Predicting the number of words a child will learn: Modeling ‘burstiness’

Nicole Beckage

data from Eliana Colunga
Word learning: an overview

Input from environment and language

Interests of child

Learning

Knowledge

Observed: Productive vocabulary
Vocabulary data

MacArthur Bates Communicative Development Inventory (CDI, MCDI)

• Closed-form checklist of nearly 700 words.

• Parents report words their child can produce.

• Normative age of acquisition information.
Longitudinal data

Colunga Lab collected longitudinal data, following specific children for about a year.

- 114 children with longitudinal CDIs (L-CDIs)
- On average 9 CDIs per child
- 905 *vocabulary learning snapshots*

<table>
<thead>
<tr>
<th></th>
<th>age</th>
<th>sex</th>
<th>...</th>
<th>voc. sz</th>
<th>dog</th>
<th>house</th>
<th>...</th>
<th>zoo</th>
</tr>
</thead>
<tbody>
<tr>
<td>kid A</td>
<td>16.2</td>
<td>F</td>
<td>...</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>17.1</td>
<td>F</td>
<td>...</td>
<td>49</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>18.9</td>
<td>F</td>
<td>...</td>
<td>132</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>1</td>
</tr>
<tr>
<td>kid B</td>
<td>19.3</td>
<td>M</td>
<td>...</td>
<td>257</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>20.5</td>
<td>M</td>
<td>...</td>
<td>345</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td>0</td>
</tr>
</tbody>
</table>
Growth processes from past work

Models and image adapted from Hills et al. 2009.
Model A: Preferential attachment

- Attractor model on subset of known words
- Known words with a high number of connections are attractors
- Unknown words learned if they attach to highly connected words
- Subset of known words plays largest role in acquisition predictions
Model A: Preferential attachment

Word specific growth parameter

• Mean in-degree of attachment nodes (Hills 2009)
Model B: Preferential acquisition

- Attractor model on subset of unknown words
- **Unknown words** with a high number of connections are attractors
- Unknown words learned if they are highly **connected** in language learning environment
- **Independent** of words child knows
Model B: **Preferential acquisition**

Word specific growth parameter

- **Sum of in-degree in full network**
Model C: **Lure of the associates**

- Intersection of previous models
- Attractor model on *unknown* words counting edges coming from subset of *known words* only
- Combines the role of the language environment and the child’s learning process
Model C: Lure of the associates

Word specific growth parameter

- Sum of in-degree if word in question was added to the known graph
Growth values

Summary statistic of **growth** under each model

• For all models, function of **in-degree**
  – **Preferential attachment**: average/sum of in-degree of the attachment node
  – **Preferential acquisition**: in-degree in full network
  – **Lure of the associates**: in-degree of vocabulary subgraph if word i was added
Growth processes from past work

Models and image adapted from Hills et al. 2009.
Network burstiness

These models have standardly been used to predict what words a child will learn next.

Does it give us information beyond what words are likely to be learned?

Can the distribution of probabilities from the model approximate learning ability?
Network burstiness

[Histograms showing density vs. observed delta value for Preferential attachment and Preferential acquisition]
Predicting future language ability

If we can **predict number of words** to be learned we can predict future language ability. This is important because some children show **early language delay**.

– most of these children grow out of it
– but some don’t
Predicting # of words learned

Use regularized Poisson regression (lasso)
Train regularization parameter using cross validation
Test on withheld 20% of data

histogram of number of words learned
Predicting # of words learned

Input features

- child specific (age, voc. size) features
- summary of distribution of growth values
  - min, 1st quartile, median, mean, 3rd quartile and max
  - as counts
  - as probabilities

Predict number of words learned

Child features only is the baseline model
# Results

**Average absolute error** contribution per subject for train and test

<table>
<thead>
<tr>
<th></th>
<th>Kid features</th>
<th>Pref Att</th>
<th>Pref Acq</th>
<th>Lure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kid features</td>
<td>15.47</td>
<td>11.85</td>
<td>13.50</td>
<td>12.26</td>
</tr>
<tr>
<td><strong>Test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kid features</td>
<td>12.08</td>
<td>9.95</td>
<td>10.86</td>
<td>10.11</td>
</tr>
</tbody>
</table>

**Average Squared error**

<table>
<thead>
<tr>
<th></th>
<th>Kid features</th>
<th>Pref Att</th>
<th>Pref Acq</th>
<th>Lure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kid features</td>
<td>426.95</td>
<td>288.09</td>
<td><strong>347.41</strong></td>
<td>305.90</td>
</tr>
<tr>
<td><strong>Test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kid features</td>
<td>245.28</td>
<td>167.95</td>
<td><strong>201.85</strong></td>
<td>176.94</td>
</tr>
</tbody>
</table>
Predictions from child feature model

Number learned vs predicted: Child features

![Graph showing number of words predicted vs index.](image-url)
Predictions from preferential attachment

Number learned vs predicted: Preferential attachment
Predictions from preferential acquisition

Number learned vs predicted: Preferential acquisition

![Graph showing number learned vs predicted]
Predictions

Number learned vs predicted: Child features

# words predicted vs index
Conclusions

All models outperform the model with only the child features \((p < 0.003)\)

The best model at predicting the number of words learned is preferential attachment

– no significant difference from Lure of the associates
– significant improvement over Preferential acquisition

We can do better than child features at predicting number of words to be learned

And our improvement varies with model choice
Future directions

We can improve the current model.

– Try more **predictive baselines**
  • additional child features, norm prediction of # of words learned

– Try additional **growth values**
  • maybe something other than in-degree would work

Can we use this to understand the presence (or absence) of the **vocabulary spur**t in children?

What other **developmental language** based phenomena can **network science** help with?
Thanks!

email: nicole.beckage@colorado.edu

Special thanks to Eliana Colunga, the DACS lab and Boulder parents for the data!
Details of the model

number of words known, age at measure and prediction were always included as features.
  – though the coefficient of known words is small
  – and age measure, age at prediction are nearly the same (with opposite signs)

1\textsuperscript{st} Quartile, mean and max were in most models

Regularization penalty near zero for all models (less than 0.1)

It doesn’t matter (much) if you represent delta values as probabilities or include both probabilities and the raw values.