Some categories such as “grow”, “Shrink”, “triangle” and “square” remained stable
- Criteria for other categories such as “big”, “small” and “slim” varied
- Smaller number of learning epochs causes smaller sample sets resulting in unstable concepts

![Graphs showing categories unstable over generations.](image)

*Fig. 7. Categories unstable over generations. The receptive field of each category was projected into the plane with dimensions sizeX; sizeY.*

(a) overspecialization – the size of the receptive field converged to zero over generations, (b) overgeneralization – a random correlation of some attributes in the sample overtook other attributes that became overgeneralized (ignored).
Meaning Bottleneck:

- The sample size for a learner’s category depends on the probability of occurrence of instances of the teacher’s category on the scene.
- The smaller the sample, the bigger the chance that it will contain random correlations that are not a part of the original meaning and a covariance-based detector would not reconstruct the original meaning properly.
- Selection Pressure: Special meanings describing situations that occur very rarely have smaller survival chances than frequently applicable general meanings.

Meaning transmission can be viewed as an evolutionary process with meanings as competing replicators.
**Condition:** the teacher could invent new meanings or extend old ones, in case it had no meanings applicable to describe some object on the current scene.

- The teacher of the first generation started with the criteria for triangle, square, small, big, and slim
- If, for some object on the scene there was no criterion returning nonzero activity for the object, a new criterion (named by a new random word) was created with the object as the first example.

**Results:**

- Category of “triangle” remained stable while category of “square” was stable for simulation runs of the experiment with more than 200 learning epochs. Categories of “big”, “small” and “slim” were instable
- Rich-get-richer dynamic: Meanings with under-threshold activity competed for selection and the meaning with highest activity was selected and extended subsequently.
Ecologically relevant categories can be constructed from sensorimotor interactions with the environment.

All action categories associated with some object category represented affordances of the object.

A linguistic instruction accompanied by a non-verbal reference can lead to cross-situational construction of the learner’s meanings.

While a high similarity between teacher’s and learner’s meanings is achieved rapidly and maintained within each generation, meanings do change throughout the generations.

Meanings constituted by simple structural relations and invariant attribute values have shown to be more stable in the iterated transmission than meanings based on interval values of uncorrelated attributes (nouns vs adjectives).

Learning in the models is based on extracting cross-situational similarities between examples of a category (learning the meaning of words across multiple exposures).

The role of social learning in the acquisition of concepts and language: Active and rich social interactions served the role of narrowing the context and reducing the noise.

categories constructed for the purpose of identification rather than discrimination are more suitable for the detached use of language (talking about things not present here and now).