Errors in word meaning acquisition

as explained by semantic typology and computational modeling

Barend Beekhuizen

University of Toronto

UCSB Linguistics colloquium
March 10, 2016

Joint work with Afsaneh Fazly and Suzanne Stevenson

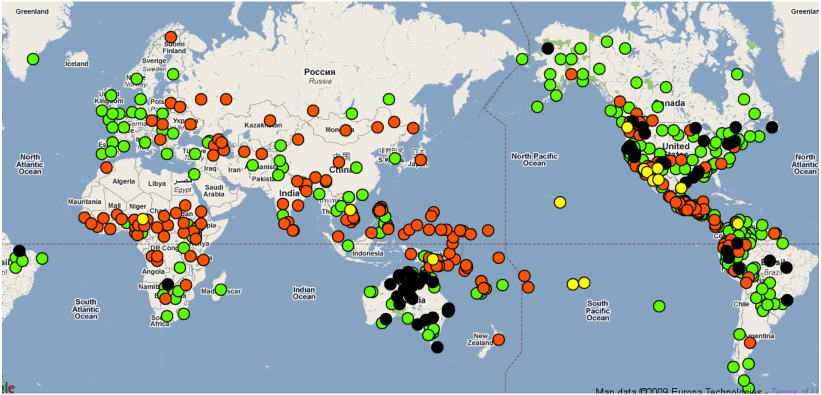
1. Introduction
   - Errors in word meaning acquisition
   - Outline of our approach

2. The approach
   - Representing word meanings
   - Simulating a learner
   - Evaluating the model's predictions

Issue 1: Errors in lexical semantic acquisition

- Word meaning acquisition is not flawless
  - Calling all round things *ball*,
  - Mixing up first and second person pronouns,
  - Using one preposition where another would be 'correct'
- Errors display patterns
- Asymmetries!
- Why do children make these errors?

Typology!


Greenland
Iceland
Suomi
Finland
Sverige
Sweden
Россия
Russia
North Atlantic Ocean
Algeria
Libya
Egypt
Saudi Arabia
Iraq
Iran
Mauritania
Mali
Niger
Angola
India
Indonesia
Indian Ocean
South Atlantic Ocean
North Pacific Ocean
South Pacific Ocean
New Zealand
Chile
Map data ©2009 Europa Technologies

Typological Prevalence Hypothesis

- Explanation: cognitive accessibility of a meaning concept.
- Bowerman & Gentner (2009): Typological prevalence reflects cognitive accessibility (simplicity, salience)
  - Many languages will use a particular meaning concept if it is easily accessible
  - Reversed: if a meaning concept is widespread, it must be cognitively accessible (all other things being equal)
- And: low cognitive accessibility leads to errors.
- In particular: overextension of high-accessible meaning concepts to words signifying low-accessible ones.
- Case study: Dutch *op* ‘surface support’ overextended to ‘tenuous support’ situations (expressed with *aan*)

Our approach

- Use crosslinguistic elicitation data to represent word meaning
- Train a computational word learning model to associate word forms with these representations.
- See if the developmental pattern of the model is similar to that of children
  - If this is indeed due to the representations used, this supports the Typological Prevalence Hypothesis

1 Introduction
- Errors in word meaning acquisition
- Outline of our approach

2 The approach
- Representing word meanings
- Simulating a learner
- Evaluating the model's predictions

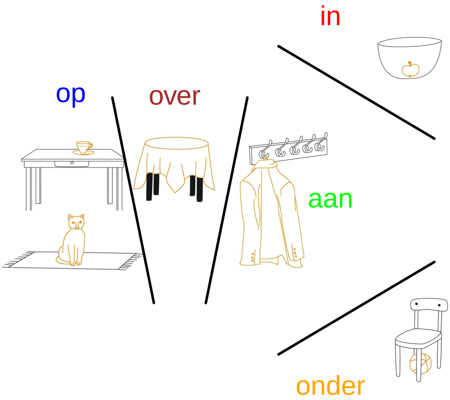


Figure: Dutch divisions of SPACE

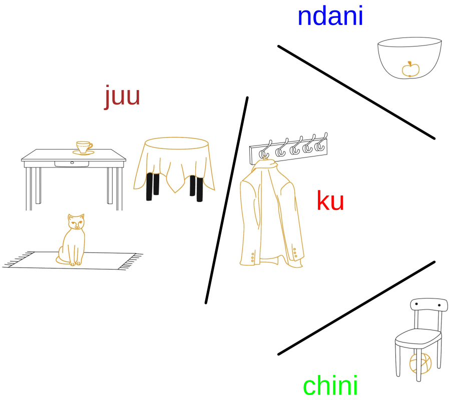


Figure: Swahili divisions of SPACE

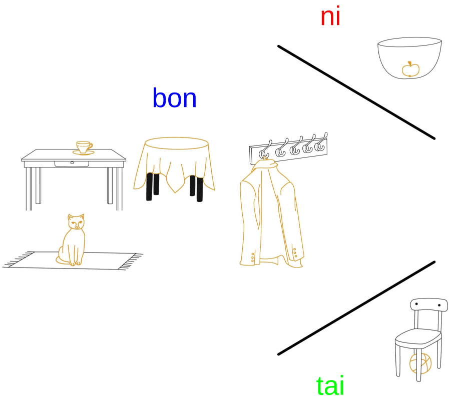


Figure: Thai divisions of SPACE

Step 1: gather count matrices

For every language, count number of labels per situation

situation	op	in	aan	onder	over
cup on table	10				
apple in bowl		10			
coat on hook			10		
ball under chair				10	
tablecloth on table	4				6
cat on mat	10				

Table: Counts of Dutch terms

situation	bon	ni	trong	fi	tai

Step 2: extract distances between situations

Per language, for every pair of situations, calculate Euclidean distance between counts. Then normalize to $[0, 1]$

	cup	apple	coat	ball	cloth	cat
cup on table	0	1	1	1	0.85	0
apple in bowl		0	1	1	1	1
coat on hook			0	1	1	1
ball under chair				0	1	1
tablecloth on table					0	0.85
cat on mat						0

Table: Distance matrix of situations for Dutch

Step 3: Global distance matrix

Sum all distance matrices. Then normalize to $[0, 1]$ again.

	cup	apple	coat	ball	cloth	cat
cup on table	0	0.97	0.70	0.98	0.22	0.07
apple in bowl		0	0.79	0.97	0.90	0.97
coat on hook			0	0.80	0.60	0.69
ball under chair				0	0.92	0.98
tablecloth on table					0	0.22
cat on mat						0

Table: Distance matrix of situations in 15 languages

Step 4: PCA

Apply PCA to matrix; extract coordinates per situation.

	PC 1	PC 2	PC 3	PC 4	PC 5
cup on table	-0.74	-0.01	-0.12	-0.09	-0.05
apple in bowl	0.88	0.70	-0.19	0	0
coat on hook	0.27	0	0.62	-0.04	0
ball under chair	0.93	-0.68	-0.19	0	0
tablecloth on table	-0.6	0.01	0	0.22	0
cat on mat	-0.74	-0.01	-0.12	-0.10	0.05

Table: Values on 5 principal components for the situations

Step 4: PCA

Apply PCA to matrix; extract coordinates per situation.

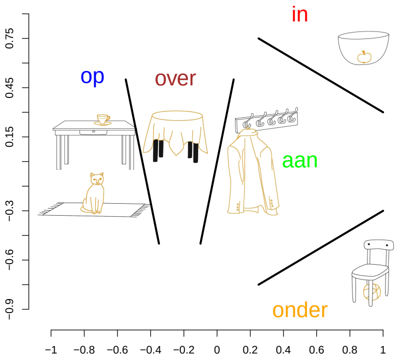


Figure: Situations in first two components, with Dutch divisions of SPACE

Learning word meanings

- Every situation is represented as a coordinate in the PCA space.
- Learning a word meaning is learning what subspace of this space should be associated with a word form.
- Note: simplistic vision of meaning – purely extensional, no distinction semantics/pragmatics, etc.

Input items

- Model learns by integrating input items in a Self-Organizing Map
- Input item consists of two parts: word form and the representation of the situation
- Word form: array of zeros with a one for the used label
  - E.g. for Dutch: $[1, 0, 0, 0, 0]$ for *op*, $[0, 1, 0, 0, 0]$ for *in*, etc.
- Situation: PCA coordinates of situation
- So: term *op* referring to situation 'cat on mat' is represented as: $[1, 0, 0, 0, 0, -0.74, -0.01, -0.12, -0.10, 0.05]$

Learning in SOM

- SOM is a grid of $m \times n$ cells.
- Each cell has the same number of values as input items
- For every input item: find cell that is most similar to input item
- Most similar cell and neighbors are then updated so that they resemble the input item more closely
- Cells start out with random values
- Over time, map comes to reflect categories of learned language
- Cells function as summary representations of input items

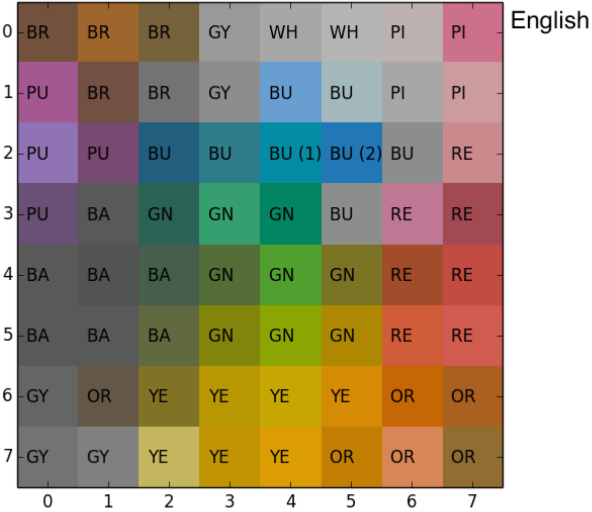


Figure: A trained SOM for English color terms

Where do input items come from?

- Children hear words with varying frequencies
- So, training data should reflect that
- At every turn, the model integrates a sampled input item (term-situation pair) into the SOM:
  - Sample term t with probability $P(t)$: the relative frequency of t in a corpus
  - Sample situation s given a term t with a probability $P(s|t)$ as observed in the elicitation data
- Possibility of controlling for frequency effects! Setting $P(t)$ to a uniform distribution does that.

Testing the model: situations without terms

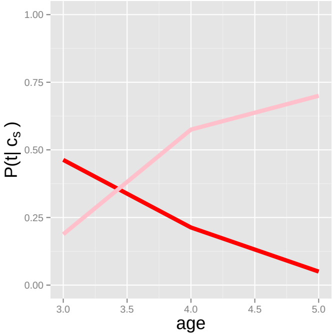
- Give model a situation without term and ask it to predict the label
- 'cat on mat': $[\cdot, \cdot, \cdot, \cdot, \cdot, -0.74, -0.01, -0.12, -0.10, 0.05]$
- Again, find most similar cell for input item (only using PCA features)
- Then: read off the term values for that cell
- E.g.: $[0.62, 0.07, 0.28, 0.02, 0.01]$
- So: *op* has a probability of 0.62, *in* of 0.07, etc.

Evaluation 1: Convergence with adult behavior

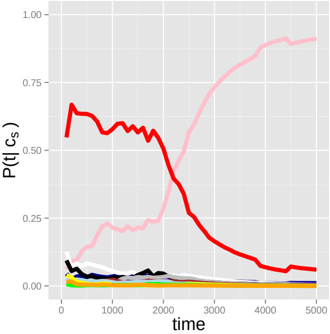
- Sanity check: does model end up behaving like adult language users?
- For all situations, predict most-likely term on basis of situation
- See how often predicted term is the same as the term most adult language users use

Evaluation 2: Matching developmental pattern

Few data points in observed data, many in model predictions



(a) Observed



(b) Predicted (avg. distribution over 30 simulations)

Aligning observed and predicted test moments

- Dynamic Time Warping: Contiguous series of bins of predicted test moments
- Bin is summary over all simulations in certain time interval
- Find bins maximizing similarity for all situations: global score of goodness of prediction

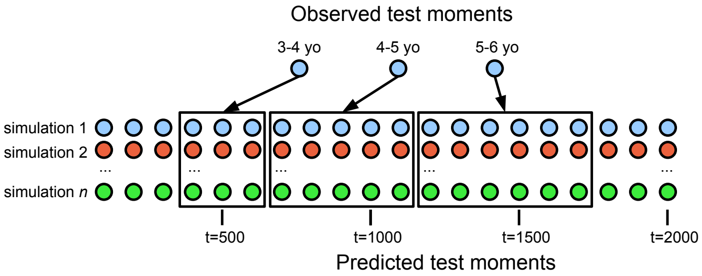
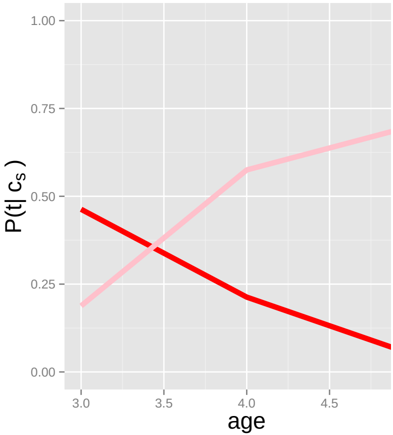
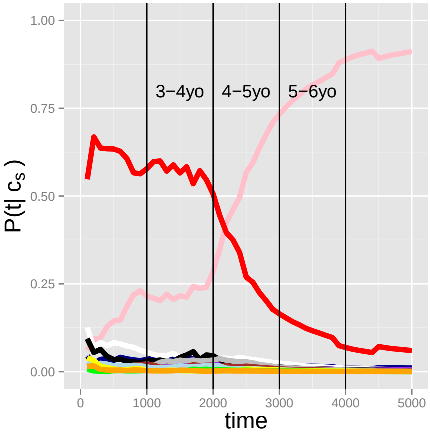


Figure: A possible alignment between predicted and observed data



(a) Observed



(b) Predicted, binned

Recap: modeling error patterns

- Obtain semantic representations with PCA over cross-linguistic data,
- Train a category learning model (Self-Organizing Map) on pairs of a word and a semantic representation,
- Evaluate match with (1) adult production, (2) observed developmental path.