Modelling the acquisition of lexical meaning from caregiver-child interaction

Getting the semantics straight

Barend Beekhuizen[1], Afsaneh Fazly[2], Aida Nematzadeh[2] & Suzanne Stevenston[2]

[1]Leiden University [2]University of Toronto

18 January 2013

Goals

Topic

Cognitive models of acquiring word-meaning mappings

Goal #1

Discuss sources of semantic data for models and present a new one

Goal #2

Show how this data can be used to re-evaluate old claims

Data?

- Cross-situational models of acquiring word meanings[1]
- Source of meaning: situational context.
- Your average CHILDES corpus does not contain that.

[1] Siskind 1996, Xu & Tenenbaum 2000, Roy & Pentland 2002, Yu & Ballard 2003, Fazly, Alishahi & Stevenson 2010

Data?

- Cross-situational models of acquiring word meanings[1]
- Source of meaning: situational context.
- Your average CHILDES corpus does not contain that.
- So: method of synthesizing semantics.
  - Every word is a semantic symbol (Fazly, Alishahi & Stevenson 2010)
  - Obtain lexical semantics from WordNet (id., 2008)
- Allows you to make large quantities of data.

[1]Siskind 1996, Xu & Tenenbaum 2000, Roy & Pentland 2002, Yu & Ballard 2003, Fazly, Alishahi & Stevenson 2010

Data?

- Cross-situational models of acquiring word meanings[1]
- Source of meaning: situational context.
- Your average CHILDES corpus does not contain that.
- So: method of synthesizing semantics.
  - Every word is a semantic symbol (Fazly, Alishahi & Stevenson 2010)
  - Obtain lexical semantics from WordNet (id., 2008)
- Allows you to make large quantities of data.
- But: quality of data?
  - Cognitive availability of meaning?
  - Situational availability? (noise, referential uncertainty)

[1] Siskind 1996, Xu & Tenenbaum 2000, Roy & Pentland 2002, Yu & Ballard 2003, Fazly, Alishahi & Stevenson 2010

Data?

- Cross-situational models of acquiring word meanings[1]
- Source of meaning: situational context.
- Your average CHILDES corpus does not contain that.
- So: method of synthesizing semantics.
  - Every word is a semantic symbol (Fazly, Alishahi & Stevenson 2010)
  - Obtain lexical semantics from WordNet (id., 2008)
- Allows you to make large quantities of data.
- But: quality of data?
  - Cognitive availability of meaning?
  - Situational availability? (noise, referential uncertainty)
- Recent method: annotating video material (Yu, Roy, Frank)
- But: either limited to basic-level objects or in the pragmatic realism (explicit labeling task).

[1] Siskind 1996, Xu & Tenenbaum 2000, Roy & Pentland 2002, Yu & Ballard 2003, Fazly, Alishahi & Stevenson 2010

Data!

Goal #1

Provide situational descriptions (of properties, objects, relations, actions) for a dataset of videotaped caregiver-child interaction that can function as a source for acquiring (first) word meanings.

Data!

Goal #1

Provide situational descriptions (of properties, objects, relations, actions) for a dataset of videotaped caregiver-child interaction that can function as a source for acquiring (first) word meanings.

- Some desiderata:
  - Children should be young enough not to know too much already.
  - Coded descriptions should be cognitively available.
  - Coded descriptions should stay close to what's observable; the coders should not have to infer too much

Data!

Goal #1

Provide situational descriptions (of properties, objects, relations, actions) for a dataset of videotaped caregiver-child interaction that can function as a source for acquiring (first) word meanings.

- Some desiderata:
  - Children should be young enough not to know too much already.
  - Coded descriptions should be cognitively available.
  - Coded descriptions should stay close to what's observable; the coders should not have to infer too much
- Realizations:
  - High-quality data can only complement high-quantity data, not replace it.
  - Little earlier work: the specifics may contain serious methodological flaws (more than happy to find out!)

The block game corpus

- $\pm$ 120 90-min videos of mother-daughter (16mo) interaction, gathered by Child Studies in Leiden
- Every dyad played a game of putting differently-shaped blocks in a bucket through corresponding holes
- 32 dyads ($\pm$ 5 min. each) were situationally coded by two coders using ELAN and transcribed by first author

The block game corpus

- $\pm$ 120 90-min videos of mother-daughter (16mo) interaction, gathered by Child Studies in Leiden
- Every dyad played a game of putting differently-shaped blocks in a bucket through corresponding holes
- 32 dyads ($\pm$ 5 min. each) were situationally coded by two coders using ELAN and transcribed by first author
- 175 minutes of material, 7842 word tokens, 2492 utterances.
- Situational coding. For every interval of 3 seconds, code:
  - simple behavior (`grab`,`move`,`position`,`letgo`),
  - changes in spatial relations (`in`,`on`,`out`,`off`,`match`),
  - objects (`block`,`bucket`,`mother`,`table`)
  - properties (`triangular`,`square`,`red`,`blue`)
- Structured: `grab(mother,(red,square,block))`
- High intra- & interannotator agreement (almost all $\kappa > 0.8$)

Example

Table: A sample of the dataset. The dash-separated abbreviations denote blocks and holes and their properties (colors & shapes)

time	type	coding/transcription
0m0s	**situation**	
	language	een. nou jij een.
	translation	"One. Now you try one."
0m3s	**situation**	position(mother, toy, on(toy, floor)) grab(child, b-ye-tr) move(child, b-ye-tr, on(b-ye-tr, floor), near(b-ye-tr, ho-ro)), mismatch(b-ye-tr, ho-ro)
	language	nee daar.
	translation	"No, there."
0m6s	**situation**	point(mother, ho-tr, child) position(child, b-ye-tr, near(b-ye-tr, ho-ro)) mismatch(b-ye-tr, ho-ro)
	language	nee lieverd hier past ie niet.
	translation	"No sweetie, it won't fit in here."

Acquiring lexical meaning

- How to learn the meaning of a word?
  - Cross-situationally observing objects, relations, events, properties.
  - Seems insufficient (esp. for relational terms; verbs, prepositions)
    - Number of possibilities is vast (Gentner 1978)
    - Many actions and relations do not take place at the moment of utterance (Gleitman 1990)
  - Bootstrapping by using linguistic structure (Gleitman 1990), intentionality (Tomasello 2003), . . .

Acquiring lexical meaning

- How to learn the meaning of a word?
  - Cross-situationally observing objects, relations, events, properties.
  - Seems insufficient (esp. for relational terms; verbs, prepositions)
    - Number of possibilities is vast (Gentner 1978)
    - Many actions and relations do not take place at the moment of utterance (Gleitman 1990)
  - Bootstrapping by using linguistic structure (Gleitman 1990), intentionality (Tomasello 2003), ...

Goal #2

Using this data set to re-evaluate the claim that relational terms are more difficult than non-relational terms.

The model

- Fazly, Alishahi & Stevenson (2010) incremental model of aligning words in utterance $U = \{w_1, \ldots, w_n\}$ with features in situation $S = \{f_1, \ldots, f_n\}$.

The model

- Fazly, Alishahi & Stevenson (2010) incremental model of aligning words in utterance $U = \{w_1, \ldots, w_n\}$ with features in situation $S = \{f_1, \ldots, f_n\}$.
- Calculating alignment on the basis of conditional probabilities:

$$a(w|f, U^{(t)}, S^{(t)}) = \frac{p^{(t-1)}(f|w)}{\sum_{w' \in U^{(t)}} p^{(t-1)}(f|w')} \quad (1)$$

The model

- Fazly, Alishahi & Stevenson (2010) incremental model of aligning words in utterance $U = \{w_1, \ldots, w_n\}$ with features in situation $S = \{f_1, \ldots, f_n\}$.
- Calculating alignment on the basis of conditional probabilities:

$$a(w|f, U^{(t)}, S^{(t)}) = \frac{p^{(t-1)}(f|w)}{\sum\limits_{w' \in U^{(t)}} p^{(t-1)}(f|w')} \qquad (1)$$

- Updating the association score (initialized at 0):

$$\text{assoc}^{(t)}(w, f) = \text{assoc}^{(t-1)}(w, f) + a(w|f, U^{(t)}, S^{(t)}) \qquad (2)$$

The model

- Fazly, Alishahi & Stevenson (2010) incremental model of aligning words in utterance $U = \{w_1, \ldots, w_n\}$ with features in situation $S = \{f_1, \ldots, f_n\}$.
- Calculating alignment on the basis of conditional probabilities:

$$a(w|f, U^{(t)}, S^{(t)}) = \frac{p^{(t-1)}(f|w)}{\sum_{w' \in U^{(t)}} p^{(t-1)}(f|w')} \quad (1)$$

- Updating the association score (initialized at 0):

$$\text{assoc}^{(t)}(w, f) = \text{assoc}^{(t-1)}(w, f) + a(w|f, U^{(t)}, S^{(t)}) \quad (2)$$

- Recalculating the conditional probabilities on the basis of the association scores:

$$p^{(t)}(f|w) = \frac{\text{assoc}^{(t)}(w, f) + \lambda}{\sum_{f' \in F} \text{assoc}^{(t)}(w, f') + \beta \times \lambda} \quad (3)$$

Data preparation

- Representations are structured, so flatten them:
  `grab(mother,(red,square,block))` $\rightarrow$
  `{grab,mother,red,square,block}`
- Take the set of all flattened representations of the situation taking place in the interval in which the utterance was beginning to be produced.
- We used lemma representations for the words

Evaluation

- No golden lexicon, so hand-built one for 'meaningful' words ($n = 41$):
  - Object labels: *blok* meaning `block`
  - Properties: *rood* meaning `red`
  - Spatial relations: *op* meaning `on`
  - Actions: *passen* meaning `match`, *stoppen* meaning {`move`,`in`}

Evaluation

- No golden lexicon, so hand-built one for ‘meaningful’ words ($n = 41$):
  - Object labels: *blok* meaning `block`
  - Properties: *rood* meaning `red`
  - Spatial relations: *op* meaning `on`
  - Actions: *passen* meaning `match`, *stoppen* meaning {`move`,`in`}
- Two (partially complementary) measures:
  - Summed Conditional Probability (*SCP*): how much probability mass is assigned to the true meanings given a word?
  - Average Precision (*AP*): how are the true meanings ranked (on conditional probability) w.r.t. the other meanings.

Results

Table: Results of experiment 1. Given are mean *SCP* and *AP* values per class

	property	object	spatial	action	**total**
SCP	0.10	0.05	0.09	0.07	**0.08**
AP	0.81	0.25	0.19	0.15	**0.31**

- Conditional probability distributions do not get very peaky in general
- Ranking is good for properties (colors, shapes), but rather bad for other classes.

Model dependence?

- Compared with one other model: Jon Stevens (2011)' hypothesis testing model.
- Same direction of results: properties > objects > spatial relations > actions

Table: Results of experiment 1

		property	object	spatial	action	**total**
FAS10	*SCP*	0.10	0.05	0.09	0.07	**0.08**
	AP	0.81	0.25	0.19	0.15	**0.31**
S11	*SCP*	0.09	0.06	0.06	0.02	**0.05**
	AP	0.28	0.20	0.13	0.09	**0.17**

Interpretation

	property	object	spatial	action	**total**
SCP	0.10	0.05	0.09	0.07	**0.08**
AP	0.81	0.25	0.19	0.15	**0.31**

- Re-evaluation corroborates Gleitman's finding:
  Properties > object labels > spatial relations and actions
- Why are the latter three harder to learn?
  1. True meaning is absent from S
  2. Foil features are structurally present in S
  3. True meaning is also present in many other Ss
- Combination of these! For properties, 2) and 3) hold as well.

- Focussing on absent true meanings
- Perhaps the temporal scope is too narrow?
- Learners may focus on situations slightly temporally displaced
- Pragmatically defined window: $S =$ all coded material in intervals between the previous utterance, $U^{(t-1)}$, and the next one, $U^{(t+1)}$.
- Variable: sometimes a large window of situations, sometimes just the time of the utterance itself.

W		prop.	object	spatial	action	**total**
$0:0$	*SCP*	0.10	0.05	0.09	0.07	**0.08**
	AP	0.81	0.25	0.19	0.15	**0.31**
$U^{(t-1)}:U^{(t+1)}$	*SCP*	0.08	0.05	0.10	0.08	**0.07**
	AP	0.79	0.41	0.22	0.20	**0.39**

- Slight increase for three less-learned categories:
  - wider context is informative, more true meanings found
  - while not producing more referential uncertainty (as expected).
- Pragmatics: people talk about what should happen, or what has happened.

- Difficulty of getting good data; perhaps more tedious than developing a realistic model.
- Manual coding of situational contexts can be done
  - to complement synthesization methods (how much noise and uncertainty is realistic for which meaning category?)
  - to perform small-scale evaluations experiments
- However, ideally: wider situational contexts

- Difficulty of getting good data; perhaps more tedious than developing a realistic model.
- Manual coding of situational contexts can be done
  - to complement synthesization methods (how much noise and uncertainty is realistic for which meaning category?)
  - to perform small-scale evaluations experiments
- However, ideally: wider situational contexts
- Verbs and Prepositions are harder to learn than Nouns, which are harder than Color & Shape terms
- A wider scope helps a bit
- Structured learning? (Bootstrapping on syntax, using structure of semantics)

- Difficulty of getting good data; perhaps more tedious than developing a realistic model.
- Manual coding of situational contexts can be done
    - to complement synthesization methods (how much noise and uncertainty is realistic for which meaning category?)
    - to perform small-scale evaluations experiments
- However, ideally: wider situational contexts
- Verbs and Prepositions are harder to learn than Nouns, which are harder than Color & Shape terms
- A wider scope helps a bit
- Structured learning? (Bootstrapping on syntax, using structure of semantics)
- Realistic data is important!