Constraining the search space in cross-situational learning:

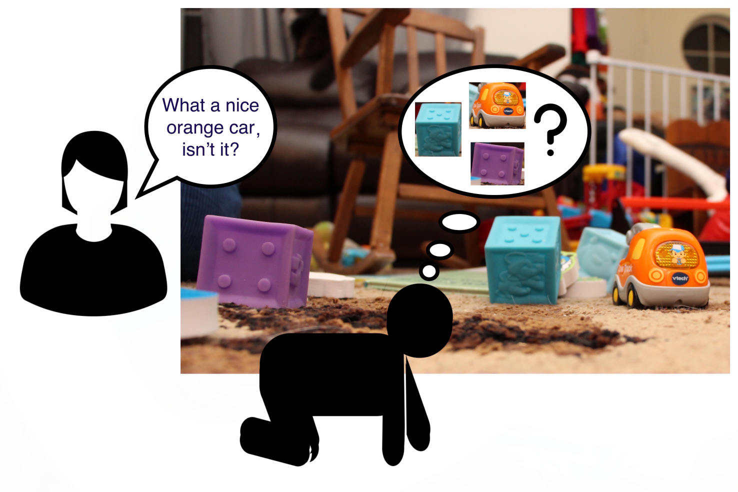
Different models make different predictions

Giovanni Cassani

27 May 2016

CLiPS
Computational Linguistics & Psycholinguistics
University of Antwerp

The blooming buzzing confusion...



Many possible referents can be mapped to utterance parts: still, children resolve this problem brilliantly. How?

…AND HOW TO MAKE SENSE OF IT

Keep track of **co-occurrences** of utterance parts and real-world referents over many different utterances and situations. If pairings are meaningful, they should occur more often than random pairings.

Many computational models try to account for the possible mechanisms behind cross-situational learning: I tested four against a single, simple set of behavioral data [2].

The successful models also learn from **missing co-occurrences**, i.e. the fact that a word and an object don't co-occur.

Behavioral data

The dataset from Ramscar et al (2013) [5]

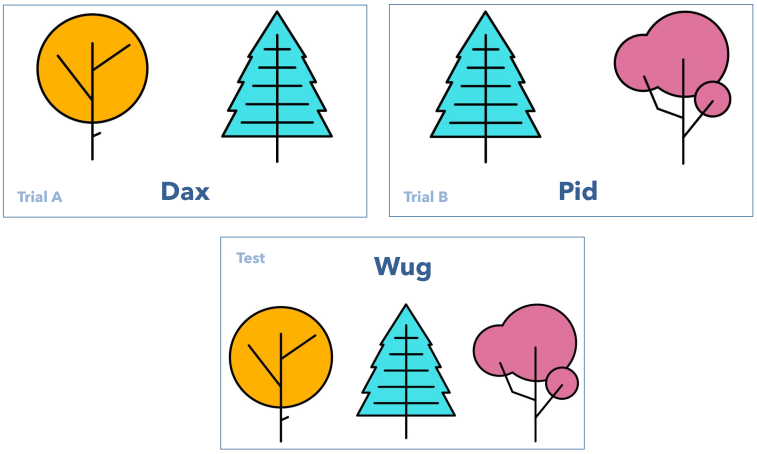


Figure 1: During training, subjects saw two objects and then heard a word. At test, they heard a word and were asked to retrieve the associated object.

Training trials summary

Table 1: Co-occurrence statistics and input to the computational models

Objects (Cues)	Words (Outcomes)	Frequency
ObjA_ObjB_Context1_ExptContext	DAX	9
ObjB_ObjC_Context2_ExptContext	PID	9

Behavioral results

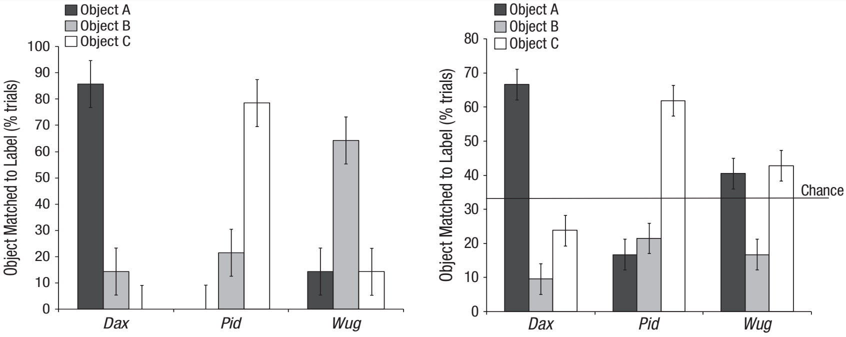


Figure 2: Undergraduates responses (left) and children responses (right). The two groups are consistent when asked about words they heard during training, but differ in the responses to the presentation of the withheld word.

Computational models

Hebbian learner [4]

$$V_{ij}^{t+1} = v^t + \Delta V_{ij}$$

$$\Delta V_{ij} = \begin{cases} k & \text{if } c_i \in t \text{ and } o_j \in t \\ 0 & \text{else} \end{cases}$$

The association between an input node (cue) i and and output node (outcome) j is incremented by a constant k every time the two **co-occur** in the same learning trial .

Code for all computational models can be found at
`https://github.com/GiovanniCassani/cross_situational_learning`

Naïve Discriminative Learning [1]

$$V_{ij}^{t+1} = v^t + \Delta V_{ij}$$

$$\Delta V_{ij} = \begin{cases} \alpha_i \beta_1 (\lambda - \sum_{c \in t} V_i) & \text{if } c_i \in t \text{ and } o_j \in t \\ \alpha_i \beta_2 (0 - \sum_{c \in t} V_i) & \text{if } c_i \in t \text{ and } o_j \notin t \\ 0 & \text{if } c_i \notin t \end{cases}$$

Cue-outcome associations are updated according to the Rescorla-Wagner equations: on a learning trial t, the model **predicts** whether an outcome is or isn't present and then check if it was right. The change in association is bigger if the prediction error is large.

Probabilistic Learner [3]

$$a(c|o, O_t, C_t) = \frac{p_{t-1}(o|c)}{\sum_{c' \in C_t} p_{t-1}(o|c')}$$

$$assoc_t(c, o) = assoc_{t-1}(c, o) + a(c|o, O_t, C_t)$$

$$p_t(o|c) = \frac{assoc_t(c, o) + \lambda}{\sum_{o' \in O} assoc_t(c, o') + \beta \cdot \lambda}$$

First computes and updates cue-outcome associations, which are then used to compute a **full probability distribution** over outcomes for each cue. The highest the probability mass allocated to an outcome, the highest the confidence that's the matching outcome.

Hypothesis Testing Model [6]

1. On the first trial, it picks a **single cue-outcome hypothesis** at random.
2. On each subsequent trials, it retrieves a cue-outcome hypothesis (with probability p and checks if it is supported by the trial.
3. If it does not, the hypothesis is dumped and a new one is formed at random. If it does, the hypothesis gets strengthened.

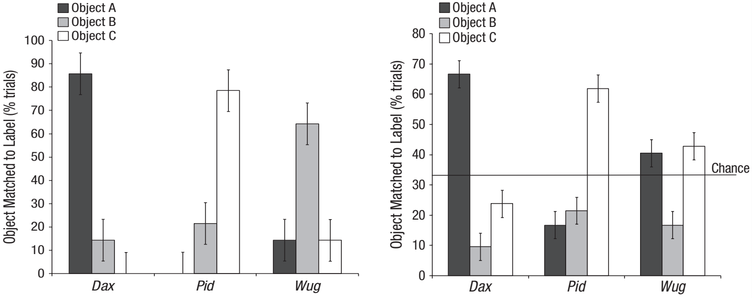
SIMULATIONS

Task definition

200 simulated learners were run on the trials faced by the human subjects in [5], randomizing the order of presentation.

We focused on the cases in which adults and children were consistent, i.e. for words presented during training.

Recap



A good model can **unambiguously pick one object** given a word presented during training. If no object-word association is higher than the others, the model would have to choose at random, unlike human subjects.

Results

Model	Cue	DAX	PID
Hebbian Learner	ObjA	9	.
	ObjB	9	9
	ObjC	.	9
NDL	ObjA	.134 ±.001	-.021 ±.005
	ObjB	.113 ±.005	.113 ±.005
	ObjC	-.021 ±.005	.134 ±.001
Probabilistic Learner	ObjA	.967 ±.003	.
	ObjB	.483 ±.082	.486 ±.082
	ObjC	.	.967 ±.003
HTM	ObjA	.455	.
	ObjB	.545	.485
	ObjC	.	.515

Conclusion

Upshot

Not all cross-situational learners are created equal: two fitted the data, two didn't.

Human learners don't care if spurious associations occur as frequently as true associations. Actually, in our dataset there are no spurious or true associations: however, the co-occurrences of ObjectB with both labels are **perceived** as spurious.

Conclusions

Human cross-situational learning doesn't depend only on words and referents co-occurences, but much more on the their systematicity: *a model needs to be able to also learn from situations where things fail to co-occur*, not simply from situations were two things co-occur.

References I

R. H. Baayen, P. Milin, D. F. Durdević, P. Hendrix, and M. Marelli.
An amorphous model for morphological processing in visual comprehension based on naive discriminative learning.
Psychological Review, 118(3):438–481, 2011.

G. Cassani, R. Grimm, S. Gillis, and W. Daelemans.
Constraining the search space in cross-situational learning: Different models make different predictions.
In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, 2016.

A. Fazly, A. Alishahi, and S. Stevenson.
A probabilistic computational model of cross-situational word learning.
Cognitive Science, 34(6):1017–1063, 2010.

References II

D. O. Hebb.
The organization of behavior.
John Wiley and Sons, New York, NY, 1949.

M. Ramscar, M. Dye, and J. Klein.
Children value informativity over logic in word learning.
Psychological Science, 24(6):1017–1023, 2013.

J. C. Trueswell, T. N. Medina, A. Hafri, and L. R. Gleitman.
Propose but verify: Fast mapping meets cross-situational word learning.
Cognitive Psychology, 66(1):126–156, 2013.

Thank you!

Questions?