Modeling language acquisition:
From phonology to meaning

Christina Bergmann (MPI Psycholinguistics, Nijmegen, NL)
christina.bergmann@mpi.nl
@chbergma

Welcome to day 2

Recap day 1

What is a model/who is a modeller?

Learning sound categories

Segmenting speech

Any open questions?

Key aspects of a model

1. **What** should be modelled? What will be excluded from the model?
   Goal
2. Which **processing abilities** will the model have?
   How
3. Which **theories and data** will the model build on?
   Prerequisites, innate knowledge
4. Which **level of abstraction** will be modelled?
5. How will the model be **evaluated**?

Example 3: Word learning

- Children learn new words under referential uncertainty ("gavagai")
- They use "fast mapping"

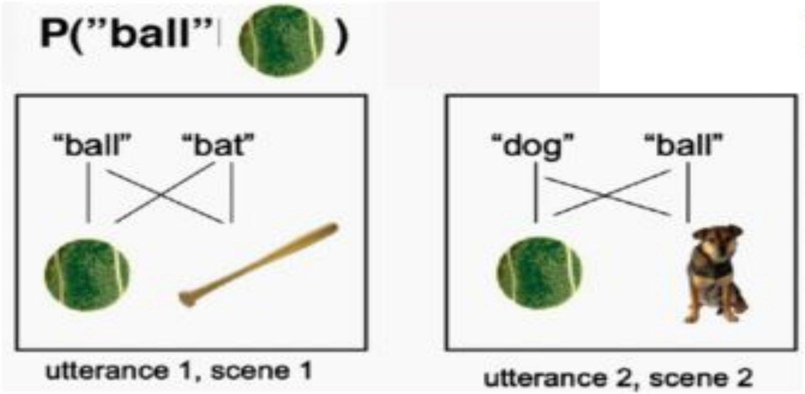
$$P(H \mid D) = \frac{P(D \mid H) * P(H)}{P(D)}$$

Posterior probability

Likelihood

Prior

Probability of observing the data, no matter what hypothesis is true



Example 3: Word learning

→ Fast mapping (Trueswell et al., 2013)

Zud = HAND?

Example 3: Word learning

→ Fast mapping (Trueswell et al., 2013)

~~Zud = HAND?~~

Zud = EYE?

Example 3: Word learning

Goal

- Input: Words and "scenes", Output: Associations between words and meanings (from those "scenes")

Processing abilities

- Calculate probabilities of word-meaning co-occurrence
- Segmentation abilities (word-level), object perception

Level of abstraction

- No direct link to biology (algorithmic)

Evaluation: Correct associations

Example 3: Word learning - Implementations

2 implementations of these specifications

1. Fazly et al.
2. Yu & Ballard

Example 3: Word learning - Implementation 1

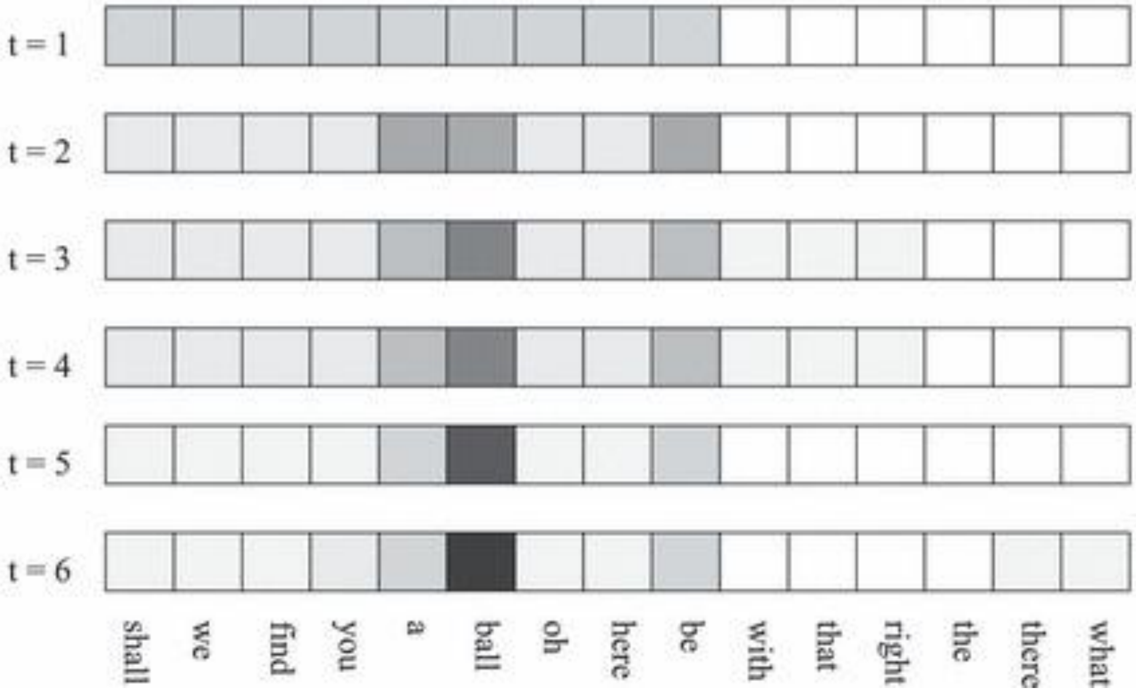
Objects → "Concepts"

Example

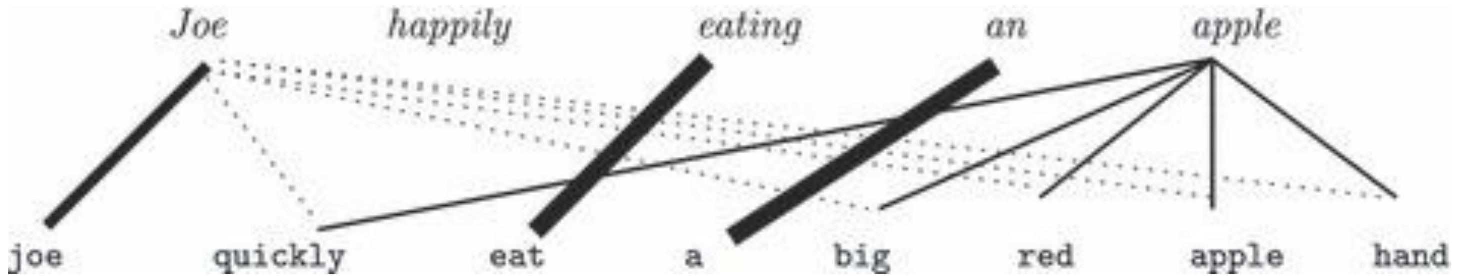
Utterance	**Scene**
shall we find you a ball	{shall, we, find, you, a, ball, oh, here, be}
with a ball	{with, a, ball, that, be, right}
the ball there	{the, ball, there, and, what, about, boat}
get your other ball under there look	{get, your, other, under, there, look, cooker}
the ball what	{the, ball, what, touch, it}
do you kick the ball	{do, you, kick, the, ball, what, else}

Example

Utterance	Scene
shall we find you a ball	{shall, we, find, you, a, ball, oh, here, be}
with a ball	{with, a, ball, that, be, right}
the ball there	{the, ball, there, and, what, about, boat}
get your other ball under there look	{get, your, other, under, there, look, cooker}
the ball what	{the, ball, what, touch, it}
do you kick the ball	{do, you, kick, the, ball, what, else}



Example 3: Word learning



Fazly, A., Alishahi, A., & Stevenson, S. (2010). A Probabilistic Computational Model of Cross-Situational Word Learning. *Cognitive Science*, *34*(6), 1017-1063.

Example 3: Word learning

Strengths

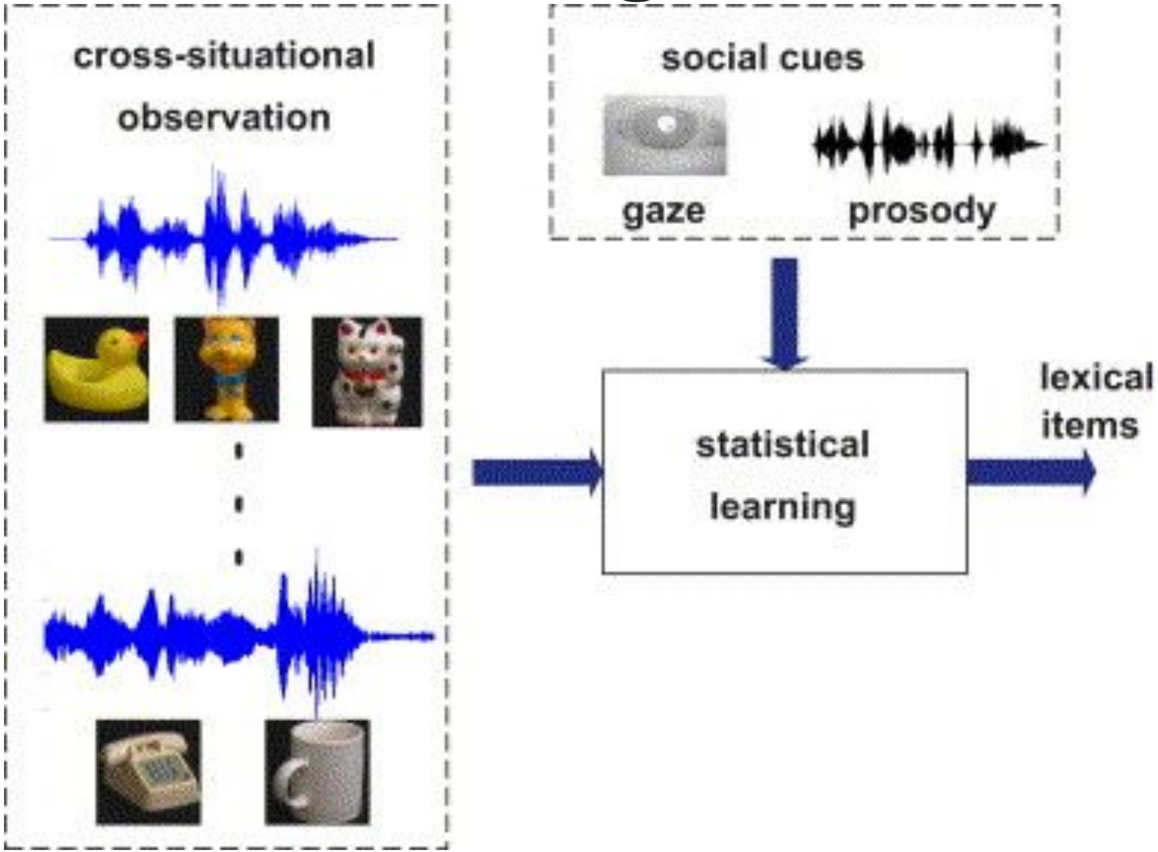
- Mutual exclusivity falls out of the model
- Graded associations, mirrors uncertainty about the presence of a referent
  - Can deal with natural situations
- Simple, incremental learning mechanism
  - Updates after every new observation
  - We can trace learning curves based on input and change input properties to investigate the consequences (frequency, referential uncertainty, noise)

Example 3: Word learning

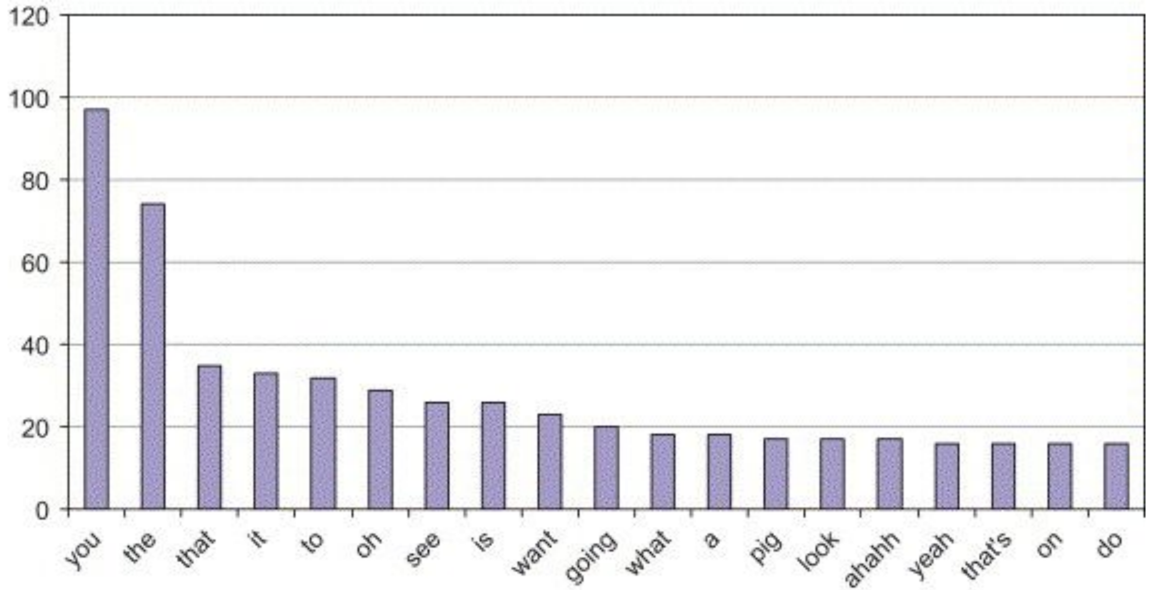
Weaknesses

- Input highly simplified
  What about prosodic, social, and other cues? No learning of grammar/syntax
  → Necessary simplification
- Acquisition of phonology and the ability to segment words are assumed
  → Children actually learn words while still acquiring those abilities
  → Word segmentation is actually aided by known words
- What about abstract words? (Cf. referent vs concept)

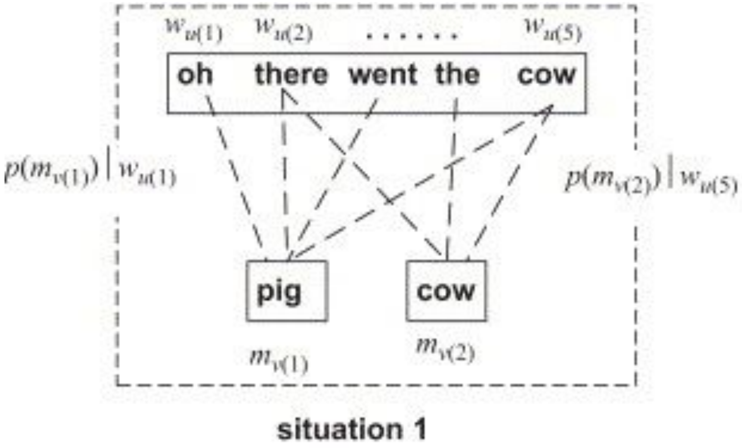
Example 3: Word learning - Implementation 2



Example 3: Word learning - Implementation 2



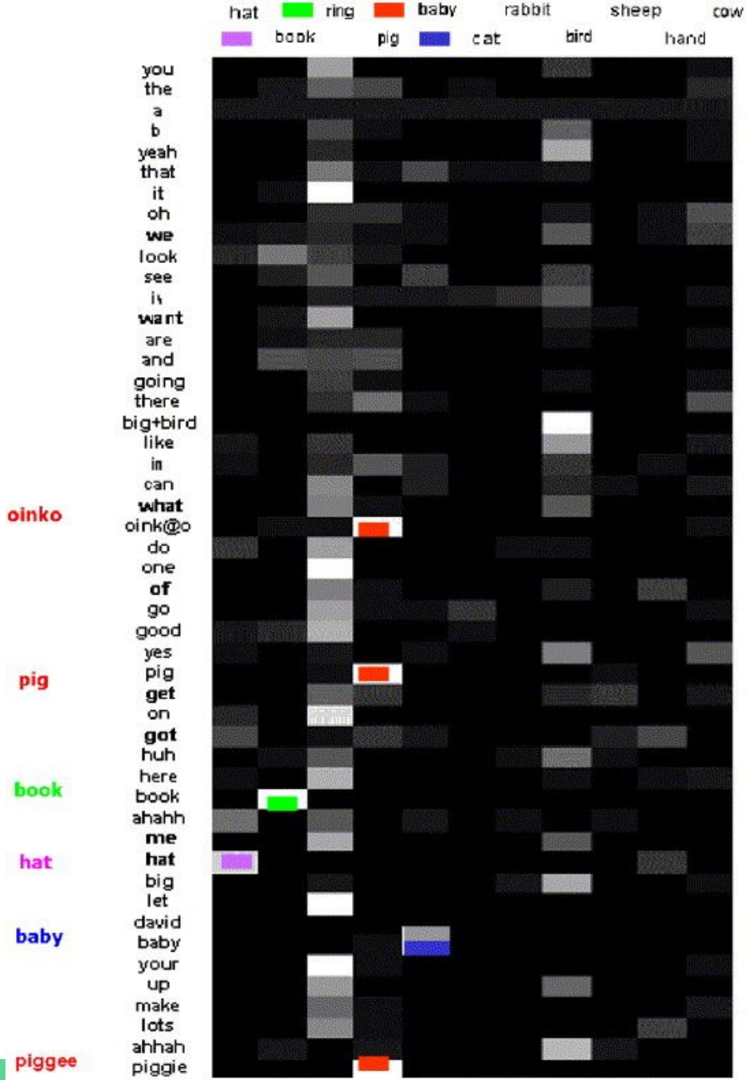
Example 3: Word learning - Implementation 2



Example 3: Word learning

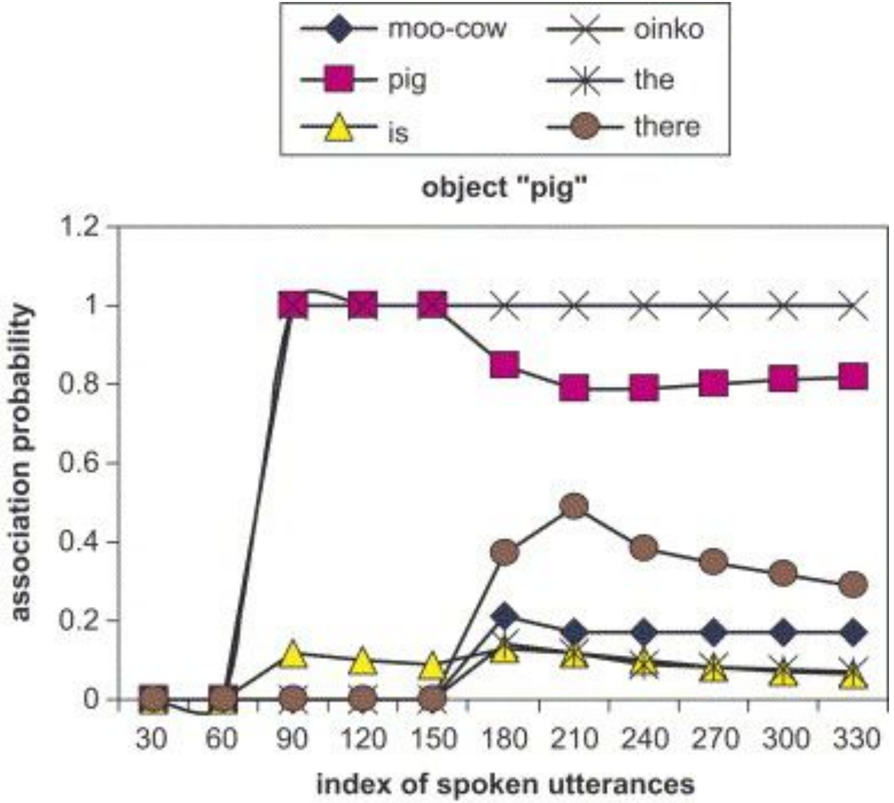
Evaluation: Association matrix

→ Is the "correct" concept linked to the right word?



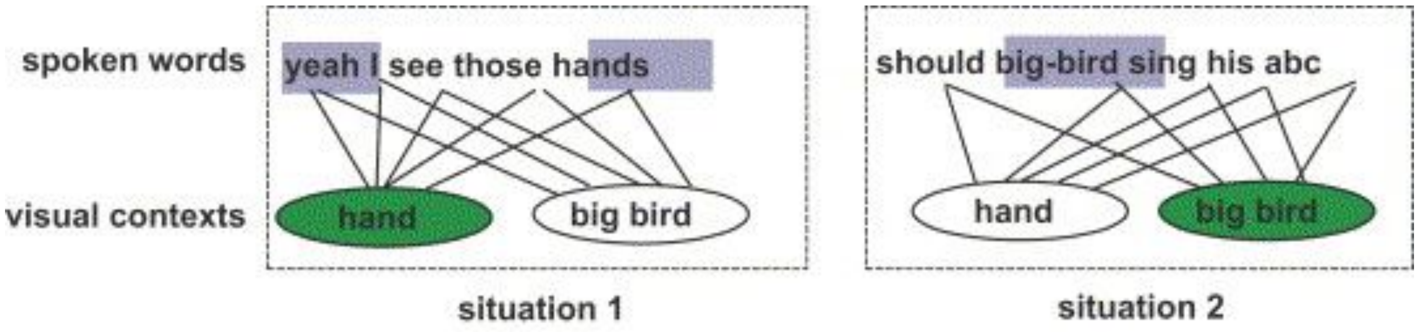
Example 3: Word learning - Implementation 2

Evaluation: Time course



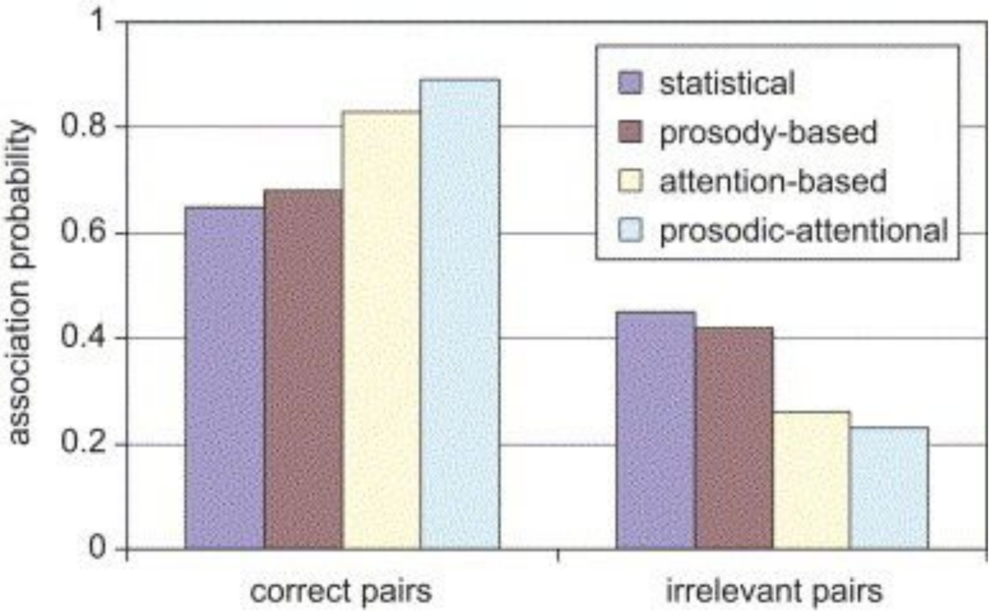
Example 3: Word learning - Implementation 2

Adding "social cues"



Example 3: Word learning - Implementation 2

Evaluation



Example 3: Word learning - Implementation 2

Strengths

- (Many shared with previous model)
- Realistic cues to meaning
- Null associations

Weaknesses

- "Batch" algorithm
- Are the gains due to prosody / attention real?
- Assumptions about segmentation / speech representation
- Can the results be compared to what children would have learned?

Example 4: Word learning from speech

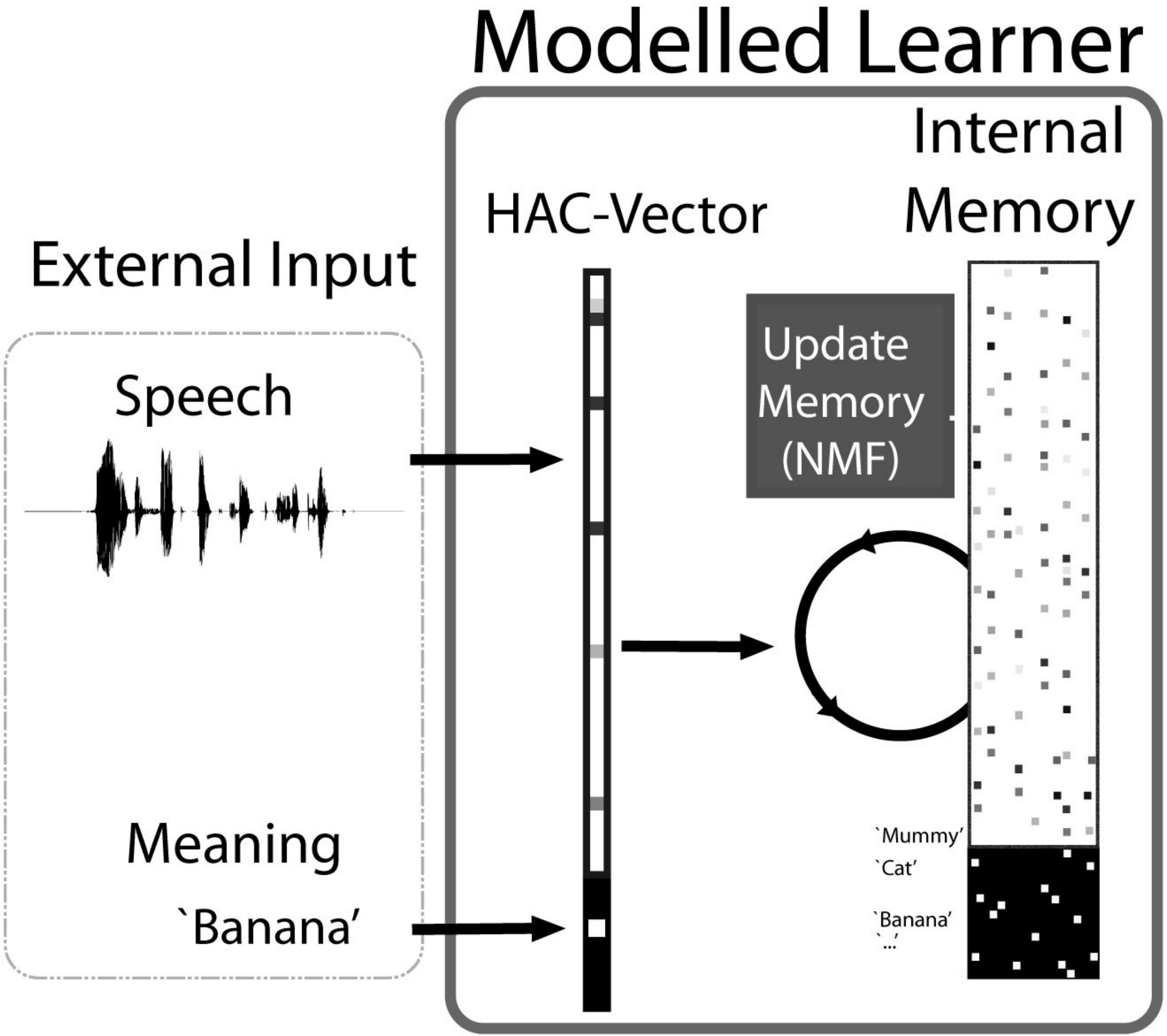
Can we build a similar model with speech?

→ Make variable-length waveforms computer readable
(often required: fixed-length representation)

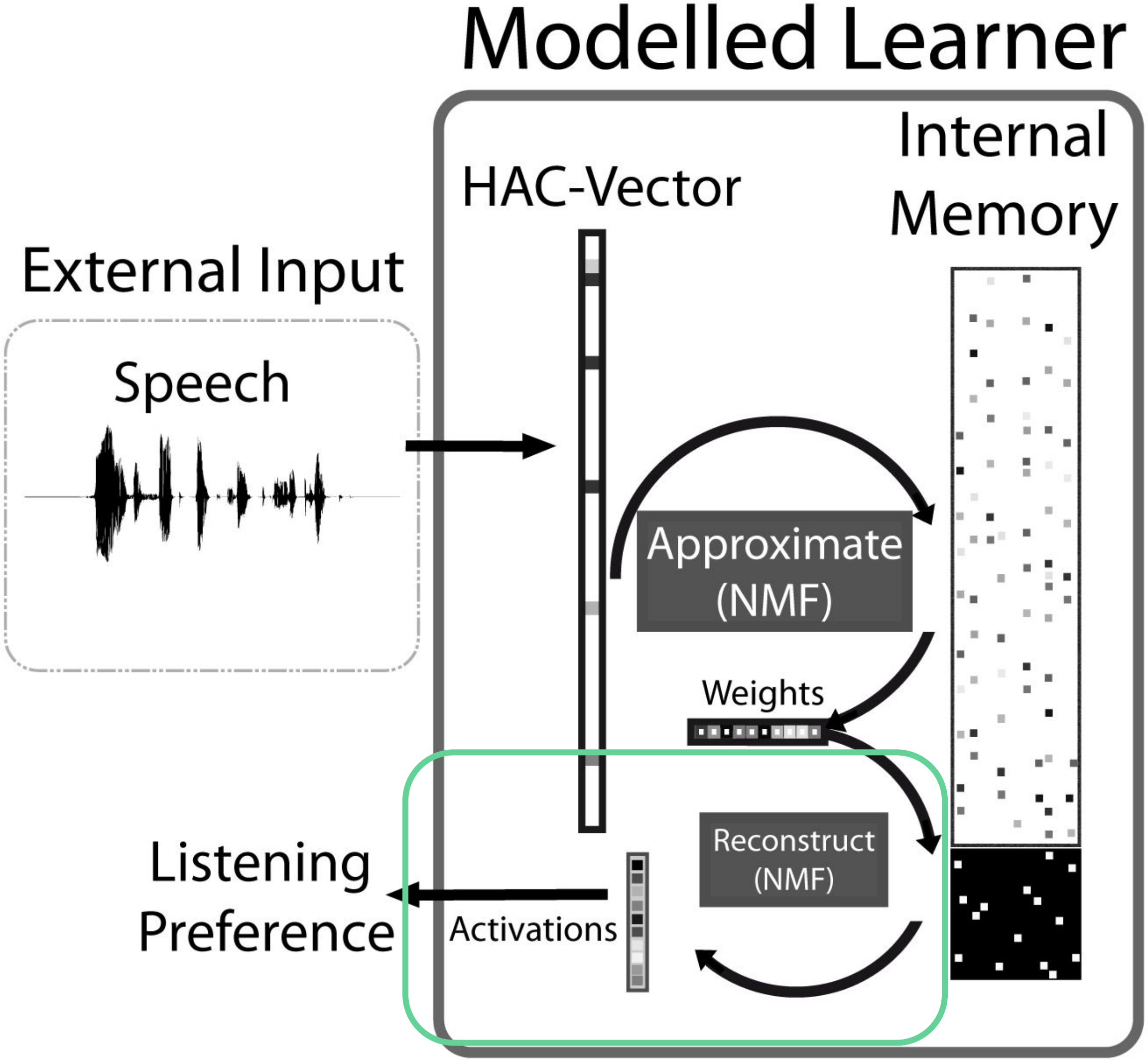
Input: Simple sentences, 9 concepts

Output: Associations between new sentences and all concepts

Example 4:

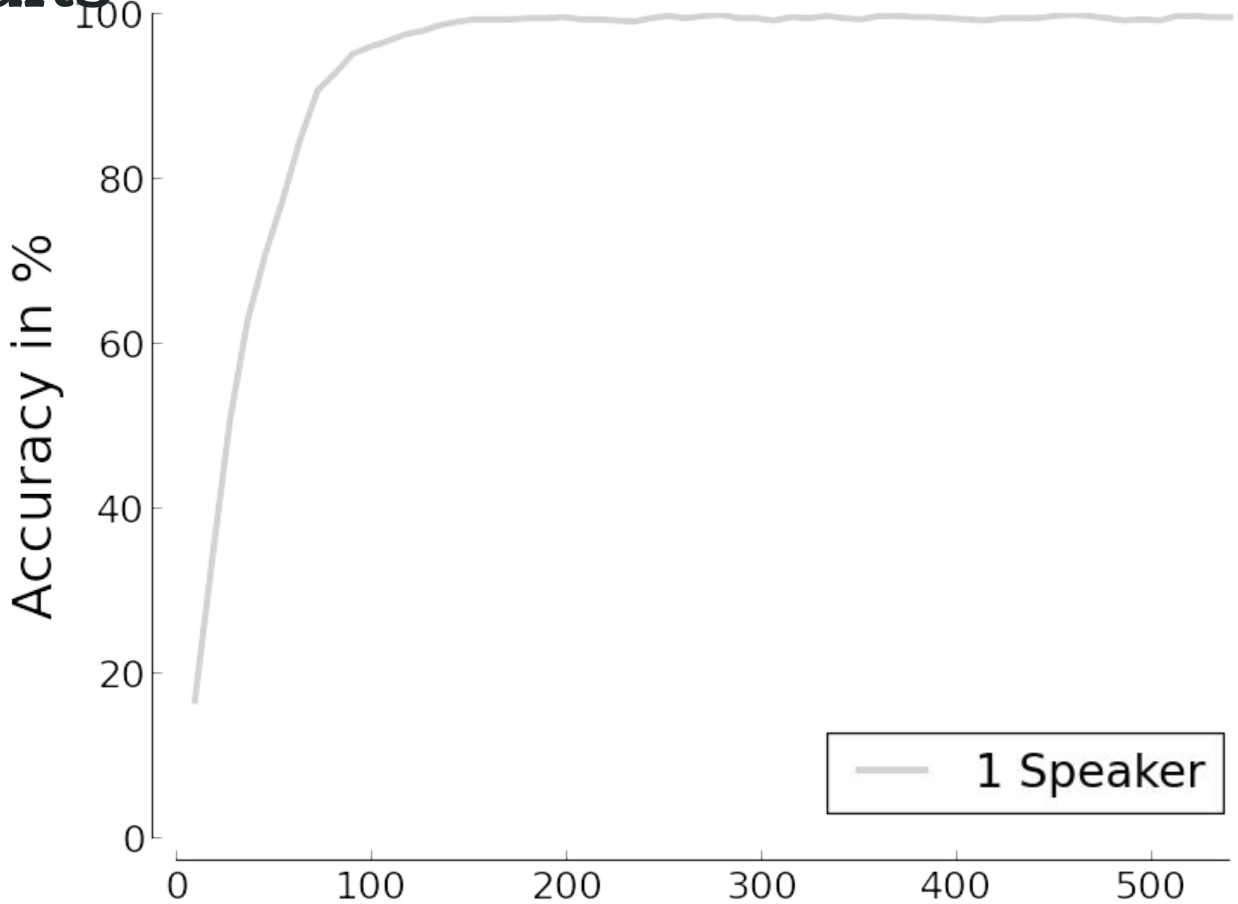

Modelled Learner
External Input
Speech
Meaning
`Banana'
HAC-Vector
Internal
Memory
Update
Memory
(NMF)
`Mummy'
`Cat'
`Banana'
`...'

Example 4:

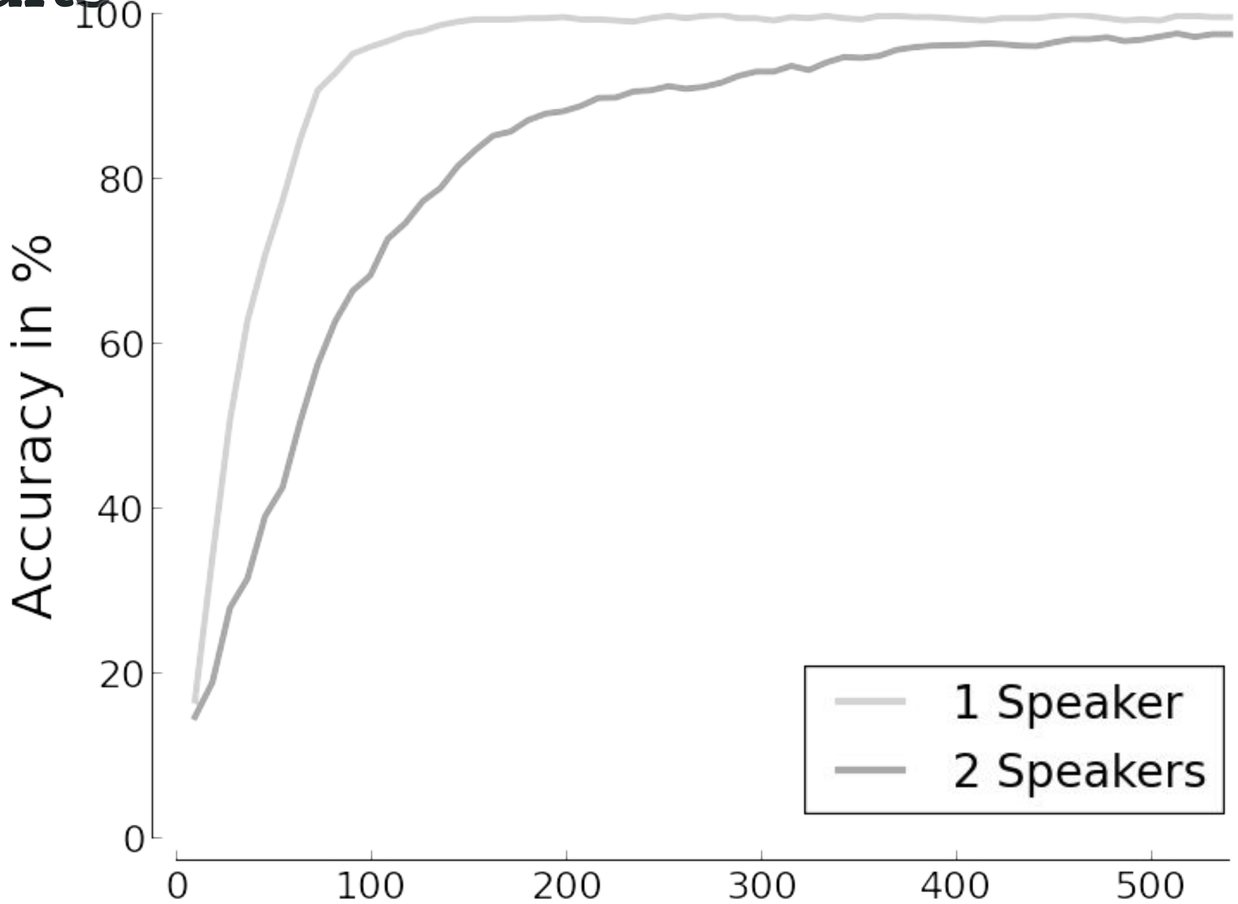


Modelled Learner
External Input
Speech
HAC-Vector
Internal
Memory
Approximate
(NMF)
Weights
Reconstruct
(NMF)
Listening
Preference
Activations

Results

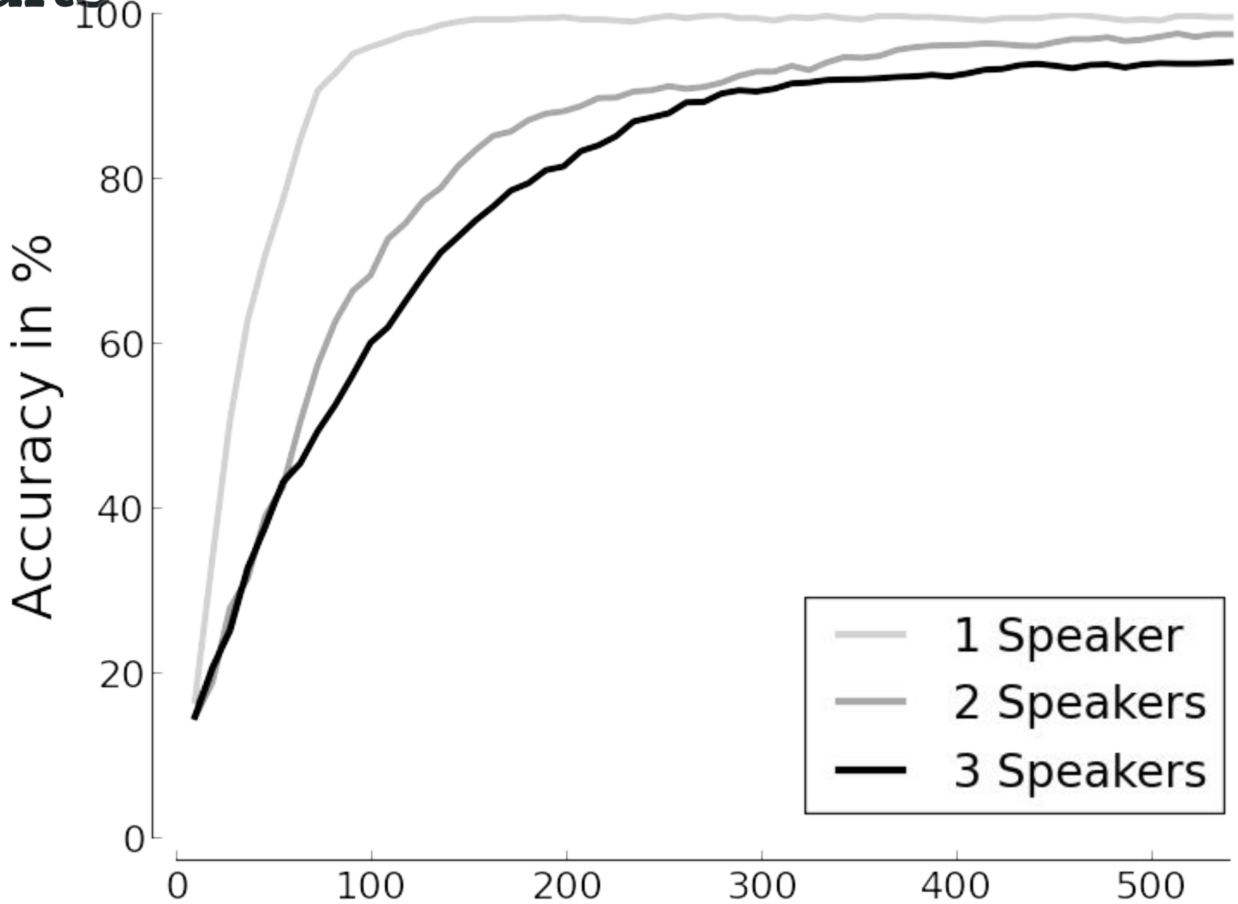


Results



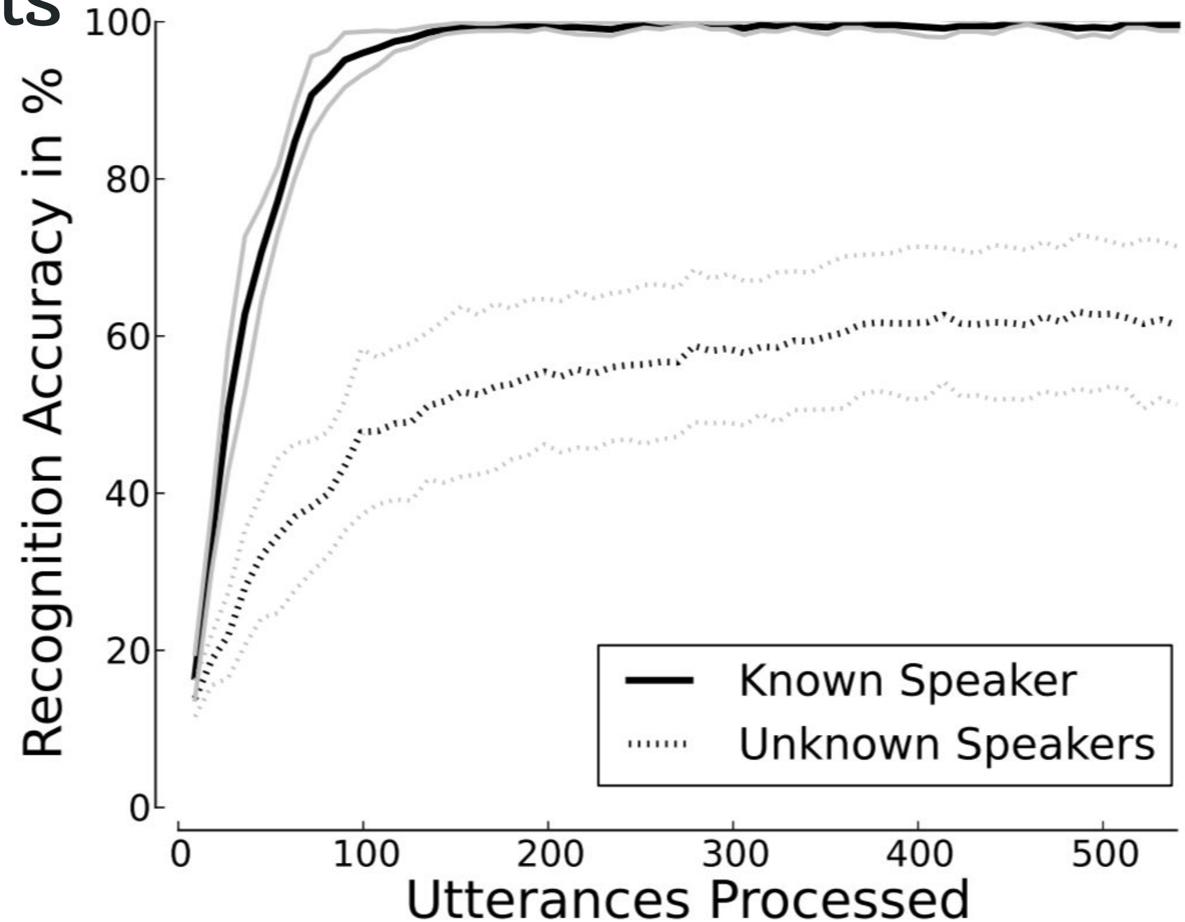
Accuracy in %
100
80
60
40
20
0
0
100
200
300
400
500
1 Speaker
2 Speakers

Results



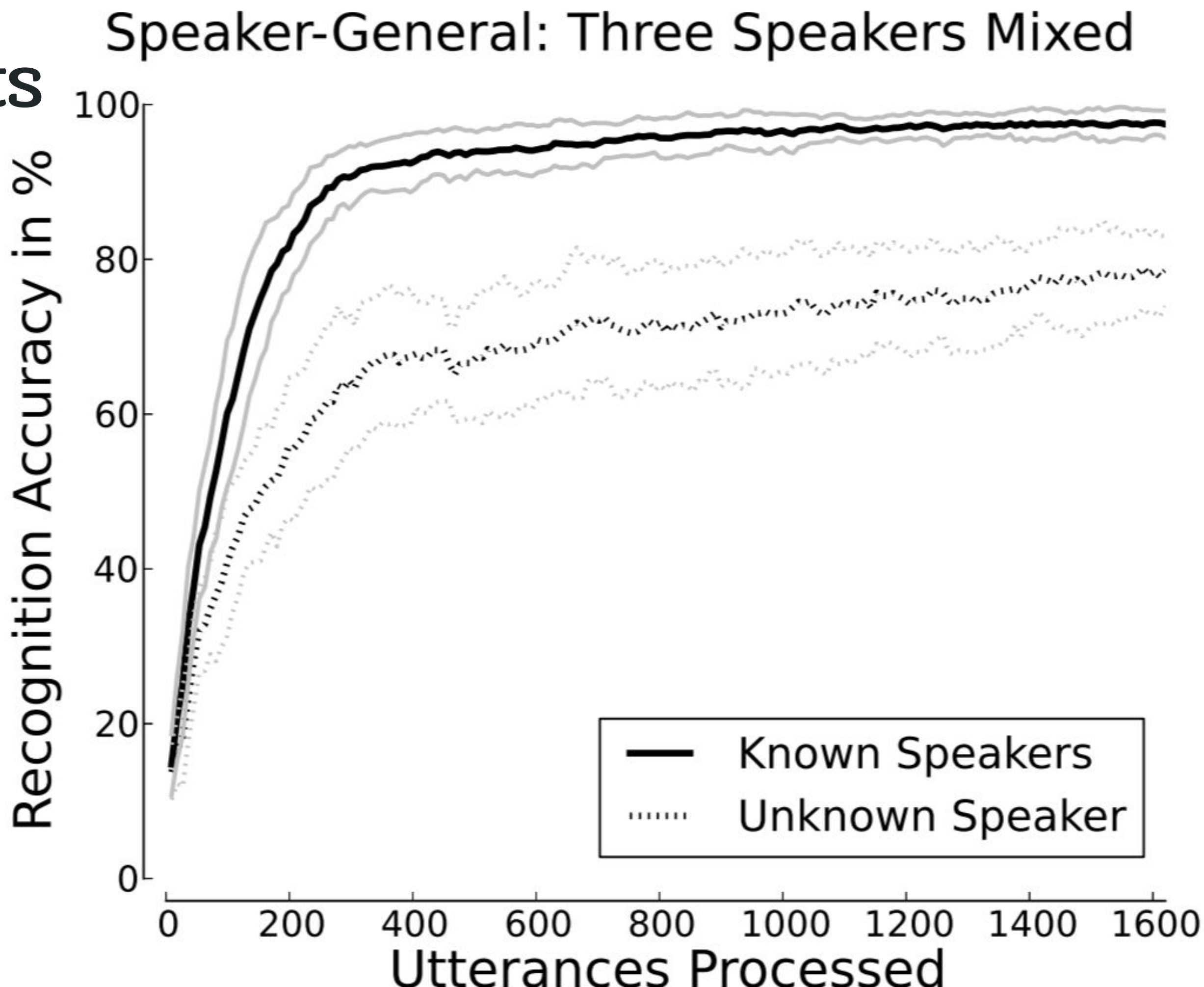
Accuracy in %
100
80
60
40
20
0
0
100
200
300
400
500
1 Speaker
2 Speakers
3 Speakers

Results



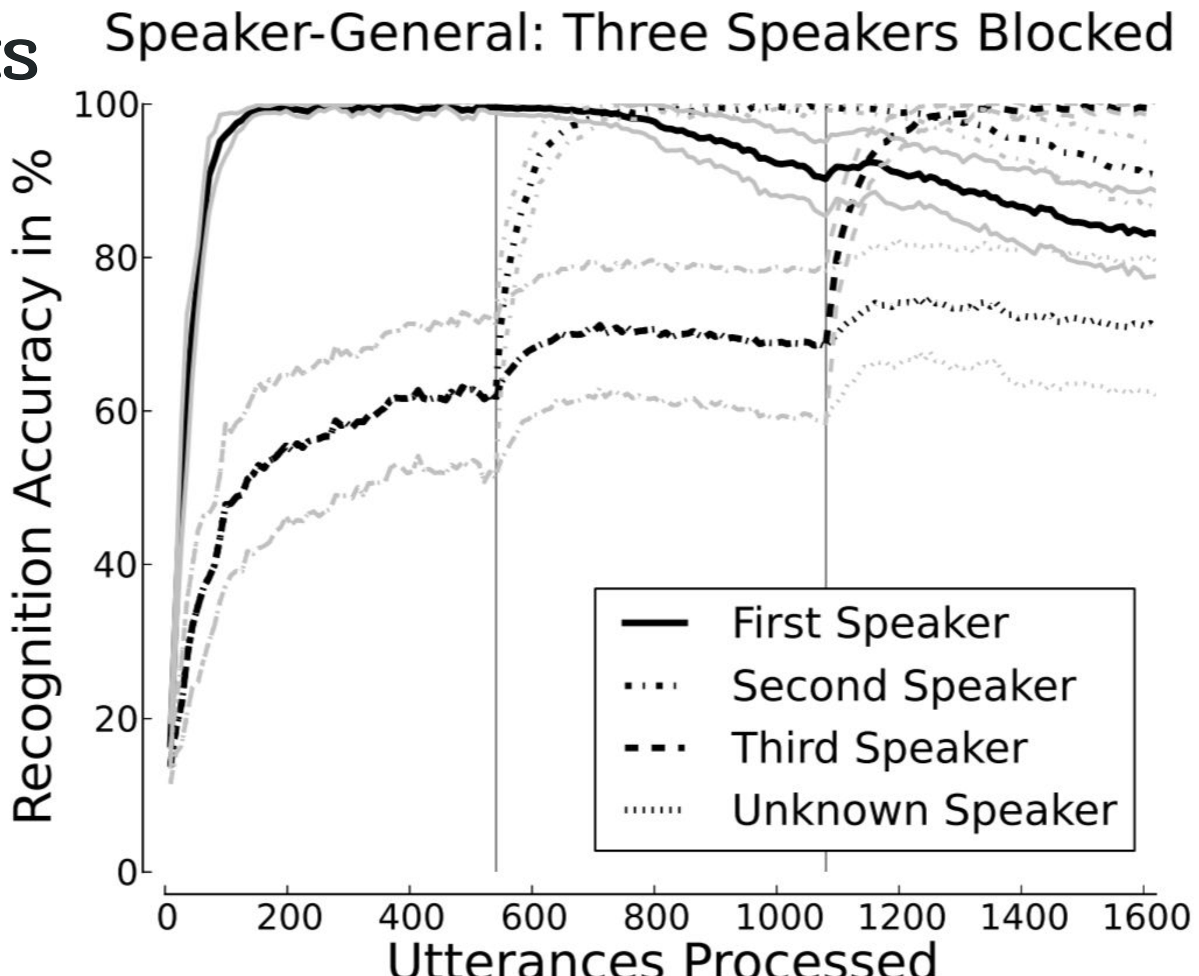
Recognition Accuracy in %
Utterances Processed
0
20
40
60
80
100
0
100
200
300
400
500
Known Speaker
Unknown Speakers

Results



Speaker-General: Three Speakers Mixed
Recognition Accuracy in %
Utterances Processed
Known Speakers
Unknown Speaker
0
20
40
60
80
100
0
200
400
600
800
1000
1200
1400
1600

Results


Speaker-General: Three Speakers Blocked
Recognition Accuracy in %
Utterances Processed
0
20
40
60
80
100
0
200
400
600
800
1000
1200
1400
1600
First Speaker
Second Speaker
Third Speaker
Unknown Speaker

Example 4: Summary

Strengths

- Incremental learning
- Can simulate impact of indexical variation
- No assumptions about segmentation, phonology

Weaknesses

- Learned only 9 words
- Indirect assessment

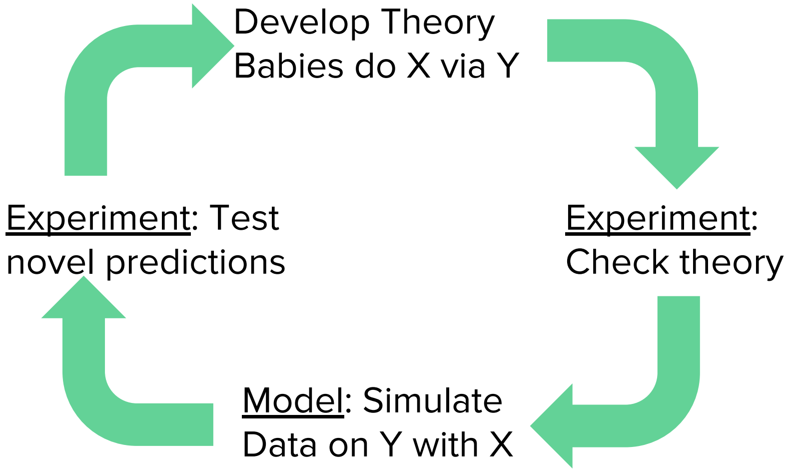
Which words and cues might work in real life?

Can we extract our own corpus statistics?

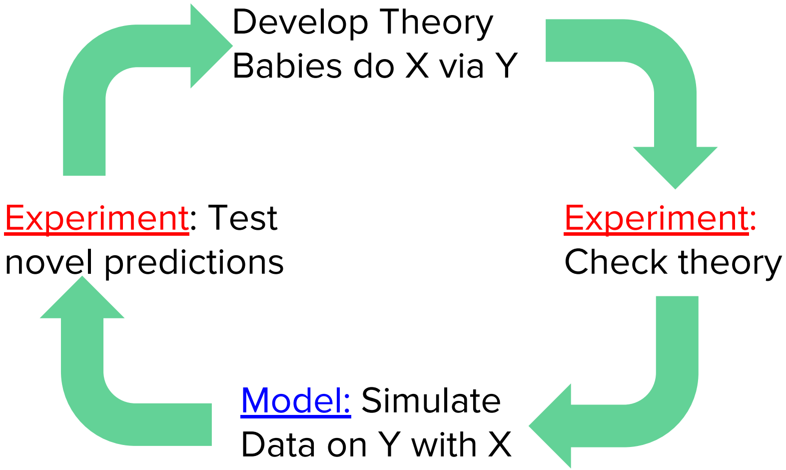
What would be a good outcome measure?

Let’s look at some code

Bridging the gap between models and infant studies



Bridging the gap between models and infant studies



Bridging the gap between models and infant studies

Data to be simulated:

- Behavior, measured by looking times, pointing, attention

  → External observation vs internal processes

Bridging the gap between models and infant studies

Data to be simulated:

- Behavior, measured by looking times, pointing, attention

  → External observation vs internal processes

Simulated data:

- Idealized experimental data (on the group- not the infant-level!)
- Internal processes and knowledge ("word recognition")

  → Sidestepping behavior and measurement

Bridging the gap between models and infant studies

Data to be simulated:

- Behavior, measured by looking times, pointing, attention

Simulated data:

- Idealized experimental data (on the group- not the infant-level!)
- Internal processes and knowledge ("word recognition")

Problems:

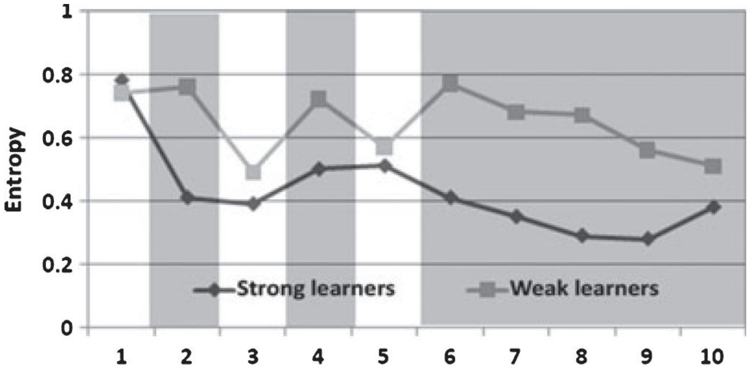
1. Assumption that internal processes == overt behavior

   → Measurement noise? False positives? (Replication crisis)
2. Differences between participants might be meaningful

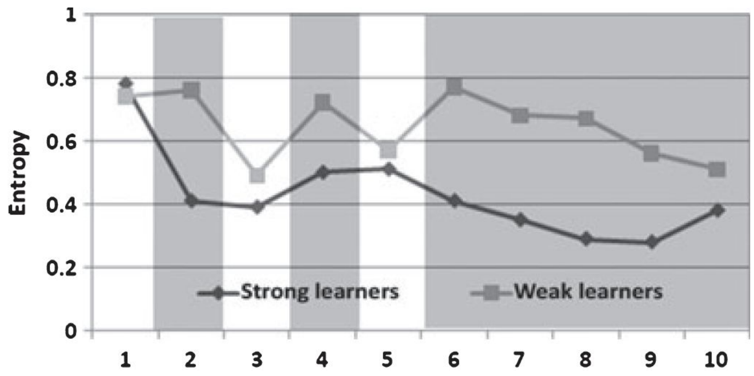
Bridging the gap between models and infant studies

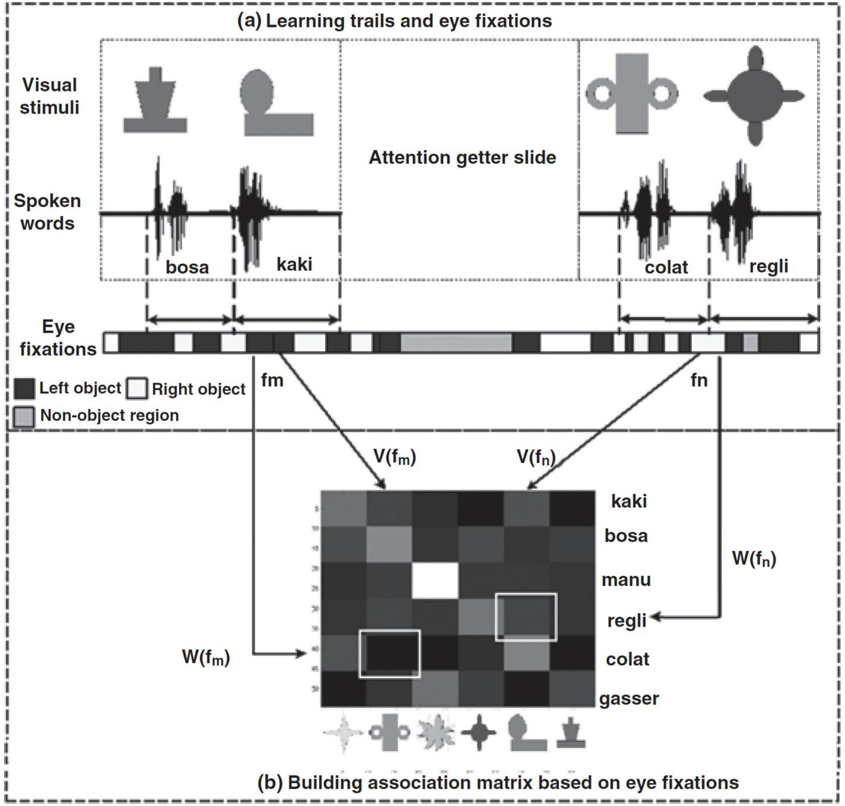
Idea: Simulate data from a specific study

Modelling a specific study



Modelling a specific study





Modelling a specific study

Note: Different time scale

Advantage:

- Direct link to child performance
- Very simple algorithm

Disadvantage:

- Difficult to predict unseen data / conditions
- Not a learning model in the narrow sense

→ Explanation, not prediction (proof of concept)

Bridging the gap between models and infant studies

Idea: Model the task along with the internal process

Example: Word segmentation from native speech

(s. Bergmann, ten Bosch, Fikkert, & Boves, 2013; http://journal.frontiersin.org/Journal/10.3389/fpsyg.2013.00676/full)

Bridging the gap between models and infant studies

Idea: Model the task along with the internal process

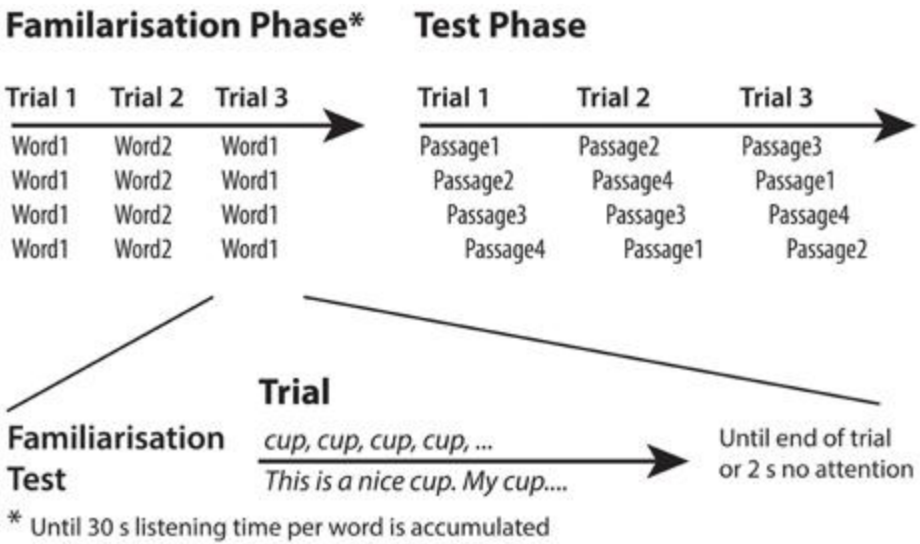
Example: Word segmentation from native speech

Part 1: internal activations (= "word recognition")

Part 2: External behavior in the task

(s. Bergmann, ten Bosch, Fikkert, & Boves, 2013; http://journal.frontiersin.org/Journal/10.3389/fpsyg.2013.00676/full)

Bridging the gap between models and infant studies



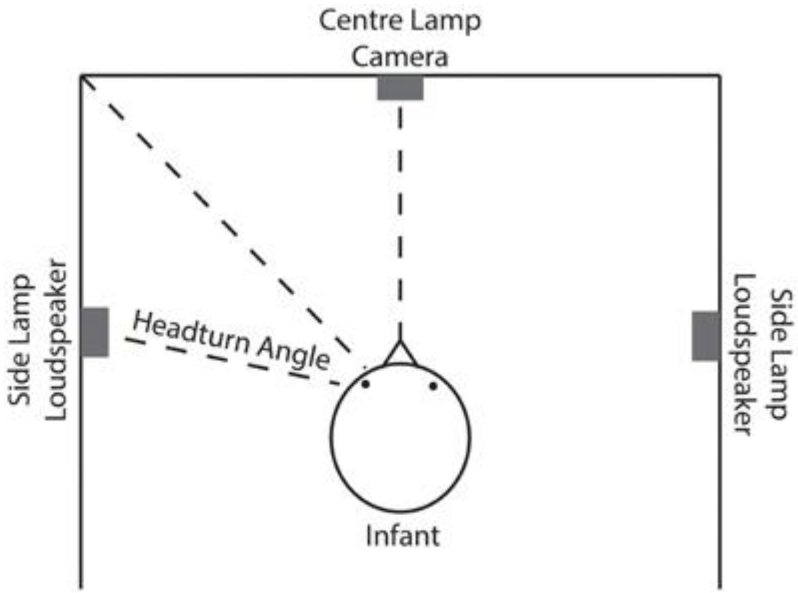
(s. Bergmann, ten Bosch, Fikkert, & Boves, 2013; http://journal.frontiersin.org/Journal/10.3389/fpsyg.2013.00676/full)

Bridging the gap between models and infant studies

Cup:	The cup was bright and shiny.
	A clown drank from the red cup.
	The other one picked up the big cup.
	His cup was filled with milk.
	Meg put her cup back on the table.
	Some milk from your cup spilled on the rug.
Dog:	The dog ran around the yard.
	The mailman called to the big dog.
	He patted his dog on the head.
	The happy red dog was very friendly.
	Her dog barked only at squirrels.
	The neighborhood kids played with your dog.
Feet:	The feet were all different sizes.
	This girl has very big feet.
	Even the toes on her feet are large.
	The shoes gave the man red feet.
	His feet get sore from standing all day.
	The doctor wants your feet to be clean.
Bike:	His bike had big black wheels.
	The girl rode her big bike.
	Her bike could go very fast.
	The bell on the bike was really loud.
	The boy had a new red bike.
	Your bike always stays in the garage.

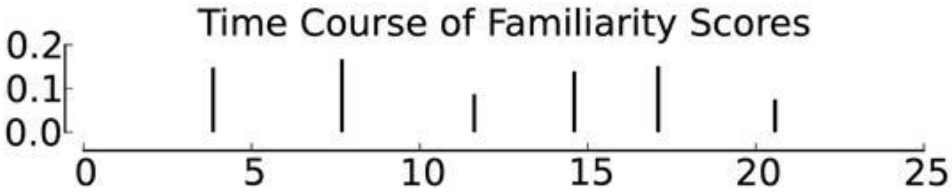
(s. Bergmann, ten Bosch, Fikkert, & Boves, 2013; http://journal.frontiersin.org/Journal/10.3389/fpsyg.2013.00676/full)

Bridging the gap between models and infant studies



(s. Bergmann, ten Bosch, Fikkert, & Boves, 2013;
http://journal.frontiersin.org/Journal/10.3389/fpsyg.2013.00676/full)

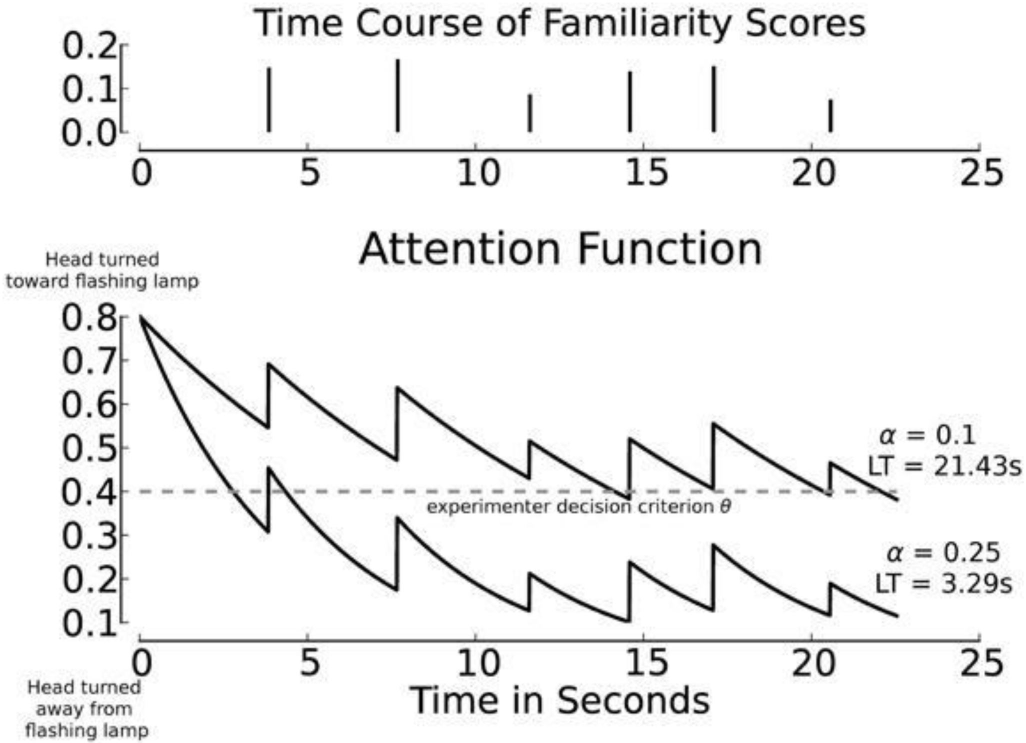
Bridging the gap between models and infant studies



Part 1: internal activations (= "word recognition")

(s. Bergmann, ten Bosch, Fikkert, & Boves, 2013; http://journal.frontiersin.org/Journal/10.3389/fpsyg.2013.00676/full)

Bridging the gap between models and infant studies

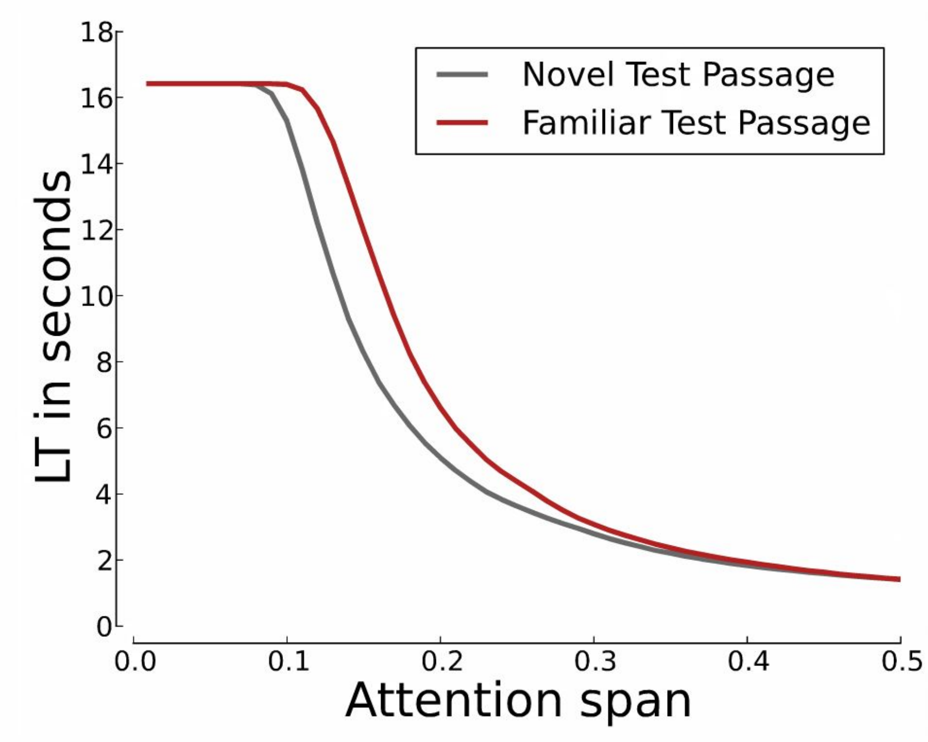


Part 1: internal activations (= "word recognition")

Part 2: External behavior in the task

(s. Bergmann, ten Bosch, Fikkert, & Boves, 2013; http://journal.frontiersin.org/Journal/10.3389/fpsyg.2013.00676/full)

Bridging the gap between models and infant studies



Simulates listening times (LT)
Models dependency on infants' attention span via α

→ Behavior is (also) influenced by non-linguistic factors which vary across children

(s. Bergmann, ten Bosch, Fikkert, & Boves, 2013; http://journal.frontiersin.org/Journal/10.3389/fpsyg.2013.00676/full)

Bridging the gap between models and infant studies

Summary: Observed behavior is influenced by language processing abilities **and** non-linguistic factors (attention span)

Note: The underlying algorithm did not *segment* words from speech, just calculated (sort of) acoustic similarities between words and sentences…
→ The proposed ability can be simulated in different ways, too

(s. Bergmann, ten Bosch, Fikkert, & Boves, 2013; http://journal.frontiersin.org/Journal/10.3389/fpsyg.2013.00676/full)

Limitations of the word learning models

1. Input a tiny fraction of what infants hear
2. What is the link to infant behavior?
   a. Think eg of learning nouns before function words
3. Assumptions about infant memory

Modelling different kinds of data

So far:

- Vocabulary norms

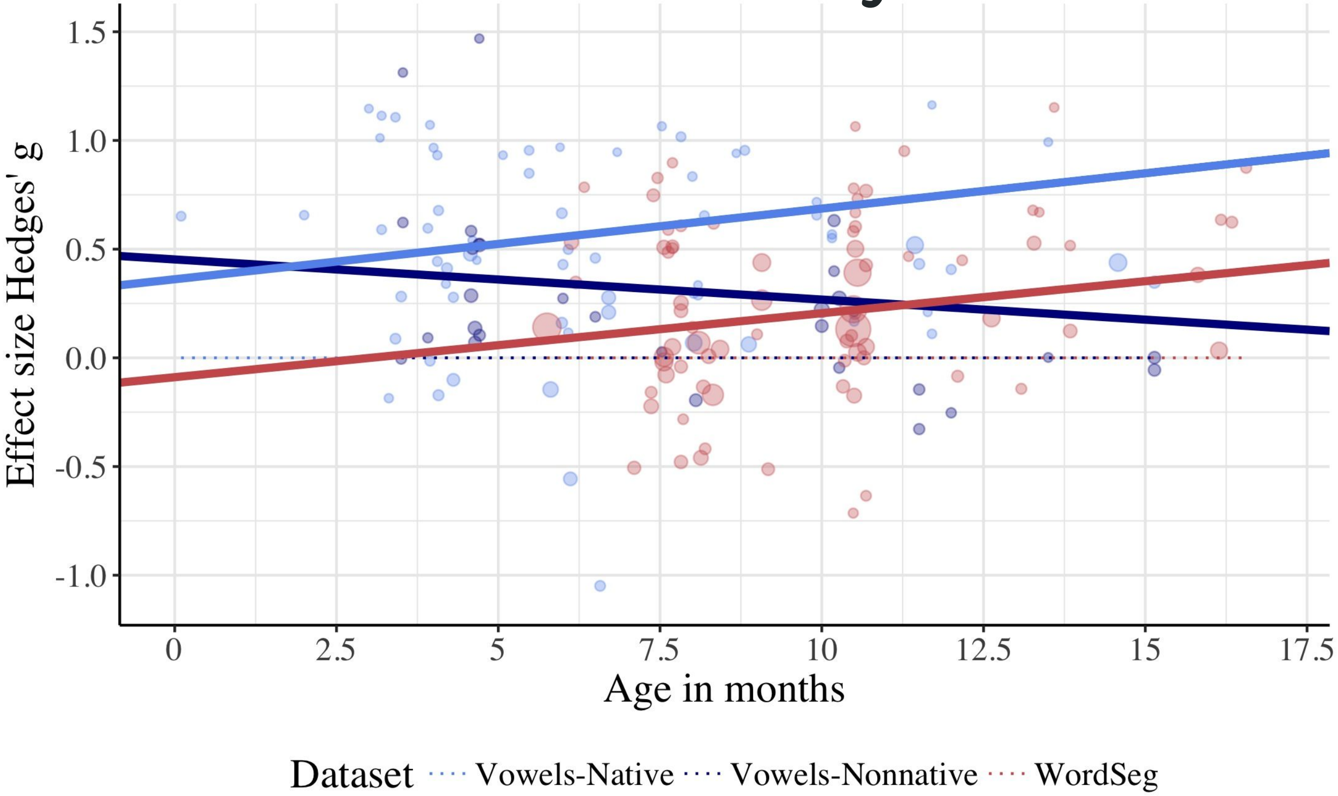
- Produced sentences

- Abstract skills

- Overt behavior

... What else could we model?

Different data 1: Meta-analyses



Bergmann, Tsuji, & Cristia (2017). DOI: [10.21437/Interspeech.2017-1443](https://doi.org/10.21437/Interspeech.2017-1443)

Different data 2: Large-scale studies

… forthcoming …

Language 0-5

Do infants solve the same problem as models?

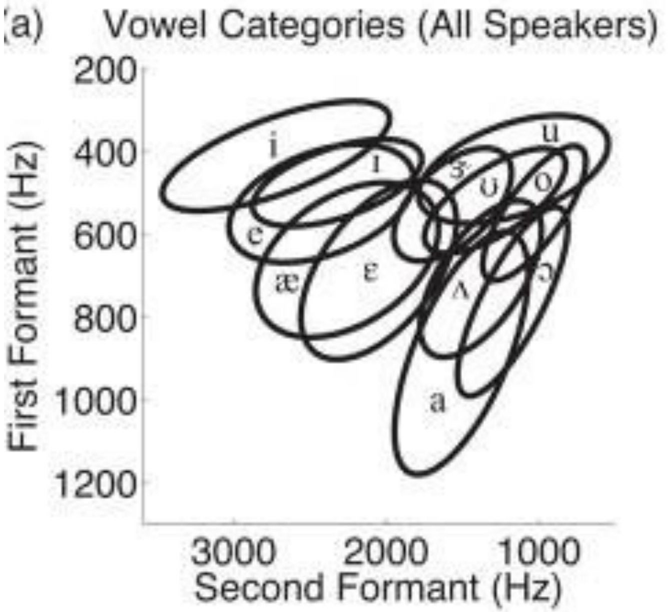
Do infants solve the same problem as models?

Are we over-simplifying?

Would information from other linguistic levels actually help?

Do infants solve the same problem as models?

The problem: adult target

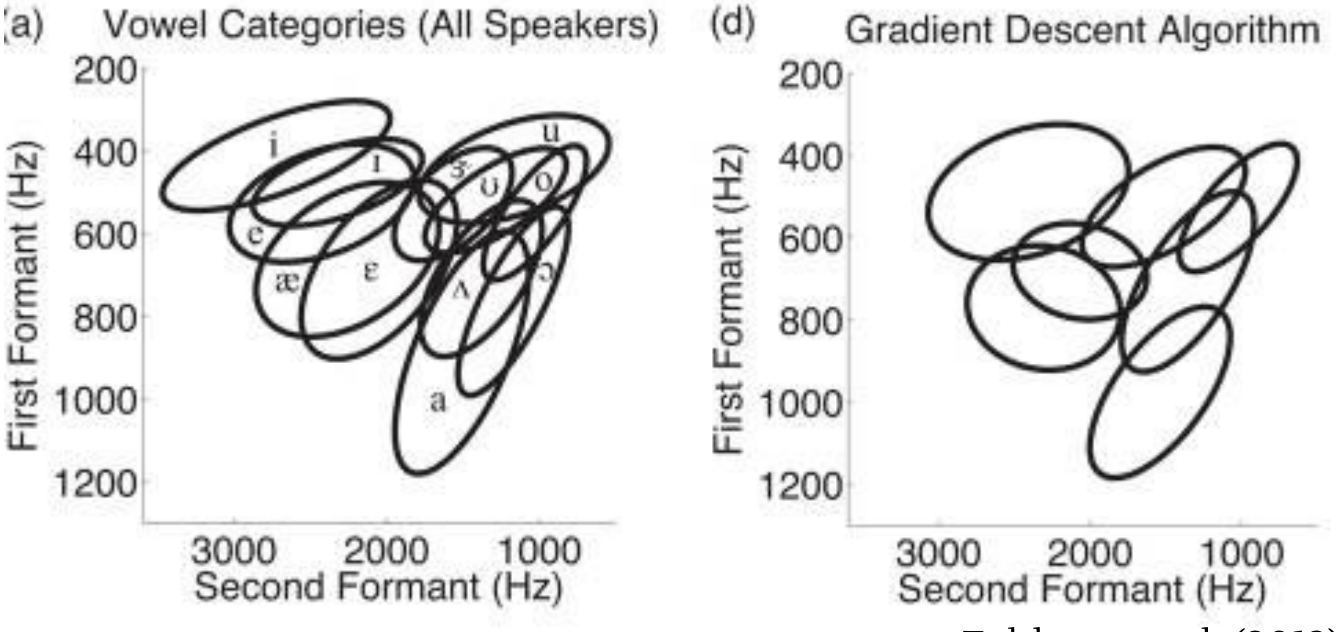


Feldman et al. (2013). DOI: [10.1037/a0034245](https://doi.org/10.1037/a0034245)

Do infants solve the same problem as models?

The problem: input

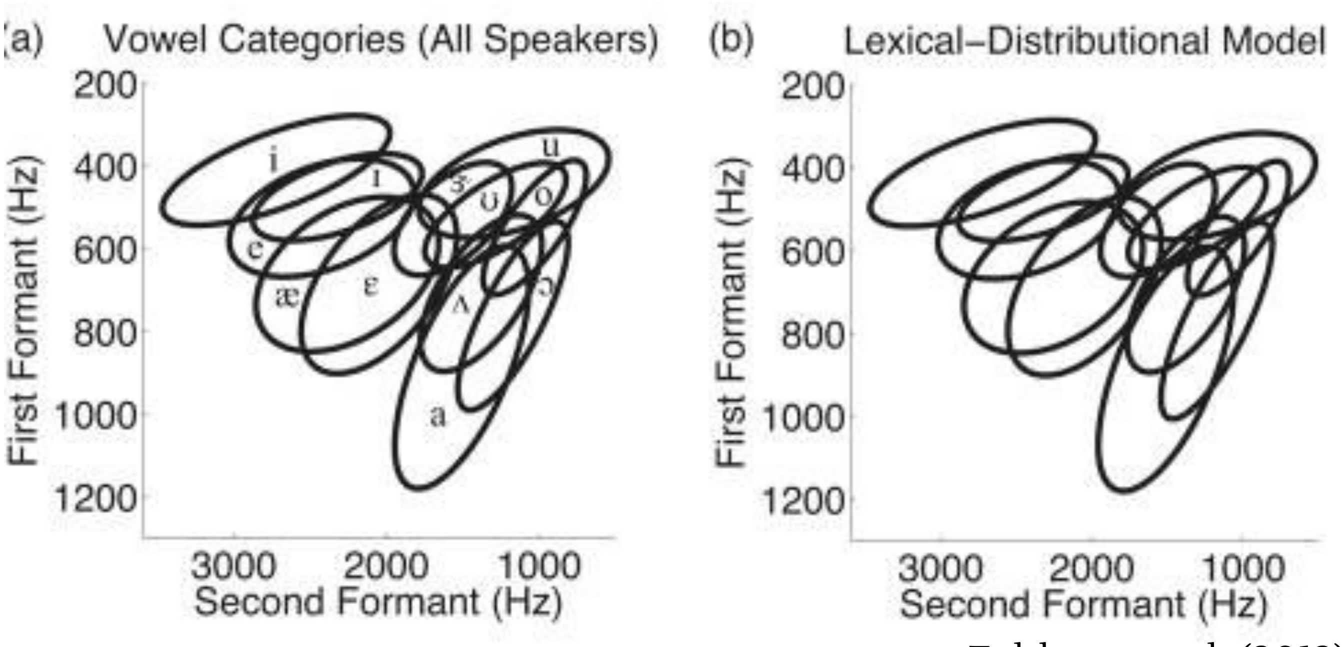
The simulated learner



Feldman et al. (2013). DOI: 10.1037/a0034245

Do infants solve the same problem as models?

A possible solution: Word knowledge



Feldman et al. (2013). DOI: 10.1037/a0034245

Summary

1. We can model all kinds of data - individual infants and groups, short-term learning and long-term development
2. How we model infants (and their abilities) and compare this to experimental data is crucial
3. Sometimes, simplification can make a problem harder

What would YOU like to model?