Modeling language acquisition:
From phonology to meaning

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

Overview

Day 1

1. What is modeling?
2. Motivation: Why model at all?
3. The difference between theories and models
4. What can you model?
5. How to evaluate models?
6. Models of language acquisition

Practical: Building a model of Saffran, Aslin, & Newport (1996)

Day 2

1. Models of language acquisition continued
2. Limitations of modeling

Practical: Predicting word acquisition order from input statistics and using model output to constrain theories

3. Bridging the gap between models and infant studies
4. Discussion: Modeling for your own research project?

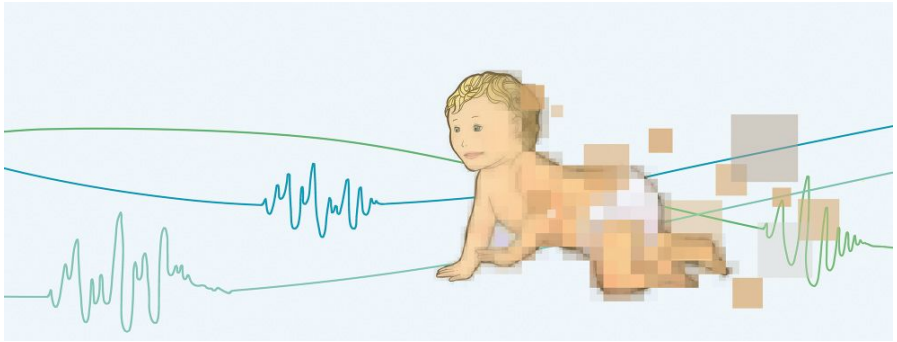
A quick bio or: Why am *I* teaching this?

Education:

- Cognitive Science in Osnabrück, DE
- Cognitive Neuroscience (specialization Psycholinguistics) in Nijmegen, NL
- PhD in Computational models of language acquisition

Then: post docs in Paris (ENS, LSCP, etcetc) and Nijmegen (MPI)

More on my website: https://sites.google.com/site/chbergma/

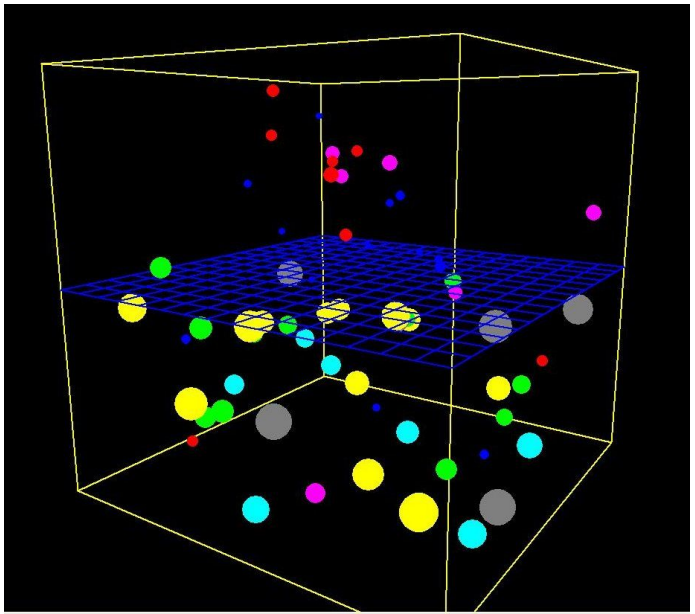


Who is a modeller?

Would you call yourself a modeller?

What is a model?

Any ideas?

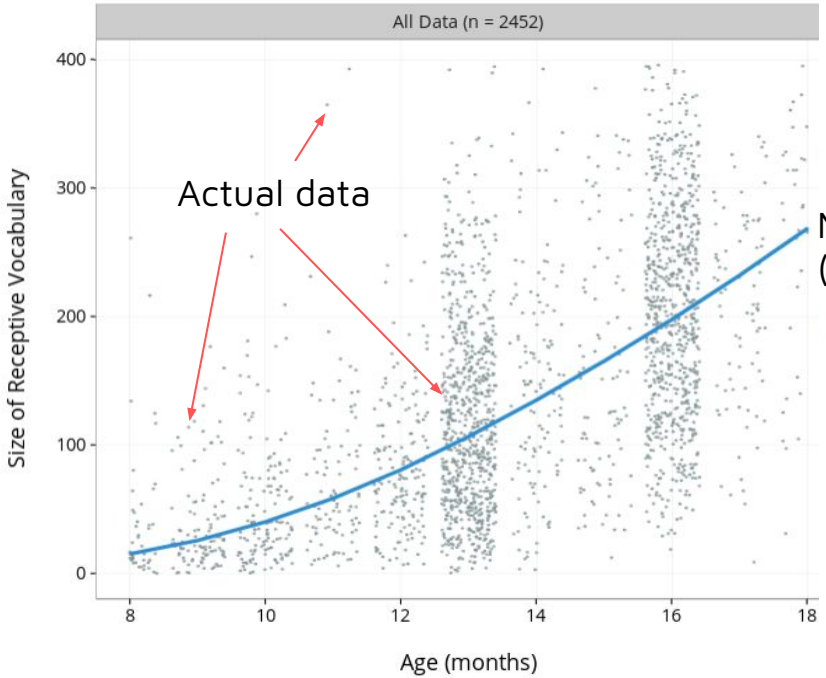


Models: From simple to complex

Fun fact: A regression is a model

Models: From simple to complex

Fun fact: A regression is a model



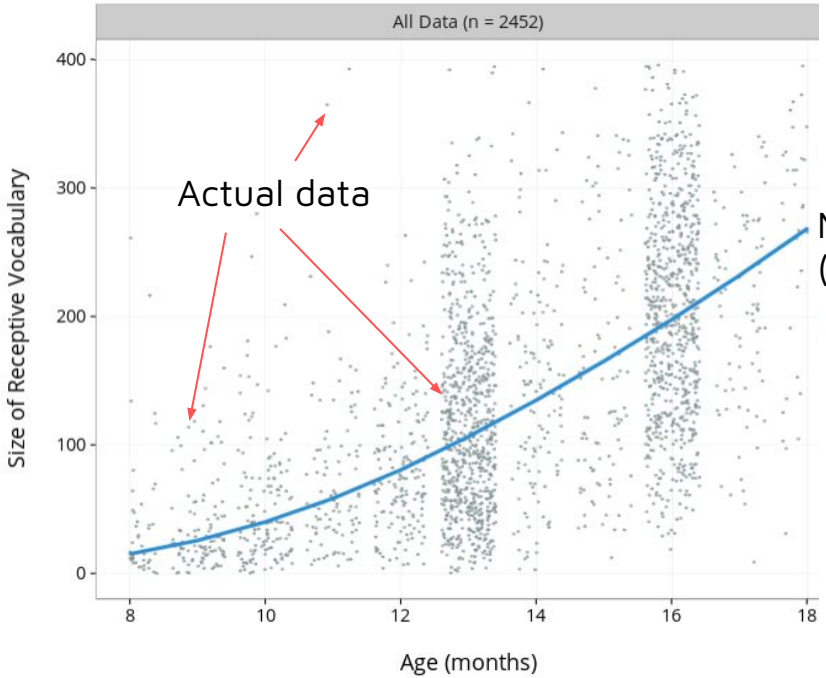
Mean receptive vocabulary size (Modelled!)

Source: Wordbank

Who is a modeller (v2)?

Did you ever compute a regression?

Models: From simple to complex



Mean receptive vocabulary size (Modelled!)

But how do things work "under the hood"?
→ **Cognitive** models

Source: Wordbank

What is computational modelling?

What is computational modelling?

1. Method in language acquisition research (a means)

    Supplement to experiments and theories

What is computational modelling?

1. Method in language acquisition research (a means)

    Supplement to experiments and theories

    → In tandem with both:

    Model-first route

Develop Theory
Babies do X via Y

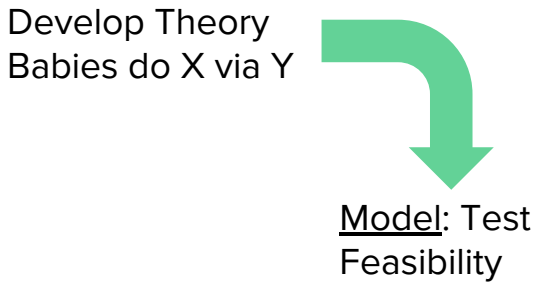
What is computational modelling?

1. Method in language acquisition research (a means)

    Supplement to experiments and theories

    → In tandem with both:

        Model-first route



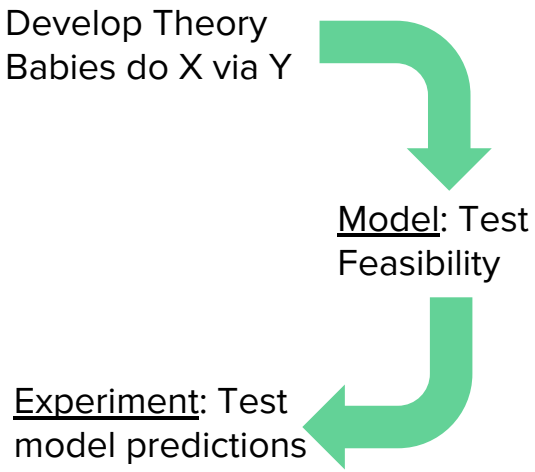
What is computational modelling?

1. Method in language acquisition research (a means)

    Supplement to experiments and theories

    → In tandem with both:

        Model-first route



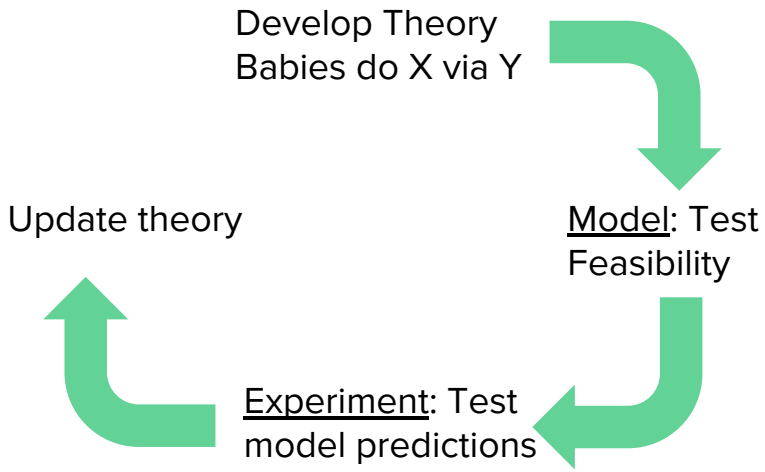
What is computational modelling?

1. Method in language acquisition research (a means)

   Supplement to experiments and theories

   → In tandem with both:

   Model-first route



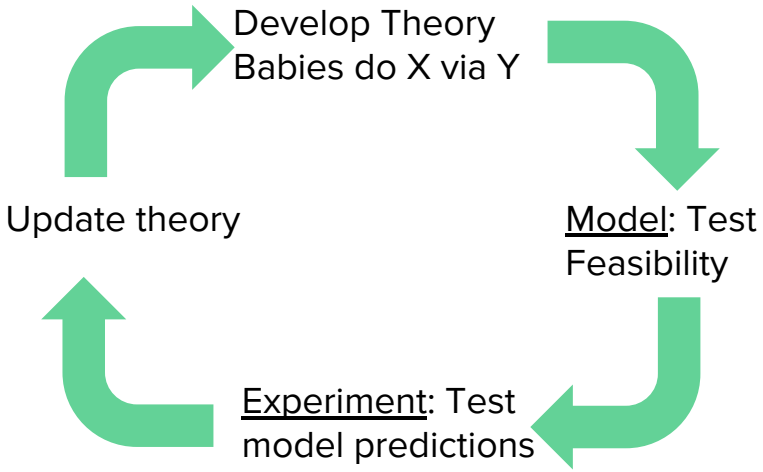
What is computational modelling?

1. Method in language acquisition research (a means)

   Supplement to experiments and theories

   → In tandem with both:

   Model-first route



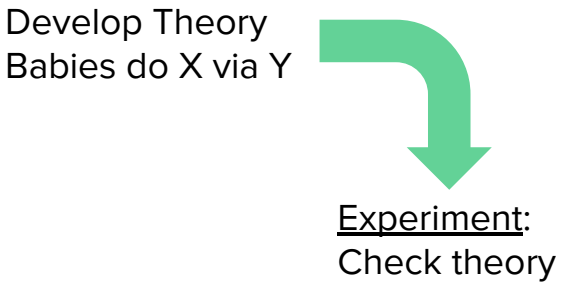
What is computational modelling?

1. Method in language acquisition research (a means)

    Supplement to experiments and theories

    → In tandem with both:

        Experiment-first route



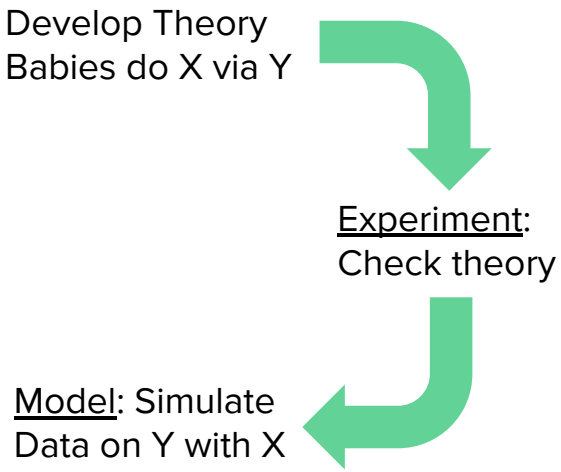
What is computational modelling?

1. Method in language acquisition research (a means)

    Supplement to experiments and theories

    → In tandem with both:

        Experiment-first route



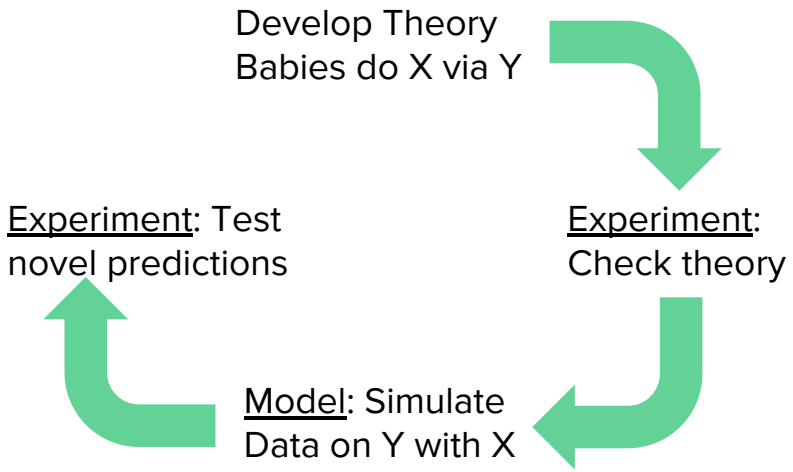
What is computational modelling?

1. Method in language acquisition research (a means)

   Supplement to experiments and theories

   → In tandem with both:

   Experiment-first route



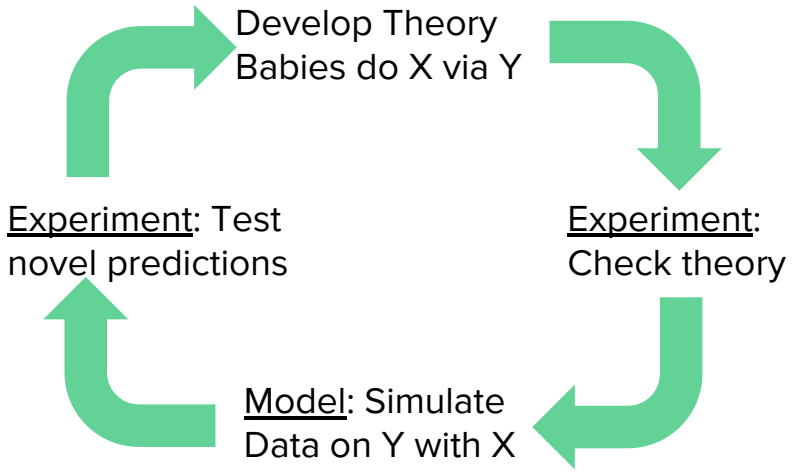
What is computational modelling?

1. Method in language acquisition research (a means)

    Supplement to experiments and theories

    → In tandem with both:

        Experiment-first route



What is computational modelling?

1. Method in language acquisition research (a means)
2. An independent line of research (a theme)
   Goal: Optimize models → Super-human performance (SIRI, Alexa, military…)
   Independent of specific infant data
   Language acquisition: inspiration, promising ways forward

What is computational modelling?

1. Method in language acquisition research (a means)
2. An independent line of research (a theme)
   Goal: Optimize models → Super-human performance (SIRI, Alexa, military…)
   Independent of specific infant data
   Language acquisition: inspiration, promising ways forward

Key questions

What do children learn?

How do they acquire the needed knowledge and abilities?

Which prerequisites do children bring to the task?

→ Use models to answer (parts of) those questions

Key questions

What does the model learn?

How does the model acquire the needed knowledge and abilities?

Which prerequisites does the model bring to the task?

Key questions

What does the model learn?

How does the model acquire the needed knowledge and abilities?

Which prerequisites does the model bring to the task?

→ All three questions *have to* be answered.

Key questions

What does the model learn?

Goal

Key questions

What does the model learn?

Goal

How does the model acquire the needed knowledge and abilities?

Process, Algorithm

Key questions

What does the model learn?

Goal

How does the model acquire the needed knowledge and abilities?

Process, Algorithm

Which prerequisites does the model bring to the task?

Assumptions about what is *innate*

Theories versus models

Abstract

- General statements
- Details vague / implicit
- Informal

Theories versus models

Abstract

- General statements
- Details vague / implicit
- Informal

Concrete

- Formalized, explicit
- Implemented:
  - All aspects have to be specified

Theories versus models

Abstract

- General statements
- Details vague / implicit
- Informal

Concrete

- Formalized, explicit
- Implemented:
  - All aspects have to be specified

Global

- Describes general relationships
- Very powerful

Theories versus models

Abstract
- General statements
- Details vague / implicit
- Informal

Global
- Describes general relationships
- Very powerful

Concrete
- Formalized, explicit
- Implemented:
  - All aspects have to be specified

Specific
- Feasibility:
  - Focus on single aspects, phenomena, mechanisms

Can yield a "proof of concept"

Marr's levels

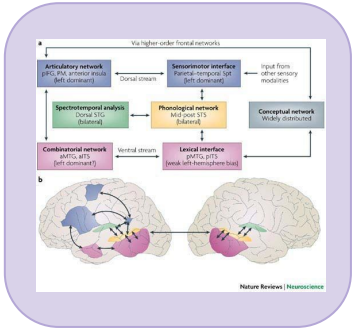
Computational level:

- Goal, structural problem
- What are input and output? Can be problem be divided into parts?

Marr's levels

Algorithmic level:

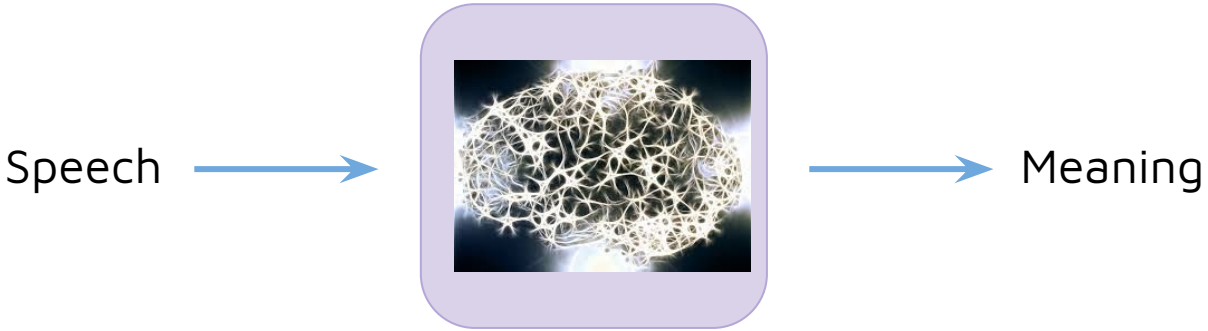
- How can the goal be reached?
- What algorithm(s) can transform the input into the output?

Speech →  → Meaning

Marr's levels

Implementational level:

- How is everything physically realized?
  As program (in R or python) or in a single brain



Marr's levels

Computational level:

- Goal, structural problem
- What are input and output? Can be problem be divided into parts?

Algorithmic level:

- How can the goal be reached?
- What algorithm(s) can transform the input into the output?

Implementational level:

- How is everything physically realized?
  As program (in R or python) or in a single brain

Marr's levels

Computational level:

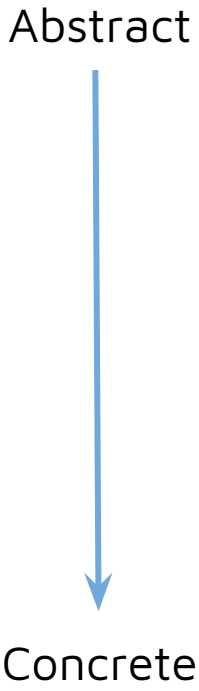
- Goal, structural problem
- What are input and output? Can be problem be divided into parts?

Algorithmic level:

- How can the goal be reached?
- What algorithm(s) can transform the input into the output?

Implementational level:

- How is everything physically realized?
  As program (in R or python) or in a single brain



Marr's levels extended

Note: More intermediate levels possible

What is a ”good” model?

What is a ”good” model?

No general consensus (yet), also because of range of model types / goals, but the following have been mentioned repeatedly

What is a ”good” model?

No general consensus (yet), also because of range of model types / goals, but the following have been mentioned repeatedly

1. Link to experimental data
   It is impossible to evaluate a model without (generally) comparable experimental data

What is a ”good” model?

No general consensus (yet), also because of range of model types / goals, but the following have been mentioned repeatedly

1. Link to experimental data
   It is impossible to evaluate a model without (generally) comparable experimental data

2. Makes plausible assumptions
   Example: Does the model have to process all input of several weeks at once and several times to learn? Can it learn incrementally?
   Note: Not always made explicit, but read closely and you will find out!

What is a ”good” model?

No general consensus (yet), also because of range of model types / goals, but the following have been mentioned repeatedly

1. Link to experimental data
   It is impossible to evaluate a model without (generally) comparable experimental data

2. Makes plausible assumptions
   Example: Does the model have to process all input of several weeks at once and several times to learn? Can it learn incrementally?
   Note: Not always made explicit, but read closely and you will find out!

3. Generates testable predictions (cf. model-experiment-theory cycle)

Creating a model

Creating a model

First, go through the following decision process:

Creating a model

1. What should be modelled? What will be excluded from the model?

   Not all aspects of a child's experience can be included (think audio / visual / tactile / proprioceptive input). Simplifications are necessary.

Creating a model

1. What should be modelled? What will be excluded from the model?
   Not all aspects of a child's experience can be included (think audio / visual / tactile / proprioceptive input). Simplifications are necessary.
2. Which processing abilities will the model have?
   Does the model learn step by step, does it compute statistics, can it forget?

Creating a model

1. What should be modelled? What will be excluded from the model?
   Not all aspects of a child's experience can be included (think audio / visual / tactile / proprioceptive input). Simplifications are necessary.
2. Which processing abilities will the model have?
   Does the model learn step by step, does it compute statistics, can it forget?
3. Which theories and data will the model build on?
   Data and theories can be contradictory, models need to take a stance.

Creating a model

1. What should be modelled? What will be excluded from the model?
   Not all aspects of a child's experience can be included (think audio / visual / tactile / proprioceptive input). Simplifications are necessary.
2. Which processing abilities will the model have?
   Does the model learn step by step, does it compute statistics, can it forget?
3. Which theories and data will the model build on?
   Data and theories can be contradictory, models need to take a stance.
4. Which level of abstraction will be modelled?
   Is this a general proof of concept or does it simulate single infants. Does it simulate minutes (in the lab), days, months... of learning?

Creating a model

1. What should be modelled? What will be excluded from the model?
   Not all aspects of a child's experience can be included (think audio / visual / tactile / proprioceptive input). Simplifications are necessary.
2. Which processing abilities will the model have?
   Does the model learn step by step, does it compute statistics, can it forget?
3. Which theories and data will the model build on?
   Data and theories can be contradictory, models need to take a stance.
4. Which level of abstraction will be modelled?
   Is this a general proof of concept or does it simulate single infants. Does it simulate minutes (in the lab), days, months... of learning?
5. How will the model be evaluated?

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 1: Phoneme acquisition

- Categorical perception from birth on
- Category boundaries move (or disappear) based on native language input
- Theory: Speech statistics move boundaries

Concretely: Kuhl's (1991) *perceptual magnet effect*

Example 1: Perceptual magnet effect

Goal

- Native language sound categorization
- Input: Sounds (Vowels, 1^{st} /2^{nd} formant), Output: Sound category membership

Example 1: Perceptual magnet effect

Goal

- Native language sound categorization
- Input: Sounds (Vowels, 1^{st} /2^{nd} formant), Output: Sound category membership

Processing abilities

- Process: Neuron(al groups) move their perceptual fields
  (Self-Organizing Map; SOM - demo: https://irath96.github.io/som-demo/)
- General statistical processing

Example 1: Perceptual magnet effect

Goal

- Native language sound categorization
- Input: Sounds (Vowels, 1^{st} /2^{nd} formant), Output: Sound category membership

Processing abilities

- Process: Neuron(al groups) move their perceptual fields
  (Self-Organizing Map; SOM - demo: https://irath96.github.io/som-demo/)
- General statistical processing

Level of abstraction

- Neural assemblies (algorithmic-implementational)

Example 1: Perceptual magnet effect

Goal

- Native language sound categorization
- Input: Sounds (Vowels, 1st /2nd formant), Output: Sound category membership

Processing abilities

- Process: Neuron(al groups) move their perceptual fields
  (Self-Organizing Map; SOM - demo: https://irath96.github.io/som-demo/)
- General statistical processing

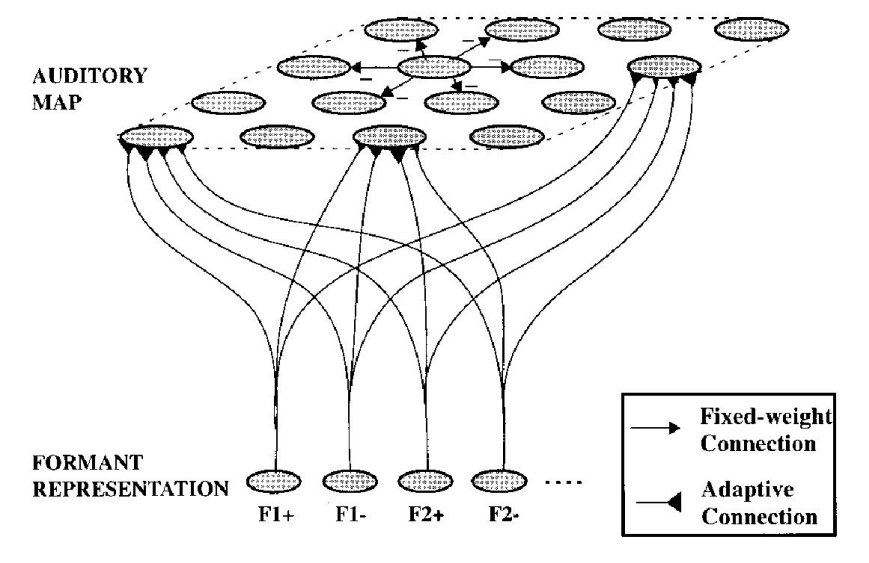
Level of abstraction

- Neural assemblies (algorithmic-implementational)

Evaluation

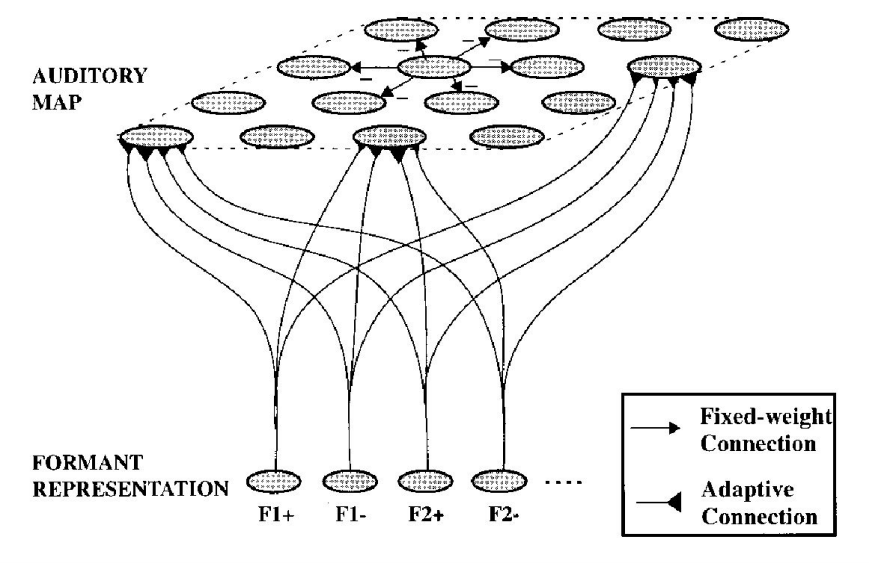
- Correct categorization performance

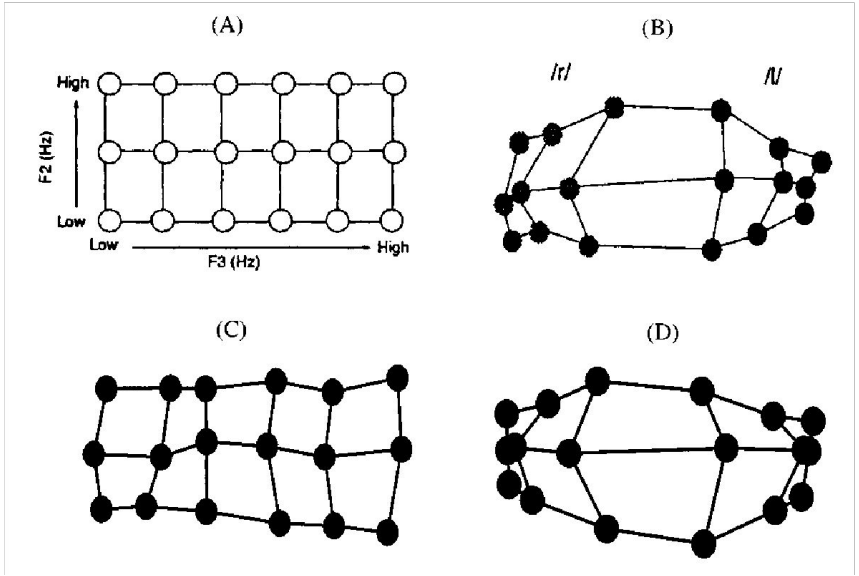
Example 1: Perceptual magnet effect



Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *The Journal of the Acoustical Society of America*, *100*(2), 1111-1121.

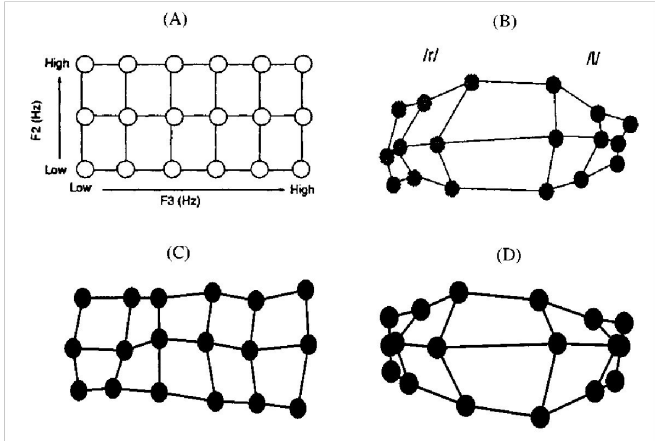
Example 1: Perceptual magnet effect

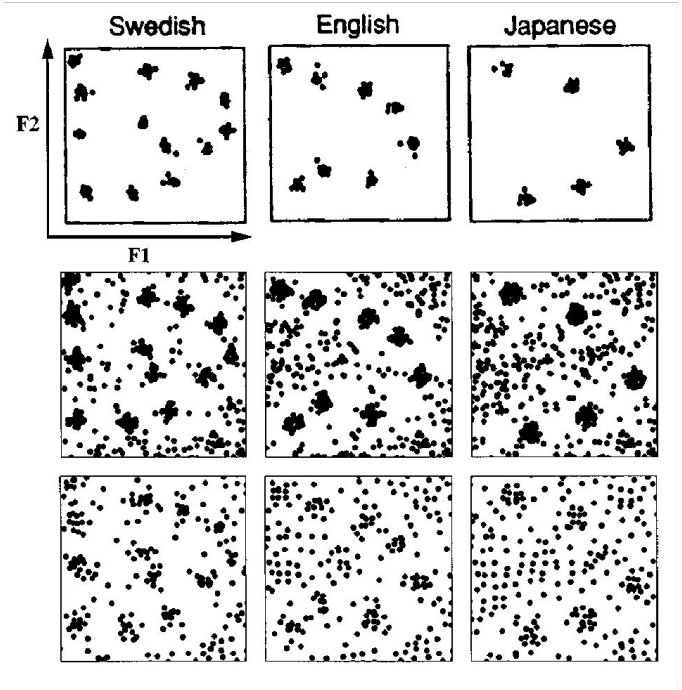




Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *The Journal of the Acoustical Society of America*, *100*(2), 1111-1121.

Example 1: Perceptual magnet effect





Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *The Journal of the Acoustical Society of America*, *100*(2), 1111-1121.

Example 1: Perceptual magnet effect

Strengths

Example 1: Perceptual magnet effect

Strengths

- Simple, intuitive
- Concrete, testable predictions: e.g., effects of sound distance / frequency

Example 1: Perceptual magnet effect

Strengths

- Simple, intuitive
- Concrete, testable predictions: e.g., effects of sound distance / frequency

Weaknesses

Example 1: Perceptual magnet effect

Strengths

- Simple, intuitive
- Concrete, testable predictions: e.g., effects of sound distance / frequency

Weaknesses

- Input highly simplified

  Are formants sufficient to learn vowel categories?

  What about consonants?

  → How do infants know what aspects of the acoustics to pay attention to?

Example 1: Perceptual magnet effect

Strengths

- Simple, intuitive
- Concrete, testable predictions: e.g., effects of sound distance / frequency

Weaknesses

- Input highly simplified

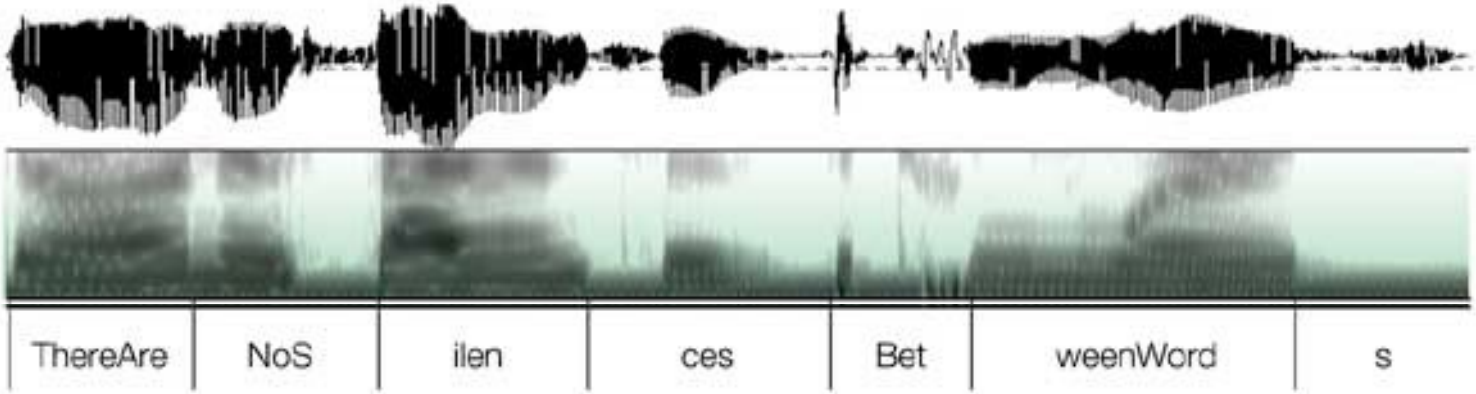
  Are formants sufficient to learn vowel categories?

  What about consonants?

  → How do infants know what aspects of the acoustics to pay attention to?

Which **theories** about speech perception does the model build on?

Example 2: Word segmentation



Example 2: Word segmentation

Transitional probability (TP)

Lookatthepretty baby.

Isthatmybabyboy?

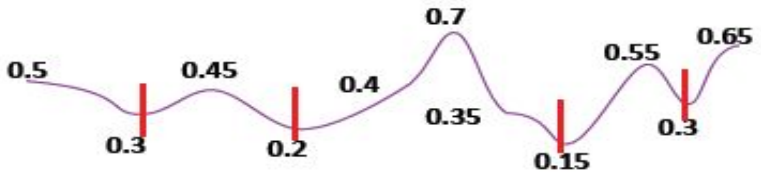
Hearthebabycrying.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Science. 1926-8.

Example 2: Word segmentation

Transitional probability (TP)

$$\textbf{TP (XY)} = P(Y \mid X) = \frac{P(XY)}{P(X)}$$



Lookatthepretttybaby.

Isthatmybabyboy?

Hearthebabycrying.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Science. 1926-8.

Example 2: Word segmentation

bidukapolimatifagerunosapolimarunosabidukatifage

bi-du-ka-po-li-ma-ti-fa-ge-ru-no-sa-po-li-ma-ru-no-sa-bi-du-ka-ti-fa-ge

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Science. 1926-8.

Example 2: Word segmentation

$$\textbf{TP (XY)} = P(Y \mid X) = \frac{P(XY)}{P(X)}$$

Transitional probability (TP)

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Science. 1926-8.

Example 2: Word segmentation

Goal

- Input: Stream of syllables, Output: Probability of a syllable sequence being a "word"

Example 2: Word segmentation

Goal

- Input: Stream of syllables, Output: Probability of a syllable sequence being a "word"

Processing abilities

- Calculate Transitional Probabilities
- General statistical processing properties: computed over pairs of syllables
- Perceive syllables

Example 2: Word segmentation

Goal

- Input: Stream of syllables, Output: Probability of a syllable sequence being a "word"

Processing abilities

- Calculate Transitional Probabilities
- General statistical processing properties: computed over pairs of syllables
- Perceive syllables

Level of abstraction

- No direct link to biology (algorithmic)

Example 2: Word segmentation

Goal

- Input: Stream of syllables, Output: Probability of a syllable sequence being a "word"

Processing abilities

- Calculate Transitional Probabilities
- General statistical processing properties: computed over pairs of syllables
- Perceive syllables

Level of abstraction

- No direct link to biology (algorithmic)

Evaluation

- Probabilities of syllable sequences

Example 2: Word segmentation

Try simplest implementation: Implement basic TP formula

$$\textbf{TP (XY)} = P(Y \mid X) = \frac{P(XY)}{P(X)}$$

Time to code!

Example 2: Word segmentation

Try simplest implementation: Implement basic TP formula

$$\mathbf{TP\ (XY)} = P(Y \mid X) = \frac{P(XY)}{P(X)}$$

Results for words (by design) 1.0

Results for non-words (by design) 0.0

Results for part words around 0.33

Example 2: Word segmentation

Strengths

- Simple, intuitive
- Very powerful process, also works for Italian (e.g., Pelucchi, Hay, & Saffran, 2009)

- Can be applied to natural language corpora to find "probable word chunks"

  As predicted by the model, children react to highly probable syllable streams similarly independent of whether they correspond to words or not! (Ngon et al., 2013)

Example 2: Word segmentation

Weaknesses

- Does not explain all experimental data
  - Children cannot learn words when word lengths are mixed (2- and 3-syllable "words") → The model is too "smart"
  - Cannot account for word stress and prosody, which are also important cues to word boundaries
  - Only looks at *pairs* of syllables. What about longer ranging dependencies? Think about rule learning, is it a different process?
- Probabilities were calculated based on the WHOLE experimental input (2 minutes)
  - Do children have such efficient memory capacities?
  - Not a learning mechanism in the narrow sense

Example 2: Word segmentation

Alternative models: PUDDLE + Chunk-based learner

→ Scale to corpora of infant-directed speech

Example 2b: Word segmentation (PUDDLE)

Goal

- Input: Stream of sounds, Output: "Chunks" (i.e. words)

Example 2b: Word segmentation (PUDDLE)

Goal

- Input: Stream of sounds, Output: "Chunks" (i.e. words)

Processing abilities

- Compute frequencies, store sequences at utterance boundaries
- Perceive phonemes (not phones, acoustics), utterance boundaries

Example 2b: Word segmentation (PUDDLE)

Goal

- Input: Stream of sounds, Output: "Chunks" (i.e. words)

Processing abilities

- Compute frequencies, store sequences at utterance boundaries
- Perceive phonemes (not phones, acoustics), utterance boundaries

Level of abstraction

- No direct link to biology (algorithmic)

Example 2b: Word segmentation (PUDDLE)

Goal

- Input: Stream of sounds, Output: "Chunks" (i.e. words)

Processing abilities

- Compute frequencies, store sequences at utterance boundaries
- Perceive phonemes (not phones, acoustics), utterance boundaries

Level of abstraction

- No direct link to biology (algorithmic)

Evaluation

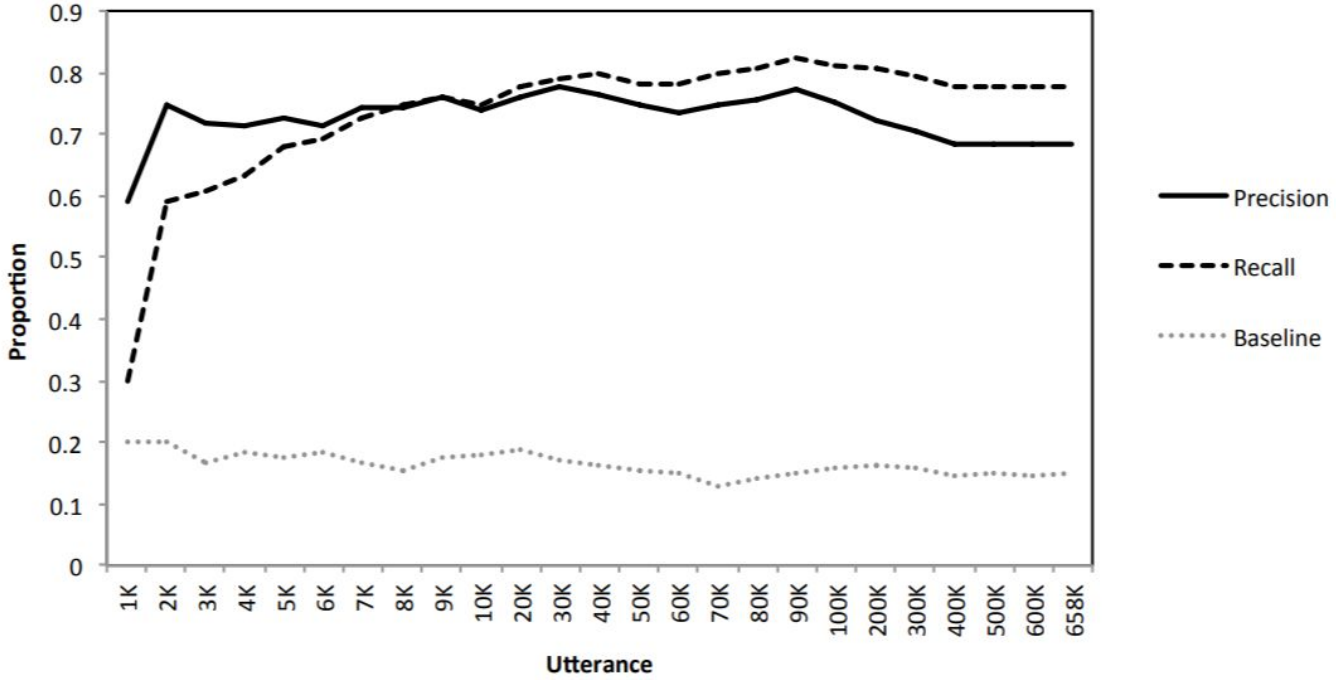
- Real word boundaries found

Examp E)

Stage	INPUT BUFFER	CHUNK INVENTORY
1	thedogbelongstopeter	the 3 peter 2
2	thedogbelongstopeter	the 3 peter 2
2	the dogbelongstopeter	**the** 4 peter 2
3	the dogbelongstopeter	the 4 peter 2
4	the dogbelongstopeter	the 4 peter 2
...	...	...
14	the dogbelongstopeter	the 4 peter 2
14	the dogbelongsto peter	the 4 **peter** 3
14	the dogbelongsto peter	the 4 peter 3 dogbelongsto 1

Example 2b: Word segmentation (PUDDLE)

Results



Example 2b: Word segmentation (PUDDLE)

Results

The twenty most highly activated items in the PUDDLE chunk inventory after 1000 and 10,000 utterances during the course of the dense corpus simulation

1000 Utterances	10000 Utterances
no	*the*
oh	*a*
this	*no*
dear	*oh*
and	*and*
hat	*you*
thomas	*this*
it	*that*
what	*ee*
grapes	*it*
ahh	*to*
there	*are*
oops	*on*
where	*dear*
two	*what*
ooh	*thomas*
look	*two*
a_hat	*is*
blue	*what's*
what's_this	*there*

Example 2b: Word segmentation (PUDDLE)

Strengths

- Stepwise processing
- Realistic memory assumptions
- Consideration of possible words / phonotactics

Weaknesses

- Lots of processing needed (comparing input with all stored items)
- What about overlapping words?
  - hand vs handy

Summary

Models are a useful tool in language (acquisition) research

Important to pay attention to implicit assumptions!

Plan for tomorrow

1. Let's look at word learning
2. Limitations
3. Bridging the gap between models and infant studies