Development of word recognition in preschoolers

Tristan Mahr

Draft of manuscript for Specific Aim 1 (2018-05-16)

Abstract

[*draft*] Vocabulary size in preschool is a robust predictor of later language development, and early language skills predict early literacy skills at school entry. 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 in toddlers 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 develops from age 3 to age 5.

Contents

[Abstract i](#_Toc514311513)

[**Prospectus**1](#_Toc514311514)

[Chapter 1: Dissertation Information 2](#_Toc514311515)

[Planned Dissertation Format 2](#_Toc514311516)

[Committee Members 2](#_Toc514311517)

[Miscellany 2](#_Toc514311518)

[Chapter 2: Specific Aims 3](#_Toc514311519)

[2.1 Specific Aim 1 (Familiar Word Recognition and Lexical Competition) 4](#_Toc514311520)

[2.2 Specific Aim 2 (Referent Selection and Mispronunciations) 4](#_Toc514311521)

[2.3 Summary 5](#_Toc514311522)

[Chapter 3: Research Hypotheses 6](#_Toc514311523)

[**Aim 1: Familiar Word Recognition and Lexical Competition** 8](#_Toc514311526)

[Chapter 4: Familiar Word Recognition 9](#_Toc514311527)

[4.1 Lexical processing dynamics 9](#_Toc514311528)

[4.2 Individual differences in word recognition 12](#_Toc514311529)

[4.3 The current study 13](#_Toc514311530)

[Chapter 5: Method 14](#_Toc514311531)

[5.1 Participants 14](#_Toc514311532)

[5.2 Visual World Paradigm 15](#_Toc514311533)

[5.3 Experiment administration 16](#_Toc514311534)

[5.4 Stimuli 18](#_Toc514311535)

[5.5 Data screening 19](#_Toc514311536)

[5.6 Model preparation 20](#_Toc514311537)

[Chapter 6: Analysis of Familiar Word Recognition 22](#_Toc514311538)

[6.1 Growth curve analysis 22](#_Toc514311539)

[6.1.1 Growth curve features as measures of word recognition performance 24](#_Toc514311540)

[6.2 Year over year changes in word recognition performance 25](#_Toc514311541)

[6.3 Exploring plausible ranges of performance over time 28](#_Toc514311542)

[6.4 Are individual differences stable over time? 32](#_Toc514311543)

[6.5 Predicting future vocabulary size 34](#_Toc514311544)

[6.6 Discussion 36](#_Toc514311545)

[Chapter 7: Effects of Phonological and Semantic Competitors 39](#_Toc514311546)

[7.1 Looks to the phonological competitor 39](#_Toc514311547)

[7.2 Looks to the semantic competitor 47](#_Toc514311548)

[7.3 Child-level differences in competitor sensitivity at age 3 50](#_Toc514311549)

[Summary. 52](#_Toc514311550)

[7.4 Discussion 52](#_Toc514311551)

[7.4.1 Immediate activation of phonological neighbors 52](#_Toc514311552)

[7.4.2 Late activation of semantic neighbors 54](#_Toc514311553)

[7.4.3 Lexical competitors and child-level predictors 56](#_Toc514311554)

[Chapter 8: General Discussion 58](#_Toc514311555)

[8.1 How to improve word recognition 58](#_Toc514311556)

[8.2 Learn words and learn connections between words 61](#_Toc514311557)

[8.3 Individual differences are most important at younger ages 61](#_Toc514311558)

[8.4 Limitations and implications 63](#_Toc514311559)

[**Appendices** 66](#_Toc514311560)

[Appendix A: Items Used in the Visual World Experiment 67](#_Toc514311561)

[Appendix B: Computational Details for Specific Aim 1 69](#_Toc514311562)

[B.1 Growth curve analyses 69](#_Toc514311563)

[B.2 Generalized additive models 72](#_Toc514311564)

[Appendix C: Related Work 74](#_Toc514311565)

[References 76](#_Toc514311566)

Prospectus

Chapter 1: Dissertation Information

## Planned Dissertation Format

The dissertation will consist of two thematically and empirically related manuscripts to be completed Summer 2018. These two manuscripts will serve as the main two *parts* of this book, and each of the conventional manuscript sections (introduction, methods, etc.) will serve as chapters within those two parts.

## Committee Members

Jan Edwards, primary advisor and chair  
Department of Hearing and Speech Sciences, University of Maryland

Susan Ellis Weismer, advisor at UW-Madison  
Department of Communication Sciences and Disorders

Margarita Kaushanskaya  
Department of Communication Sciences and Disorders

Jenny Saffran  
Department of Psychology

David Kaplan  
Department of Educational Psychology

* Bob McMurray  
  Department of Psychological and Brain Sciences, University of Iowa

## Miscellany

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

Web version of the dissertation: <https://tjmahr.github.io/dissertation/>

* Research compendium with data, scripts and source documents: <https://github.com/tjmahr/dissertation>

Chapter 2: 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 & Fernald, [2013](#ref-Weisleder2013)).

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 a 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 other child-level measures (such as vocabulary and speech perception).

## 2.1 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 develops year over year. Of interest is how individual differences at age 3 persist into age 5. I will also 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.

## 2.2 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 (Law & Edwards, [2015](#ref-MPPaper); based on White & Morgan, [2008](#ref-WhiteMorgan2008)), 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 the child’s phonological categories by showing 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, Horst, & Samuelson, [2012](#ref-McMurray2012)). Finally, I will examine how individual differences in vocabulary and speech perception predict responses to mispronunciations and novel words.

## 2.3 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 longitudinally 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.

Chapter 3: 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.

## 3.1 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 age 3 remain relatively fast at age 4 and age 5.

However, the magnitude of these individual differences diminishes over time, as children converge on a mature level of performance for this paradigm.

Consequently, 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 or age 5.

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.

## 3.2 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 predict 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).

Aim 1: Familiar Word Recognition and Lexical Competition

Chapter 4: Familiar Word Recognition

## 4.1 Lexical processing dynamics

Mature listeners recognize spoken words by continuously evaluating incoming speech for possible word matches. The first part of a word activates multiple candidate words in parallel. These candidates compete as more of the speech signal enters the system, and the best-fitting word is the favored interpretation. For example, the onset “bee” might activate phonologically compatible candidates like *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—for instance, where *castle* activates the phonologically similar *candy* which in turn activates the semantically related *sweet*—have been demonstrated (Marslen-Wilson & Zwitserlood, [1989](#ref-Marslen-Wilson1989)). Both low-level phonetic cues and high-level grammatical, semantic and pragmatic information can influence this process, but this *continuous processing of multiple competing candidates* is the essential dynamic underlying word recognition in adults (Magnuson, Mirman, & Myers, [2013](#ref-Magnuson2013)).

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, Pinto, & Fernald, [1999](#ref-Swingley1999)), match truncated words to their intended referents (Fernald, Swingley, & Pinto, [2001](#ref-Fernald2001)), and use information from neighboring words in a sentence to facilitate word recognition. This information can be high-level grammatical or semantic cues. Lew-Williams and Fernald ([2007](#ref-Lew-Williams2007)) 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. Borovsky, Elman, and Fernald ([2012](#ref-Borovsky2012)) showed that children can use semantic information from an agent and a verb (e.g., *the dog chased*) to anticipate a plausible noun (*the cat*). The information can also be low-level phonetic variation: We found that toddlers look earlier to a named image when the coarticulatory formant cues on word *the* predicted the noun of the sentence, compared to tokens with neutral coarticulation (Mahr, McMillan, Saffran, Ellis Weismer, & Edwards, [2015](#ref-Mahr2015)).

There is some evidence for lexical competition where children are sensitive to phonological and semantic similarities among words. Ellis Weismer, Haebig, Edwards, Saffran, and Venker ([2016](#ref-EllisWeismer2016)) showed that toddlers (14–29 months old) look less reliably to a named image when the onscreen competitor was a semantically related word or perceptually similar image. Huang and Snedeker ([2011](#ref-Huang2011)) presented evidence of cascading semantic-phonological activation in five-year-olds such that for a target word like *log*, the 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*.

Priming studies also reveal that children are sensitive to phonological similarities among words. Mani and Plunkett ([2010](#ref-Mani2010)) 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 (cup and shoe) were presented, one of which was named (*cup*). Children on average looked more to the target word (*cup*) when it was primed by an image of a phonological neighbor (*cat*), and the children performed at chance when the prime was not related to the named word. Mani, Durrant, and Floccia ([2012](#ref-Mani2012)) found a similar result for cascading phonological-semantic priming with 24-month-olds: Children looked more to a target (e.g., *shoe*) compared to a distractor (*door*) when primed by an image (*clock*, assumed to activate *sock* which primed *shoe*).[[1]](#footnote-1)

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 develops within children.

One open question is how lexical competition develops in young listeners. For example, how and when do phonological or semantically similar words exert their influence on word recognition? 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 they inhibit one another, so that a word is truly “learned” (integrated into the lexicon) when it can influence the processing of other words (a line of reasoning reviewed by Kapnoula, Packard, Gupta, & McMurray, [2015](#ref-Kapnoula2015)). Increasing sensitivity to similar sounding or similar meaning words over time would reveal that children more thoroughly learn familiar words with age.

## 4.2 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 and prospectively.

Many studies highlight the predictive power of word recognition ability. Marchman and Fernald ([2008](#ref-MarchmanFernald2008)) found that vocabulary size and lexical processing efficiency at age 2 jointly predicted working memory scores and expressive language scores at age 8. Fernald and Marchman ([2012](#ref-Fernald2012)) 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](#ref-Weisleder2013)) found that lexical processing and language input at 19 months predicted vocabulary size at 25 months and that lexical processing mediated the effect of language input—the basic idea being that rich language input builds up word recognition ability which in turn supports word learning. Lany ([2017](#ref-Lany2017)) found a direct link between lexical processing and word learning: 18-month-olds and 30-month-olds who were faster at recognizing familiar words were also more accurate at recognizing novel words in a word-learning task. Thus, children who are better at recognizing words learn more words over time and perform better at word-learning tasks.

Word recognition performance predicts future language outcomes, so we conclude that individual differences in word recognition are important. But we do not know how word recognition develops within children, so we have no context for evaluating these individual differences. Are these differences in lexical processing persistent over development? Is word recognition a skill where most children catch up and converge on a mature range of performance by a certain age?

## 4.3 The current study

In the previous two sections, I outlined two gaps in knowledge. The first is that we do not have a clear understanding of how the mechanisms underlying word recognition change in early childhood. We know that children show plenty of adult-like features of word recognition, but each of these findings is an isolated fact. What we need is a coherent set of facts that show how specific features of word recognition change with age. The second gap is that although we know that individual differences in word recognition are predictive of later outcomes, we do not have a developmental picture of these individual differences.

In this study, I tackle these two lines of research: The development of lexical competition effects and individual differences in familiar word recognition. I report the results of a longitudinal study of word recognition in preschoolers at age 3, age 4, and age 5. The study is described in detail in [Chapter](#aim1-method) [5](#aim1-method). Briefly stated, this experiment tested word recognition by presenting prompts like “find the horse” and recording children’s looks to an array of four images. The array of images included the target, a phonological competitor, a semantic competitor, and an unrelated image. In [Chapter](#fam-rec) [6](#fam-rec), I analyze the development patterns of familiar word recognition (looks to the target) and how individual differences change over time. In [Chapter](#lex-competitors) [7](#lex-competitors), I study how the phonological and semantic competitors influence word recognition and how the influence of the competitors changes over time. Finally, in [Chapter](#aim1-discussion) [8](#aim1-discussion), I link these two lines of research together and describe both sets of results in terms of lexical processing dynamics.

Chapter 5: Method

## 5.1 Participants

The data were collected as part of a three-year longitudinal study. For convenience, I refer to the three years as Age 3, Age 4, and Age 5, although the participants on average were three months younger than those nominal ages. In particular, the participants were 28–39 months-old at Age 3, 39–52 at Age 4, and 51–65 at Age 5. Approximately, 180 children participated at Age 3, 170 at Age 4, and 160 at Age 5. Of these children, approximately 20 were identified by their parents as late talkers. Prospective families were interviewed over telephone before participating in the study. Children were not scheduled for testing if a parent reported language problems, vision problems, developmental delays, or an individualized education program for the child. 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 [5.1](#tab:participant-info) summarizes the cohort of children in each year of testing. The numbers and summary statistics here are general, describing children who participated at each year, but whose data may have been 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. I disclose all data exclusions following guidelines by the Center for Open Science (Nosek et al., [2014](#ref-OSF_Statement)).

|  |  |  |  |
| --- | --- | --- | --- |
|  | Year 1 (Age 3) | Year 2 (Age 4) | Year 3 (Age 5) |
| 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) |

Table 5.1: 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.

## 5.2 Visual World Paradigm

This experiment used a version of the Visual World Paradigm for word recognition experiments (Law, Mahr, Schneeberg, & Edwards, [2016](#ref-RWLPaper)). 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 [5.1](#fig:sample-vw-screen) shows an example of a trial’s items. 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.

![Example display for the target bell with the semantic foil drum, the phonological foil bee, and the unrelated swing.](data:image/png;base64;base64,)

Figure 5.1: Example display for the target *bell* with the semantic foil *drum*, the phonological foil *bee*, and the unrelated *swing*.

## 5.3 Experiment administration

Children participating in the study were tested over two lab visits (on different dates). The first portion of each visit involved “watching movies”—that is, performing two blocks of 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 this experiment and 38 for a two-image task—see Chapter X). Children received two different blocks of each experiment. The blocks for an experiment differed in trial ordering and other features. 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 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. First, the images appeared in silence on screen for 2 s as a familiarization period. The experiment then checked whether the child’s gaze was being recorded. If the experiment could continuously track the child’s gaze for 300 ms, the child’s gaze was verified and the trial continued. If the experiment could not verify the gaze after 10 s, the trial continued. This procedure guaranteed that for most trials, the child was looking to the display before presenting the carrier phrase and that the experiment was ready to record the child’s response to the carrier. During year 1 (age 3) and year 2 (age 4), 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 (age 5) to streamline the experiment for older listeners.

## 5.4 Stimuli

The four images on each trial consisted of a target noun, a phonological foil, a semantic foil, and an unrelated word. The phonological competitors shared a syllable onset (e.g., *flag*–*fly*, *bell*–*bee*), shared an initial consonant (*bread*–*bear*, *swing*–*spoon*), had a similar phonetically similar consonant onset (*kite*–*gift*), or shared a syllable rime (*van*–*pan*). The semantic competitors included words from the same category (e.g., *shirt*–*dress*, *horse*–*bear*), words that were perceptually similar (*sword*–*pen*, *flag*–*kite*), and words with less obvious relationships (*van*–*horse*, *swan*-*bee*). A complete list of the items used in the experiment in [Appendix](#vw-experiment-items) [A](#vw-experiment-items).

The stimuli were recorded in both Mainstream American English (MAE) and African American English (AAE), so that the experiment could accommodate the child’s home dialect. Prior to the lab visit, we made a preliminary guess about the child’s home dialect, based on the recruitment channel, address, among other factors. If we expected the dialect to be AAE, then the lab visit was led by an examiner who natively spoke AAE and could fluently dialect-shift between AAE and MAE. At the beginning of the lab visit, the examiner listened to the interactions between the child and caregiver in order to confirm the child’s home dialect. 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 recorded in the frame “find/see the egg” and cross-spliced with the target nouns to minimize coarticulatory cues on the determiner “the”. The stimuli were re-recorded after the first year of the study with the same speakers so that the average duration of the two dialect versions were more similar.

The images used in the experiment consisted of color photographs on gray backgrounds. These images were piloted with 30 children from two preschool classrooms to ensure that children consistently used the same label for familiar objects. The two preschool classrooms differed in their students’ SES demographics: One classroom (13 piloting students) was part of a university research center which predominantly serves higher-SES families, and the other classroom (17 piloting students) was part of Head Start center which predominantly serves lower-SES families. The images were tested by presenting four images (a target, a phonological foil, a semantic foil and an unrelated word) and having the student point to the named image. The pictures had to be recognized by at least 80% of students in each classroom.

## 5.5 Data screening

To process the eyetracking data, I first mapped gaze *x*-*y* coordinates onto the onscreen images. I next performed *deblinking*. I interpolated short runs of missing gaze data (up to 150 ms) if the same image was fixated before and after the missing data run. Put differently, I classified a window of missing data as a blink if the window was brief and the gaze remained on the same image before and after the blink. I interpolated missing data from blinks using the fixated image.

After mapping the gaze coordinates onto the onscreen images, I performed data screening. I considered the time window from 0 to 2000 ms after target noun onset. I identified a trial as *unreliable* if at least 50% of the looks were missing during the time window. I excluded an entire block of trials if it had fewer than 12 reliable trials.

Table [5.2](#tab:screening-counts) shows the numbers of participants and trials at each year before and after data screening. There were more children in the second year than the first due to a timing error in the initial version of this experiment, leading to the exclusion of 27 participants from the first year.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Study | Children | Blocks | Trials |
| Raw | Age 3 | 178 | 332 | 7967 |
|  | Age 4 | 180 | 347 | 8327 |
|  | Age 5 | 163 | 322 | 7724 |
| Screened | Age 3 | 163 | 291 | 5951 |
|  | Age 4 | 165 | 305 | 6421 |
|  | Age 5 | 156 | 295 | 6483 |
| Raw − Screened | Age 3 | 15 | 41 | 2016 |
|  | Age 4 | 15 | 42 | 1906 |
|  | Age 5 | 7 | 27 | 1241 |

Table 5.2: Eyetracking data before and after data screening. For convenience, the number of exclusions is included as Raw − Screened.

## 5.6 Model preparation

To prepare the data for modeling, I downsampled the data into 50-ms (3-frame) bins, reducing the eyetracker’s effective sampling rate to 20 Hz. Eye movements have durations on the order of 100 or 200 ms, so capturing data every 16.67 ms oversamples eye movements and can introduce high-frequency noise into the signal. Binning together data from neighboring frames can smooth out this noise. I modeled the looks from 250 to 1500 ms. Lastly, I aggregated looks by child, study and time, and created orthogonal polynomials to use as time features for the model.

Figure [5.2](#fig:aim1-spaghetti) depicts the dataset following the data screening and model preparation steps. The lines start around .25 which is chance performance on four-alternative forced choice task. The lines rise as the word unfolds, and they peak and plateau around 1400 ms.

![Empirical word recognition growth curves from each year of the study. Each line represents an individual child’s proportion of looks to the target image over time. The heavy lines are the averages of the lines for each year.](data:image/png;base64;base64,)

Figure 5.2: Empirical word recognition growth curves from each year of the study. Each line represents an individual child’s proportion of looks to the target image over time. The heavy lines are the averages of the lines for each year.

Chapter 6: Analysis of Familiar Word Recognition

## 6.1 Growth curve analysis

Looks to the familiar image were analyzed using Bayesian mixed effects logistic regression. I used *logistic* regression because the outcome measurement is a probability (the log-odds of looking to the target image versus a distractor). I used *mixed-effects* models 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 used *Bayesian* techniques to study a generative model of the data. Instead of reporting and describing a single, best-fitting model of some data, Bayesian methods 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, one 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).[[2]](#footnote-2)

The eyetracking growth curves were fit using an orthogonal cubic polynomial function of time (a now-conventional approach; see Mirman, [2014](#ref-Mirman2014)). Put differently, I modeled the probability of looking to the target during an eyetracking task as:

That the time terms are *orthogonal* means that Time1, Time2 and Time3 are transformed so that they are uncorrelated. Under this formulation, the parameters *β*0 and *β*1 have a direct 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 the intercept as a measure 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 the linear time term as a measure of *processing efficiency*, because growth curves with stronger linear features exhibit steeper frame-by-frame increases in looking probability.[[3]](#footnote-3)

To study how word recognition changes over time, I modeled how the growth curves change over developmental time. This amounted to studying how the growth curve parameters changes year over year. I included dummy-coded indicators for Age 3, Age 4, and Age 5 and allowed these indicators to interact with the growth curve parameters. These year-by-growth-curve-feature terms captured how the shape of the growth curves changed each year. The model also included random effects to represent by-child and by-child-by-year effects to estimate a general growth curve for each child and to estimate how each child’s growth curve changed each year.

The models were fit in R (vers. 3.4.3) with the RStanARM package (vers. 2.16.3). [Appendix](#aim1-gca-models) [B](#aim1-gca-models) contains the R code used to fit the model along with a description of the model specifications represented in the model syntax.

## 6.1.1 Growth curve features as measures of word recognition performance

As mentioned above, two of the model’s growth curve features have straightforward interpretations in terms of lexical processing performance: The model’s intercept parameter corresponds to the average proportion or probability of looking to the named image over the trial window, and the linear time parameter corresponds to slope of the growth curve or lexical processing efficiency. I also was interested in *peak* proportion of looks to the target. I derived this value by computing the growth curves from the model and taking the median of the five highest points on the curve. Figure [6.1](#fig:curve-features) shows three simulated growth curves and how each of these growth curve features relate to word recognition performance.

![Illustration of the three growth curve features and how they describe lexical processing performance. The three curves used are simulations of new participants at Age 4.](data:image/png;base64;base64,)

Figure 6.1: Illustration of the three growth curve features and how they describe lexical processing performance. The three curves used are simulations of new participants at Age 4.

## 6.2 Year over year changes in word recognition performance

The mixed-effects model estimated a population-average growth curve (“fixed” effects) and how individual children deviated from average (“random” effects). Figure [6.2](#fig:average-growth-curves) shows 200 posterior samples of the average growth curves for each study. On average, the growth curves become steeper and achieve higher looking probabilities with each year of the study.

![The model estimated an average word recognition growth for each study, and the colored lines represent 200 posterior samples of these growth curves. The thick dark lines represent the observed average growth curve in each study.](data:image/png;base64;base64,)

Figure 6.2: The model estimated an average word recognition growth for each study, and the colored lines represent 200 posterior samples of these growth curves. The thick dark lines represent the observed average growth curve in each study.

Figure [6.3](#fig:effects2) depicts uncertainty intervals with the model’s average effects of each timepoint on the growth curve features. The intercept and linear time effects increased each year, confirming that children become more reliable and faster at recognizing words as they grow older. The peak probability also increased each year. For each effect, the change from age 3 to age 4 is approximately the same as the change from age 4 to age 5, as illustrated in Figure [6.4](#fig:pairwise-effects).

![Uncertainty intervals for the effects of study years on growth curve features. The intercept and peak features were converted from log-odds to proportions to ease interpretation.](data:image/png;base64;base64,)

Figure 6.3: Uncertainty intervals for the effects of study years on growth curve features. The intercept and peak features were converted from log-odds to proportions to ease interpretation.

![Uncertainty intervals for the differences between study timepoints. Again, the intercept and peak features were converted to proportions.](data:image/png;base64;base64,)

Figure 6.4: Uncertainty intervals for the differences between study timepoints. Again, the intercept and peak features were converted to proportions.

The average looking probability (intercept feature) was 0.38 [90% UI: 0.37–0.40] at age 3, 0.49 [0.47–0.50] at age 4, and 0.56 [0.54–0.57] at age 5. The averages increased by 0.10 [0.09–0.11] from age 3 to age 4 and by 0.07 [0.06–0.09] from age 4 to age 5. The peak looking probability was 0.55 [0.53–0.57] at age 3, 0.68 [0.67–0.70] at age 4, and 0.77 [0.76–0.78] at age 5. The peak values increased by 0.13 [0.11–0.16] from age 3 to age 4 and by 0.09 [0.07–0.10] from age 4 to age 5. These results numerically confirm the hypothesis that children would improve in their word recognition reliability, both in terms of average looking and in terms of peak looking, each year. The changes in peak probability were also rather large: children’s probability fixating on the target increased by approximately .1 each year. These growths indicate the task scaled with children’s development because they had room to improve each year.

**Summary.** The average growth curve features increased year over year, so that children looked to the target more quickly and more reliably as they grew older.

## 6.3 Exploring plausible ranges of performance over time

Bayesian models are generative; they describe how the data could have been generated. This model assumed that each child’s growth curve was drawn from a population of related growth curves, and it tried to infer the parameters over that distribution. These two aspects—a generative model and learning about the population of growth curves—allow the model to simulate new samples from that distribution of growth curves. That is, we can predict a set of growth curves for a hypothetical, unobserved child drawn from the same distribution as the 195 children observed in this study. This procedure of studying model implications by having the model generate new data is called *posterior predictive inference*, and in this case, it allows one to explore the plausible degrees of variability in performance at each age.

Figure [6.5](#fig:new-participants) shows the posterior predictions for 1,000 simulated participants, and it demonstrates how the model expects new participants to improve longitudinally but also exhibit stable individual features over time. Figure [6.6](#fig:new-participants-intervals) shows uncertainty intervals for these simulations. The model learned to predict less accurate and more variable performance at age 3 with improving accuracy and narrowing variability at age 4 and age 5.

![Posterior predictions for hypothetical unobserved participants. Each line represents the predicted performance for a new participant. The three dark lines highlight predictions from one single simulated participant. The simulated participant shows both longitudinal improvement in word recognition and similar relative performance compared to other simulations each year, indicating that the model would predict new children to improve year over year and show stable individual differences over time.](data:image/png;base64;base64,)

Figure 6.5: Posterior predictions for hypothetical *unobserved* participants. Each line represents the predicted performance for a new participant. The three dark lines highlight predictions from one single simulated participant. The simulated participant shows both longitudinal improvement in word recognition and similar relative performance compared to other simulations each year, indicating that the model would predict new children to improve year over year and show stable individual differences over time.

![Uncertainty intervals for the simulated participants. Variability is widest at age 3 and narrowest at age 5, consistent with the prediction that children become less variable as they grow older.](data:image/png;base64;base64,)

Figure 6.6: Uncertainty intervals for the simulated participants. Variability is widest at age 3 and narrowest at age 5, consistent with the prediction that children become less variable as they grow older.

I hypothesized that children would become less variable as they grew older and converged on a mature level of performance. To address this question, I inspected the ranges of predictions for the simulated participants. The claim that children become less variable would imply that the range of predictions should be narrower for age 5 than for age 4 and narrower for age 4 than for age 3. Figure [6.7](#fig:new-ranges) depicts the range of the predictions, both in terms of the 90 percentile range (i.e., the range of the middle 90% of the data) and in terms of the 50 percentile (interquartile) range. The ranges of performance decrease from age 3 to age 4 to age 5, consistent with the hypothesized reduction in variability.

![Ranges of predictions for simulated participants over the course of a trial. The ranges are most similar during the first half of the trial when participants are at chance performance, and the ranges are most different at the end of the trial as children reliably fixate on the target image. The ranges of performance decreases with each year of the study as children show less variability.](data:image/png;base64;base64,)

Figure 6.7: Ranges of predictions for simulated participants over the course of a trial. The ranges are most similar during the first half of the trial when participants are at chance performance, and the ranges are most different at the end of the trial as children reliably fixate on the target image. The ranges of performance decreases with each year of the study as children show less variability.

The developmental pattern of increasing reliability and decreasing variability was also observed for the growth curve peaks. For the synthetic participants, the model predicted that individual peak probabilities will increase each year, peak3 = 0.55 [90% UI: 0.35–0.77], peak4 = 0.69 [0.48–0.86], peak5 = 0.78 [0.59–0.91]. Moreover, the range of plausible values for the individual peaks narrowed each for the simulated data. For instance, the difference between the 95th and 5th percentiles was 0.43 for age 3, 0.38 for age 4, and 0.32 for age 5.

**Summary**. I used the model’s random effects estimates to simulate growth curves from 1,000 hypothetical, unobserved participants. The simulated dataset showed increasing looking probability and decreasing variability with each year of the study. These simulations confirmed the hypothesis that variability would diminish as children began to demonstrate a mature degree of performance for this task.

## 6.4 Are individual differences stable over time?

I predicted that children would show stable individual differences such that children who are faster and more reliable at recognizing words at age 3 remain relatively faster and more reliable at age 5. To evaluate this hypothesis, I used Kendall’s *W* (the coefficient of correspondence or concordance). This nonparametric statistic measures the degree of agreement among *J* judges who are rating *I* items. For these purposes, the items are the 123 children who provided reliable eyetracking for all three years of the study. (That is, I excluded children who only had reliable eyetracking data for one or two years.) The judges are the sets of growth curve parameters from each year of study. For example, the intercept term provides three sets of ratings: The participants’ intercept terms from year 1 are one set of ratings and the terms from years 2 and 3 provide two more sets of ratings. These three ratings are the “judges” used to compute the intercept’s *W*. Thus, I computed five groups of *W* coefficients, one for each set of growth curve features: intercept, Time1, Time2, Time3, and peak looking probability.

Because I used a Bayesian model, there is a distribution of ratings and thus a distribution of concordance statistics. Each sample of the posterior distribution fits a growth curve for each child in each study, so each posterior sample provides a set of ratings for concordance coefficients. This distribution of *W*’s lets us quantify our uncertainty because we can compute *W*’s for each of the 4000 samples from the posterior distribution.

One final matter is how to assess whether a concordance statistic is meaningful. To tackle this question, I also included a “null rater”, a fake parameter that assigned each child in each year a random number. I use the distribution of *W*’s generated by randomly rating children as a benchmark for assessing whether the other concordance statistics differ meaningfully from chance.

![Uncertainty intervals for the Kendall’s coefficient of concordance. Random ratings provide a baseline of null W statistics. The peak, intercept and linear time features are decisively non-null, indicating a significant degree of correspondence in children’s relative word recognition reliability and efficiency over the three years of the study.](data:image/png;base64;base64,)

Figure 6.8: Uncertainty intervals for the Kendall’s coefficient of concordance. Random ratings provide a baseline of null *W* statistics. The peak, intercept and linear time features are decisively non-null, indicating a significant degree of correspondence in children’s relative word recognition reliability and efficiency over the three years of the study.

I used the kendall() function in the irr R package (vers. 0.84; Gamer, Lemon, & Singh, [2012](#ref-irr)) to compute concordance statistics. Figure [6.8](#fig:kendall-stats) depicts uncertainty intervals for the Kendall *W*’s for these growth curve features. The 90% uncertainty interval of *W* statistics from random ratings, [.28–.39], subsumes the intervals for the Time2 effect [.30–.35] and the Time3 effect [.28–.35], indicating that these values do not differentiate children in a longitudinally stable way. Earlier, I claimed that only the intercept, linear time, and peak features have psychologically meaningful interpretations and that the higher-order time features mainly act to capture the curvature of the data. These null concordance statistics support that claim because the Time2 and Time3 features differentiate children across studies as well as random numbers.

Concordance is strongest for the peak feature, *W* = .59 [.57–.60] and the intercept term, *W* = .58 [.57–.60], followed by the linear time term, *W* = .50 [.48–.52]. Because these values are far removed from the statistics for random ratings, I conclude that there is a credible degree of correspondence across studies when ranking children using their peak looking probability, average look probability (the intercept) or their growth curve slope (linear time).

**Summary.** Growth curve features measured individual differences in word recognition performance. By using Kendall’s *W* to measure the degree of concordance among growth curve features over time, I tested whether individual differences in lexical processing persisted over development. I found that the peak looking probability, average looking probability and linear time features were stable over time. Children who relatively fast (or reliable) at word recognition at one age were also relatively fast (or reliable) at other ages too.

## 6.5 Predicting future vocabulary size

I hypothesized that individual differences in word recognition at age 3 will be more discriminating and predictive of future language outcomes than differences at age 4 or age 5. To test this hypothesis, I calculated the correlations of growth curve features with age 5 expressive vocabulary size and age 4 receptive vocabulary. (The receptive test was not administered during the last year of the study for logistical reasons.) As with the concordance analysis, I computed each of the correlations for each sample of the posterior distribution to obtain a distribution of correlations.

Figure [6.9](#fig:evt2-gca-cors) shows the correlations of the peak looking probability, average looking probability and linear time features with expressive vocabulary size at age 5, and Figure [6.10](#fig:ppvt4-gca-cors) shows analogous correlations for the receptive vocabulary at age 4. For all cases, the strongest correlations were found between the growth curve features at age 3.

Growth curve peaks from age 3 correlated with age 5 vocabulary with *r* = .52 [90% UI .50–.54], but the concurrent peaks from age 5 showed a correlation of just *r* = .31 [.29–.33], a difference between age-3 and age-5 correlations of *r*3−5 = .21 [.18–.24]. A similar pattern held for lexical processing efficiency values. Linear time features from age 3 correlated with age 5 vocabulary with *r* = .41 [.39–.44], whereas the concurrent lexical processing values from age 5 only showed a correlation of *r* = .28 [.26–.31], a difference of *r*3−5 = .13 [.10–.16]. For the average looking probabilities, the correlation for age 3, *r* = .39 [.39–.44], was probably only slightly greater than the correlation for age 4, *r*3−4 = .02 [−.01–.04] but considerably greater than the concurrent correlation at age 5, *r*3−5 = .08 [.05–.10].

![Uncertainty intervals for the correlations of growth curve features at each timepoint with age-5 expressive vocabulary (EVT-2 standard scores). The bottom rows provide intervals for the pairwise differences in correlations between timepoints. For example, the top row of the left panel is the correlation between age-3 peak probability and age-5 expressive vocabulary.](data:image/png;base64;base64,)

Figure 6.9: Uncertainty intervals for the correlations of growth curve features at each timepoint with age-5 expressive vocabulary (EVT-2 standard scores). The bottom rows provide intervals for the pairwise differences in correlations between timepoints. For example, the top row of the left panel is the correlation between age-3 peak probability and age-5 expressive vocabulary.

Peak looking probabilities from age 3 were strongly correlated with age 4 receptive vocabulary, *r* = .62 [.61–.64], and this correlation was much greater than the correlation observed for the age 4 growth curve peaks, *r*3−4 = .26 [.23–.29]. The correlation for age 3 average looking probabilities, *r* = .45 [.44–.47], was greater than the age 4 correlation, *r*3−4 = .08 [.06–.11], and the correlation for age 3 linear time features, *r* = .51 [.49–.54], was likewise greater, *r*3−4 = .22 [.19–.26].

![Uncertainty intervals for the correlations of growth curve features from age 43 and age 44 with age-4 receptive vocabulary (PPVT-4 standard scores). The bottom row shows pairwise differences between the age-3 and age-4 correlations.](data:image/png;base64;base64,)

Figure 6.10: Uncertainty intervals for the correlations of growth curve features from age 43 and age 44 with age-4 receptive vocabulary (PPVT-4 standard scores). The bottom row shows pairwise differences between the age-3 and age-4 correlations.

**Summary.** Although individual differences in word recognition were stable over time, early differences were more significant than later ones. The strongest predictors of future vocabulary size were the growth curve features from age 3. Of these features, correlations were strongest for peak looking probabilities.

## 6.6 Discussion

In the preceding analyses, I examined many aspects of children’s recognition of familiar words. First, I modeled how children’s looking patterns *on average* changed year over year. Children’s word recognition improved each year: The growth curves grew steeper, reached higher peaks, and increased in their overall average value each year. This result was unsurprising, but it was valuable because it confirmed that this word recognition task scaled with development. The task was simple enough that children could recognize words at age 3 but challenging enough for children’s performance to improve each year.

After establishing how the averages changed each year, I next asked how variability changed each year. To tackle this question, I used posterior predictive inference to have the model simulate samples of data, and in particular, to simulate new participants. The range of performance narrowed each year, so that children were most variable at age 3 and least variable at age 5. This result is consistent with a model of development children vary widely early on and converge on a more mature level of performance. From this perspective, word recognition is a skill where children “grow out” of immature and highly variable performance patterns. An alternative outcome would have been concerning: Word recognition differences that expanded with age with some children falling behind their peers.

Although the range of individual differences decreased with age, individual differences did not disappear over time. When children at each age were ranked using growth curve features, I found a high degree of correspondence among these ratings. Children who were faster or more accurate at age 3 remained relatively fast or accurate at age 5. Thus, differences in word recognition were longitudinally stable over the preschool years. Extrapolating forwards in time, these differences likely become smaller and smaller and become irrelevant for everyday listening situations. It is plausible, however, that under adverse listening conditions, individual differences might re-emerge and differentiate children’s word recognition performance.

Lastly, I analyzed how individual differences in word recognition features correlated with future vocabulary outcomes. The peak looking probabilities and growth curve slopes from age 3 showed the strongest correlations with future vocabulary scores. This finding was remarkable: Expressive vocabulary scores at age 5, for example, were more strongly correlated with word recognition data collected two years earlier than word recognition data collected during the same week.

We can understand the predictive value of age-3 word recognition performance from two perspectives. The first interpretation is statistical. Differences in children’s word recognition performance were greatest at age 3, so word recognition features at age 3 provide more variance and more information about the children and their future vocabulary size. The second interpretation is conceptual. Correlations were strongest for the growth curve peaks. We can think of this feature as measuring children’s maximum word recognition certainty. A child with a peak of .5, for example, looked the target image half of the time when they were most certain about the word. Although all of the words used were familiar to preschoolers, children with higher peaks knew those words *better*. These children had a stronger foundation for word-learning than children who show more uncertainty during word recognition, and as a result, these children had developed larger vocabularies two years later.

Chapter 7: Effects of Phonological and Semantic Competitors

## 7.1 Looks to the phonological competitor

The next question I asked was how children’s sensitivity to the phonological competitors changed over developmental time. Following our approach in Law et al. ([2016](#ref-RWLPaper)), I only examined trials for which the phonological foil and the noun shared the same syllable onset. For example, this criterion included trials with *dress*–*drum*, *fly*–*flag*, or *horse*–*heart*, but it excluded trials *kite*–*gift* (phonetic feature difference), *bear*–*bread* (onset difference), and *ring*–*swing* (rimes). I kept 13 of the 24 trials. [Appendix A](#vw-experiment-items) provides a complete list of trials used.

The outcome measure for these analyses was the log-odds of fixating on the phonological competitor versus the unrelated word. Because children looked more to the target word with each year of the study, they necessarily looked less to the distractors each year. Figure [7.1](#fig:declining-phon-props) illustrates how the proportions of looks to the phonological foils declined each year. Therefore, I examined the effect of the phonological foil in comparison to the unrelated foil. For example, on the trials where the target is *fly*, we can study the effect of the phonological foil *flag* by looking at when and to what degree the children fixate on *flag* more than the unrelated image *pen*. If a window of time of shows a consistent advantage for the phonological foil over the unrelated image, we conclude that the children were sensitive to the phonological foil during that window. By studying the time course of fixations to the phonological competitor versus the unrelated word, we can identify when the phonological competitor affected word recognition most significantly.

![Because children looked more to the target as they grew older, they numerically looked less the foils too. This effect is why I evaluated the phonological and semantic foils by comparing them against the unrelated image.](data:image/png;base64;base64,)

Figure 7.1: Because children looked more to the target as they grew older, they numerically looked less the foils too. This effect is why I evaluated the phonological and semantic foils by comparing them against the unrelated image.

As in the models from the previous chapter, I downsampled the data into 50-ms (3-frame) bins in order to smooth the data. For these trials, I modeled the looks from 0 to 1500 ms, and I aggregated looks by child, study and time bin. To account for the sparseness of the data, I used the empirical log-odds (or empirical logit) transformation (Barr, [2008](#ref-Barr2008)). This transformation adds .5 to the looking counts. For example, a time-frame with 4 looks to the phonological foil and 1 look to the unrelated image has a conventional log-odds of log(4/1) = 1.39 and empirical log-odds of log(4.5/1.5) = 1.10. This transformation fills in bins with 0 looks with .5/.5 (avoiding 0/0 problems), and it dampens the extremeness of some probabilities that arise in sparse count data.

To model these data, I fit a generalized additive model with fast restricted maximum likelihood estimation (see Sóskuthy, [2017](#ref-Soskuthy2017) for a tutorial for linguists; Winter & Wieling, [2016](#ref-Winter2016); Wood, [2017](#ref-Wood2017)). Box 1 provides a brief overview of these models. I used the mgcv R package (vers. 1.8.23; Wood, [2017](#ref-Wood2017)) with support from the tools in the itsadug R package (vers. 2.3; van Rij, Wieling, Baayen, & van Rijn, [2017](#ref-itsadug)).[[4]](#footnote-4) [Appendix B](#aim1-gca-models) contains the R code used to fit these models along with a description of the specifications represented by the model syntax.

**Box 1: Intuition behind generalized additive models**.

In these analyses, the outcome of interest is a value that changes over time in a nonlinear way. We model these time series by building a set of features to represent time values. In the growth curve analyses of familiar word recognition, I used a set of polynomial features which expressed time as the weighted sum of a linear trend, a quadratic trend and cubic trend. That is:

But another way to think about the polynomial terms is as *basis functions*: A set of features that combine to approximate some nonlinear function of time. Under this framework, the model can be expressed as:

This is the idea behind generalized additive models and their *smooth terms*. These smooths fit nonlinear functions of data by weighting and adding simple functions together. The figures below show 9 basis functions from a “thin-plate spline” and how they can be weighted and summed to fit a growth curve.

![](data:image/png;base64;base64,)

Each of these basis functions is weighted by a model coefficient, but the individual basis functions are not a priori meaningful. Rather, it is the whole set of functions that approximate the curvature of the data—i.e., *f*(Time)—so we statistically evaluate the whole batch of coefficients simultaneously. This joint testing is similar to how one might test a batch of effects in an ANOVA. If the batch of effects jointly improve model fit, we infer that there is a significant smooth or shape effect at play.

Smooth terms come with an estimated degrees of freedom (EDF). These values provide a sense of how many degrees of freedom the smooth consumed. An EDF of 1 is a perfectly straight line, indicating no smoothing. Higher EDF values indicate that the smooth term captured more curvature from the data.

The model included main effects of study year. These *parametric* terms work like conventional regression effects and determined the growth curve’s average values. The model used age 4 as the reference year, so the intercept represented the average looking probability at age 4. The year effects represented differences between age 4 vs. age 3 and age 4 vs. age 5.

The model also included *smooth* terms to represent the time course of the data. As with the parametric effects, age 4 served as the reference year. The model estimated a smooth for age 4 and it estimated *difference smooths* to capture how the curvature at age 3 and age 5 differed from the age-4 curvature. Each of these study-level smooths used 10 knots (9 basis functions). I also included child-level *random smooths* to represent child-level variation in growth curve shapes. Because there is much as less data at the child level than at the study level, these random smooths only included 5 knots (4 basis functions). We can think of these simpler splines as coarse adjustments in growth curve shape to capture child-level variation from limited data. Altogether, the model contained the following terms:

The model’s fitted values are shown in Figure [7.2](#fig:phon-vs-unre-fits). These are the average empirical log-odds of fixating on the phonological foil versus the unrelated image for each year of the study. The model captured the trend for increased looks to the competitor image with each year of the study. At age 4 and age 5, the shape rises from a baseline to the peak around 800 ms. These curves slope downwards and eventually fall beneath the initial baseline. The shape at age 3 does not have a steady rise from baseline and shows a small peak around 800 ms. The peak proportions of looks to the phonological competitor versus the unrelated word were .57 at 800 ms for age 3, .61 at 750 ms for age 4, and .64 at 750 ms for age 5.

![With each year of the study, children looked more to the phonological competitor (relative to the unrelated image) during and after the target noun. Both figures show means for each year estimated by the generalized additive model. The left panel compares model estimates to observed means and standard errors, and the right panel visualizes estimated means and their 95% confidence intervals.](data:image/png;base64;base64,)

Figure 7.2: With each year of the study, children looked more to the phonological competitor (relative to the unrelated image) during and after the target noun. Both figures show means for each year estimated by the generalized additive model. The left panel compares model estimates to observed means and standard errors, and the right panel visualizes estimated means and their 95% confidence intervals.

The early peaks occur when one would expect if children are acting on partial phonological information. The similarity between the phonological competitor and the target noun occurs early on in the trial. Suppose a child acts on the first 400 ms of the phonological competitor. Assuming a 200–300 ms overhead to execute an eye movement in response to speech, the child would reach the phonological foil around 600–700 ms. This window is slightly before the observed peaks at 750–800 ms, but the age 4 and age 5 curves both are on the rise away from baseline during this window.

The average looks to the phonological foil over the unrelated image for age 4 was 0.16 emp. log-odds, .54 proportion units. The averages for age 3 and age 4 did not significantly differ, *p* = .85, but the average value was significantly greater at age 5, 0.31 emp. log-odds, .58 proportion units, *p* < .001. Visually, this effect shows up in the almost constant height difference between the age-4 and the age-5 curves.

There was a significant smooth term for time at age 4, estimated degrees of freedom (EDF) = 7.28, *p* < .001. Figure [7.3](#fig:phon-diff-curves) visualizes how and when the smooths from other studies differed from the age-4 smooth.

![Differences in the average looks to the phonological competitor versus the unrelated image between age 4 and the other ages. Plotted line is estimated difference and the shaded region is the 95% confidence interval around that difference. Boxes highlight regions where the 95% interval excludes zero. From age 3 to age 4, children become more sensitive to the phonological foil during and after the target noun. The linear difference curve for age 4 versus age 5 indicates that the two years largely have the same curvature, but they steadily diverge over the course of the trial.](data:image/png;base64;base64,)

Figure 7.3: Differences in the average looks to the phonological competitor versus the unrelated image between age 4 and the other ages. Plotted line is estimated difference and the shaded region is the 95% confidence interval around that difference. Boxes highlight regions where the 95% interval excludes zero. From age 3 to age 4, children become more sensitive to the phonological foil during and after the target noun. The linear difference curve for age 4 versus age 5 indicates that the two years largely have the same curvature, but they steadily diverge over the course of the trial.

The age-3 and age-4 curves significantly differed, EDF = 5.48, *p* < .001. In particular, the curves are significantly different from 500 to 1050 ms. This result confirms that the looks to the phonological foil increased from age 3 and age 4 during the time window immediately following presentation of the noun and that children became more sensitive to the phonological similarities between the competitor and the target from age 3 to age 4.

The age-3 and age-4 curves also differed significantly after 1250 ms, so that at age 4 children looked less to the competitor compared to age 3. The effect reflects how the looks to phonological competitor decrease as a trial progresses. After an incorrect look to the foil, the children on average corrected their gaze and looked even less to the phonological foil. We do not observe this degree of correction during age 3, because children at age 3 looked less overall to the phonological foil early on.

The age-4 and age-5 smooths also significantly differed, EDF = 1.00, *p* < .001, although the low EDF values indicates that the shape of the difference was a flat line. Thus, the difference between the age-4 and age-5 smooths is driven primarily by the intercept difference and a linear diverging trend—that is, the distance between the two grows slowly over time. The same general curvature was observed for the two studies, suggesting the same general looking behavior at both time points: Children showed an early increase in looks to the phonological foil relative to the unrelated image but after receiving disqualifying information from the rest of the word, the looks to the phonological foil rapidly decrease. The primary difference between age-4 and age-5 is that the competitor effect becomes more pronounced at age 5.

**Summary**. Children looked more to the phonological competitor than the unrelated image early on in the trials. The advantage of the phonological competitor peaked on average around 800 ms after target onset, and the early timing indicates that children were shifting their gaze in response to the fleeting phonological similarity of the competitor to the target noun. The peak was small at age 3 but increased in height with each year of the study. Children became more sensitive to the phonological cohort competitors as they grew older.

## 7.2 Looks to the semantic competitor

I asked how children’s sensitivity to the semantic competitor changed as they grew older. As in Law et al. ([2016](#ref-RWLPaper)), I only examined trials for which the semantic foil and the noun were part of the same category. For example, I included trials with *bee*–*fly*, *shirt*–*dress*, and *spoon*–*pan*, but I excluded trials where the similarity was perceptual (*sword*–*pen*) or too abstract (*swan*–*bee*). This criterion kept 13 of the 24 trials. [Appendix A](#vw-experiment-items) provides a complete list of trials used.

For these trials, I used the same modeling technique as the one used for phonological competitors: Generalized additive models with study effects and a time smooth, time-by-study difference smooths, and time-by-child random smooths. I modeled the looks from 250 to 1800 ms. This window was 300 ms longer than the one used for the phonological competitors in order to capture late-occurring semantic effects.

The model’s fitted values are shown in Figure [7.4](#fig:semy-vs-unre-fits). The average empirical log-odds of fixating on the semantic competitor versus the unrelated word increased with each year of the study. All three years show the same general time course of effects: Looks begin to increase from a baseline around 750 ms and peak around 1300 ms. The peak proportions of looks to the semantic competitor versus the unrelated word increased as children grew older: The peaks were .65 at 1400 ms for age 3, .68 at 1400 ms for age 4, and .71 at 1350 ms for age 5. Moreover, the semantic competitor shows a decisive advantage over the unrelated image at age 3, in contrast to the limited advantage of the phonological competitor at age 3.

![With each year of the study, children looked more to the semantic foil (relative to the unrelated image) with peak looking occurring after the target noun. Both figures show means for each year estimated by the generalized additive model. The left panel compares model estimates to observed means and standard errors, and the right panel visualizes estimated means and their 95% confidence intervals.](data:image/png;base64;base64,)

Figure 7.4: With each year of the study, children looked more to the semantic foil (relative to the unrelated image) with peak looking occurring after the target noun. Both figures show means for each year estimated by the generalized additive model. The left panel compares model estimates to observed means and standard errors, and the right panel visualizes estimated means and their 95% confidence intervals.

The average looks to the semantic foil over the unrelated image for age 4 was 0.44 emp. log-odds, .61 proportion units. Children looked significantly less to the semantic foil on average at age 3, 0.30 emp. log-odds, .57 proportion units, *p* < .001, and they looked significantly more to the semantic foil at age 5, 0.50 emp. log-odds, .62 proportion units, *p* < .001.

There was a significant smooth term for time at age 4, estimated degrees of freedom (EDF) = 7.04, *p* < .001. Figure [7.5](#fig:semy-diff-curves) visualizes the time course of the differences between the smooths from each study.

![Differences in the average looks to the semantic competitor versus the unrelated word between age 4 and the other ages. Plotted line is estimated difference and the shaded region is the 95% confidence interval around that difference. Boxes highlight regions where the 95% interval excludes zero. The flat line on the left reflects how the shape of the growth curves remained the same from age 3 to age 4 and only differed in average height. From age 4 to age 5, the lines quickly diverge and the age-5 curve reaches a higher peak value.](data:image/png;base64;base64,)

Figure 7.5: Differences in the average looks to the semantic competitor versus the unrelated word between age 4 and the other ages. Plotted line is estimated difference and the shaded region is the 95% confidence interval around that difference. Boxes highlight regions where the 95% interval excludes zero. The flat line on the left reflects how the shape of the growth curves remained the same from age 3 to age 4 and only differed in average height. From age 4 to age 5, the lines quickly diverge and the age-5 curve reaches a higher peak value.

The shapes of the age-3 and age-4 curves did not significantly differ, EDF = 1.00, *p* = .535. The age-3 curve begins to rise about 100 ms later, and it reaches a shallower peak value than the age-4 curve. These two features create a nearly constant height difference between the two curves, and thus the two curves show the same overall shape.

The age-4 and age-5 smooths significantly differed, EDF = 3.74, *p* < .001. The differences are greatest after the end of the target noun, in the window from 750 to 1500 ms. The two curves start from a similar baseline but quickly diverge as the age-5 curve reaches a higher peak value. After 1500 ms, the age-5 curve turns downwards to overlap with the age-4 curve. Children looked more to the semantic foil relative to the unrelated image, but they were also quicker to correct and look away from it.

**Summary.** Children became more sensitive to the semantic competitor, compared to the unrelated word, with each year of the study. The semantic foils clearly influenced looking patterns at age 3, in contrast to the muted effect observed for the phonological foils. The semantic effect also occurred when we would expect: After the end of the target noun, following the lexical activation of the target noun and its semantic neighbors.

## 7.3 Child-level differences in competitor sensitivity at age 3

Next, I asked whether children differed reliably in their sensitivity to the phonological and semantic foils based on speech perception and vocabulary measures collected at age 3

As a measure of speech perception, I used scores from a minimal pair discrimination experiment administered during the first year of the study. The task is essentially an ABX discrimination task: A picture of a familiar object is shown and labeled (e.g., “car”), another object is shown and labeled (“jar”), and then both images are shown and one of the two is named. The child then indicated which word they heard by tapping on the image on a touch-screen.

I derived speech perception scores by fitting a hierarchical item-response model. This logistic regression model estimates the probability of child *i* correctly choosing word *j* on word-pair *k*. The equation below provides a term-by-term description of the model. The model’s intercept term represents the average participant’s probability of correctly answering for an average item. By-child random intercepts capture a child’s deviation from the overall average, so they estimate the child’s *ability*. By-word and by-word-in-pair random intercepts capture the relative *difficulty* of particular items on the experiment. The by-word-in-pair effects were necessary because four words appeared in more than one word pair (e.g., *juice*–*goose* and *juice*–*moose*). The model also controlled for the children’s ages and receptive vocabulary scores (PPVT-4 growth scale values). These predictors were transformed to have mean 0 and standard deviation 1, so the model’s intercept reflected a child of an average age and an average vocabulary level. Therefore, the by-child intercepts reflect a child’s ability after controlling for age and receptive vocabulary.

I tested whether phonemic discrimination ability at age 3 predicted looks to the phonological competitor over the unrelated image by modifying the generalized additive model from earlier. In particular, I included a smooth term for the phonemic discrimination ability score and a “smooth interaction” between the smooth of time and phonemic ability. These smooth interaction terms are analogous to interaction terms in linear models. In this case, the interaction term allows the ability score to change the shape of the time trend. The additive model was therefore:

The model included data from 144 participants; these were children with eyetracking data, receptive vocabulary and phonemic discrimination data at age 3. There was not a significant smooth effect for discrimination ability, EDF = 1.00, *p* = .551 or for an interaction smooth between time and ability, EDF = 8.37, *p* = .303.

To test the role of receptive vocabulary, I also fit analogous models using growth scale value scores from the PPVT-4, a receptive vocabulary test. I first adjusted these scores in a regression model to control for–that is, to partial out the effects of—age and predicted accuracy on the discrimination task. There was not a significant smooth effect for receptive vocabulary, EDF = 1.00, *p* = .868, or a significant interaction smooth between time and receptive vocabulary, EDF = 5.57, *p* = .610. Receptive vocabulary therefore was not related to looks to the phonological foil at age 3.

I tested the same two predictors on looks to the semantic foil at age 3. These child-level factors did not show any significant parametric effects, smooth effects or smooth interactions with time. Thus, children’s looks to the semantic foil were not reliably related to phonemic discrimination or receptive vocabulary.

Summary. These models tested whether two child-level factors—minimal-pair discrimination ability and receptive vocabulary—predicted looks to the phonological and semantic competitors at age 3. No significant effects were observed for all cases.

## 7.4 Discussion

In the preceding analyses, I examined children’s fixation patterns to the phonological and semantic competitors and how these fixation patterns changed over developmental time. With each year of the study, children looked more to the target overall, so they consequently looked less to the competitor images each year. To account for this fact, these analyses examined the ratio of looks to the competitors versus the unrelated word. This ratio measured the relative advantage of a competitor over the unrelated word.

### 7.4.1 Immediate activation of phonological neighbors

Developmentally, children became more sensitive to the phonological competitors with each year of the study. These words shared the same syllable onset as the target noun—for example, the pairs *dress*–*drum* or *fly*–*flag*. The competitors affected word recognition early on, with relative looks to the phonological foils peaking around 800 ms. The target nouns were approximately 800 ms in duration at age 3 and 550–800 ms at later ages. Assuming a 200–300 ms overhead for executing an eye movement in response to speech, this timing indicates that children shifted their gaze immediately, based on partial information. Moreover, the tendency to act on partial information became stronger with age, because the early advantage of the phonological competitor increased with each year of the study.

When children looked to the phonological competitor, they fixated on the wrong image and had to revise their interpretation of the noun. At ages 4 and 5, the early peaks of looks to the phonological competitor were followed by a steep, monotonic decrease in looks: Children rejected their initial interpretation of the word and considered other images. At age 3, the average pattern showed more wiggliness, suggesting that children were less decisive in rejecting the phonological competitor. The shapes of the looking patterns at age 4 and age 5 were essentially the same. In particular, the older children were not any faster in the rejecting the phonological competitor on average.

We can interpret these findings in terms of lexical processing dynamics. Under this view, incoming speech activates phonetic and phonemic and lexical representations. The word with the strongest activation is the favored interpretation and the object of the child’s fixations. The early looks to the phonological competitors reflect immediate activation of lexical units: Children activate words on the basis of partial acoustic information. This result is a hallmark of spoken word recognition. The activation of phonologically plausible words becomes stronger with age, as reflected in children’s increasing sensitivity to the phonological competitors. Some mechanisms that may explain this developmental pattern include changes in lexical organization so that neighborhoods of phonologically similar words coactivate and changes in lexical representations so that partial information can more eagerly activate compatible words.[[5]](#footnote-5)

Children at age 4 and age 5 did not show any changes in how quickly they rejected the phonological foil, and this result suggests that lexical inhibition may not change over the preschool years. The reasoning is as follows: If children developed stronger lexical inhibition with age, so that lexical competition resolves more quickly, then we would expect activation of the phonological competitors to decay more quickly and for children to reject the phonological competitor more quickly. But this pattern is not what we observed in the growth curve analyses.[[6]](#footnote-6) The developmental trajectory here is one of increased activation, of children learning words and learning similarities among them so that phonological similar words participate in word recognition.

### 7.4.2 Late activation of semantic neighbors

The semantic competitors were from the same category as the target noun: for example, *bee*–*fly* or *shirt*–*dress*. Children showed year-over-year increases in their sensitivity to the semantic competitor, compared to the unrelated image. Looks the semantic foils started rising steadily 500–700 ms after target onset and peaked late in the trial, around 1300 ms. This time-course is more protracted than the immediate peaks observed for the phonological competitor.

In terms of lexical processing, this late timing is consistent with cascading activation: Spoken words immediately activate phonological neighborhoods with activation cascading onto semantically related words. As a particular word is favored, its semantic relatives receive more secondary activation. For example, children hear “find the shirt”, activate the target *shirt*, but also activate other pieces of clothing including *dress*. The late timing of looks to the semantic competitor therefore reflects late, secondary activation of the spoken word’s semantic relatives. In other words, the activation of a semantic neighbor (like *dress*) is greatest when the activation of the spoken word (*shirt*) is greatest which happens relatively late, once the competition among phonological alternatives resolves.

Under this account, children hear a word, activate it, and become increasingly likely to fixate on the semantic competitor, compared to the unrelated image. The late looks probably reflect a combination of behaviors: children considering the semantically related image to check their initial interpretation as well as children looking to the wrong image because of confusion, lack of knowledge, overriding activation from the semantic competitor, or lack of interest in the target.

Initially, I had subscribed to a confusion or lack-of-knowledge interpretation of the semantic competitor’s advantage. That is, children look to the semantic competitor because they do not know the difference between the target and the semantic competitor. After all, my thinking went, these were young children and decisions like *bee* vs. *fly* or *goat* vs. *sheep* can be difficult. But there are two objections to that line of reasoning. First, our lab piloted the set of words in preschool classrooms, so we confirmed that children could reliably and correctly point to *bee* even when *fly* is an alternative. Second, we would a priori expect that children’s confusion among words to be greatest when they are youngest and have much less experience with these semantic categories. (Indeed, children at age 3 looked less to the target overall, so in general, they were less successful at recognizing the target word.)

The late looks to the semantic competitor, relative to the unrelated image, however, were greatest at age 5. Children’s looks became more selective with age: They looked more to the semantic competitor because they had discovered the semantic connections among words. They had learned the similarity between *bee* and *fly* or *shirt* and *dress*. Put another way, to demonstrate confusion between two choices, children must learn some association that connects the two; they must use or activate some information that induces warranted uncertainty. Rather than confusion about the meaning of nouns, the late looks likely reflect a confirmatory behavior where children give some consideration to the semantic alternative. This is especially the case at age 5, where the advantage of semantic competitor quickly decreases after its peak, indicating rejection of the semantic competitor.

### 7.4.3 Lexical competitors and child-level predictors

I asked whether offline child-level measures predicted sensitivity to the phonological and semantic competitors at age 3. I used children’s ability scores from a minimal-pair discrimination task as a measure of phonemic speech perception, and I also used scores from a receptive vocabulary test. For the phonological competitor, I expected that children with better phonemic discrimination would show increased looks to the phonological competitor because they had more detailed phonemic representations that would activate phonological neighborhoods more quickly. For the semantic competitor, I likewise expected children with larger receptive vocabularies to show increased looks to the semantic competitor because these children knew more words and likely developed more semantic connections among the words. I tested these effects by using the scores as parametric effects to see if they predicted average looks to the competitor, and alternatively, by using the scores for smooth effects to see if they influenced the time course of looks to the foils.

None of these expectations held: Neither of the child-level measures predicted average sensitivity to the phonological or semantic competitors at age 3. Part of the result may be artifactual: The data—looks to a subset of images on a subset of trials—may be too limited at the individual level for the models to pick up on child-level effects. Part of the result may be developmental too: Children were least sensitive to the competitors at age 3, so individual differences may be too small for the data or models to capture. Further work, with different experimental designs, may elaborate on whether offline measures can reliably detect differences in sensitivity to lexical competitors during word recognition.

Chapter 8: General Discussion

This study examined the development of familiar word recognition over the preschool years. The word recognition data came from a visual-world eyetracking experiment which recorded children’s fixations to images in response to prompts like *see the bear*. The trials featured a target noun (e.g., *bear*) along with a phonological competitor (*bell*), a semantic competitor (*horse*), and an unrelated image (*ring*). To describe children’s word recognition ability, I analyzed how the probability of fixating on the target image changed over the time course of a trial. The presence of the competitor images also allowed additional analyses about children’s sensitivities to the phonological and semantic competitors. The experiment was conducted as part of a three-year longitudinal study; children were 28–39 months-old at the Age 3 visit, 39–52 at Age 4, and 51–65 at Age 5. The longitudinal design allowed me to describe developmental changes in word recognition.

## 8.1 How to improve word recognition

Children showed year-over-year improvements in word recognition, as measured by average looking probabilities, peak looking probabilities, and the rate of change in looking probabilities. Children became more reliable, less uncertain, and faster at recognizing familiar words as they grew older. At the same time, children also became more sensitive to the phonological and semantic competitors, compared to the unrelated images. With each year, children looked more to the target image, but when they erred, they were more likely to err on a lexically relevant word.

We can interpret these developmental patterns in terms of lexical activation and processing dynamics. In this task, children hear a stream of speech and activate some phonetic, phonological, lexical, and semantic representations that match the speech input. As they hear more of a word, the activation builds until a particular word is favored, and children shift their gaze onto the named image. Let’s imagine that we have to engineer this system. To make word recognition more efficient, we have to find ways to increase the relative activation of the correct word. In particular, we can boost the strength of connections so that activation can propagate more quickly through the system, and we can also allow inhibition among competing words so that the correct word can win out over its competitors more quickly.

The results from these studies indicate that children become more efficient at activating the target word *and related words* over the preschool years. As they grew older, children were faster to look at a named image and more likely to fixate on the phonological competitor (compared to the unrelated image). These two findings reflect changes in how partial acoustic information can propagate to activate phonologically plausible words. The phonological competitors shared the same syllable onset as the target noun (e.g., *dress*–*drum*), so the early part of the word matched both words. That children became more sensitive to the phonological competitor means that they learned and somehow encoded the phonological similarities among words because part of a word could activate a neighborhood of phonologically plausible matches. This developmental change supports faster word recognition because the listener can channel activation to relevant words more quickly. A similar line of reasoning applies to the semantic competitors: Relative looks to the semantic competitors increased with age, suggesting that children had learned semantic connections among words and activated semantically related words during word recognition.

The other mechanism we might tune to improve word recognition is inhibition. Children’s looks to the phonological or semantic competitors were temporary: Looks increase to some peak level and then quickly decrease. Behaviorally, the drop in looking probability reflects the rejection of an interpretation: for example, a child hears “dr”, shifts looks to *dress*, but hears “um”, revises the interpretation and jumps to *drum*. We can read these corrections as evidence for an inhibitory process: Corrections indicate a change in relative activation where a different word overrides an initial interpretation. But the evidence for *developmental changes* in lexical inhibition from these data was scant. The rate of rejection of the phonological competitor—that is, how quickly looks fall from their peak value—did not change from age 4 to age 5, although the rate did increase for the semantic competitor from age 4 to age 5. Preschoolers did demonstrate inhibition by revising their interpretations of nouns, but there were no clear developmental changes in inhibition.

Previous simulation work can help identify more specific mechanisms at play. McMurray, Samelson, Lee, and Tomblin ([2010](#ref-McMurray2010)) used the TRACE model of word recognition (McClelland & Elman, [1986](#ref-TRACE)) to simulate looks to a target and phonological competitors (cohorts and rimes) in adolescents with specific language impairment. The authors tuned a number of model parameters and analyzed how those changes affected simulated looks to the target and competitors. In the current dataset, I observed a developmental trend where the relative looks to the phonological competitors peak higher each year. In those TRACE simulations, looks to the cohort competitor peak higher if 1) the rate of lexical activation increased, 2) the rate of lexical decay decreased, or 3) strength of lexical inhibition decreased. Of these options, the growth curve for the decrease in lexical inhibition best matches the shape of the current data. The similarity does not mean that children inhibited words any less as they grew older. That would be too simplistic: Developmental changes in preschoolers are the result of simultaneous changes in many mechanisms. But those simulation results suggest that an *increase* in lexical inhibition is *not* one of the key developmental changes in preschoolers’ word recognition.

## 8.2 Learn words and learn connections between words

Preschoolers showed increased activation of the target noun and semantically and phonologically related words but little developmental change in lexical inhibition. Paired with the findings from older children, these results lead to a compelling developmental story. Rigler et al. ([2015](#ref-Rigler2015)) compared 9- and 16-year-olds on a visual-world word recognition experiment with phonological (cohort and rime) competitors. The younger children were slower to look to the target image and showed more looks to the competitors. The implications are that children’s word recognition is still developing in late childhood and that in particular, children’s inhibition of lexical competitors became stronger with age.

The current study with 3-, 4-, and 5-year-olds followed a different pattern: Relative looks to the competitor images increased with age. Taken together, these two studies suggest an interesting progression for the development of lexical processing. During the preschool years, children learn many, many words, and they establish phonological and semantic connections between these words. These connections support the immediate activation of neighborhoods of related words. Later childhood, based on the Rigler et al. ([2015](#ref-Rigler2015)) findings, then is a time for refinement of those connections so that sensitivity to the competitors decreases. This refinement could follow from more selective activation channels, increased lexical inhibition, changes in resting activation (to favor more frequent words), or likely a combination of these factors.

## 8.3 Individual differences are most important at younger ages

Another dimension of this study concerned individual differences in word recognition. Some children were faster or more accurate during word recognition, and these children also were more likely to be faster or more accurate at later ages. The magnitude of these differences diminished over time, as children approached a more mature level of performance.

In terms of lexical processing dynamics, we might think of early differences as reflecting early differences in the burgeoning lexicon. Children may have different numbers of words, different degrees of experience with some words, less established connections among words, and at a lower level, different phonetic and speech perception abilities, given the links between speech perception in infancy and early vocabulary development (Cristia, Seidl, Junge, Soderstrom, & Hagoort, [2014](#ref-Cristia2014_Review)). Differences in word recognition are greatest early on in development because this is when the differences among children’s lexicons are greatest. The task of learning new words, and more importantly, of developing representations and associations to organize words normalizes the early differences among children’s lexicons. That pressure would make the overall variability among children decrease over time while still preserving a relative ordering among children.

We can also interpret the predictive power of word recognition measures in terms of lexical processing and lexical organization. Correlations between word recognition performance and future vocabulary were strongest for the age-3 growth curve features, particularly for the peak probability of looking to the target word. The peak probability measures the overall certainty in word recognition and how strongly the target word is activated. Children with more efficient representations of familiar words at age 3 have a stronger foundation for encoding and integrating future words, and as a result, they showed larger vocabularies at age 4 and age 5.

Initially, I had expected processing *speed*—as approximated by growth curve slopes—to be the most predictive measure of vocabulary growth. Children who can more quickly recognize words, the reasoning goes, can take in information more quickly and devote extra processing resources towards learning.[[7]](#footnote-7) Processing speed was indeed correlated with future vocabulary size, yet peak probability was a stronger predictor of future vocabulary size. Granted, these two processing measures are highly related; to hit a higher peak by time *x*, a growth curve needs to start from higher baseline or have steeper slope. The idea of uncertainty suggests an alternative explanation of the predictive power of word recognition: Children who are more accurate (or less uncertain) during word recognition can extract and activate *more information* from the speech signal.

## 8.4 Limitations and implications

The discussion of processing speed and word recognition certainty highlights one limitation of this research: The experiment’s four-image, eyetracking-based design meant that a clean measure of processing speed was not feasible. Other eyetracking studies with two images can use the latency of how long it takes the child to shift between images as a measure like reaction time. This approach does not translate to the four-image design, as children can visit multiple images on their way to the target. Visual world studies with older participants can obtain an explicit reaction time measure by means of a mouse click or tap on a touchscreen, but those additional task demands may not translate to young children like those in this study. Thus, this study could not address directly whether the predictive power of word recognition performance reflects a more developed lexicon, a general reaction-time-like speed advantage, or both.

The experimental design included semantic and phonological competitors on every trial, so isolating out the semantic and phonological competition effects required some subtlety. A more direct design would compare different types of a trials: for example, trials with a target vs. three unrelated images intermixed with trials with a target vs. a competitor vs. two unrelated images. The trials also used different kinds of phonological and semantic competitors. For example, two of the phonological competitors rhymed with the target, so they could not be included the analysis of phonological competitors (which focused on just competitors with the same onset as the target). The current design limited the number of trials that could be used in the analyses of the competitors and weakened the power of the analyses.

A final limitation includes the changes in the experiment procedure over the course of the longitudinal study. From age 3 to age 4, we re-recorded the stimuli (with the same original speakers) so that the noun durations between the two different dialect versions of the experiment were similar. From age 4 to age 5, we also shortened the duration of the trials by removing attention-getting prompts (e.g., *this is fun!*) from the ends of the trials. These small procedural changes mean that year-to-year differences do not reflect *pure* development differences. It is implausible, however, that the robust year-over-changes owe more to procedural changes than a year of learning and language development.

The findings from this study have implications for our understanding of word recognition and word learning. The first is the overall developmental narrative. Preschool children become better at recognizing words by learning similarities among words and using those similarities to activate neighborhoods of lexically relevant words. Rather than just measuring vocabulary size, word recognition reveals how well words have been integrated into the lexicon. The developmental trends here show that familiar words become more integrated and more connected over the preschool years. Even if a child knows a word at age 3 well enough to recognize or express it, their knowledge of the word will strengthen over time as the word develops connections to other similar words.

From this perspective, we can think of individual differences in word recognition as differences in lexical development. Variability in word recognition diminishes over time, so that differences are more predictive and discriminating at younger ages. Thus, if we wanted to intervene on word recognition, these results indicate that early intervention is better and that intervention should build connections among words and should target words that build onto existing semantic and phonological networks. The natural closing of gaps in word recognition performance with age, however, suggests that word recognition in and of itself may not be an important intervention target. Rather, word recognition measures should serve to supplement other vocabulary measures as an indicator of lexical processing and lexical integration.

Appendices

Appendix A: Items Used in the Visual World Experiment

Table [A.1](#tab:aim1-all-items) lists the items used for the visual world experiment. Each row of the table represents a set of four images used in a trial. There were two blocks of trials with different images and trial orderings. For the two unrelated foils with more than one word listed, the two foils were used in different blocks.

|  |  |  |  |
| --- | --- | --- | --- |
| 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 |

Table A.1: Sets of four images used for the Visual World Experiment.

For the analysis of phonological competitors, I only used trials where the target and the phonological foil shared the same syllable onset (Table [A.2](#tab:phonological-competitor-items)). For the analysis of semantic competitors, I only used trials where the target and the semantic foil belonged to the same category (Table [A.3](#tab:semantic-competitor-items)).

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

Table A.2: Items used for the analysis of phonological versus unrelated competitors.

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

Table A.3: Items used for the analysis of semantic versus unrelated competitors.

Appendix B: Computational Details for Specific Aim 1

## B.1 Growth curve analyses

These models were fit in R (vers. 3.4.3; R Core Team, [2018](#ref-R-base)) with the RStanARM package (vers. 2.16.3; Gabry & Goodrich, [2018](#ref-R-rstanarm)).

When I computed the orthogonal polynomial features for Time, they were scaled so that the linear feature ranged from −.5 to .5. Under this scaling a unit change in Time1 was equal to change from the start to the end of the analysis window. Table [B.1](#tab:poly-feature-range) shows the ranges of the time features.

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Min | Max | Range |
| Time1 | −0.50 | 0.50 | 1.00 |
| Time2 | −0.33 | 0.60 | 0.93 |
| Time3 | −0.63 | 0.63 | 1.26 |
| Trial window (ms) | 250.00 | 1500.00 | 1250.00 |

Table B.1: Ranges of the polynomial time features.

It took approximately 24 hours to run the model on four Monte Carlo sampling chains with 1000 warm-up iterations and 1000 sampling iterations. Warm-up iterations are discarded, so the model comprises 4000 samples from the posterior distribution.

The code used to fit the model with RStanARM is printed below. The variables ot1, ot2, and ot3 are the polynomial time features, ResearchID identifies children, and Study identifies the age/year of the study. Primary counts the number of looks to the target image at each time bin; Others counts looks to the other three images. cbind(Primary, Others) is used to package both counts together for a logistic regression.

library(rstanarm)  
  
# Run chains on different cores  
options(mc.cores = parallel::detectCores())  
  
m <- stan\_glmer(  
 cbind(Primary, Others) ~  
 (ot1 + ot2 + ot3) \* Study +  
 (ot1 + ot2 + ot3 | ResearchID/Study),  
 family = binomial,  
 prior = normal(0, 1),  
 prior\_intercept = normal(0, 5),  
 prior\_covariance = decov(1, 1, 1),  
 data = d\_m)  
  
# Save the output  
readr::write\_rds(m, "./data/stan\_aim1\_cubic\_model.rds.gz")

The code cbind(Primary, Others) ~ (ot1 + ot2 + ot3) \* Study fits a cubic growth curve for each year of the study. It uses R’s formula syntax to regress the looking counts onto an intercept term (implicitly included by default), ot1, ot2, ot3 along with the interactions of the Study variable with the intercept, ot1, ot2, and ot3.

The line (ot1 + ot2 + ot3 | ResearchID/Study) describes the random-effect structure of the model with the / indicating that data from each Study is nested within each ResearchID. Thus, for each child, we have a general intercept and general effects for Time1, Time2, and Time3. These child-level effects are further adjusted using Study:ResearchID effects. The effects in each level are allowed to correlate. For example, I would expect that participants with low average looking probabilities (low intercepts) to have flatter growth curves (low Time1 effects), and this relationship would be captured by one of the random-effect correlation terms. All told, the random effect structure and point estimates for these effects were:

#> Error terms:  
#> Groups Name Std.Dev. Corr   
#> Study:ResearchID (Intercept) 0.305   
#> ot1 0.691 0.20   
#> ot2 0.437 -0.11 0.02   
#> ot3 0.294 -0.11 -0.44 -0.06  
#> ResearchID (Intercept) 0.264   
#> ot1 0.423 0.78   
#> ot2 0.125 -0.75 -0.56   
#> ot3 0.058 -0.23 -0.31 0.19  
#> Num. levels: Study:ResearchID 484, ResearchID 195

The model used the following priors:

prior\_summary(m)  
#> Priors for model 'm'   
#> ------  
#> Intercept (after predictors centered)  
#> ~ normal(location = 0, scale = 5)  
#>   
#> Coefficients  
#> ~ normal(location = [0,0,0,...], scale = [1,1,1,...])  
#> \*\*adjusted scale = [3.33,3.33,3.33,...]  
#>   
#> Covariance  
#> ~ decov(reg. = 2, conc. = 1, shape = 1, scale = 1)  
#> ------  
#> See help('prior\_summary.stanreg') for more details

The priors for the intercept and regression coefficients are wide, very weakly informative normal distributions. These distributions are centered at 0, so negative and positive effects are equally likely. The intercept distribution as a standard deviation of 5, and the coefficients have a distribution around 3. On the log-odds scale, 95% looking to target would be 2.94, so effects of this magnitude are easily accommodated by distributions like Normal(0 [mean], 3 [SD]) and Normal(0, 5).

For the random-effect part of the model, I used RStanARM’s decov() prior which simultaneously sets a prior on the variances and correlations of the model’s random effect terms. I used the default prior for the variance terms and applied a weakly informative LKJ(2) prior on the random-effect correlations. Figure [B.1](#fig:lkj-prior) shows samples from the prior distribution of two dummy models fit with the default LKJ(1) prior and the weakly informative LKJ(2) prior used here. Under LKJ(2), extreme correlations are less plausible; the prior shifts the probability mass away from the ±1 boundaries towards the center. The motivation for this kind of prior was *regularization*: I give the model a small amount of information to nudge it away from extreme, degenerate values.

![Samples of correlation effects drawn from LKJ(1) and LKJ(2) priors.](data:image/png;base64;base64,)

Figure B.1: Samples of correlation effects drawn from LKJ(1) and LKJ(2) priors.

## B.2 Generalized additive models

To model the looks to the competitor images, I used generalized additive (mixed) models. The models were fit in R (vers. 3.4.3) using the mgcv R package (vers. 1.8.23; Wood, [2017](#ref-Wood2017)) with support from tools in the itsadug R package (vers. 2.3; van Rij et al., [2017](#ref-itsadug)).

I will briefly walk through the code used to fit one of these models in order to articulate the modeling decisions at play. I first convert the categorical variables into the right types, so that the model can fit difference smooths.

# Create a Study dummy variable with Age 4 as the reference level  
phon\_d$S <- factor(phon\_d$Study, c("TimePoint2", "TimePoint1", "TimePoint3"))  
  
# Convert the ResearchID into a factor  
phon\_d$R <- as.factor(phon\_d$ResearchID)  
  
# Convert the Study factor (phon\_d$S) into an ordered factor.  
# This step is needed for the ti model estimate difference smooths.  
phon\_d$S2 <- as.ordered(phon\_d$S)  
contrasts(phon\_d$S2) <- "contr.treatment"

I fit the generalized additive model with the code below. The outcome elog is the empirical log-odds of looking to the phonological competitor relative to the unrelated word.

phon\_gam <- bam(  
 elog ~ S2 +  
 s(Time) + s(Time, by = S2) +  
 s(Time, R, bs = "fs", m = 1, k = 5),  
 data = phon\_d)  
  
# Save the output  
readr::write\_rds(phon\_gam, "./data/aim1-phon-random-smooths.rds.gz")

There is just one parametric term: S2. The term computes the average effect of each study with Age 4 serving as the reference condition (and as the model intercept).

Next come the smooth terms. s(Time) fits the shape of Time for the reference condition (Age 4). s(Time, by = S2) fits the difference smooths for Age 3 versus Age 4 and Age 5 versus Age 4. s(Time, R, bs = "fs", m = 1, k = 5) fits a smooth for each participant (R). bs = "fs" means that the model should use a factor smooth (fs) basis (bs)—that is, a “random effect” smooth for each participant. m = 1 changes the smoothness penalty so that the random effects are pulled towards the group average; Winter and Wieling ([2016](#ref-Winter2016)) and Baayen, Rij, Cat, and Wood ([2016](#ref-Baayen2016)) suggest using this option. k = 5 means to use 5 knots (k) for the basis function. The other smooths use the default number of knots (10). I used fewer knots for the by-child smooths because of limited data. As a result, these smooths capture by-child variation by making coarse adjustments to study-level growth curves.

Appendix 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 et al. ([2016](#ref-RWLPaper)) 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 included data from 23 participants from year 1 of the longitudinal study (i.e., what I refer to as age 3) in order to enrich parts of the sample demographics. For this manuscript, we analyzed how vocabulary and maternal education predicted looking patterns, including relative looks to the semantic and phonological foils, but with conventional polynomial growth curves. The use of generalized additive models is an innovation I developed for my dissertation.

Law and Edwards ([2015](#ref-MPPaper)) 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.

Mahr and Edwards (in press) was the manuscript I originally authored for my preliminary examinations. The paper analyzes the same kinds of relations as Weisleder and Fernald ([2013](#ref-Weisleder2013)) which showed that lexical processing efficiency mediated the effect of language input on future vocabulary size. Specifically, 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 (age 3) predicted vocabulary size at year 2 (age 4). 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. In short, we found that receptive vocabulary was more sensitive to variability in lexical processing and home language input than expressive vocabulary.

References

Arias-Trejo, N., & Plunkett, K. (2009). Lexical-semantic priming effects during infancy. *Philosophical Transactions of the Royal Society of London*, *364*, 3633–3647. doi:[10.1098/rstb.2009.0146](https://doi.org/10.1098/rstb.2009.0146)

Baayen, R. H., Rij, J. van, Cat, C. de, & Wood, S. N. (2016). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models. Retrieved from <http://arxiv.org/abs/1601.02043>

Barr, D. J. (2008). Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. doi:[10.1016/j.jml.2007.09.002](https://doi.org/10.1016/j.jml.2007.09.002)

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one’s age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436. doi:[10.1016/j.jecp.2012.01.005](https://doi.org/10.1016/j.jecp.2012.01.005)

Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child Development*, *85*(4), 1330–1345. doi:[10.1111/cdev.12193](https://doi.org/10.1111/cdev.12193)

Ellis Weismer, S., Haebig, E., Edwards, J. R., Saffran, J. R., & 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. doi:[10.1007/s10803-016-2926-y](https://doi.org/10.1007/s10803-016-2926-y)

Fernald, A., & 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. doi:[10.1111/j.1467-8624.2011.01692.x](https://doi.org/10.1111/j.1467-8624.2011.01692.x)

Fernald, A., Swingley, D., & 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. doi:[10.1111/1467-8624.00331](https://doi.org/10.1111/1467-8624.00331)

Gabry, J., & Goodrich, B. (2018). *Rstanarm: Bayesian applied regression modeling via stan*. Retrieved from <https://CRAN.R-project.org/package=rstanarm>

Gamer, M., Lemon, J., & Singh, I. F. P. (2012). irr: Various coefficients of interrater reliability and agreement. Retrieved from <https://CRAN.R-project.org/package=irr>

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Huang, Y. T., & Snedeker, J. (2011). Cascading activation across levels of representation in children’s lexical processing. *Journal of Child Language*, *38*(03), 644–661. doi:[10.1017/S0305000910000206](https://doi.org/10.1017/S0305000910000206)

Kapnoula, E. C., Packard, S., Gupta, P., & McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition*, *134*, 85–99. doi:[10.1016/j.cognition.2014.09.007](https://doi.org/10.1016/j.cognition.2014.09.007)

Lany, J. (2017). Lexical-processing efficiency leverages novel word learning in infants and toddlers. *Developmental Science*. doi:[10.1111/desc.12569](https://doi.org/10.1111/desc.12569)

Law, F., II, & Edwards, J. R. (2015). Effects of vocabulary size on online lexical processing by preschoolers. *Language Learning and Development*, *11*(4), 331–355. doi:[10.1080/15475441.2014.961066](https://doi.org/10.1080/15475441.2014.961066)

Law, F., II, Mahr, T., Schneeberg, A., & Edwards, J. R. (2016). Vocabulary size and auditory word recognition in preschool children. *Applied Psycholinguistics*. doi:[10.1017/S0142716416000126](https://doi.org/10.1017/S0142716416000126)

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–8. doi:[10.1111/j.1467-9280.2007.01871.x](https://doi.org/10.1111/j.1467-9280.2007.01871.x)

Magnuson, J. S., Mirman, D., & Myers, E. (2013). Spoken word recognition. In D. Reisberg (Ed.), *The oxford handbook of cognitive psychology*. Oxford University Press. doi:[10.1093/oxfordhb/9780195376746.013.0027](https://doi.org/10.1093/oxfordhb/9780195376746.013.0027)

Mahr, T., McMillan, B. T. M., Saffran, J. R., Ellis Weismer, S., & Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, *142*, 345–350. doi:[10.1016/j.cognition.2015.05.009](https://doi.org/10.1016/j.cognition.2015.05.009)

Mani, N., & Plunkett, K. (2010). In the infant’s mind’s ear: Evidence for implicit naming in 18-month-olds. *Psychological Science*, *21*(7), 908–913. doi:[10.1177/0956797610373371](https://doi.org/10.1177/0956797610373371)

Mani, N., Durrant, S., & Floccia, C. (2012). Activation of phonological and semantic codes in toddlers. *Journal of Memory and Language*, *66*(4), 612–622. doi:[10.1016/j.jml.2012.03.003](https://doi.org/10.1016/j.jml.2012.03.003)

Marchman, V. A., & 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. doi:[10.1111/j.1467-7687.2008.00671.x](https://doi.org/10.1111/j.1467-7687.2008.00671.x)

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 576–585. doi:[10.1037/0096-1523.15.3.576](https://doi.org/10.1037/0096-1523.15.3.576)

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. doi:[10.1016/0010-0285(86)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton, FL: CRC Press/Taylor & Francis Group.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. doi:[10.1037/a0029872](https://doi.org/10.1037/a0029872)

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39. doi:[10.1016/j.cogpsych.2009.06.003](https://doi.org/10.1016/j.cogpsych.2009.06.003)

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton, FL: Chapman & Hall/CRC.

Nosek, B. A., Simonsohn, U., Moore, D. A., Nelson, L. D., Simmons, J. P., Sallans, A., & LeBel, E. P. (2014, February). Standard reviewer statement for disclosure of sample, conditions, measures, and exclusions. Open Science Framework. Retrieved from <https://osf.io/hadz3>

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & McMurray, B. (2015). The slow developmental time course of real-time spoken word recognition. *Developmental Psychology*, *51*(12), 1690–1703. doi:[10.1037/dev0000044](https://doi.org/10.1037/dev0000044)

Sóskuthy, M. (2017). Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction. Retrieved from <http://arxiv.org/abs/1703.05339>

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73–108. doi:[10.1016/S0010-0277(99)00021-9](https://doi.org/10.1016/S0010-0277(99)00021-9)

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2017). itsadug: Interpreting time series and autocorrelated data using GAMMs.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–52. doi:[10.1177/0956797613488145](https://doi.org/10.1177/0956797613488145)

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, *59*(1), 114–132. doi:[10.1016/j.jml.2008.03.001](https://doi.org/10.1016/j.jml.2008.03.001)

Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, growth curve analysis and generalized additive modeling. *Journal of Language Evolution*, *1*(1), 7–18. doi:[10.1093/jole/lzv003](https://doi.org/10.1093/jole/lzv003)

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). CRC Press.

1. Arias-Trejo and Plunkett ([2009](#ref-Arias-Trejo2009)) is commonly cited as evidence of semantic priming effects. Toddlers heard sentences like “I saw a cat… dog”. During the word *dog*, two images (dog and door) are presented. The idea is that *cat* should prime looks to its semantic neighbor *dog*. The unnatural stimulus order (a sentence followed by an isolated single word) and a condition effect where 18 month-olds children outperformed 21 month-olds make me skeptical that semantic priming is the most plausible explanation of those results. [↑](#footnote-ref-1)
2. My goals in using this method were simply to estimate model effects and quantify the uncertainty about those effects. This pragmatic, estimation-based approach of Bayesian statistics is illustrated in texts by Gelman and Hill ([2007](#ref-GelmanHill)) and McElreath ([2016](#ref-RethinkingBook)). [↑](#footnote-ref-2)
3. The polynomial other terms are less important—or rather, they 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. [↑](#footnote-ref-3)
4. Initially, I tried to use Bayesian polynomial growth curve models, as in the earlier analysis of the looks to the target image. These models however did not converge, even when strong priors were placed on the parameters. In principle, I could have used Bayesian generalized additive models, but the software ecosystem and available tools for model criticism and inference are currently rather limited. [↑](#footnote-ref-4)
5. I am not too committed to any particular mechanisms of *representation* or *organization*. Under a connectionist framework with distributed representations, for instance, a word is represented as a pattern of activation distributed over many shared units. (I think of numbers on a digital clock where seven lines turn on or off to make ten digits but exponentially more complicated.) In that case, representation and organization are inseparable, and it would make more sense to talk about the strength and number of connections instead. My point here is that the lexical mechanisms involved should be ones that enable stronger immediate activations as a result of learning more words. [↑](#footnote-ref-5)
6. Granted, there might be some subtle nonlinear effect at play where higher peak activations require a greater degree of inhibition to overcome, so changes in inhibition could be a plausible part of the developmental story. But there is no compelling reason from the data to make that assertion. [↑](#footnote-ref-6)
7. “The infant who identifies familiar words more quickly has more resources available for attending to subsequent words, with advantages for learning new words later in the sentence, as well as for tracking distributional information about relations among words… Being slow to identify the referent of a familiar word could interfere with lexical activation and impede success in tracking distributional regularities and managing attentional resources in real time (Evans, Saffran, & Robe-Torres, 2009)” (Fernald & Marchman, [2012](#ref-Fernald2012), p. 217). [↑](#footnote-ref-7)