

TRISTAN MAHR

MY DISSERTATION

Contents

<i>Development of word recognition in preschoolers</i>	1
--	---

<i>Updates</i>	1
----------------	---

<i>1 Scratch paper</i>	3
------------------------	---

<i>1.1 Bookdown cheatsheet</i>	3
--------------------------------	---

<i>1.2 Debug info</i>	5
-----------------------	---

<i>Prospectus</i>	11
-------------------	----

<i>2 Front Matter</i>	11
-----------------------	----

<i>3 Specific Aims</i>	13
------------------------	----

<i>3.1 Specific Aim 1 (Familiar Word Recognition and Lexical Competition)</i>	14
---	----

<i>3.2 Specific Aim 2 (Referent Selection and Mispronunciations)</i>	14
--	----

<i>3.3 Specific Aim 3 (Computational Modeling)</i>	15
--	----

<i>3.4 Summary</i>	15
--------------------	----

<i>4 Significance</i>	17
-----------------------	----

<i>4.1 Public Health Significance</i>	17
---------------------------------------	----

<i>4.2 Scientific Significance</i>	17
------------------------------------	----

5	<i>Research Hypotheses</i>	25
5.1	<i>Specific Aim 1 (Familiar Word Recognition and Lexical Competition)</i>	25
5.2	<i>Specific Aim 2 (Referent Selection and Mispronunciations)</i>	26
5.3	<i>Specific Aim 3 (Computational Modeling)</i>	26
6	<i>Methods</i>	29
6.1	<i>General Research Design: Participants</i>	29
6.2	<i>General Eyetracking Procedure</i>	30
6.3	<i>Specific Procedure: Aim 1 (Familiar Word Recognition and Lexical Competition)</i>	32
6.4	<i>Specific Procedure: Aim 2 (Referent Selection and Mispronunciations)</i>	33
7	<i>Outcome Measures</i>	35
7.1	<i>Sample Data</i>	36
8	<i>Analysis</i>	39
8.1	<i>Growth Curve Analysis</i>	39
8.2	<i>Aim 1 (Familiar Word Recognition and Lexical Competition)</i>	41
8.3	<i>Aim 2 (Referent Selection and Mispronunciations)</i>	42
8.4	<i>Aim 3 (Computational Modeling)</i>	45
A	<i>Items used in the mispronunciation experiment</i>	49
B	<i>Items used in the visual world experiment</i>	51
C	<i>Related Work</i>	53
D	<i>Bibliography</i>	55

Development of word recognition in preschoolers

This book, when finished, will contain my dissertation research.

Updates

- 2017-08-18: I migrated what I wrote for my dissertation proposal. In doing so, I worked all the kinks required to generate a nice web version¹, a lovely pdf version², and a... uh... *functional* Word document version³.
- Last compiled: 2017-08-21 10:40:03

¹ <https://tjmahr.github.io/dissertation/>

² <https://tjmahr.github.io/dissertation/dissertation.pdf>

³ <https://tjmahr.github.io/dissertation/dissertation.docx>

1

Scratch paper

This book is made with bookdown, an R package/tool-chain for creating a books in multiple formats. This chapter is just a placeholder section and some scratch-paper so that I have some examples on-hand of how to use bookdown's syntax and features.

This is *a book* written in **Markdown**. You can use anything that Pandoc's Markdown supports, e.g., a math equation $a^2 + b^2 = c^2$.

For now, you have to install the development versions of **bookdown** from Github:

```
devtools::install_github("rstudio/bookdown")
```

Code settings:

```
knitr::opts_chunk$set(  
  tidy = FALSE,  
  collapse = TRUE,  
  comment = "#>",  
  out.width = 80  
)
```

```
options(width = 80)
```

1.1 Bookdown cheatsheet

1.1.1 Cross-references to sections

The headings above were written with the following markdown:

```
## Bookdown cheatsheet
```

```
### Cross-references to sections {#manual-section-label-demo}
```

The first one gets an implicit label. The second one has an explicit section label.

I can refer to Section `\@ref(bookdown-cheatsheet)` and `[link to it](#bookdown-cheatsheet)` with its implicit label.

I can refer to Section `\@ref(manual-section-label-demo)` and `[link to it](#manual-section-label-demo)` with its explicit label.

I can refer to Section 1.1 and link to it with its implicit label.

I can refer to Section 1.1.1 and link to it with its explicit label.

1.1.2 Cross-references to appendices

The sample principles apply to appendices.

This is a reference to `[an appendix](#mp-experiment-items)`

The number of that appendix `\@ref(mp-experiment-items)`. I hope.

Both: See `[Appendix \@ref(mp-experiment-items)](#mp-experiment-items)`

This is a reference to an appendix

The number of that appendix A. I hope.

Both: See Appendix A

1.1.3 Cross-references to tables

The chunk label ``table-single`` provides an implicit label for Table `\@ref(tab:table-single)`.

```
``{r table-single, echo = FALSE}
knitr::kable(
  head(mtcars[, 1:5], 5), booktabs = TRUE,
  caption = 'A table of the first 5 rows of the mtcars data.'
)
...

```

The chunk label `table-single` provides an implicit label for Table 1.1.

1.1.4 Figure references and using text references as captions

The caption for Figure `\@ref(fig:cat)` is defined as a `_text` reference below and passed to the ``fig.cap`` chunk option.

`(ref:cat-cap)` This is a happy cat.

	mpg	cyl	disp	hp	drat
Mazda RX4	21.0	6	160	110	3.90
Mazda RX4 Wag	21.0	6	160	110	3.90
Datsun 710	22.8	4	108	93	3.85
Hornet 4 Drive	21.4	6	258	110	3.08
Hornet Sportabout	18.7	8	360	175	3.15

Table 1.1: A table of the first 5 rows of the mtcars data.

```
```{r cat, fig.cap = "(ref:cat-cap)", out.width = "30%", fig.show = "hold"}
knitr::include_graphics(
 rep("./misc/happy-cat-grooming-itself-vector-file.png", 2)
)
```
```

The caption for Figure 1.1 is defined as a *text reference* below and passed to the `fig.cap` chunk option.

```
knitr::include_graphics(
  rep("./misc/happy-cat-grooming-itself-vector-file.png", 2)
)
```

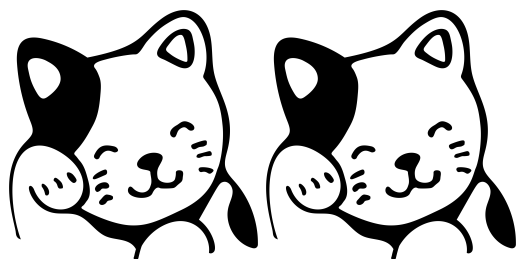


Figure 1.1: This is a happy cat.

1.2 Debug info

```
str(list(html = is_html_output(), latex = is_latex_output(),
        word = is_word_output(), width = options("width")[[1]]))
#> List of 4
#> $ html : logi FALSE
#> $ latex: logi TRUE
#> $ word : logi FALSE
#> $ width: int 80
```

```
devtools::session_info()
#> Session info -----
#> setting value
```

```

#> version R version 3.4.0 (2017-04-21)
#> system x86_64, mingw32
#> ui RTerm
#> language (EN)
#> collate English_United_States.1252
#> tz America/Chicago
#> date 2017-08-21
#> Packages -----
#> package * version date source
#> backports 1.1.0 2017-05-22 CRAN (R 3.4.0)
#> base * 3.4.0 2017-04-21 local
#> bookdown 0.5 2017-08-20 CRAN (R 3.4.1)
#> compiler 3.4.0 2017-04-21 local
#> datasets * 3.4.0 2017-04-21 local
#> devtools 1.13.3 2017-08-02 CRAN (R 3.4.0)
#> digest 0.6.12 2017-01-27 CRAN (R 3.3.2)
#> evaluate 0.10.1 2017-06-24 CRAN (R 3.4.1)
#> formatR 1.5 2017-04-25 CRAN (R 3.4.0)
#> git2r 0.19.0 2017-07-19 CRAN (R 3.4.1)
#> graphics * 3.4.0 2017-04-21 local
#> grDevices * 3.4.0 2017-04-21 local
#> highr 0.6 2016-05-09 CRAN (R 3.2.3)
#> htmltools 0.3.6 2017-04-28 CRAN (R 3.4.0)
#> knitr 1.17 2017-08-10 CRAN (R 3.4.0)
#> magrittr 1.5 2014-11-22 CRAN (R 3.1.2)
#> memoise 1.1.0 2017-04-21 CRAN (R 3.3.2)
#> methods * 3.4.0 2017-04-21 local
#> Rcpp 0.12.12 2017-07-15 CRAN (R 3.4.1)
#> rmarkdown 1.6 2017-06-15 CRAN (R 3.4.0)
#> rprojroot 1.2 2017-01-16 CRAN (R 3.3.2)
#> rstudioapi 0.6 2016-06-27 CRAN (R 3.3.0)
#> stats * 3.4.0 2017-04-21 local
#> stringi 1.1.5 2017-04-07 CRAN (R 3.3.3)
#> stringr 1.2.0 2017-02-18 CRAN (R 3.3.2)
#> tools 3.4.0 2017-04-21 local
#> tufte 0.2 2016-02-07 CRAN (R 3.3.1)
#> utils * 3.4.0 2017-04-21 local
#> withr 2.0.0 2017-07-28 CRAN (R 3.4.1)
#> yaml 2.1.14 2016-11-12 CRAN (R 3.3.2)

last_four_commits <- git2r::commits(git2r::repository("."), n = 4)
msgs <- lapply(last_four_commits, show)
#> [108a3b5] 2017-08-21: notes on migration
#> [ad8640d] 2017-08-18: include related work appendix

```

#> [e8f2b63] 2017-08-18: update published version

#> [3d6ae6f] 2017-08-18: print hyperlinks as footnotes in pdf

Prospectus

2

Front Matter

About This Document

This document outlines the research questions, data, and methods for my dissertation. This proposal started out as a grant-writing project, so it has some of the touchstones of NIH F31 grant (Specific Aims, Significance, Approach), but these sections have been expanded considerably.

Date of oral presentation of dissertation proposal: April 3, 2017.

Dissertation Committee Members

- Jan Edwards, primary mentor and chair, Department of Hearing and Speech Sciences, University of Maryland
- Susan Ellis Weismer, official advisor at UW-Madison, Department of Communication Sciences and Disorders
- Margarita Kaushanskaya, Department of Communication Sciences and Disorders
- Audra Sterling, Department of Communication Sciences and Disorders
- David Kaplan, Department of Educational Psychology
- Bob McMurray, Department of Psychological and Brain Sciences, University of Iowa

Planned Dissertation Format

Three thematically related manuscripts, one for each specific aim, to be completed by Summer 2018.

Specific Aims

Individual differences in language ability are apparent as soon as children start talking, but it is difficult to identify children at risk for language delay or disorder. Recent work suggests word recognition efficiency—that is, how well children map incoming speech to words—may help identify early differences in children’s language trajectories. Children learn spoken language by listening to caregivers, so children who are faster at recognizing words have an advantage for word learning. This view is borne out by some studies suggesting that children who are faster at processing words show greater vocabulary gains months later (e.g., Weisleder and Fernald, 2013).

We do not know, however, how word recognition itself develops over time within a child. This is an important open question because word recognition may provide a key mechanism for understanding how individual differences emerge in word learning and persist into early language development. Without a developmental account of word recognition, we lack the context for understanding individual differences in lexical processing. Thus, even the big-picture questions are unclear: Do early differences persist over time so that faster processors remain relatively fast later in childhood? Or, is such a question ill-posed because the magnitude of the differences among children shrink with age? I plan to address this gap in knowledge by analyzing three years of word recognition data collected in recently completed longitudinal study of 180 children.

In particular, I will examine the development of *familiar word recognition*, *lexical competition*, and *fast referent selection* (the ability to map novel words to novel objects in the moment). Through these analyses, I will develop a fine-grained description of how the dynamics of word recognition change year over year, and I will study how differences in word recognition performance relate to child-level measures (such as vocabulary and speech perception). I will complement these empirical analyses with computational cognitive models. With these models, I will simulate the word recognition data from each year and

study how the models need to change to adapt to children’s developing word recognition abilities. These simulations can identify plausible psychological mechanisms that underlie changes in word recognition behavior.

3.1 Specific Aim 1 (Familiar Word Recognition and Lexical Competition)

To characterize the development of familiar word recognition and lexical competition, I will analyze data from a visual world paradigm experiment, conducted at age 3, age 4, and age 5.

In these eyetracking experiments, children were presented with four images of familiar objects and heard a prompt to view one of the images. The four images included a target word (e.g., *bell*), a semantically related word (*drum*), a phonologically similar word (*bee*), and an unrelated word (*swing*). I will use a series of growth curve analyses to describe how children’s familiar word recognition develop year over year. Of interest is how individual differences at Year 1 persist into Year 3. I will also analyze how expressive vocabulary and lexical processing develop together over time. Lastly, I will examine the children’s looks to the distractors to study the developmental course of lexical competition from similar sounding and similar meaning words. Changes in sensitivity to competing words can reveal how lexical competition emerges as a byproduct of learning new words.

3.2 Specific Aim 2 (Referent Selection and Mispronunciations)

To characterize how fast referent selection develops longitudinally, I will analyze data from a looking-while-listening mispronunciation experiment, conducted at age 3, age 4, and age 5.

In these eyetracking experiments (based on White and Morgan, 2008; Law and Edwards, 2015), children saw an image of a familiar object and an unfamiliar object, and they heard either a correct production of the familiar object (e.g., *soup*), a one-feature mispronunciation of the familiar object (*shoop*), or a novel word unrelated to either image (*cheem*). The correct productions test familiar word recognition and the nonwords test fast referent selection. The mispronunciations test a child’s phonological categories (whether the child permits, rejects, or equivocates about mispronunciations).

I will use growth curve analyses to study how children’s responses to the three word types change over time. I will examine familiar word recognition and fast referent selection to determine which feature

of lexical processing better predicts vocabulary growth. I plan to examine dissociations or asymmetries in these forms of processing within children as a way to empirically assess the claim that “novel word processing (referent selection) is not distinct from familiar word recognition” (McMurray et al., 2012). Finally, I will examine how individual differences in vocabulary and speech perception predict responses to mispronunciations and novel words.

3.3 Specific Aim 3 (Computational Modeling)

To identify plausible psychological mechanisms underlying the development of word recognition, I will simulate the word recognition data using cognitive computation models.

The TRACE model of word recognition (McClelland and Elman, 1986) has been used to simulate word recognition data from adults (Alloppenna et al., 1998), adults with aphasia (Mirman et al., 2011), toddlers (Mayor and Plunkett, 2014), and adolescents with language impairments (McMurray et al., 2010). In this model, incoming acoustic information activates perceptual units which in turn activate phoneme units which in turn activate word units. Connections are interactive, so the model can accommodate top-down processing effects and competition among units through inhibition. The model is controlled by psychological parameters like inhibition strength, activation decay rates, and lexicon size and composition. The advantage of using this model is how one can map different behavioral patterns onto changes in model parameterizations to develop a plausible psychological account of developmental changes. I will simulate the Year 1 word recognition data in TRACE, and I will study how the model’s parameters need to change in order to accommodate data from Year 2 and Year 3. I will examine how different model parameters map onto individual differences in word recognition. For instance, under what modeling conditions are mispronunciations of familiar words accepted and do these conditions correspond to child-level differences in vocabulary or speech perception? These simulations will provide a psychological account for the developmental trends and individual variation in word recognition.

3.4 Summary

This project investigates how word recognition develops during the preschool years. There has been no research studying word recognition longitudinally after age two. Findings will show how individual differences in lexical processing change over time and can reveal how low-level mechanisms underlying word recognition mature longitu-

dinally in children. These findings will have translational value by studying processing abilities that subserve word learning and by assessing the predictive relationships between early word recognition ability and later language outcomes.

4

Significance

4.1 Public Health Significance

Vocabulary size in preschool is a robust predictor of later language development, and early language skills predict early literacy skills at school entry (Morgan et al., 2015). By studying the mechanisms that shape word learning, we can understand how individual differences in language ability arise and identify strategies for closing language gaps between children. Word recognition—the process of mapping incoming speech sounds to known or novel words—has been shown to predict later language outcomes. We do not know how this ability develops over time, and we do not know when word recognition is most predictive of future outcomes. This project will provide an integrated account of how word recognition and its relationship with vocabulary size change from age 3 to age 5.

4.2 Scientific Significance

4.2.1 Lexical Processing Dynamics

Mature listeners recognize words by continuously evaluating incoming speech input for possible word matches through lexical competition. The first part of a word activates multiple candidate words in parallel, and these candidates compete so that the best-fitting word is recognized. For example, the onset “bee” might activate the candidates *bee*, *beam*, *beetle*, *beak*, *beaker*, *beginning*, and so on, but an additional “m” would narrow the candidates to just *beam*. Semantic relationships also influence lexical processing, and cascading phonological-semantic effects—e.g., where *castle* activates the phonologically similar *candy* which in turn activates the semantically related *sweet*—have been demonstrated (Marslen-Wilson and Zwitserlood, 1989). Both low-level phonetic cues and high-level grammatical, semantic and pragmatic information can influence this process, but the *continuous processing*

of multiple competing candidates is the essential dynamic underlying word recognition (Magnuson et al., 2013).

What about young children who know considerably fewer words? Eyetracking studies with toddlers have suggested a developmental continuity between toddlers and adult listeners. Children recognize words incrementally (Swingley et al., 1999), match truncated words to their intended referents (Fernald et al., 2001), and use information from neighboring words in a sentence to facilitate word recognition. This information can be grammatical: Lew-Williams and Fernald (2007) found that Spanish-acquiring preschoolers can use grammatical gender on determiners (*el* or *la*) to anticipate the word named in a two-object word recognition task. The information can also be subcategorical phonetic variation: We found that English-acquiring toddlers look earlier to a named image when the coarticulatory formant cues on word *the* predict the noun of the sentence, compared to tokens with neutral coarticulation (Mahr et al., 2015).

There is some evidence for lexical competition where children are sensitive to phonological and semantic similarities among words. Ellis Weismer et al. (2016) showed that toddlers (14–29 months old) looked less reliably to a named image when the onscreen competitor was a semantically related word or perceptually similar image. In Law et al. (2016), preschoolers (28–60 months old) demonstrated sensitivity to semantic and phonological competitors in a four-image eyetracking task. Huang and Snedeker (2011) presented evidence of cascading semantic-phonological activation in five-year-olds such that for a target word like *log*, children looked more to an indirect phonological competitor like *key* (competing through its activation of *lock*) than they looked to an unrelated image like *carrot*. In contrast to these studies which all demonstrate interference from similar words, Mani and Plunkett (2010) demonstrated cross-modal phonological priming effects in 18-month-olds. In this study, a picture of prime word (e.g., cat or teeth) was presented in silence; then two images (e.g., cup and shoe) were presented, one of which was named (*cup*). Children on average looked more to the target word (like *cup*) when it was primed by an image of a phonological neighbor (like *cat*), and the children performed at chance when the prime was not related to the named word. Mani et al. (2012) found a similar result for cascading phonological-semantic priming with 24-month-olds: Children looked more to a target *shoe* compared to a distractor *door* when primed by an image of *clock*, assumed to activate *sock* which primed *shoe*.

The above studies involved young children of different ages tested under different procedures, sometimes in different dialects and languages. Averaging these results together, so to speak, the studies suggest that early word recognition demonstrates some hallmarks of adult

behavior: Continuous processing of words, integration of information from different levels of representation, and the influence of similar, unspoken words on recognition of a word. Nevertheless, we only have a fragmented view of how familiar word recognition and lexical competition develop within children.

One open question is how lexical competition develops within children. For example, do phonological similar words exert more interference during word recognition as children grow older? As a guiding hypothesis, we can think of word learning as a gradual process where familiarity with a word moves from shallow receptive knowledge to deeper expressive knowledge. In adult listeners, words compete and inhibit one another, so that a word is truly “learned” and integrated into the lexicon when it can influence the processing of other words (a line of reasoning reviewed by Kapnoula et al., 2015). Increasing sensitivity to similar sounding words over time would reveal that children improve their ability to consider multiple candidates in parallel. By studying how sensitivity to similar-sounding and similar-meaning words develop over time and within ever-growing vocabularies, this project can reveal how children come to process words efficiently.

Another avenue for studying word recognition is to examine how listeners respond to unfamiliar or novel stimuli. A productive line of research has found that children are sensitive to mispronunciations during word recognition (e.g., Swingley and Aslin, 2000, 2002). White and Morgan (2008) presented toddlers with images of a familiar and novel object, and children heard a correct production of the familiar object, mispronunciations of the familiar object of varying severity, or an unrelated nonword. Toddlers looked less to a familiar word when the first segment was mispronounced. Moreover, they demonstrated graded sensitivity such that a 1-feature mispronunciation yielded more looks to an image than a 2-feature mispronunciation, and a 2-feature mispronunciation yielded more looks than a 3-feature one. Finally, in the nonword condition, the children looked more to the novel object than the familiar one, demonstrating *fast referent selection* as they associated novel words to novel objects in the moment. A similar pattern of effects was observed in the mispronunciation study by Law and Edwards (2015) with preschoolers mapping real words to familiar objects, nonwords to novel objects, and equivocating about mispronunciations of familiar words.

As with lexical competition, it is unclear how children’s responses to mispronunciations and novel words change over time or how individual differences among children change over time. For example, do children become more forgiving of mispronunciations as they mature and learn more words? We might expect so, as children become more experienced at listening to noisy, degraded, or misspoken speech.

Another open question involves the development of fast referent selection. At face value, we might expect a child’s ability to associate new words with unfamiliar objects to be more direct measure of word-learning capacity than a child’s ability to process known words. Under this assumption, we would expect individual differences in fast referent selection to be highly correlated with vocabulary growth. But McMurray et al. (2012) propose that the same basic process is at play in both recognition of familiar words and fast association of nonwords. In experiments, the observed behaviors are the same: Children hear a word and direct their attention to an appropriate referent. This project can tackle these questions by describing how mispronunciations are processed as children grow older and by examining whether familiar word recognition and fast referent selection dissociate and which one is a better predictor of vocabulary growth.

4.2.2 *Individual Differences in Word Recognition*

We have a rough understanding of the development of word recognition, and these gaps in knowledge matter because young children differ in their word recognition abilities. These differences are usually measured using *accuracy* (a probability of recognizing to a word) or *efficiency* (a reaction time or some measure of how quickly accuracy changes over time). These differences are consequential too, as word recognition differences correlate with other language measures concurrently, retrospectively, and prospectively.

The best predictor of lexical processing efficiency is concurrent vocabulary size: Children who know more words look more quickly and reliably to a named word (e.g., Law and Edwards, 2015).

This fact deserves a brief reflection: Suppose the information processing mechanism behind word recognition were just a naïve table search. Then this finding is somewhat puzzling: Children with larger lexicons have to find a needle in a larger haystack—yet this apparent liability is an advantage. That is why the search analogy is naïve. One explanation follows from the earlier described idea about graded word learning: Children become better at recognizing words as they learn more words because they extract regularities and discover similarities among words and develop more efficient lexical representations—the haystack develops regularity and becomes easier to search.

Although it is a robust predictor of word recognition, vocabulary size is nonspecific. For lexical processing dynamics, vocabulary size can be considered an indicator for the organization and efficiency of a child’s lexicon, but it also correlates with other (meaningful) differences. Vocabulary is related to differences in speech perception (Cristia et al., 2014) and environmental factors like language input

(e.g., Hart and Risley, 1995; Hoff, 2003). For instance, measures of speech perception at 6–8 months predict vocabulary size at 24 months (e.g., Tsao et al., 2004; Kuhl et al., 2008), so processing predicts future vocabulary predicts concurrent processing.

A related complication is the apparent predictive validity of word recognition measures. Marchman and Fernald (2008) found that vocabulary size and lexical processing efficiency at age 2 jointly predicted working memory scores and expressive language scores at age 8. This result would suggest domain-general processing advantages influence word learning. Fernald and Marchman (2012) also found that late talkers who looked more quickly to a named word at 18 months showed larger gains in vocabulary by 30 months compared to late-talkers who looked more slowly at 18 months. Weisleder and Fernald (2013) found that lexical processing and language input at 19 months predict vocabulary size at 25 months and that lexical processing mediated the effect of language input.

Word recognition efficiency and vocabulary size are interconnected measures with concurrent and predictive associations. This project can clarify this relationship by examining the co-development of word recognition, vocabulary size, and speech perception. In particular, I will ask how individual differences in word recognition change over time alongside differences in vocabulary. I can also which features of word recognition (fast referent selection, lexical competition, etc.) are most predictive of vocabulary outcomes at age 5. The additional measures of speech perception can also help clarify the specific effects of vocabulary size on word recognition.

4.2.3 *Computational Modeling*

One way to bolster a theory of word recognition is to implement the theory in a computational model and simulate human behavior with the model. If the model can produce responses like those of human listeners, then the behavioral data support the model and the model’s underlying theory. Models have adjustable parameters, and these generally correspond to plausible psychological constructs like the inhibition strength among competing units or a learning rate for modifying connections. Part of the simulation process therefore is to ask under what conditions a model simulates human behavior, and then interpret the simulations in terms of psychological mechanisms.

Here is a hypothetical example of this modeling strategy: Suppose we want to investigate the finding that a larger vocabulary size predicts more efficient word recognition. We would ask under which parameters (i.e., model settings) a model with a large lexicon outperforms one with a smaller lexicon at word recognition. It might be the

case that these results occur only when parameters are set so that the representations of speech sounds are comparatively noisier in a model with fewer words. In this scenario, the models provide a plausible psychological interpretation for the empirical findings: Children develop better representations of speech sounds as they learn words, and these better representations enable more efficient word recognition. Granted, this example is just a hypothetical case, but it illustrates how I intend to use computational models as a way to describe word recognition trends and variation in terms of psychological mechanisms and processing parameters.

The TRACE model of speech perception and word recognition (McClelland and Elman, 1986) is well suited for this kind of simulation work. TRACE can simulate a dozen or so empirical results from speech perception literature (Strauss et al., 2007, Table 1), and it has been used to simulate word-recognition data from adults (Allopenna et al., 1998), adults with aphasia (Mirman et al., 2011), and toddlers (Mayor and Plunkett, 2014).

McMurray et al. (2010) used TRACE to simulate word recognition results from adolescents with specific language impairment (SLI). They mapped certain theories about SLI onto different model parameters. To test the theory that listeners with SLI have impairments in acoustic perception, for example, they varied three of the model’s parameters: amount of noise added to the model’s mock-speech input (acoustic resolution), temporal spread of acoustic features in the input (temporal resolution), and rate of decay in the model’s acoustic feature detectors (perceptual memory). Other theories of SLI (and other model parameters) provided a closer fit to the observed data than the perceptual deficit theory. Specifically, lexical decay—“the ability to maintain words in memory” (p. 23)—was the most important model parameter, implying that individual differences in word recognition for listeners with SLI are rooted in lexical processes, as opposed to perceptual or phonological ones.

This example shows how simulations with TRACE recast word recognition performance in terms of psychological processes. For this project, I will use TRACE to describe how cognitive processes and representations need to change to simulate the development of word recognition in preschoolers.

4.2.4 *Summary*

This project studies word recognition in children over three years, so it will provide the first longitudinal study of word recognition in preschoolers. Children in this cohort cover a range of vocabulary scores at Time 1, and this variability allows one to investigate individual dif-

ferences in vocabulary and word recognition over time and assess the predictive value of these measures. Furthermore, this project studies word recognition in two experimental tasks that can tap into different aspects of word recognition. Specifically, a four-image experiment with semantic and phonological foils allows me to study how lexical competition develops, and a two-image experiment with nonwords and mispronunciations enables me to study how fast referent selection develops over time as well. I will use mixed effects modeling to study not just gross measures of accuracy or interference from distractors, but the time course and lexical dynamics of word recognition using growth curve analyses. These empirical models of the longitudinal word recognition data will be supported by computational models, so that the developmental changes can be described by plausible psychological mechanisms of word recognition.

5

Research Hypotheses

In this section, I outline the main hypotheses I plan to study for each specific aim. This section is intended to preregister the main analyses for this project, so I can cleanly separate confirmatory and exploratory results.

5.1 Specific Aim 1 (Familiar Word Recognition and Lexical Competition)

- Children's accuracy and efficiency of recognizing words will improve each year.
- There are stable individual differences in lexical processing of familiar words such that children who are relatively fast at Year 1 remain relatively fast at Year 2 and Year 3.
- However, the magnitude of these individual differences diminishes over time, as children converge on a mature level of performance for this paradigm.
- As a consequence, individual differences in word recognition at age 3, for example, will be more discriminating and predictive of age 5 language outcomes than differences at age 4.
- Vocabulary size and lexical processing will be tightly correlated such that large year-over-year gains in one measure will predict large year-over-years gains in the other measure.
- Children will become more sensitive to lexical competitors as they age, based on the hypothesis that children discover similarities among words as a consequence of learning more and more words.
- Children will differ in their sensitivity to lexical competitors, and these individual differences will correlate with other child-level measures.

5.2 *Specific Aim 2 (Referent Selection and Mispronunciations)*

- Children’s accuracy and efficiency of recognizing real words and fast-associating nonwords will improve each year.
- Performance in real word recognition and fast association of nonwords will be highly correlated, based on the hypothesis that the same process (referent selection) operates in both situations.
- Under the alternative hypothesis, real word recognition and fast referent selection reflect different skills with different developmental trajectories. Thus, if there is any dissociation between recognition of real words and nonwords, it will be observed in younger children.
- Although these two measures will be correlated, I predict performance in the nonword condition will be a better predictor of future vocabulary growth than performance in the real word condition. This hypothesis is based on the idea that fast referent selection is a more relevant skill for learning new words than recognition of known words.
- For the mispronunciations, I predict children with larger vocabularies (that is, older children) will be more likely to tolerate a mispronunciation as a production of familiar word compared to children with smaller vocabularies.
- Mispronunciations that feature later-mastered sounds (e.g., *rice/wice*) will be more likely to be associated to novel objects than earlier-mastered sounds (*duck/guck*).

5.3 *Specific Aim 3 (Computational Modeling)*

The research questions for this aim are more exploratory than confirmatory. I plan to use the well-studied TRACE model of spoken word recognition (McClelland and Elman, 1986). The TRACE model has no built-in semantic representations, so it will be used to model data from the two-image mispronunciation experiment (Aim 2) and a “semantic-less” subset of the data from the four-image experiment (Aim 1). Research questions include:

- What parameterizations of the TRACE model of word recognition account Year 1 word recognition data?
- If the default model of TRACE represents “default” adult listeners, how the children are Year 1 different than adults.
- How do these parameterizations have to change year over year to accommodate the data from older children at Year 2 and again

at Year 3? What is the developmental story for these changing parameterizations?

- How does the model handle mispronunciations? What changes are required to accommodate mispronunciations and novel words? Mayor and Plunkett (2014) have simulated mispronunciation experiments like White and Morgan (2008), although it is not clear how the model can accommodate the current mispronunciation experiment which presents mispronunciations and nonwords.
- How does the model's vocabulary parameters (namely the size and composition of the lexicon used for the simulations) correlate with the vocabulary size of participants?
- Can the model simulations incorporate other child-level measures? For example, do children who demonstrate poorer speech perception in a non-eyetracking speech discrimination task require lower levels of phoneme inhibition in their word recognition models?

6

Methods

6.1 General Research Design: Participants

Word recognition data and vocabulary data, among other measures, were collected over a three-year longitudinal study (R01DC002932; the Learning to Talk project). Children were 28–39 months-old at Time 1, 39–52 at Time 2, and 51–65 at Time 3. Approximately, 180 children participated at Time 1, 170 at Time 2, and 160 at Time 3. Of these children, approximately 20 were identified by their parents as late talkers. Prospective families were interviewed over telephone before participating in the study, and “children with an individualized education program or any parent-reported visual problems, language problems, or developmental delays were not scheduled for testing” (Law et al., 2016).¹ Recruitment and data collection occurred at two Learning to Talk lab sites—one at the University of Wisconsin–Madison and the other at the University of Minnesota.

Table 6.1 summarizes the cohort of children in each year of testing. The numbers and summary statistics here are approximate, describing children who participated at each year, but whose data may still be excluded from the analyses. Some potential reasons for exclusion include: excessive missing data during eyetracking, experiment or technology error, developmental concerns not identified until later in study, or a failed hearing screening. Final sample sizes will depend on the measures needed for an analysis and the results from data screening checks. For each project aim, I will disclose all measurements and data exclusions following guidelines by the Center for Open Science (Nosek et al., 2014).

¹ Appendix C describes how this dissertation relates to other work from our lab.

Table 6.1: Participant characteristics. Education levels: *Low*: less than high school, or high school; *Middle*: trade school, technical or associates degree, some college, or college degree; and *High*: graduate degree.

| | Year 1 | Year 2 | Year 3 |
|---------------------------------------|----------------|----------------|----------------|
| N | 184 | 175 | 160 |
| Boys, Girls | 94, 90 | 89, 86 | 82, 78 |
| Maternal education: Low, Middle, High | 15, 98, 71 | 12, 92, 71 | 6, 90, 64 |
| Dialect: MAE, AAE | 171, 13 | 163, 12 | 153, 7 |
| Parent-identified late talkers | 20 | 19 | 16 |
| Age (months): Mean (SD) [Range] | 33 (3) [28–39] | 45 (4) [39–52] | 57 (4) [51–66] |
| EVT-2 standard score: Mean (SD) | 115 (18) | 118 (16) | 118 (14) |
| PPVT-4 standard score: Mean (SD) | 113 (17) | 120 (16) | — |
| GFTA-2 standard score: Mean (SD) | 92 (13) | — | 91 (13) |

6.2 General Eyetracking Procedure

Two eyetracking experiments were performed each year of the longitudinal study. These experiments followed the same essential procedure: During each trial, photographs of images appeared on a computer screen for a few seconds followed by a prompt to view one of the images (e.g., *find the dog*). This procedure measures a child’s real-time comprehension of words by capturing how the child’s gaze location changes over time in response to speech.

6.2.1 Experiment Administration

Children participating in the study were tested over two lab visits (i.e., on different dates). The first portion of each visit involved “watching movies”—that is, performing two blocks of the eyetracking experiments. A play break or hearing screening occurred between the two eyetracking blocks, depending on the visit.

Each eyetracking experiment was administered as a block of trials (24 for the four-image task and 38 for the two-image task). Children received two different blocks of each experiment. The blocks for an experiment differed in trial ordering and other features (see Specific Procedures). Experiment order and block selection were counterbalanced over children and visits. For example, a child might have received Exp. 1 Block A and Exp. 2 Block B on Visit 1 and next received Exp. 2 Block A and Exp. 1 Block B on Visit 2. The purpose of this presentation was to control possible ordering effects where a particular experiment or block benefited from consistently occurring first or second.

Experiments were administered using E-Prime 2.0 and a Tobii T60XL eyetracker which recorded gaze location at a rate of 60 Hz. The experiments were conducted by two examiners, one “behind the scenes” who controlled the computer running the experiment and another “onstage” who guided the child through the experiment. At the beginning of each block, the child was positioned so the child’s eyes were approximately 60 cm from the screen. The examiners calibrated the eyetracker to the child’s eyes using a five-point calibration procedure (center of screen and centers of four screen quadrants). The examiners would repeated this calibration procedure if one of the five calibration points for one of the eyes did not calibrate successfully. During the experiment, the behind-the-scenes examiner monitored the child’s distance from the screen and whether the eyetracker was capturing the child’s gaze. The onstage examiner coached the child to stay fixated on the screen and repositioned the child as needed to ensure the child’s eyes were being tracked. Every six or seven trials in a block of an experiment, the experiment briefly paused with a reinforcing animation or activity. During these breaks, the onstage examiner could reposition the child if necessary before resuming the experiment.

We used a gaze-contingent stimulus presentation. “After 2 s of familiarization time with the images in silence, the experiment paused to verify that the child’s gaze was being tracked. After 300 ms of continuous gaze tracking, the trial advanced. Otherwise, if the gaze could not be verified after 10 s, the trial advanced. This step ensured that for nearly every trial, the gaze was being tracked before playing the carrier phrase, or in other words, that the child was ready to hear the carrier stimuli” (Mahr & Edwards, in revision). During Year 1 and Year 2, an attention-getter (e.g., *check it out!*) played 1 s following the end of the target noun. These reinforcers were dropped in Year 3 to streamline the experiment for older listeners.

6.2.2 Stimuli

For both experiments, “stimuli were presented in children’s home dialect, either Mainstream American English (MAE) or African American English (AAE). We made an initial guess about what the home dialect was likely to be based on a number of factors, including the recruitment source and the child’s address. For most children, the home dialect was MAE. If we thought the home dialect might be AAE, a native AAE speaker who was a fluent dialect-shifter between AAE and MAE was scheduled for the lab visit, and she confirmed the home dialect by listening to the caregiver interact with the child during the consent procedure at the beginning of the visit” (Mahr & Edwards, in revision). Prompts to view the target image of a trial (e.g., *find the*

girl) used the carrier phrases “find the” and “see the”. These carriers were recording in the frame “find/see the egg” and cross-spliced with the target nouns to minimize coarticulatory cues on the determiner “the”.

The images used in each experiment consisted of color photographs on gray backgrounds. These images were piloted in a preschool classroom to ensure that children consistently used the same label for familiar objects and did not consistently use the same label for novel/unfamiliar objects.

6.2.3 Data Preparation

Data from both experiments were prepared using the same procedure. “We mapped the gaze x - y coordinates onto the images onscreen. We performed *deblinking* by interpolating short windows of missing data (up to 150 ms) if the child fixated on the same image before and after a missing data window. In other words, if the gaze did not shift to another image, and if the missing data window was short enough, that window was classified as a blink and interpolated, using the fixated image as the imputed value” (Mahr & Edwards, in revision). Next, we performed trial-level cleaning. “We examined eyetracking data in the 2-s window following the onset of the target word. A trial was considered unreliable if at least 50% of the eyetracking data during the 2-s window was missing. These trials were not reliable because the child did not look at the display for the majority of the analysis window” (Mahr & Edwards, in revision). If more than half of a child’s trials, combined across blocks, were unreliable, that child was excluded from analysis. “Finally, we downsampled our data into 50-ms bins, reducing the eyetracking sampling rate from 60 Hz to 20 Hz. This procedure smoothed out high-frequency noise in the data by pooling together data from adjacent frames” (Mahr & Edwards, in revision).

6.3 Specific Procedure: Aim 1 (*Familiar Word Recognition and Lexical Competition*)

Visual World Paradigm Task. In eyetracking studies with toddlers, two familiar images are usually presented: a target and a distractor. This experiment is a four-image eyetracking task that was designed to provide a more demanding word recognition task for preschoolers. In this procedure, four familiar images are presented onscreen followed by a prompt to view one of the images (e.g., *find the bell!*). The four images include the target word (e.g., *bell*), a semantically related word (*drum*), a phonologically similar word (*bee*), and an unrelated word (*swing*).

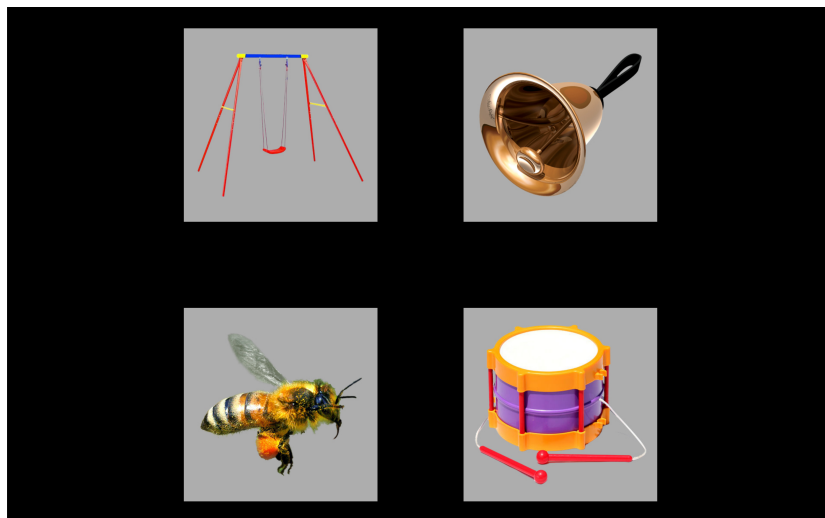


Figure 6.1: Example display for the target *bell* with the semantic foil *drum*, the phonological foil *bee*, and the unrelated *swing*.

6.4 Specific Procedure: Aim 2 (Referent Selection and Mispronunciations)

Mispronunciation Task. This experiment is an adaptation of the mispronunciation detection task by White and Morgan (2008) and Law and Edwards (2015). In this experiment, two images are presented onscreen—a familiar object and an unfamiliar object—and the child hears a prompt to view one of the images. In the *correct pronunciation* (or *real word*) and *mispronunciation* conditions, the child hears either the familiar word (e.g., *soup*) or a one-feature mispronunciation of the first consonant of the target word (*[sh]oup*). Importantly, within a block of trials, the child never hears both the correct and mispronounced forms of the word. These conditions are designed to test whether children map mispronunciations to novel words. To encourage fast referent selection, there were also trials in a *nonword* condition where the label was an unambiguous novel word (e.g., *cheem* presented with images of a bed and a novel-looking pastry mixer). Each nonword was constructed to match the phonotactic probability of one of the mispronunciations.

In a block of trials, there were 12 trials each from the nonword condition, correct production condition, and mispronunciation conditions, and children received two blocks of the experiment. A complete list of the items used in the experiment over the three years of the study is included in Appendix A.



Figure 6.2: Example display for a trial in which *duck* is mispronounced as “guck”.

Outcome Measures

The primary outcome measure for this project is word recognition performance, and I will study how it changes over time, under different experimental conditions, across children, and so on. I will measure word recognition performance using eyetracking growth curves. In this technique, I aggregate looking locations across trials to compute the proportion of looks to the target object (versus the distractor images) at each time point after the onset of the target word. This growth curve measures how the probability of fixating on a named object changes over time. Thus, children who have steeper accuracy growth curves demonstrate faster word recognition because the probability of looking at a named object increases more quickly over time.

The other primary child-level measurements of interest for this project are vocabulary and speech perception. Expressive vocabulary was measured at all three years using the Expressive Vocabulary Test (EVT-2, Williams, 2007). Receptive vocabulary was measured at Year 1 and Year 2 using the Peabody Picture Vocabulary Test (PPVT-4, Dunn and Dunn, 2007). These tests will provide child-level measures of vocabulary and vocabulary growth. At Year 1 and Year 2, we measured speech perception using a minimal-pair discrimination task (as in Baylis et al., 2008), and Year 2 and Year 3, we measured speech perception with a speech sound judgment task (SAILS as in Rvachew, 2006). Performance on these tasks will quantify individual differences in speech perception.

Two other relevant measures include an articulation test administered at Year 1 and Year 3 (GFTA-2, Goldman and Fristoe, 2000) and phonological awareness given at Year 2 and Year 3 (CTOPP-2, Wagner et al., 2013). They are relevant because they draw on a child's phonological knowledge of words, but I limit their consideration for only post-hoc, exploratory analyses because they measure a child's speech production performance.

7.1 Sample Data

Figure 7.1 shows the development of word recognition on the four-image task over the three years of the longitudinal study for two participants in this data set. To limit data-peeking, I only considered at data from the first 20 participants and then selected two patterns of development with clean individual differences.

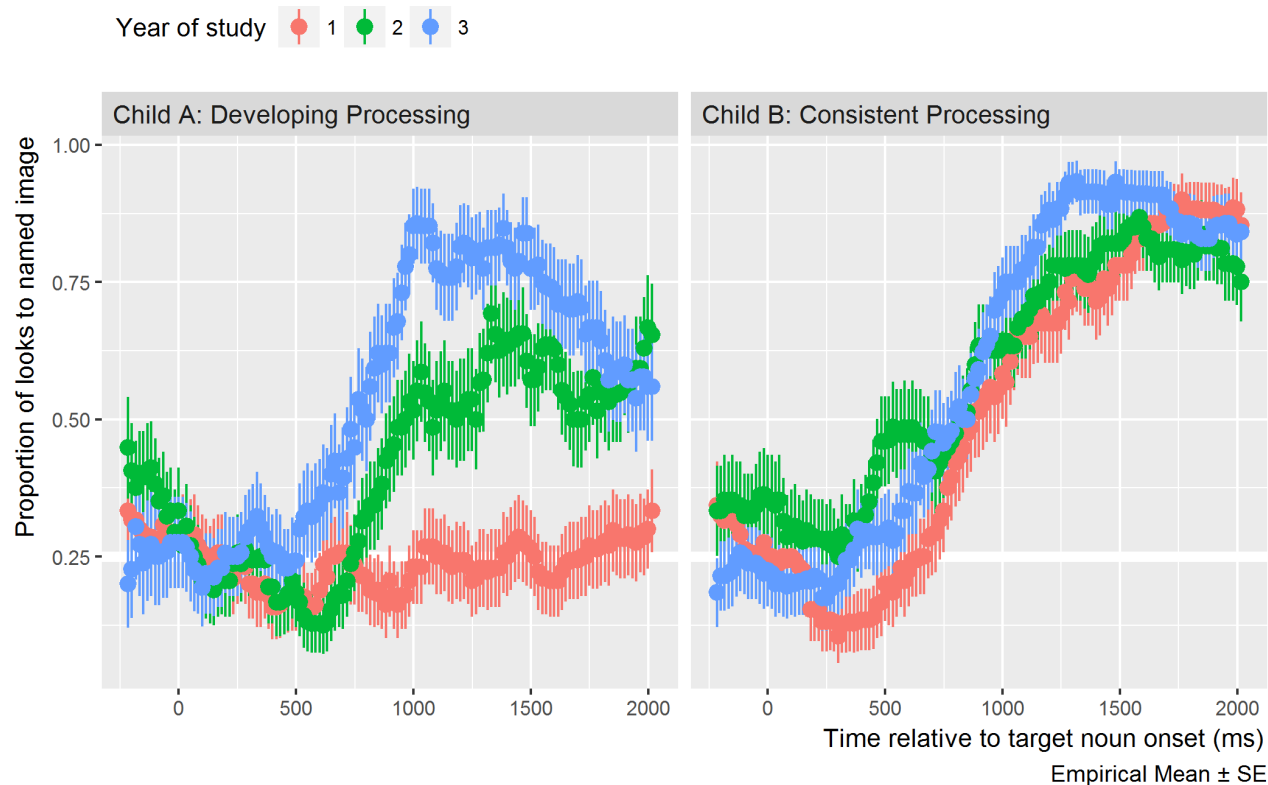


Figure 7.1: Data for two children on the four-image task over the three years of the study.

The participant on the left showed substantial gains in processing ability from year to year, whereas the child on the right showed approximately the same level of performance each year, with some small gains in peak accuracy and rate of change. The child with substantial changes in word recognition also showed gains in expressive vocabulary with improved *standard* scores on the EVT-2: 83 at Year 1, 95 at Year 2, and 103 at Year 3. In contrast, the child with consistent processing had an above-average standard expressive vocabulary score during each year of the study: 133 at Year 1, 138 at Year 2, and 130 at Year 3.

These two children give a glimpse of the many possible developmental trajectories for this task and how they relate to expressive

vocabulary. The main work of this project will be describing and characterizing the trajectories of many children on different aspects of lexical processing.

8

Analysis

8.1 Growth Curve Analysis

Eyetracking growth curves will be analyzed using **Bayesian mixed effects logistic regression**. I will use *logistic* regression because the outcome measurement is a probability (the log-odds of looking to the target image versus a distractor). I will use *mixed-effects* models because I want to estimate a separate growth curve for each child (to measure individual differences in word recognition) but also treat each child's individual growth curve as a draw from a distribution of related curves.

I plan to use *Bayesian* techniques to study a generative model of the data. Instead of reporting and describing a single, best-fitting model of some data, Bayesian techniques consider an entire distribution of plausible models that are consistent with the data and any prior information we have about the models. By using this approach, I can explicitly quantify uncertainty about statistical effects and draw inferences using estimates of uncertainty (instead of using statistical significance—which is not a straightforward matter for mixed-effects models).¹

The eyetracking growth curves will be fit using an orthogonal cubic polynomial function of time (a now-conventional approach; see Mirman, 2014). Put differently, I will model the probability of looking to the target during an eyetracking task as:

$$\text{log-odds}(\textit{looking}) = \beta_0 + \beta_1 * \textit{Time}^1 + \beta_2 * \textit{Time}^2 + \beta_3 * \textit{Time}^3$$

That the time terms are *orthogonal* means that \textit{Time}^1 , \textit{Time}^2 and \textit{Time}^3 are transformed so that they are uncorrelated. Under this formulation, the parameters β_0 and β_1 have a clear interpretation in terms of lexical processing performance. The intercept, β_0 , measures the area under the growth curve—or the probability of fixating on the target word averaged over the whole window. We can think of

¹ It is tempting to further justify this approach by comparing Bayesian versus classical/frequentist statistics, but my goals in using this method are simple: To estimate statistical effects and quantify uncertainty about those effects. This pragmatic brand of Bayesian statistics is illustrated in texts by Gelman and Hill (2007) and McElreath (2016).

β_0 as a measure of average accuracy or of *word recognition reliability*. The linear time parameter, β_1 , estimates the steepness of the growth curve—or how the probability of fixating changes from frame to frame. We can think of β_1 as a measure of *processing efficiency*, because growth curves with stronger linear features exhibit steeper frame-by-frame increases in looking probability.²

For each experimental task, I will study how word recognition changes over time by modeling how growth curves change over developmental time. This amounts to study how the growth curve parameters changes year over year in the study. I can model the data for an eye-tracking task by including dummy-coded indicators for Year 1, Year 2, and Year 3 and having these indicators interact with the growth curve parameters. In such a model, Year 2 would be the reference year, so the Year 1 parameters would estimate how the word-recognition-curves change from Year 2 to Year 1, and Year 3 parameters would be interpreted similarly.

Year 2 Growth Curve:

$$\log\text{-odds}(\textit{looking}) = \beta_0 + \beta_1 * \textit{Time}^1 + \beta_2 * \textit{Time}^2 + \beta_3 * \textit{Time}^3$$

Adjustments to Year 2:

$$\beta_i = \gamma_{i:2} + \gamma_{i:1} * \textit{Year1} + \gamma_{i:3} * \textit{Year3}$$

Thus, the interaction effects for the intercept term ($\gamma_{0:1}$, $\gamma_{0:3}$) describe how overall accuracy changed between years, and interaction effects for the linear-time terms ($\gamma_{1:1}$, $\gamma_{1:3}$) describe changes in overall processing efficiency between years.³

Lastly, a brief comment about priors. Bayesian models require prior information (“priors”). For these models, I will use weakly to moderately informative priors. For example, suppose x and y are scaled to mean 0 and standard deviation 1. A weakly informative prior for the effect of x on y might be $\text{Normal}(0, 5)$ —a normal distribution with mean 0 and standard deviation 5. If we fit a regression model and observed an effect size of 12 SD units, our first assumption would be that something went wrong with our software. The weakly informative prior captures this level of prior information. A moderately informative prior would be $\text{Normal}(0, 1)$. This prior information captures our disciplinary experience that effect sizes greater than ± 1 relatively uncommon in child language research. A strongly informative prior for this effect might be something like $\text{Normal}(.4, .1)$ which says that our model should be very skeptical of negative effects and of effects larger than .8. For this project, I will default to the first two levels of prior information.

² The polynomial other terms are less important—or rather, they have do not map as neatly onto behavioral descriptions as the accuracy and efficiency parameters. The primary purpose of quadratic and cubic terms is to ensure that the estimated growth curve adequately fits the data. In this kind of data, there is a steady baseline at chance probability before the child hears the word, followed a window of increasing probability of fixating on the target as the child recognizes the word, followed by a period of plateauing and then diminishing looks to target. The cubic polynomial allows the growth curve to be fit with two inflection points: the point when the looks to target start to increase from baseline and the point when the looks to target stops increasing.

³ In the case that these omnibus multi-year interaction models do not converge or are otherwise computationally ill-behaved, I will model each year separately, extract each participant’s growth curve parameters, and compare those measures in a second stage of modeling.

8.2 Aim 1 (*Familiar Word Recognition and Lexical Competition*)

In the four-image task, I will model the development of familiar word recognition by studying how looks to the target image change year over year, as described in Growth Curve Analysis.

I predict that children will be more sensitive to the phonological foil and semantic foils in this task as they age and learn more words. This hypothesis is based on the idea that children discover similarities among words as they learn word and integrate them into their lexicon. To test this hypothesis, I will study how the probability of fixating on the foils changes over trial-time and how these growth curves change from year to year. In the conventional model of eyetracking data, the outcome is a binomial choice—Target versus Distractor—and we can estimate the log-odds of fixating on the target image relative to the distractors. To study the specific effect of the Phonological foil in this task, Law et al. (2016) treated the Unrelated foil as a reference distractor and compared two separate binomial growth curves (see Figure 8.1): Target versus Phonological, and Target versus Unrelated. The same technique was used on the Semantic foil as well. With this approach, shown below, we observed an early negative effect of the phonological foil and a late negative effect of the semantic foil.

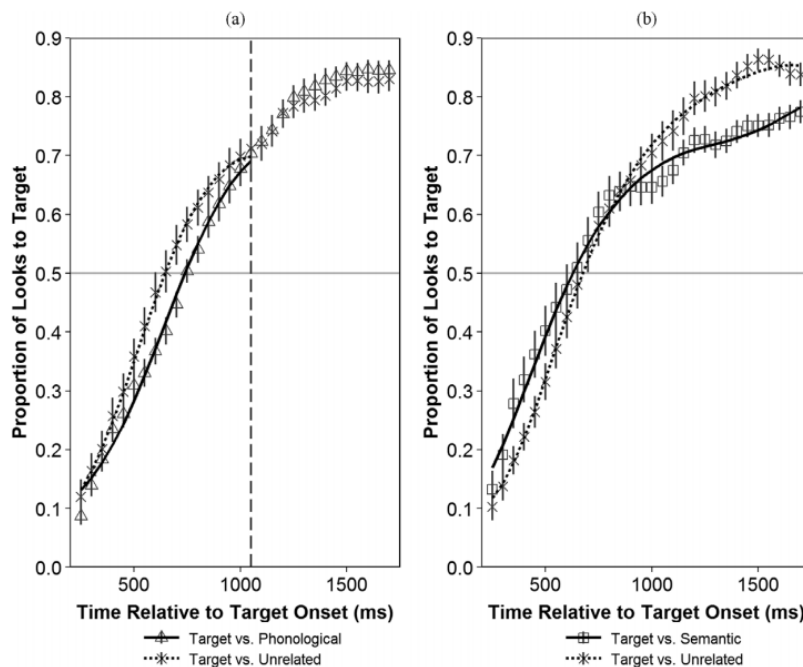


Figure 8.1: Reprint of Figure 4 from Law et al. (2016) to illustrate the strategy of examining lexical competition effects where the semantic foil and phonological foil are compared to the unrelated image.

I plan to employ this technique for studying lexical competition

effects and their development from Year 1 to Year 2 to Year 3.⁴ Increased lexical completion would be reflected in greater interference from the foils compared to the unrelated image.

A more comprehensive statistical model for this experiment would capture the fact that the data are multinomial: Target versus Phonological versus Semantic versus Unrelated. However, multinomial mixed effects growth curves have not been used for eyetracking data. They are not estimable with the standard classical modeling software (lme4). I plan to examine whether such a model is feasible with Bayesian techniques, but it may prove to be too unstable. In that case, I will fall back to the above described strategy.

I will also examine whether individual differences in lexical processing are stable over time. For an initial analytic step, I will identify how frequently children change quartiles or deciles. The idea here is that stable individual differences in processing should preserve some relative rankings—fast-processing children should remain relatively fast compared to their peers. If children make large swings in their rankings, e.g., changing from the bottom 55th to the 30th percentile, then we have evidence that these rankings are unstable. I am also interested in how the magnitude of individual differences change over time. The differences among children could diminish over time so that the rankings are unstable and reflect small variations among children.

Another question concerns the relationship between the development of lexical processing and the development of vocabulary size. Specifically, I will ask how age-adjusted lexical processing measures (accuracy and efficiency) correlate with age-adjusted vocabulary size each year. Moreover, I will also examine whether children who make large gains in vocabulary size also show large concurrent changes in their lexical processing measures.

8.3 Aim 2 (*Referent Selection and Mispronunciations*)

For this task, I will model how the looks to the familiar image differ in each condition (real words, mispronunciations, nonwords) and how the growth curves for each condition change year over year. This model will use growth curve model described in Growth Curve Analysis but augmented with Condition effects.

I will examine whether and when any dissociation is observed for word recognition in the real word and nonword conditions. McMurray et al. (2012) argue that familiar word recognition and fast association for novel words reflect the same cognitive process: referent selection. Data from this task would support with this hypothesis when the growth curves for looks to the familiar image are symmetrical for the real word and nonword conditions. Figure 8.2, showing data from Law

⁴ Most of the phonological foils for this task are cohorts or words that shared onset consonants (*bear-bell*, *flag-fly*). Only trials with this type of phonological foil will be used to measure phonological interference effects. The other phonological foils included some rhymes (like *ring-swing*) and some words with onsets that differed by a phonetic feature (*kite-gift*). Effects from those foils will be considered separately.

and Edwards (2015, $n = 34$ children, 30-46 months old), shows some symmetry for the real word and nonword conditions.

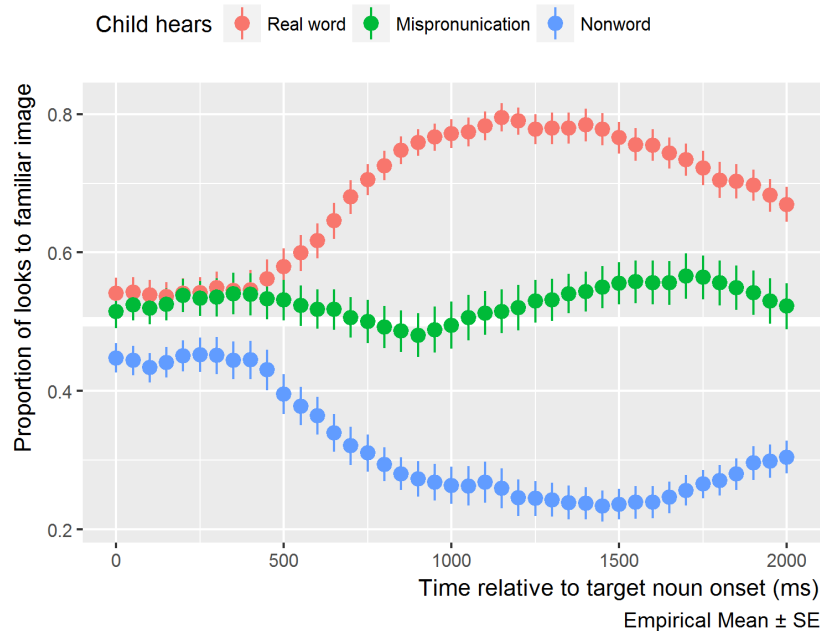


Figure 8.2: Condition averages for data described by Law and Edwards (2015). Compare to Figure 2 in the original manuscript.

I will test whether the two measures ever dissociate by computing the posterior predicted difference between the growth curves. This approach is similar to the bootstrap-based divergence analyses used in some word recognition experiments (e.g., Oleson et al., 2015; Dink and Ferguson, 2016). The essential question is when—at which specific time points—do two growth curves differ significantly from one another. The bootstrap approach uses resampling to get an estimate, whereas I will use posterior predicted samples to estimate these differences. (I have not seen my approach used yet in the literature, so it is a small innovation.)

Specifically, I will compute the posterior-predicted looks to the familiar object in the real word condition, $P(\text{Familiar} \mid \text{Real Word}, \text{Time } t, \text{Child } i)$ and the analogous looks to the unfamiliar object in the nonword condition, $P(\text{Unfamiliar} \mid \text{Nonword}, \text{Time } t, \text{Child } i)$. The difference between these two probabilities estimates how the time course of word recognition differs between these two conditions, and I can use 50% and 90% uncertainty intervals to determine during which time points the curves credibly differ from each other. Figure 8.3 shows this calculation performed on data from Law and Edwards (2015). If feasible, I will also examine whether these measures dissociate *within* children and examine which child-level factors are associated with

these kinds of listeners.

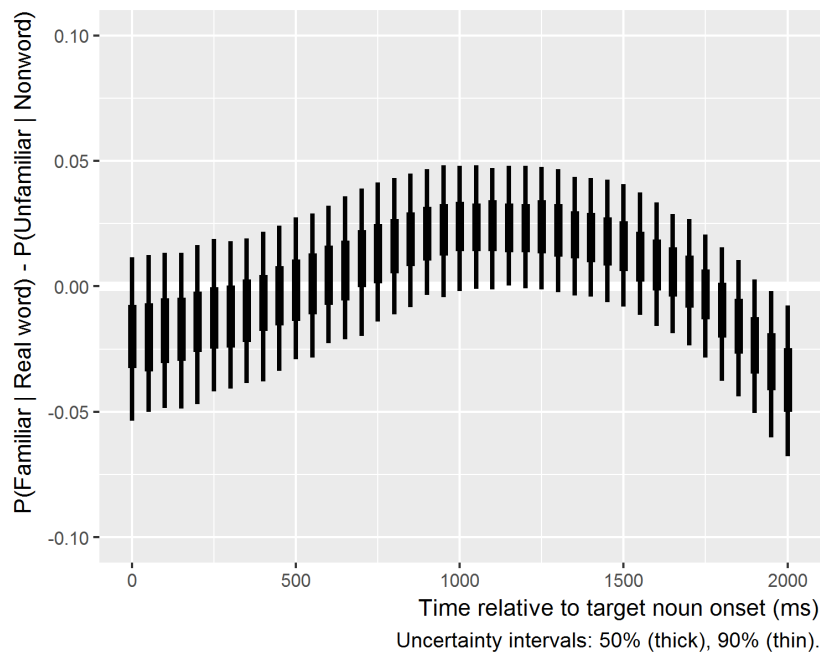


Figure 8.3: Demonstration of posterior difference technique on data from Law and Edwards (2015).

Even though performance on the real word and nonword conditions might be highly correlated, one might intuitively hypothesize that that performance on the nonword condition to be a better predictor of concurrent or future vocabulary size. The rationale would be that referent selection for novel words is a more transparent test of the word learner’s basic task of associating new labels with objects. Therefore, I will examine how each of these measures relates to vocabulary growth.

I will describe how looking behavior in the mispronunciation condition changes over time and changes for specific mispronunciation patterns. Overall, I predict that children will be more tolerant of mispronunciations as they age, because older children know more words and have more implicit knowledge about the similarities among words. As for specific mispronunciation items, let us (safely) suppose that speech perception improves with age, especially for later mastered sounds. Then we should expect that looking patterns for the *rice-wice* trials change significantly between Year 2 and Year 3, at least compared to looking patterns on trials with mispronunciations of earlier acquired sounds (e.g., *girl-dirl* or *duck-guck*). Therefore, I will examine individual mispronunciation effects and how they are associated with child-level measures, including speech perception.

8.4 Aim 3 (Computational Modeling)

8.4.1 TRACE Model Architecture

TRACE (McClelland and Elman, 1986) is an interactive activation model, and it interprets an input pattern by spreading energy (activation) through a network of processing units. The pattern of activation over the network is its interpretation of the input signal, so that more active units represent more likely interpretations. Over many processing cycles, the network propagates energy among its connections until it settles into a stable pattern of activation.

The input for TRACE is a mock-speech signal that activates perceptual feature-detectors. These units respond to phonetic features like voicing or vocalic resonance. The perceptual units activate phoneme units, and the phoneme units activate lexical word units, as shown in Figure 8.4.

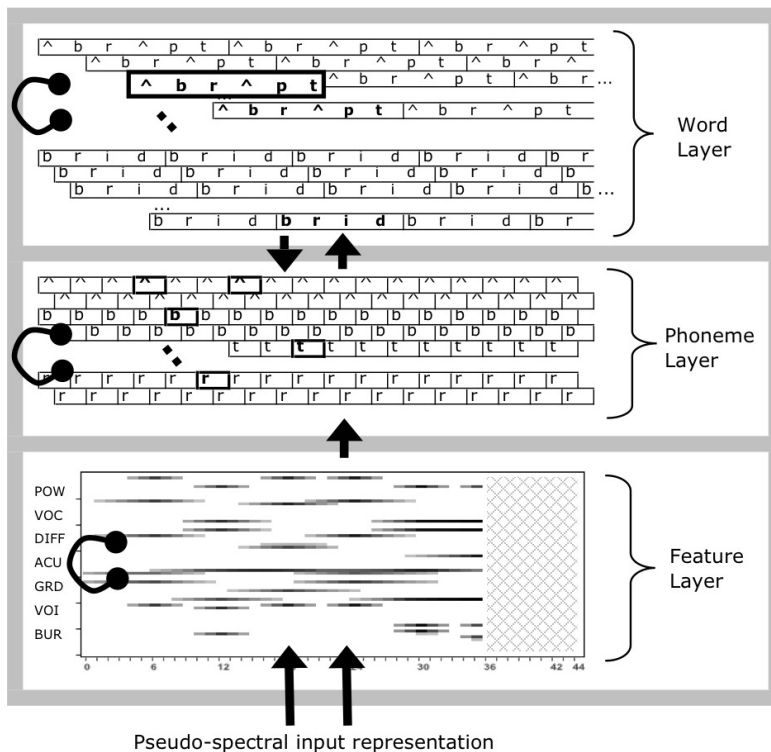


Figure 8.4: TRACE model architecture. Thick arrows indicate excitatory connections between layers, including a top-down connections from words onto phoneme units. Lines with points at the end reflect inhibitory connections among competing units within a layer. This image is a lightly modified public-domain version of Figure 1 in Strauss et al. (2007): https://en.wikipedia.org/wiki/File:TRACE_architecture.jpg

Units within a level (phonetic, phonemic, lexical) compete through lateral inhibition, so that more active units can suppress less active units. This inhibition allows the network to rule out possible units and narrow its interpretations over time. There are also top-down connections so that word units in the lexical layer can reinforce the sound

units that make up those words. One consequence of this feature is that the network can resolve ambiguous phonemes, as in *Xift* where *X* is a sound between /k/ and /g/ and top-down influence supports *gift* rather than *kift*. Lastly, activation in units gradually decays over time, and the network will eventually “forget” its input pattern to return to a resting state.

The model parameters that govern how activation propagates through the network map onto psychological processing constructs. For instance, the phoneme inhibition strength parameter controls how decisively (or categorically) the network interprets speech sounds. Phoneme-to-word activation strength reflects how quickly speech sounds begin to activate words. Decay parameters control the model’s temporary memory for different kinds of representations.

8.4.2 Modeling Looking Data

To simulate behavioral data, we need a *linking hypothesis* for translating between human behavior and network behavior (Magnuson et al., 2012). During each network cycle, some words are activated, some more activated than others, so that the unit with the highest activation is the preferred interpretation of the input. We can convert these activations into probabilities using the *softmax* function:

$$P(\text{word}_w) = \text{softmax}(\text{activation}_w) = \frac{\exp(k * \text{activation}_w)}{\sum \exp(k * \text{activation}_{[\text{words to choose from}]})}$$

In other words, scale the activation using some parameter k , exponentiate the scaled activation values of all relevant choices, and the proportion of the total exponentiated activation belonging to word w is the probability of fixating on word w .⁵ The scaling value k is manipulated to help the model-simulated probabilities match the observed looking probabilities.

The next step, and this process will be rather open-ended and iterative, is to simulate the behavioral data with the model. As a first step, I will need to create a developmentally appropriate lexicon. I will use developmental norms and databases to determine an appropriate set of items. I will first try to simulate the mispronunciation experiment. Mayor and Plunkett (2014) successfully simulated some mispronunciation data from experiments with data, so I will borrow some of their strategies for modeling these kinds of experiments. After I get the model to simulate Year 1 data, I will explore which parameters need to change to account for Year 2 and then Year 3 data. There are many degrees of freedom (model parameters) for how to replicate these development changes, so I will plan to systematically report the model fitting and the development consequences of the model fit.

⁵ Some terminology trivia: This linking function is usually ascribed to Luce (2008; 1959): “The Luce choice rule is the standard method of linking model activations to behavioral responses (e.g., McClelland & Elman, 1986), including fixation behavior measured in human participants” (Mirman et al., 2011, p. 62). From my reading, it seems that the mathematical function is the softmax, but the Luce rule is a trick that let us ignore irrelevant choices: “An important property of the softmax function [...] is known as independence from irrelevant attributes (Luce, 1959, 2008). The model implies that the ratio of probabilities of two outcomes is the same regardless of what other possible outcomes are included in the set” (Kruschke, 2015, p. 654). In other words, to model a 4-alternative forced choice task, we only need to consider the activation of the four relevant alternatives.

The four-image experiment is slightly trickier, because it incorporates semantically related information, and TRACE has no built-in semantic representations. Two options for modeling the experiment are 1) to ignore looks to the semantic foil, following the Luce choice rule (2008; 1959) and 2) to try to incorporate semantic information into the model by modifying connections between semantically related words—although it is unclear whether either is tractable. At any rate, part of this project will be to explore how to reconcile these data with this apparent limitation with the TRACE model.

Finally, I would like to account for individual differences among children and incorporate child-level information into the simulations. Specifically, I would like to test how vocabulary differences are associated with large lexicons for model simulations. Moreover, I would like to whether children’s speech perception abilities systematically relate to the model’s phonological activation parameters. Such a correspondence would further validate the models. These simulations would further validate these child-level measures by describing how they specifically affect word recognition.

For these research questions, I will following the modeling heuristics described by Magnuson et al. (2012). For example, for any modeling failures, I will assess the failure in terms of the computational consequences: Is the failure a problem with the theory, implementation, parameters or link between human data and model activity? Similarly, can the model fit specific item effects or just condition effects? The authors provide other heuristics, and these guidelines heuristics will help formalize the assessment and comparison of computational models.

A

Items used in the mispronunciation experiment

The stimuli changed between Year 1 and Year 2 so that dog/tog was replaced with rice/wice.

| Time Points | Word Group | Condition | Audio (IPA) | Familiar Object | Unfamiliar Object |
|-------------|------------|--------------------|-------------|-----------------|-------------------|
| 1 | dog | Correct Production | /d g/ | dog | wombat |
| | | Mispronunciation | /t g/ | dog | wombat |
| | | Nonword | /vef/ | ball | sextant |
| 1, 2, 3 | cake | Correct Production | /kek/ | cake | horned melon |
| | | Mispronunciation | /gek/ | cake | horned melon |
| | | Nonword | /p m/ | book | churn |
| 1, 2, 3 | duck | Correct Production | /d k/ | duck | toy creature |
| | | Mispronunciation | /g k/ | duck | toy creature |
| | | Nonword | / æn/ | cup | reed |
| 1, 2, 3 | girl | Correct Production | /g l/ | girl | marmoset |
| | | Mispronunciation | /d l/ | girl | marmoset |
| | | Nonword | /ned / | car | work holder |
| 1, 2, 3 | shoes | Correct Production | / uz/ | shoes | flasks |
| | | Mispronunciation | /suz/ | shoes | flasks |
| | | Nonword | /giv/ | sock | trolley |
| 1, 2, 3 | soup | Correct Production | /sup/ | soup | steamer |
| | | Mispronunciation | / up/ | soup | steamer |
| | | Nonword | /t im/ | bed | pastry mixer |
| 2, 3 | rice | Correct Production | / a s/ | rice | anise |
| | | Mispronunciation | /wa s/ | rice | anise |
| | | Nonword | /bep/ | ball | sextant |

B

Items used in the visual world experiment

Each row of the table represents a set of four images used in a trial for the experiment. There were two blocks of trials with different images and trial orderings. For the two unrelated foils with more than one word listed, the first word was used in block one and the second in block 2.

| Target | Phonological | Semantic | Unrelated |
|--------|--------------|----------|-----------|
| bear | bell | horse | ring |
| bee | bear | fly | heart |
| bell | bee | drum | swing |
| bread | bear | cheese | vase |
| cheese | shirt | bread | van |
| dress | drum | shirt | swing |
| drum | dress | bell | sword |
| flag | fly | kite | pear |
| fly | flag | bee | pen |
| gift | kite | vase | bread |
| heart | horse | ring | bread/pan |
| horse | heart | bear | pan |
| kite | gift | flag | shirt |
| pan | pear | spoon | vase |
| pear | pen | cheese | ring/vase |
| pen | pear | sword | van |
| ring | swing | dress | flag |
| shirt | cheese | dress | fly |
| spoon | swan | pan | drum |
| swan | spoon | bee | bell |
| swing | spoon | kite | heart |
| sword | swan | pen | gift |
| van | pan | horse | sword |
| vase | van | gift | swan |

C

Related Work

In this section, I clarify relationships between this project and other word recognition research reported from our lab. In short, our lab has reported results about the two-image and four-image experiments from cross-sectional samples, describing child-level measures that predict performance in these tasks. In contrast, my dissertation 1) focuses on the longitudinal development of word recognition and 2) engages with the fine-grained details of lexical processing.

Law and Edwards (2015) analyzed a different version of the mispronunciation experiment on a different sample of children ($n = 34$, 30-46 months old). This earlier version included both real word and the mispronunciation of the real word in the same block of trial. For example, a child would hear “dog” and “tog” during the same session of the experiment. This design might subtly temper the effect of mispronounced stimuli by allowing the listener to compare the mispronunciation to its correctly produced counterpart. The version of the experiment in Specific Aim 2 separates the real words and mispronunciations so that a child never hears a familiar word and its mispronunciation during the same block of trials. With this design, there is no explicit point of comparison for the mispronunciation, and the child has to rely on his or her own lexical representations when processing these words.

Law et al. (2016) analyzed data from the four-image experiment in Specific Aim 1. This study featured a diverse cross-sectional sample of 60 children, half of whom received the experiment in African American English and half received it in Mainstream American English. The sample ranged in age from 28 to 60 months. The study “borrowed” data from 23 participants from Year 1 of the longitudinal study to enrich parts of the samples demographics. For this manuscript, we analyzed how vocabulary and maternal education predicted looking patterns, including relative looks to the semantic and phonological foils.

Mahr and Edwards (under review) was the manuscript I originally

authored for my preliminary examinations. The paper analyzes the same kinds of relations as Weisleder and Fernald (2013) which showed that lexical processing efficiency mediated the effect of language input on future vocabulary size. In particular, I asked whether word recognition performance on the four-image task of Specific Aim 1, vocabulary size, and home language input data from Year 1 predicted vocabulary size at Year 2. The paper only examined looks to the familiar image from one year of the study, so it did not analyze any lexical competition effects or the development of word recognition within children.

D

Bibliography

- Alloppenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4):419–439.
- Baylis, A. L., Munson, B., and Moller, K. T. (2008). Factors affecting articulation skills in children with velocardiofacial syndrome and children with cleft palate or velopharyngeal dysfunction: A preliminary report. *The Cleft Palate-Craniofacial Journal*, 45(2):193–207.
- Cristia, A., Seidl, A., Junge, C., Soderstrom, M., and Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child Development*, 85(4):1330–1345.
- Dink, J. and Ferguson, B. (2016). *eyetrackingR*. R package version 0.1.6.
- Dunn, L. M. and Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition*. Pearson Education, San Antonio, TX.
- Ellis Weismer, S., Haebig, E., Edwards, J. R., Saffran, J. R., and Venker, C. E. (2016). Lexical processing in toddlers with ASD: Does weak central coherence play a role? *Journal of Autism and Developmental Disorders*, 46(12):3755–3769.
- Fernald, A. and Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development*, 83(1):203–222.
- Fernald, A., Swingle, D., and Pinto, J. P. (2001). When half a word is enough: infants can recognize spoken words using partial phonetic information. *Child development*, 72(4):1003–15.

- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York.
- Goldman, R. and Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation, Second Edition*. Pearson, Minneapolis, MN.
- Hart, B. and Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. P.H. Brookes, Baltimore.
- Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74(5):1368–1378.
- Huang, Y. T. and Snedeker, J. (2011). Cascading activation across levels of representation in children’s lexical processing. *Journal of Child Language*, 38(03):644–661.
- Kapnoula, E. C., Packard, S., Gupta, P., and McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition*, 134:85–99.
- Kruschke, J. K. (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press, Amsterdam, 2 edition.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493):979–1000.
- Law, II, F. and Edwards, J. R. (2015). Effects of vocabulary size on online lexical processing by preschoolers. *Language Learning and Development*, 11(4):331–355.
- Law, II, F., Mahr, T., Schneeberg, A., and Edwards, J. R. (2016). Vocabulary size and auditory word recognition in preschool children. *Applied Psycholinguistics*.
- Lew-Williams, C. and Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18(3):193–8.
- Luce, R. D. (1959). *Individual choice behavior*. Wiley, New York.
- Luce, R. D. (2008). Luce’s choice axiom. *Scholarpedia*, 3(12):8077. revision #121550.

- Magnuson, J. S., Mirman, D., and Harris, H. D. (2012). Computational models of spoken word recognition. In Spivey, M. J., McRae, K., and Joanisse, M. F., editors, *The Cambridge Handbook of Psycholinguistics*, Cambridge Handbooks in Psychology, chapter 5, pages 76–103. Cambridge University Press, Cambridge.
- Magnuson, J. S., Mirman, D., and Myers, E. (2013). Spoken word recognition. In Reisberg, D., editor, *The Oxford Handbook of Cognitive Psychology*. Oxford University Press.
- Mahr, T., McMillan, B. T. M., Saffran, J. R., Ellis Weismer, S., and Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, 142:345–350.
- Mani, N., Durrant, S., and Floccia, C. (2012). Activation of phonological and semantic codes in toddlers. *Journal of Memory and Language*, 66(4):612–622.
- Mani, N. and Plunkett, K. (2010). In the infant’s mind’s ear: evidence for implicit naming in 18-month-olds. *Psychological Science*, 21(7):908–913.
- Marchman, V. A. and Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, 11(3):F9–16.
- Marslen-Wilson, W. D. and Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3):576–585.
- Mayor, J. and Plunkett, K. (2014). Infant word recognition: Insights from TRACE simulations. *Journal of Memory and Language*, 71(1):89–123.
- McClelland, J. L. and Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1):1–86.
- McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press/Taylor & Francis Group, Boca Raton, FL.
- McMurray, B., Horst, J. S., and Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, 119(4):831–877.
- McMurray, B., Samelson, V. M., Lee, S. H., and Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, 60(1):1–39.

- Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Chapman & Hall/CRC, Boca Raton, FL.
- Mirman, D., Yee, E., Blumstein, S. E., and Magnuson, J. S. (2011). Theories of spoken word recognition deficits in Aphasia: Evidence from eye-tracking and computational modeling. *Brain and Language*, 117(2):53–68.
- Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., and Maczuga, S. (2015). 24-Month-Old Children With Larger Oral Vocabularies Display Greater Academic and Behavioral Functioning at Kindergarten Entry. *Child Development*, 86(5):1351–1370.
- Nosek, B. A., Simonsohn, U., Moore, D. A., Nelson, L. D., Simmons, J. P., Sallans, A., and LeBel, E. P. (2014). Standard reviewer statement for disclosure of sample, conditions, measures, and exclusions.
- Oleson, J. J., Cavanaugh, J. E., McMurray, B., and Brown, G. (2015). Detecting time-specific differences between temporal nonlinear curves: Analyzing data from the visual world paradigm. *Statistical methods in medical research*.
- Rvachew, S. (2006). Longitudinal predictors of implicit phonological awareness skills. *American Journal of Speech-Language Pathology*, 15(2):165–176.
- Strauss, T. J., Harris, H. D., and Magnuson, J. S. (2007). jTRACE: A reimplement and extension of the TRACE model of speech perception and spoken word recognition. *Behavior Research Methods*, 39(1):19–30.
- Swingle, D. and Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76(2):147–66.
- Swingle, D. and Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13(5):480–484.
- Swingle, D., Pinto, J. P., and Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, 71(2):73–108.
- Tsao, F.-M., Liu, H.-M., and Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Development*, 75(4):1067–1084.
- Wagner, R., Torgesen, J., Rashotte, C., and Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing, Second Edition*. PRO-ED, Austin, TX.

- Weisleder, A. and Fernald, A. (2013). Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychological science*, 24(11):2143–52.
- White, K. S. and Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, 59(1):114–132.
- Williams, K. T. (2007). *Expressive Vocabulary Test, Second Edition*. Pearson Education, San Antonio, TX.