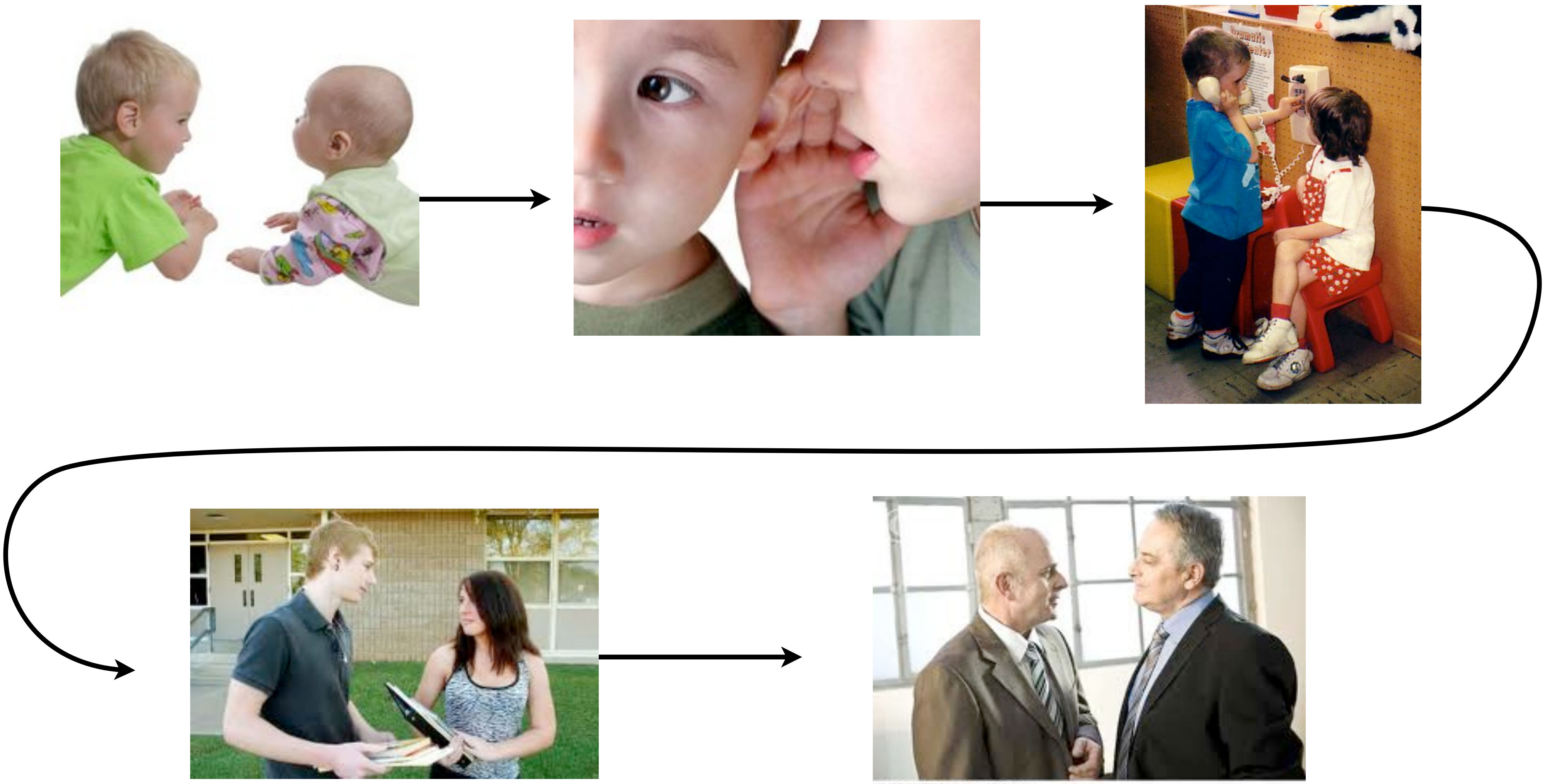
Automatically Learning Measures of Child Language Development

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Sam Sahakian and Benjamin Snyder

How does this happen?



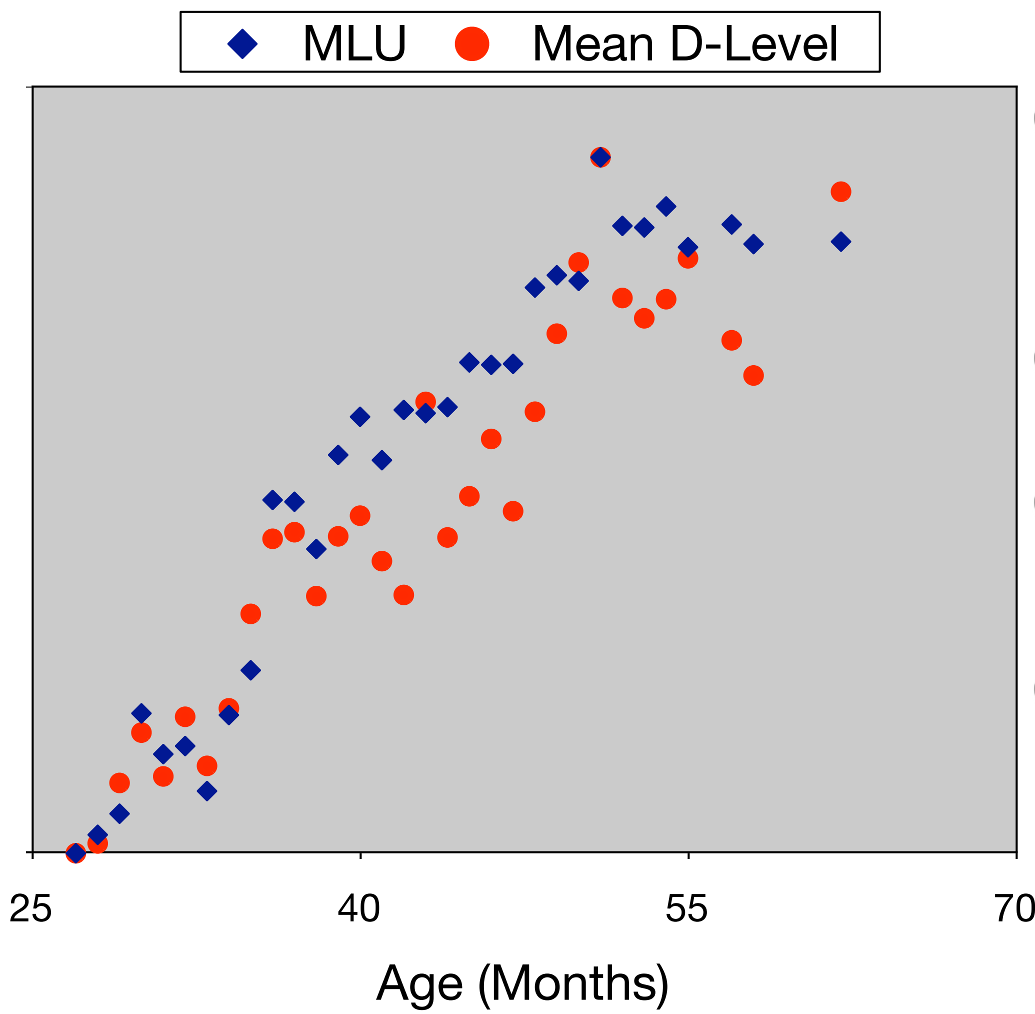
When does this happen?

6 months

2 years

4 years

Language Development Metrics



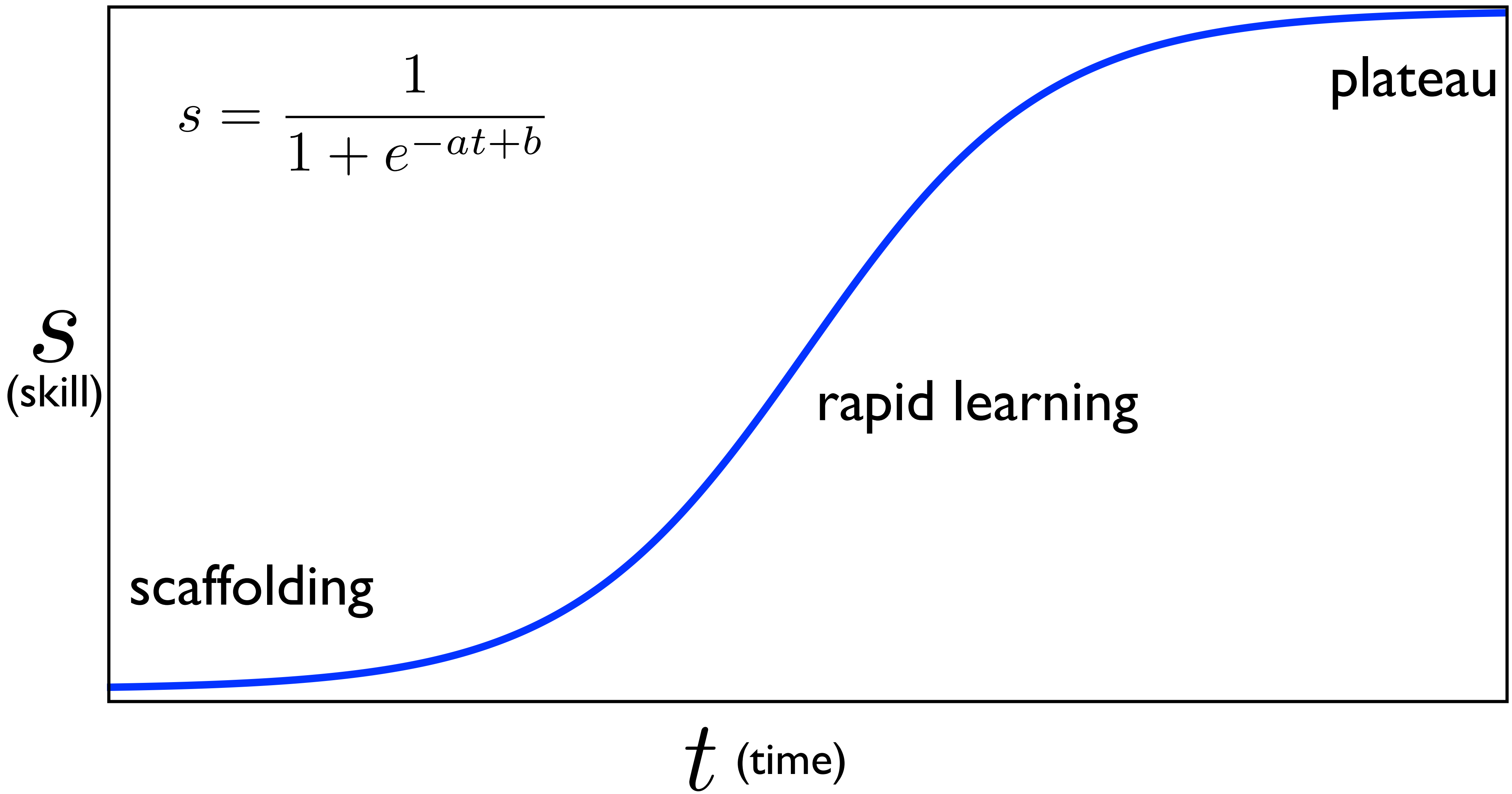
- MLU (Mean Length of Utterance) [Brown '73]
- Parse depth [Yngve, '60]
- D-Level [Rosenberg et al., '87; Covington et al., '06] [Lu, '09]
- IPSYN [Scarborough, 1990] [Sagae, '05]

Language Development Metrics

- Drawbacks of previous metrics:
  - Coarse and ad-hoc
  - Questionable validity
  - Accuracy degrades with age

<u>Question 1</u>: Can we induce a more accurate metric using statistical learning methods?

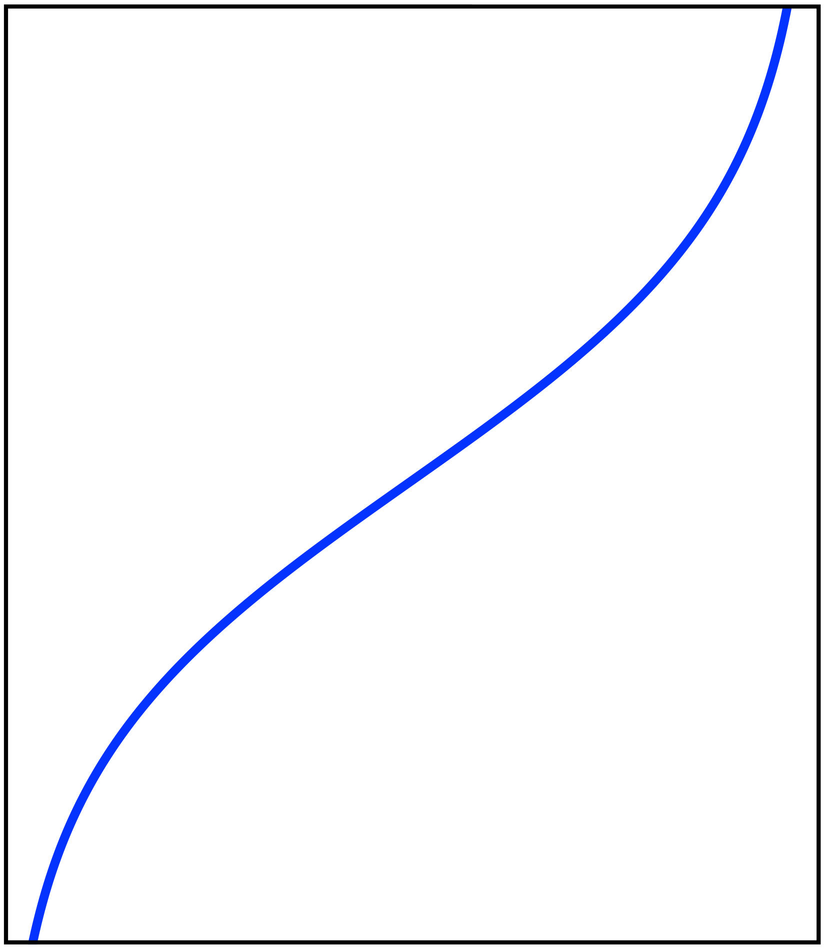
Skill as function of time



- Skill acquisition follows sigmoidal curve [Hodgetts '91]

Time as Ground Truth

t
(time)



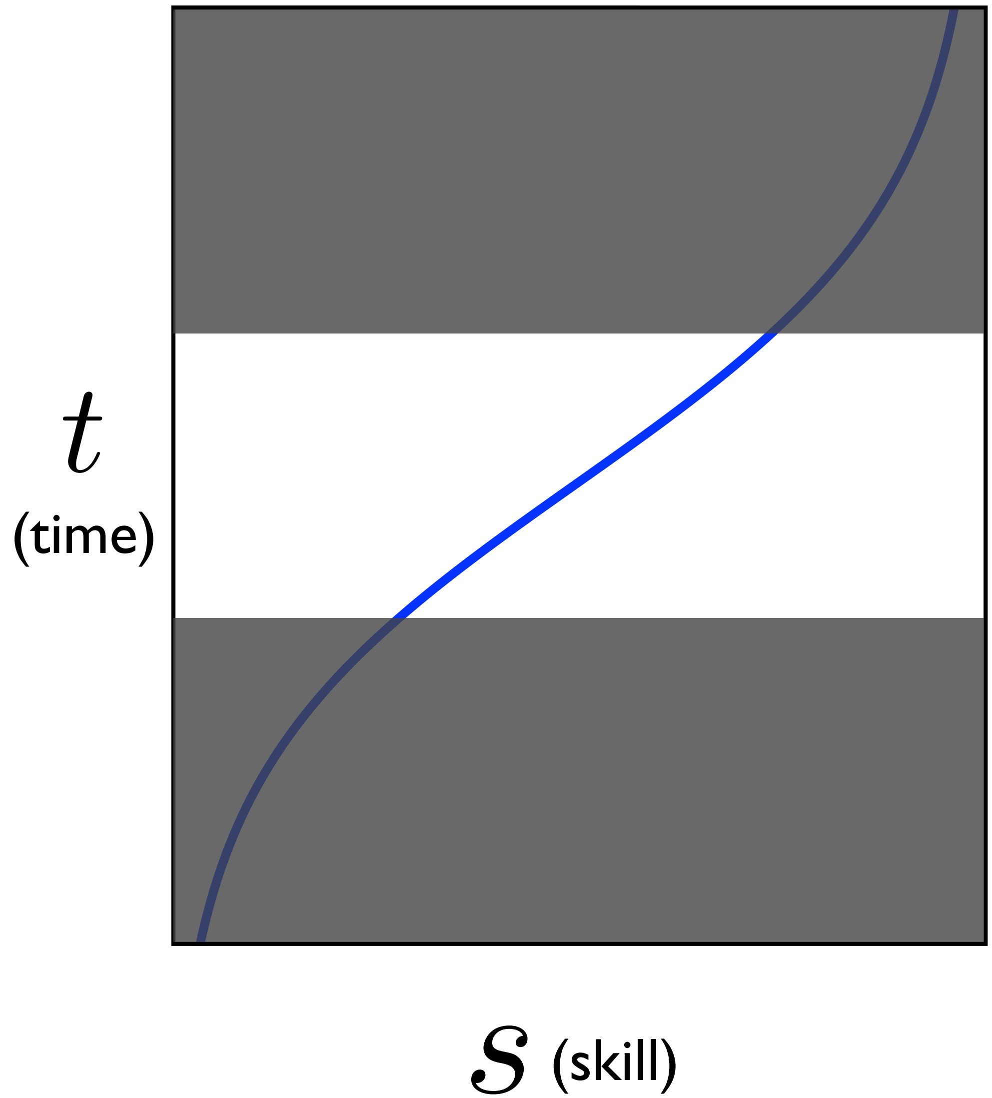
$$s \approx \beta \cdot \mathbf{x}$$

skill parameters features

$$t = \frac{b - \ln\left(\frac{1}{s} - 1\right)}{a}$$

- Invert sigmoid
- Skill as combination of features
- Evaluate learned metric via age prediction error

Age Prediction Model



$$t = a(\beta \cdot \mathbf{x}) + b$$

- Age window at linear part of sigmoid
- Predict age as linear function of skill

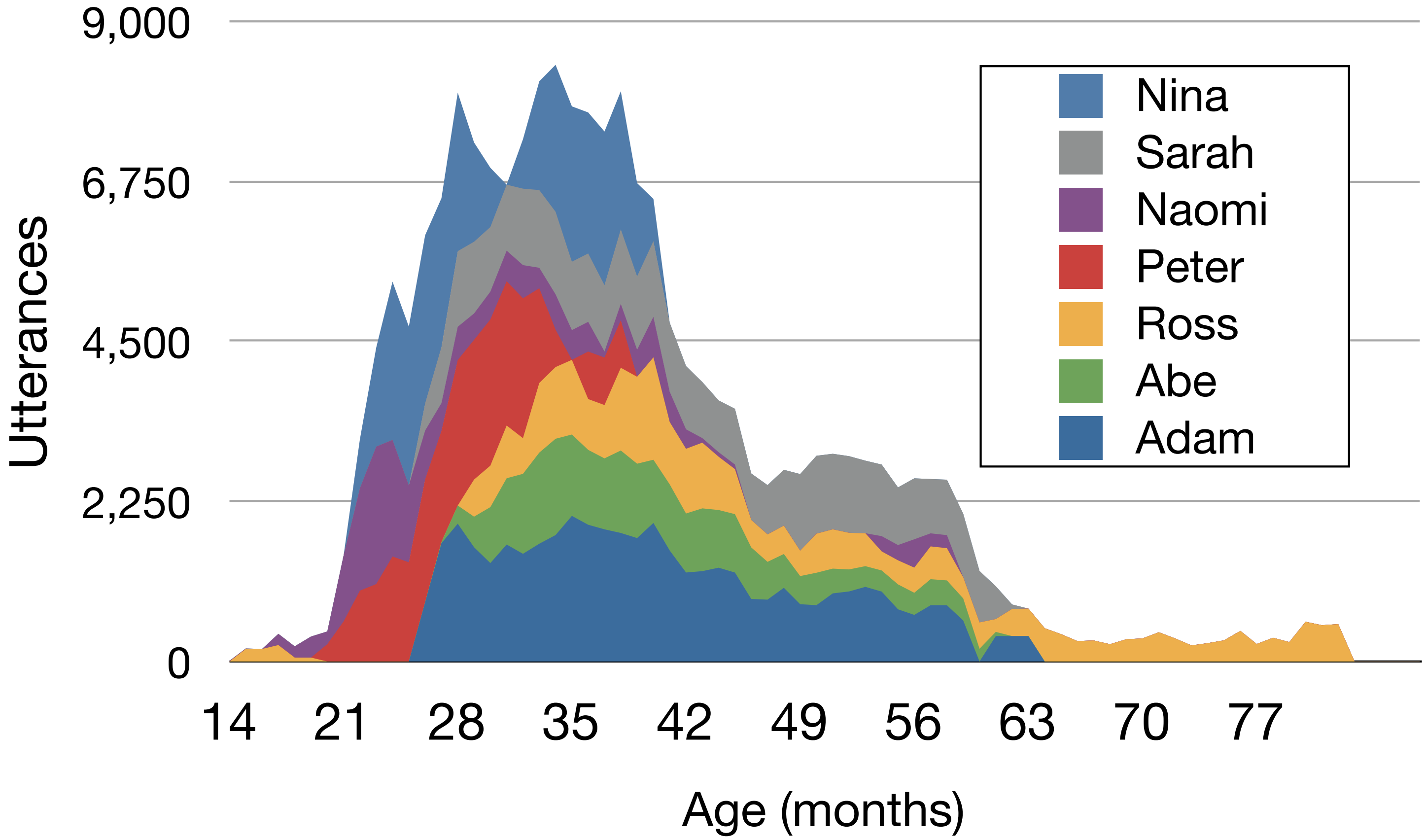
Features

- Pre-defined metrics:
  - MLU
  - Parse Depth
  - D-Level
- Novel features
  - Preposition counts
  - "Be" verb counts
  - Article counts
  - Word frequency
  - Function to content word ratio

Data

- Child speech from transcribed conversations in CHILDES database [MacWhinney, '00]

- Longitudinal studies of 7 children



- Learn via linear regression -- Separately for each child.

Results
(lower is better)

	D-Level	Depth	MLU	All Features
Mean	63.795	66.327	64.578	**54.041**

Mean squared error of age prediction in months

Results
(lower is better)

	D-Level	Depth	MLU	All Features
Adam	14.037	14.149	**11.128**	14.175
Abe	34.69	44.701	**34.509**	39.931
Ross	329.64	336.612	345.046	**244.071**
Peter	23.58	13.045	**8.245**	24.128
Naomi	**24.458**	28.426	34.956	45.036
Sarah	12.503	20.878	13.905	**6.989**
Nina	7.654	6.477	4.255	**3.96**
Mean	63.795	66.327	64.578	**54.041**

Mean squared error of age prediction in months

Results
(lower is better)

	D-Level	Depth	MLU	All Features
Adam	14.037	14.149	**11.128**	14.175
Abe	34.69	44.701	**34.509**	39.931
Ross	329.64	336.612	345.046	**244.071**
Peter	23.58	13.045	**8.245**	24.128
Naomi	**24.458**	28.426	34.956	45.036
Sarah	12.503	20.878	13.905	**6.989**
Nina	7.654	6.477	4.255	**3.96**
Mean	63.795	66.327	64.578	**54.041**

Mean squared error of age prediction in months

Question 2: Can we learn a metric that generalizes across children?

- Task: Train on a set of children, evaluate on a held-out child.
- Children learn at different rates, so must predict relative mastery, not absolute age.

Ordering Model

$$y = \sum_i \beta_i \mathbf{x}_i + \sum_{i,j} \gamma_{ij} \mathbf{x}_i \mathbf{x}_j$$

Score used for ranking

Sum over features

Sum over feature pairs

- Each iteration trains on 6 children, tests on held-out child
- Score each sample as weighted combination of features and feature pairs
- Rank speech samples in order of ascending score

Evaluation: Kendall's τ

$$\tau = \frac{(\text{num. concordant pairs}) - (\text{num. discordant pairs})}{\frac{1}{2}n(n-1)}$$

- Kendall's rank correlation coefficient
  - ~ Measures similarity between 2 orderings over a set
  - ~ Identical orderings yield +1, independent orderings yield 0

Parameter Estimation

$$(\beta^*, \gamma^*) = \underset{\beta,\gamma}{\operatorname{argmax}} \sum_{k \in kids} \tau(k, \beta, \gamma)$$

- $\tau(k, \beta, \gamma) \equiv$ Kendall τ between model ordering and true chronological order for child k.
- Find best parameters via Nelder-Mead [Nelder and Mead, '65]
  - ~ Gradient-free hill climbing search that shifts parameter values until reaching a local optimum.

Results

(higher is better)

MLU	All Features	MLU & Fn. / Content
0.7456	0.7457	**0.7780**

Average Kendall τ of model orderings versus true chronological orderings.

Contributions

- New method of inducing language development metrics
- Methodology for validating these metrics
- Increased performance over hand-crafted baseline metrics