ABSTRACT

Title of dissertation: HOW GRAMMARS GROW: ARGUMENT STRUCTURE AND THE ACQUISITION OF NON-BASIC SYNTAX

Laurel Perkins
Doctor of Philosophy, 2019

Dissertation directed by: Professor Jeffrey Lidz
Department of Linguistics

This dissertation examines the acquisition of argument structure as a window into the role of development in grammar learning. The way that children represent the data for language acquisition depends on the grammatical knowledge they have at any given point in development. Children use their immature grammatical knowledge, together with other non-linguistic conceptual, pragmatic, and cognitive abilities, to parse and interpret their input. But until children have fully acquired the target grammar, these input representations will be incomplete and potentially inaccurate. Our learning theory must take into account how learning can operate over input representations that change over the course of development. What allows learners to acquire new knowledge from partial and noisy representations of their data, one step at a time, and still converge on the right grammar?

The case study in this dissertation points towards one way to characterize the role of development in grammar acquisition by probing more deeply into the resources that learners bring to their learning task. I consider two types of resources.

The first is representational: learners need resources for representing their input in useful ways, even early in development. In two behavioral studies, I ask what resources infants in their second year of life use to represent their input for argument structure acquisition. I show that English learners differentiate the grammatical and thematic relations of clause arguments, and that they recognize local argument relations before they recognize non-local predicate-argument dependencies. The second type of resource includes mechanisms for learning from input representations even when they are incomplete or inaccurate early in development. In two computational experiments, I investigate how learners could in principle use a combination of domain-specific linguistic knowledge and domain-general cognitive abilities in order to draw accurate inferences about verb argument structure from messy data, and to identify the forms that argument movement can take in their language.

By investigating some of the earliest steps of syntax acquisition in infancy, this work aims to provide a fuller picture of what portion of the input is useful to an individual child at any single point in development, how the child perceives that portion of the input given her current grammatical knowledge, and what internal mechanisms enable the child to generalize beyond her input in inferring the grammar of her language. This work has implications not only for theories of language learning, but also for learning in general, by offering a new perspective on the use of data in the acquisition of knowledge.

HOW GRAMMARS GROW: ARGUMENT STRUCTURE AND THE ACQUISITION OF NON-BASIC SYNTAX

by

Laurel Perkins

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Jeffrey Lidz, Chair
Professor Naomi H. Feldman
Professor Colin Phillips
Professor Alexander Williams
Professor Jan Edwards, Dean's Representative

© Copyright by
Laurel Perkins
2019

Preface

Chapter 2 discusses joint work with Jeffrey Lidz and Alexander Williams. Chapter 3 discusses joint work with Jeffrey Lidz. Chapters 4 and 5 discuss joint work with Jeffrey Lidz and Naomi H. Feldman. Portions of Chapter 4 have been published in Perkins, Feldman, and Lidz (2017). This work was supported by the National Science Foundation (#BCS-1551629, Doctoral Dissertation Improvement grant #BCS-1827709, and NRT award #DGE-1449815).

Acknowledgments

I have been extraordinarily lucky to be part of this particular community at this particular time. My gratitude to the people who have enriched my life at Maryland goes far beyond what I am able to write here.

First, my deepest thanks to Jeff Lidz. I'm not sure what warranted his belief that I'd be able to manage the crazy risks that I took in my graduate research. That belief somehow made it possible. His encouragement to persist in thinking carefully through hard problems, and to treat failures and successes as equal opportunities for learning, have made me a better scientist. I am immensely grateful for Jeff's mentorship and friendship, and for his ability to listen with an open mind and say just the right thing to offer support. He's the sort of person who will help you discover that even if you fall into the deep end, you have the resources to swim.

Somehow I also lucked into two unofficial advisors. Thanks to Naomi Feldman for being willing to take on a novice computational modeller, and for coaching me through the process of discovering how these methods could enrich my research. Naomi's enthusiasm for good science across disciplinary boundaries is infectious, and she provided invaluable support in figuring out how to connect with bigger-picture questions. Thanks to Alexander Williams for his dedication to developing this work and for believing in me even when I didn't really understand much at all. I'm hugely grateful for the ways in which Alexander pushed me to become a more precise thinker, speaker, and writer, without losing the forest for the trees.

To the rest of my committee: thanks to Colin Phillips for his ever insightful questions, and for his energetic support of the directions that my research has taken within the broader Maryland Language Science community. Jan Edwards has taught me a great deal about not only communicating my work to different audiences, but also about how scientists can give back to the communities that we are part of.

Thanks as well to the many other faculty and staff at Maryland who supported me throughout my graduate program. Ellen Lau, Valentine Hacquard, Tonia Bleam, Philip Resnik, Omer Preminger, Norbert Hornstein, Paul Pietroski, Bill Idsardi, Howard Lasnik, Masha Polinsky, and Andrea Zukowski provided frequent and thoughtful feedback in many different venues— courses, lab meetings, practice talks, and personal discussions. I am grateful for the chance to hear so many different perspectives, for being challenged to articulate the broader relevance of these research questions, and for the warm lunchtime company. My thanks as well to Tonia Bleam and Peggy Antonisse for their support in developing my skills as a teacher and mentor. I have learned a huge amount from their examples in engaging with students inside and outside the classroom, and am thankful for the hours they spent helping me develop my teaching materials. And no one in the department knows better than Kim Kwok about how to navigate this graduate program. Kim's guidance saved me many times over.

I also give my deepest thanks to Tonia for her friendship through the tricky business of juggling personal and professional lives. Tonia provided support through some of the hardest parts of this process, and offered help when I didn't know I

needed it. For that I am tremendously grateful. And for both her sage advice and wicked sense of humor, which she breaks out at exactly the right moments.

Thanks to Tara Mease for her thoughtfulness and calm leadership, and for seeing the humor in the messy business of conducting research on infant development. Thanks to the entire Project on Children's Language Learning for making this research possible. I am grateful for the community I found there and the labmates I have been lucky to work with: Mina Hirzel, Tyler Knowlton, Adam Liter, Yu'an Yang, Hisao Kurokami, Rachel Dudley, Mike Fetters, Sirri Björnsdóttir, Julie Gerard, Angela He, Aaron Steven White, Kate Harrigan, and Naho Orita. Special thanks to Mina, Tyler, Rachel, Angela, and Sirri for being wonderful collaborators on these research projects. And huge thanks to the research assistants whose contributions to this work have been invaluable: Alexander Shushunov, Erika Arrigo, Lillianna Righter, Gbemileke Soyoye, Abigail Kameny, William Simpson, John-Paul Teti, and Jordan Schneider. I have learned so much from working with each of you.

My profound thanks to Mina for always keeping me on my toes intellectually, and for a friendship that has sustained me through challenges both personal and professional. Mina's artistry goes far beyond uncanny sketches of my experimental stimuli. I am so grateful for her patience and ability to see things that others miss.

Thanks to Tyler for being an energetic and thoughtful collaborator, for introducing me to the Cinnaminson Special, and for forgiving me for taking his toy truck over hours of stimuli filming. Thanks as well to Zoe Ovans for her insightful feedback and formidable board game skills.

Thanks to my cohort and officemates— Kasia Hitczenko, Nick Huang, Gesoel Mendes, Sirri Björnsdóttir, and Phoebe Gaston— for listening to my half-baked ideas, giving me the encouragement I needed to work through problems, and putting up with my playlists when I forgot to plug my headphones into my computer.

I have learned a great deal from the members of the Bias in Linguistics Working Group: Adam Liter, Bethany Dickerson, Jackie Nelligan, Phoebe Gaston, Hanna Muller, Mina Hirzel, Kasia Hitczenko, Paulina Lyskawa, Maxime Papillon, Margaret Kandel, and Karthik Durvasula. I am grateful for the efforts of everyone working towards improving representation and equity in our field, and to the people at Maryland and in the broader community who have supported those efforts.

One such person is Virginia Valian, who not only has been hugely influential in how I understand issues of gender equity, but was also the reason I decided to pursue language acquisition research in the first place. I don't know why Virginia invited me to join her lab with almost no relevant background, but I am exceedingly grateful for that generosity and for her ongoing mentorship. I did not understand at the time how much I was learning from Virginia's approach as a scientist and advisor. But I now appreciate how much that approach has shaped both the type of research I do and how I see my role in the wider academic community.

Thanks to the Maryland Language Science Center and to the audiences who provided feedback on various versions of this work, including the Cognitive Neuroscience of Language Lab, S-Lab, and the ProbMod Reading Group.

Finally, thanks to my parents and the rest of my family. Your support has meant so much to me.

Table of Contents

Preface ii

Acknowledgements iii

List of Tables viii

List of Figures x

1 Introduction 1
- 1.1 Input Representations and Grammar Learning in Development 1
- 1.2 The Role of Transitivity in Bootstrapping 5
  - 1.2.1 Semantic and Syntactic Bootstrapping 6
  - 1.2.2 Which Bootstrapping Inference? 9
- 1.3 Recognizing Transitivity in Non-Basic Clauses 12
  - 1.3.1 Representing Wh-Dependencies 15
  - 1.3.2 Filtering Non-Basic Clauses for Verb Learning 19
  - 1.3.3 Modeling Non-Basic Clause Development 21
- 1.4 Overview 24

2 Bootstrapping from Transitivity 27
- 2.1 Background 27
  - 2.1.1 Bootstrapping from Argument Number 29
  - 2.1.2 Bootstrapping from Thematic Content 36
  - 2.1.3 Diagnosing Event Representations 45
- 2.2 Experiment 1: Transitive Frame 52
  - 2.2.1 Method 52
  - 2.2.2 Results 60
  - 2.2.3 Discussion 63
- 2.3 Experiment 2: Intransitive Frame 67
  - 2.3.1 Results 71
  - 2.3.2 Discussion 73
- 2.4 General Discussion 76

3 Representing Transitivity in Non-Basic Clauses 83
3.1 Background . . . 83
3.1.1 Acquiring Wh-Dependencies . . . 87
3.1.2 Hypothesis: Gap-driven Learning . . . 89
3.1.3 Prior Experimental Results . . . 92
3.2 Experiment 1: 14- and 15-month-olds . . . 96
3.2.1 Method . . . 96
3.2.2 Results . . . 103
3.2.3 Discussion . . . 105
3.3 Experiment 2: 17- and 18-month-olds . . . 108
3.3.1 Method . . . 108
3.3.2 Results . . . 110
3.3.3 Discussion . . . 112
3.4 General Discussion . . . 115

4 Filtering Input for Transitivity Acquisition 121
4.1 Background . . . 121
4.1.1 Filtering . . . 122
4.1.2 Computational Models of Verb Learning . . . 125
4.2 Model . . . 128
4.2.1 Generative Model . . . 130
4.2.2 Joint Inference . . . 133
4.3 Simulation 1 . . . 133
4.3.1 Data . . . 134
4.3.2 Results . . . 136
4.3.3 Model Comparisons . . . 140
4.3.4 Discussion . . . 147
4.4 Simulation 2 . . . 151
4.4.1 Data . . . 152
4.4.2 Results . . . 152
4.4.3 Model Comparisons . . . 156
4.4.4 Discussion . . . 158
4.5 General Discussion . . . 160

5 Learning the Surface Forms of Argument Movement 169
5.1 Background . . . 169
5.1.1 Syntactically-Informed Distributional Learning . . . 171
5.2 Model . . . 179
5.2.1 Generative Model . . . 180
5.2.2 Joint Inference . . . 186
5.3 Simulations . . . 187
5.3.1 Data . . . 187
5.3.2 Results . . . 189
5.3.3 Model Comparisons . . . 197
5.3.4 Interim Summary . . . 203

5.3.5 Features of Argument-Gap Categories 204
5.4 Discussion 212
5.4.1 Which Syntactic Dependencies? 217
5.4.2 Conclusion 228

6 Conclusion 230
6.1 Summary of Findings 233
6.2 Future Directions and Broader Implications 238

A Details of Gibbs Sampling for Verb Transitivity Learner 244

B Details of Gibbs Sampling for Argument Gap Learner 248
B.1 Initialization 248
B.2 Sampling c 251
B.3 Sampling e 254
B.4 Sampling with Annealing 257

C Description of Argument-Gap and Other Movement Categories 259

D Odds Ratios for Features of Argument Gap Categories 263

References 268

List of Tables

2.1 Structure of Experimental Trial, Experiment 1 54
2.2 Training Trials, Experiments 1 and 2 58
2.3 Structure of Experimental Trial, Experiment 2 69

3.1 Sample Test Sentences, Experiments 1 and 2 97

4.1 Corpora of Child-Directed Speech 134
4.2 Dataset: Uses with Overt Direct Objects (DO) of 50 Verbs 136
4.3 Percentages of Verbs Categorized Correctly, Simulation 1 137
4.4 Percentages of Verbs Categorized Correctly by Threshold Models . . 146
4.5 Percentages of Verbs Categorized Correctly, Simulations 1 and 2 . . . 153

5.1 Morphosyntactic Features (F) 182
5.2 Distribution of Clause Types in Dataset 189
5.3 Sample Sentences from Model's Argument-Gap and Other Movement Categories 192
5.4 Accuracy on Identifying Sentences with Movement 194
5.5 % Sentences Identified as Having Gaps by Joint Inference Model, by Movement and Verb Type 195
5.6 % Object Movement Sentences Identified by No-Category Baseline . . 198
5.7 Features with Significantly Higher Odds in Argument-Gap Categories 206
5.8 Distributions of Distinctive Features from Table 5.7 in Non-Argument Gap Categories 209

D.1 Distinctive Features of Category 3 (Basic) 264
D.2 Distinctive Features of Category 9 (Basic) 264
D.3 Distinctive Features of Category 11 (Wh-Questions) 264
D.4 Distinctive Features of Category 12 (Wh-Questions) 264
D.5 Distinctive Features of Category 14 (Wh-Questions) 265
D.6 Distinctive Features of Category 16 (Wh-Questions) 265
D.7 Distinctive Features of Category 17 (Wh-Questions) 265
D.8 Distinctive Features of Category 21 (Other Questions) 265
D.9 Distinctive Features of Category 23 (Passives) 266

D.10 Distinctive Features of Category 24 (Passives) 266
D.11 Distinctive Features of Category 25 (Passives) 266
D.12 Distinctive Features of Category 26 (Passives) 266
D.13 Distinctive Features of Category 30 (Embedded Clauses) 267
D.14 Distinctive Features of Category 33 (Embedded Clauses) 267
D.15 Distinctive Features of Category 35 (Imperatives) 267

List of Figures

1.1 Traditional Model of Language Acquisition 2
1.2 Developmental Model of Language Acquisition 3

2.1 Image of Experimental Test Videos, TAKING (right) vs. MOVING (left) 57
2.2 Timecourse of Proportion Looks to TAKING Video, Exp. 1 61
2.3 Timecourse of Proportion Looks to TAKING Video, Exp. 2 72

3.1 Looking Time by Trial Type, Condition, and Age, Exp. 1 103
3.2 Preference for Object Gap Trials by Condition and Age, Exp. 1 . . . 104
3.3 Looking Time by Trial Type, Condition, and Age, Exp. 2 110
3.4 Preference for Object Gap Trials by Condition and Age, Exp. 2 . . . 111

4.1 Graphical Model 131
4.2 Posterior Distributions over Verb Categories (T), Joint Inference Model 137
4.3 Posterior Distributions over ϵ and δ, Simulation 1 140
4.4 Posterior Distributions over Verb Categories (T), Oracle Model . . . 141
4.5 % Verbs Categorized Correctly by Varying Values of ϵ and δ 142
4.6 Posterior Distributions over Verb Categories (T), Simulation 2a . . . 152
4.7 Posterior Distributions over Verb Categories (T), Simulation 2b . . . 153
4.8 Posterior Distributions over ϵ and δ, Simulation 2a 155
4.9 Posterior Distributions over ϵ and δ, Simulation 2b 156
4.10 Accuracy by Prior Probability on Alternating Verbs 158

5.1 Graphical Model 181
5.2 Proportions of Clause Types in Sentence Categories (c) Inferred by Joint Inference Model 190
5.3 Proportions of Clause Types in Sentence Categories (c) Inferred by Baseline Distributional Learner 201

6.1 Developmental Model of Language Acquisition 231

Chapter 1: Introduction

1.1 Input Representations and Grammar Learning in Development

Children acquire grammars for languages in their community. Each grammar is abstract in relation to the child's specific experiences, which are always consistent with many possible grammars. Yet children in the same linguistic community arrive at remarkably similar grammars. This is the problem of language acquisition (Chomsky, 1965).

Linguistic theories since Chomsky (1965) and Wexler and Culicover (1980) have traditionally studied this problem by modelling language acquisition along the lines of Fig. 1.1. In this model, children start with a set of data: a corpus of primarily linguistic input containing the sentences that they hear others utter around them. They then apply their general cognitive abilities, plus knowledge of the types of properties that grammars can have, in order to infer the specific properties of the grammar they are acquiring. This model idealizes language acquisition as an instantaneous process that maps a corpus of input onto a grammar, abstracting away from time and resource constraints. Some modern approaches make a similar idealization in asking whether the data as a whole support grammar selection (Perfors, Tenenbaum, & Regier, 2006; Yang, 2002). These approaches are not intended to model

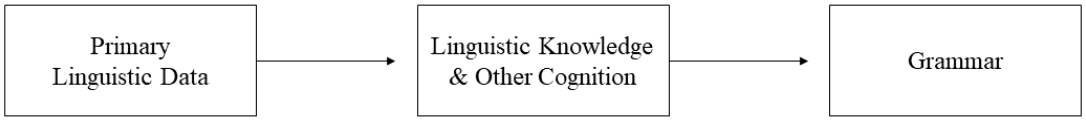

Figure 1.1: Traditional Model of Language Acquisition

learning over time; instead, they illuminate the problem of language learnability from a global perspective, asking what combination of data, linguistic knowledge, and other cognitive abilities enables grammar learning to converge in the limit. But there is an important dimension that is missed in this abstraction: the role that development plays in language acquisition.

Language learning, like other forms of learning, is incremental: what we can learn depends on what we already know. Just as a child who can't count cannot learn arithmetic, a child who can't segment words cannot learn whether her language has verb raising. Learning in any domain depends on how the data for learning are represented; as input representations change over the course of development, this impacts how learners form categories and draw generalizations. Thus, the way that learners perceive and use the input to grammar learning is not static— it changes as they learn more about their language.

This means that language acquisition, as it unfolds over development, looks more like the model in Fig. 1.2. In this model, children take in their linguistic input and represent it using the knowledge of their language that they currently have available— their developing grammar— together with their developing conceptual and cognitive abilities. These input representations are immature and incomplete at early stages of development, but they then become the data that children use to

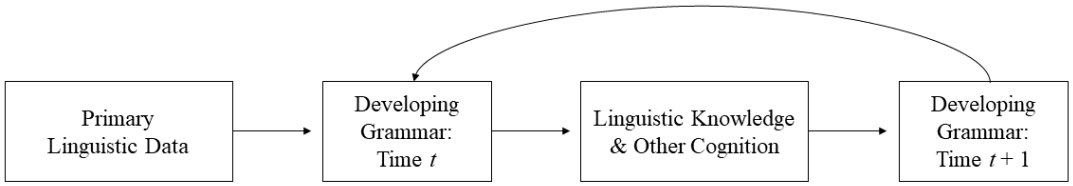

Figure 1.2: Developmental Model of Language Acquisition

gain new grammatical knowledge (Lidz & Gagliardi, 2015). Children apply their general cognitive abilities, plus knowledge of the types of properties that grammars can have, to draw new inferences about their grammar on the basis of these data. And this process then iterates: children's updated grammars will allow them to represent their input more richly and completely, enabling further inferences about the grammar they are acquiring, until they incrementally arrive at the target state.

In this dissertation, I ask how we might enrich the traditional model of language acquisition in order to account for the role of development, and what we gain from doing so. Moving to a developmental model means we must solve an apparent paradox. At any given point in development, the way that children represent their input depends on their current knowledge of their language, and these representations determine what further inferences children can draw about their target grammar. Learning cannot wait until children can completely and accurately represent their input, or there would be nothing further to learn (Fodor, 1998; Valian, 1990). But if children's representations of their input are incomplete or inaccurate, how do they avoid faulty inferences, or even learn from it at all? When we take seriously the incremental nature of learning in development, we run into difficult questions of how grammar learning gets started, and how it manages to converge.

I propose that the answers to these questions will enrich our understanding of the resources that children bring with them to the language learning task. First, children need resources for representing their input in useful ways for further learning, even very early in grammatical development. For example, in order to learn syntax, it might be helpful for a child to be able to represent her input in a format that encodes some amount of syntactic structure, even if it is partial and immature, rather than only encoding string properties. Second, children need resources for learning from their input representations when they are partial and immature; they need learning mechanisms that will allow them to draw accurate inferences about their grammar on the basis of input that they cannot yet parse completely or veridically. How do learners extract signal from data that will contain a great degree of noise early in the learning process?

As a case study, I examine these questions within the domain of argument structure, one of the most basic syntactic properties to be acquired in infancy. Infants at very early stages of grammatical development draw inferences about syntax and meaning on the basis of clause transitivity, by observing how verbs distribute with subjects and objects in a clause (e.g Fisher, Gertner, Scott, & Yuan, 2010; Lidz, White, & Baier, 2017; Gagliardi, Mease, & Lidz, 2016). Here, I ask how this learning unfolds over development, focusing on a narrow timeslice within an infant's second year of life. I examine this developmental window from the two angles of representation and learning.

First, what representational resources do young infants have for representing transitivity when drawing inferences about verb meanings and syntax? Do they

represent clause subjects and objects as a set of two noun phrases, or differentiate the grammatical and thematic relations of these arguments in some further way? Second, how do learners come to recognize the predicate-argument relations in so-called "non-basic" clauses, in which transformations may displace clause arguments from their canonical positions? If learners cannot yet recognize argument movement when it has occurred, how do they cope with the messy data that non-basic clauses contribute to their input, and avoid drawing faulty inferences from their incomplete representations of these sentences? Through this case study, I show that language acquisition theories can be enriched by taking into account the way that learners represent and learn from their input at each stage of development, enabling a deeper understanding of how and why learners generalize in the way that they do.

To motivate these research questions, it will be helpful to consider the role that transitivity has played in bootstrapping theories of argument structure and clause structure acquisition.

1.2 The Role of Transitivity in Bootstrapping

Bootstrapping (Gleitman, 1990; Grimshaw, 1981; Pinker, 1984) offers a way to bridge the gap between learners' input and the abstract grammatical representations they need to acquire. The data for grammar acquisition consist of both immature representations of the sentences that learners hear, and representations of the contexts in which those sentences are uttered. If the syntactic categories that generated those sentences are related in a systematic way to the conceptual cate-

gories under which learners represent their contexts of use, learners may be able to exploit those correlations in order to gain an initial foothold into the grammatical system of their language. Because transitivity is robustly correlated with clause meaning cross-linguistically (Fisher et al., 2010; Gleitman, 1990; Hopper & Thompson, 1980; Lidz & Gleitman, 2004a; Naigles, 1990), it provides a particularly useful cue for the first stages of bootstrapping into the target grammar. Learners might be able to use clause meaning to identify subjects and objects in sentences they hear (Pinker, 1984), or use a verb's distribution with subjects and objects to draw inferences about its meaning and argument structure (Fisher et al., 2010; Gleitman, 1990; Lidz & Gleitman, 2004a; Naigles, 1990). A key question is how learners represent transitivity in their input in order to enable these inferences, and what inferences they draw on the basis of those representations.

1.2.1 Semantic and Syntactic Bootstrapping

Semantic and syntactic bootstrapping describe the two different directions in which learners might draw inferences by relating syntactic and conceptual structure. If the syntactic environments in which verbs distribute are related in a systematic way to conceptual categories of events they describe, then learners can use evidence about one of these properties (syntactic or conceptual) to draw inferences about the other.

In semantic bootstrapping, a child who represents an event under a particular conceptual description might be able to use these correspondence relations to draw

inferences about the syntax of the clause describing that event (Grimshaw, 1981; Pinker, 1989, 1984). For example, a child who perceives an event as involving an agent and a patient, and furthermore knows that subjects of active transitive clauses tend to name agents and objects tend to name patients, could then infer which argument is the subject and which is the object in a clause describing that event. This may lead to further inferences about how subjects and objects are realized in the language: for instance, how subjects and objects tend to be ordered with respect to the verb, and what surface forms indicate agreement relations between a verb and these arguments. Conversely, in syntactic bootstrapping, a child who represents a clause under a particular linguistic structure might be able to use these correspondence relations in the opposite direction to draw inferences about which event the clause describes (Fisher et al., 2010; Gleitman, 1990; Landau & Gleitman, 1985; Lasnik, 1989). For example, a child who hears an unknown verb in a clause that she represents as transitive could then infer that this clause describes an event she perceives as having an agent and a patient. This will allow her to narrow down the range of events that the new verb might describe, restricting the conceptual space she is considering for that verb's meaning.

Experimental evidence shows that infants are able to use these meaning-distribution correspondence relations in learning. In preferential looking tasks, English-learning infants as young as 17 months can use the canonical subject-verb-object word order of English to identify that the individual named by the subject of a transitive clause is the agent of an event, and the individual named by the object is the patient (Hirsh-Pasek & Golinkoff, 1996; Gertner, Fisher, & Eisengart, 2006).

Learners also draw inferences about verb meanings by observing their distributions across different clause structures. By the age of 19 months, infants reliably infer a causative meaning for a novel verb in a transitive vs. an intransitive clause, and do so under the right circumstances at 15 months as well (Arunachalam & Waxman, 2010; Arunachalam, Escovar, Hansen, & Waxman, 2013; Jin & Fisher, 2014; Messenger, Yuan, & Fisher, 2015; Naigles, 1990; Yuan & Fisher, 2009; Yuan, Fisher, & Snedeker, 2012).

Children draw even finer-grained inferences on the basis of hearing novel verbs participate in particular transitive-intransitive alternations. The subject of an intransitive clause can name either an agent (e.g. *John baked*) or a patient (e.g. *The bread rose*). How a verb distributes in intransitive clauses is related to its meaning: intransitives whose subjects name agents tend to describe activities of those agents, whereas intransitives whose subjects name patients tend to describe changes undergone by those patients (Fillmore, 1968, 1970; Levin & Hovav, 2005; Williams, 2015). Another line of experimental work has found that two-year-olds are sensitive to these distinctions (Bunger & Lidz, 2004, 2008; Naigles, 1996; Scott & Fisher, 2009).

In summary, bootstrapping allows young learners to narrow down the range of meanings for new verbs by using information about how those verbs distribute in transitive and intransitive clauses. Infants use subjects and objects in clauses to identify whether a new verb labels a causative event, which participant roles individuals bear in relation to that event, and whether that event involves a change or an activity. But these results leave open a crucial question: how exactly do

learners represent clause transitivity when drawing these inferences? Different clause representations will enable different types of bootstrapping inferences, and hence different generalizations about the target grammar.

1.2.2 Which Bootstrapping Inference?

The experimental results surveyed above do not show us how young verb-learning infants represent transitivity: whether as some type of asymmetrical relation between a subject and an object, or as merely a set of two noun phrases. If learners cannot yet reliably identify the subject and object of a clause, then they must have a heuristic for inferring verb meanings from a flatter clause representation. One influential hypothesis proposes that infants use a heuristic based only on the number of noun phrases in a clause, taking those to be the arguments and assuming that they match one-to-one the number of participants in an event the clause describes, as we naturally perceive it (Fisher et al., 2010; Lidz & Gleitman, 2004a; Naigles, 1990). Thus, a clause with two arguments should describe an event perceived with exactly two participants, and a clause with one argument should describe a one-participant event. This "**Participant-to-Argument Matching**" strategy (**PAM**) is attractive because it provides constraints on verb learning on the basis of very little linguistic knowledge. Because PAM is consistent with infants' behavior in prior experimental work, those results have been taken to support PAM as the primary bootstrapping strategy that infants use at the onset of verb learning (Fisher et al., 2010).

However, prior experimental results do not differentiate PAM from a more sophisticated bootstrapping strategy. If infants do differentiate subjects and objects in a clause, their behavior in previous studies may be driven by finer-grained inferences about the participant relations of those arguments, rather than argument number. Following Pinker, we may call this strategy "**Thematic Linking**": infants may expect particular argument positions to link to particular thematic relations, e.g. transitive subject to Agent and transitive object to Patient (Pinker, 1984; Williams, 2015; Lidz et al., 2017). The work reviewed above suggests that infants are sensitive to many of these linking principles, at least by the age of two years (Bunger & Lidz, 2004, 2008; Hirsh-Pasek & Golinkoff, 1996; Gertner et al., 2006; Naigles, 1996; Scott & Fisher, 2009). In order for infants to use this strategy to enable their earliest verb-learning inferences, they must be able to identify subjects and objects in their language with sufficient regularity. Their ability to do so at the relevant age is an empirical question.

Conceptually, there are two primary reasons why Thematic Linking is an attractive hypothesis. PAM does not generally characterize the relation between clause structure and meaning in adult grammars, either within a language or across languages, and thus could only be an initial heuristic to be abandoned later in development (Williams, 2015). But because many linking principles between grammatical and thematic relations are robustly attested cross-linguistically, children would not need to abandon them over the course of development (Baker, 1988; Dowty, 1991; Fillmore, 1968, 1970; Jackendoff, 1972; Levin & Hovav, 2005; Pinker, 1984; Williams, 2015). Moreover, PAM raises a difficult developmental question: how

do children learn to abandon a heuristic based on counting clause arguments, and acquire a grammar that encodes structure rather than number? Thematic Linking implies a simpler developmental story, in which children's learning is informed by generalizations that are consistent with the grammar they end up acquiring.

Thus, it is an open question whether infants at the onset of verb learning bootstrap primarily from the number of clause arguments, as proposed under PAM, or from the thematic content of those arguments, as proposed under Thematic Linking. When the two strategies would lead to different bootstrapping inferences, which strategy will infants use? In Chapter 2, I present new experimental evidence differentiating these hypotheses with 19- to 22-month-old infants, the youngest age range commonly tested in prior literature. Teasing apart these strategies has implications not only for our theories of the mechanisms that guide learning from input. It also helps us understand the types of input representations that infants learn from: specifically, whether infants initially represent clause arguments as a string of noun phrases, or whether they differentiate the grammatical relations of those arguments. If infants' early verb-learning inferences are primarily driven by grammatical and thematic relations rather than the number of arguments, it may be that their clause representations privilege categories like 'subject' and 'object,' and argument number may only be a proxy for recovering those categories in their input. This invites further investigation into how richly those categories are represented— are they represented *qua* subjects and objects, in an adult-like hierarchical constituent structure? And if so, how do those representations arise in development? Although I will not answer all of these questions here, this work provides a foundation for

future research to do so.

1.3 Recognizing Transitivity in Non-Basic Clauses

In order to bootstrap verb meanings and argument structure from their distributions in transitive and intransitive clauses, learners must reliably recognize clause transitivity in their input. This is the case regardless of whether transitivity is represented in terms of argument number or argument relations. But recognizing transitivity is not trivial, as transitive and intransitive clauses can be realized in highly variable ways within and across languages. Consider the following:

(1) Amy fixed her bicycle.

(2) *Amy fixed.

(3) What did Amy fix?

Recognizing that the wh-object question in (1) is underlyingly transitive depends on knowing that *what* acts as the verb's object, despite not being realized in an argument position. An infant who does not yet know that *what* is a wh-word might fail to represent it as an argument and treat the verb as intransitive. This could lead to faulty inferences about the argument structure and meaning of *fix*: a learner might think that this verb can freely occur without a direct object, with consequences for what it means.

More generally, if learners do not know how particular transformations are realized in their language, they might bootstrap from inaccurate sentence repre-

sentations[1]. For this reason, clauses that we might call "non-basic" have been recognized as problematic for both syntactic and semantic bootstrapping theories (Gleitman, 1990; Lidz & Gleitman, 2004a; Pinker, 1984). Following Keenan (1976), Pinker (1984) notes that relations like "subjects tend to name agents" hold only in "'basic sentences': roughly, those that are simple, active, affirmative, declarative, pragmatically neutral, and minimally presuppositional. In nonbasic sentences, these properties may not hold." Consider the case of the wh-object question in (1), repeated here as (4), as well as other non-basic clause types such as relative clauses (5) and passives (6):

(4) What did Amy fix?

(5) I like the bicycle that Amy fixed.

(6) The bicycle was fixed (by Amy).

In each of these examples, a syntactic transformation has applied such that the argument acting as the object of the verb no longer surfaces in canonical object position. If a child is not aware of these transformations, she may be misled when she relates the linguistic structure she (mis-)perceives in these clauses with her conceptual representations of events. For example, a semantic bootstrapper who takes (6) to be a description of an event in which she perceives Amy to be the agent and the bicycle to be the patient might construe *Amy* as the subject and *the bicycle*

[1]Note that this problem is not unique to transformation-based grammatical theories. Under theories in which transitive clauses, wh-object questions, and passives are separate "constructions" (Fillmore, Kay, & O'Connor, 1988; Goldberg, 1995; Langacker, 1999), the learner must still ultimately recognize that only verbs that occur in transitives can also occur in wh-object questions and passives. Whether this is encoded transformationally or via a construction hierarchy, the same logical problem holds.

as the object, resulting in a parse that is not only erroneous but also implies that English has object-verb-subject (OVS) word order. Likewise, the fronted arguments (*what* and *the bicycle*) in (4) and (5), if recognized as arguments, might be taken as evidence for optional OSV word order in English rather than as evidence for the wh-movement that actually produced this non-canonical word order. And if these phrases are not recognized as arguments of *fix*, a variety of other inaccurate parses would be available for these sentences. These include ones in which *fix* takes a null or implicit object, meaning either "Amy fixed stuff" or "Amy fixed the thing."

Conversely, a syntactic bootstrapper who is not aware of the transformations in these sentences may draw faulty inferences about which events in the world they describe. Because direct objects are not realized in their canonical post-verbal position, a child may not recognize that these clauses are underlyingly transitive, and thus may not infer that they describe causative events. In this case, an event in which she perceives Amy to be the agent and the bicycle to be the patient may no longer count for her as a possible "fixing." The problem is not necessarily solved as soon as she observes *fix* in a basic clause that she can recognize as transitive. In that case, she may infer that *fix* belongs some class of verbs that can alternate between transitive and intransitive uses, like *eat* or *rise*, leading to inaccurate inferences about both its syntactic and semantic properties.

In summary, non-basic clauses are problematic for early grammar acquisition. Infants who do not know which sentences have undergone particular transformations might mis-perceive the structure of these sentences, disrupting their attempts to put syntactic and conceptual categories into correspondence for grammar learning.

In particular, if learners fail to recognize transitivity when arguments have been moved, they may draw faulty inferences about verb syntax and meanings, as well as the broader clause structure properties of their language. Understanding the extent of this problem requires understanding the developmental trajectory of non-basic clause acquisition. When and how do learners come to identify the ways that movement is realized in their language? And how does this interact with their acquisition of verb argument structure? I will now provide a brief overview of what we currently know about these topics, and the open questions that I will be addressing in this work.

1.3.1 Representing Wh-Dependencies

To examine the developmental trajectory of non-basic clause acquisition, I focus on wh-dependencies, which are present in both both wh-questions and relative clauses, repeated here as (7) and (8). These are among the most common non-basic clause types that infants hear: wh-questions make up about 15% of the input to English-learning children even before their second birthday, and the majority of these questions contain non-canonical word orders (Newport, Gleitman, & Gleitman, 1977; Stromswold, 1995).

(7) What did Amy fix?

(8) I like the bicycle that Amy fixed.

These examples contain a particular type of non-local predicate-argument dependency, which can hold across arbitrarily long distances but cannot cross cer-

tain structures that are "islands" for dependency formation (Chomsky, 1977; Ross, 1967). This dependency can be realized in variable ways within and across languages, as evidenced by the different surface forms of (7) and (8). In "wh-in-situ" languages like Chinese, Japanese, and Korean, wh-dependencies appear on the surface to be quite local. The wh-phrase in (9) is pronounced in the argument position where the predicate normally assigns its thematic relation, although under many accounts it still moves covertly to the edge of the interrogative clause (e.g. Aoun, Hornstein, & Sportiche, 1981; Huang, 1982).

(9) Hufei mai-le shenme (Mandarin Chinese; Cheng, 2003)

Hufei buy-PERF what

'What did Hufei buy?'

Given the variety of forms that wh-dependencies can take, it is not a trivial task for learners to recognize when they are present in sentences they hear. And doing so has consequences for further learning. Recognizing that (7), (8), and (9) contain dependencies of a particular type is necessary for learners to assign them a correct parse, which they will use to make further generalizations about their grammar— inferring, for instance, whether wh-phrases in their language undergo overt displacement or are pronounced in their thematic positions. English-speaking infants begin to produce their first wh-questions in their own speech around the age of 20 months (Stromswold, 1995), so they may be acquiring the relevant grammatical knowledge for parsing and comprehending them in the months leading up to this age. This is moreover an age at which substantial verb learning has already taken

place, motivating a deeper look into how argument structure and wh-dependency acquisition interact.

Prior experimental results suggest that the acquisition of these two phenomena is closely related. Based on a complex pattern of results in a preferential looking task, Gagliardi et al. (2016) argue that infants do not represent the wh-dependencies in wh-questions and relative clauses until the age of 20 months, around the same age that they begin producing wh-questions themselves. Apparent comprehension of these sentences by 15-month-old learners, also found for subject wh-questions in earlier work (Seidl, Hollich, & Jusczyk, 2003), may be due to a "gap-driven" interpretation heuristic driven by argument structure knowledge. For example, infants who know that *hug* is transitive might detect that an expected object is missing after the verb in a sentence like *Which monkey is the frog hugging?* They may then infer that the missing argument is the desired answer to the question, allowing them to succeed in a preferential looking task without representing a fronted phrase as that missing argument. Thus, infants' performance at 15 months may reflect their developing knowledge of verb transitivity, rather than their ability to recognize and parse wh-dependencies.

This account receives some support from the finding that 15-month-olds' performance in a similar preferential looking task depends on vocabulary, a likely indicator of verb knowledge (Perkins & Lidz, under review). This account is also consistent with independent evidence for verb transitivity knowledge emerging around this age. Jin and Fisher (2014) found that 15-month-olds were able to draw inferences about the meaning of a novel verb on the basis of hearing it in a transitive frame, and

Lidz et al. (2017) found that high-vocabulary 16-month-olds predicted an upcoming direct object for a known transitive verb during online sentence processing.

In summary, previous findings suggest the following developmental trajectory. Infants may first acquire some knowledge of verb transitivity around the age of 15-16 months, which enables them to predict arguments of known verbs and detect when they are unexpectedly missing in a wh-question or relative clause. This ability may be developmentally prior to their ability to identify the full structure of wh-dependencies, which by hypothesis emerges around 20 months. However, the evidence from these preferential looking studies is highly indirect. These data demonstrate whether infants can identify a particular image as an answer to a question, but do not diagnose how infants represent the structure of that question in order to arrive at their interpretation. Therefore, these results cannot directly demonstrate that infants' representations of wh-dependencies are immature at early stages of learning, and develop over time.

In Chapter 3, I present a new behavioral experiment that more directly probes infants' syntactic representations of wh-dependencies. I measure whether 14-19-month-olds detect ungrammaticality in wh-questions and declaratives with and without object gaps, in the absence of referential context. This allows us to directly test whether the ability to detect a local transitivity violation is developmentally prior to the ability to recognize a fronted wh-phrase as an argument. In doing so, I bring new empirical evidence to bear on the broader question of how learners represent non-basic clauses in infancy, and what learning mechanisms enable those representations to develop. I now turn to a discussion of those mechanisms.

1.3.2 Filtering Non-Basic Clauses for Verb Learning

If children's acquisition of non-basic clause syntax develops in tandem with their acquisition of verb argument structure, this introduces a chicken-and-egg problem. By hypothesis, identifying the structure of non-basic clauses developmentally follows the acquisition of some core argument structure properties of the language, such as verb transitivity. But if learners cannot accurately identify the structure of non-basic clauses in their language, these clause types will interfere with their attempts to bootstrap into those argument structure properties. How do learners avoid being misled by non-basic clauses in their input, and still arrive at some stable knowledge of verb transitivity?

The solution proposed in the semantic and syntactic bootstrapping literature is that learners' input must be "filtered" in such a way as to boost the signal from basic clauses, in which the correspondence relations between syntax and meaning will hold more reliably (Gleitman, 1990; Lidz & Gleitman, 2004a, 2004b; Pinker, 1984, 1989). In other words, non-basic clauses are somehow filtered out of the data that children use in these bootstrapping inferences. Pinker (1984) proposes two ways that this filtering might happen. The first is for parents to do the filtering and avoid producing these sentences in their children's presence. Parental filtering doesn't seem to occur, as shown by the high rate of wh-questions in the input to young infants (Newport et al., 1977; Stromswold, 1995). The second option is for children to internally filter out non-basic clauses themselves. This approach implicitly assumes that children know which sentences to filter out. But this re-introduces

our chicken-and-egg problem: how do learners identify which clauses are non-basic, if they cannot yet accurately represent the structure of those sentences— if they are still in the process of bootstrapping the basic clause structure and argument structure properties of their language?

In Chapter 4, I resolve this apparent paradox computationally. I present a Bayesian model that learns to filter its input to infer verb transitivity, without knowing what types of sentences it should filter out. Our model does so under the assumption that it occasionally parses sentences erroneously, and it learns which parses to treat as signal and which to treat as noise for the purposes of verb learning. This allows the learner to avoid drawing faulty inferences from non-basic clauses, without having to know which clauses are non-basic. Under this approach, children might filter non-basic clauses from the data they use for verb learning without knowing that they are non-basic clauses, and without needing to infer what the features of non-basic clauses are.

This filtering mechanism fills in a necessary missing piece for understanding how argument structure and non-basic clause acquisition develop in tandem. By inferring a filter on their input, learners may be able to arrive at a stable percept of verbs' syntactic distributions, even though they do not know when those distributions are disrupted on the surface by argument-displacing transformations. Learners can then use that percept for further bootstrapping inferences about the argument structure and clause structure properties of their language, and avoid generalizing in the wrong way from misleading data. More broadly, this mechanism provides a new solution for the problem of learning from immature input representations in

development. By inferring a filter on their input, learners may be able to avoid drawing faulty inferences from input that they cannot yet represent veridically.

1.3.3 Modeling Non-Basic Clause Development

Filtering makes it possible in principle for infants to arrive at a stable representation of verbs' syntactic distributions, which they can then use to bootstrap into other grammatical properties of their language. This puts us in the position to ask: how do infants bootstrap into the properties of non-basic clauses in their language, on the basis of their early syntactic representations? I pursue the hypothesis that the "gap-driven" interpretation heuristic, proposed to characterize infants' early behavior with wh-dependencies, may arise from a more general learning mechanism for non-basic clause syntax (Gagliardi et al., 2016; Perkins & Lidz, under review). On this "**Gap-Driven Learning**" Hypothesis, argument structure acquisition developmentally precedes non-basic clause acquisition in English because the two are causally related. Infants use argument structure knowledge to identify forms of (overt) argument displacement in their language: they use the signal from predicted but unexpectedly missing arguments of verbs to identify when sentences contain non-local predicate-argument dependencies, and to infer what those dependencies are.

Consider again the case of wh-dependencies in English:

(10) What did Amy fix?

(11) I like the bicycle that Amy fixed.

By hypothesis, learners who know that *fix* requires a direct object could detect when it is unexpectedly missing after the verb in sentences like (1) and (2), and infants as young as 15 months may use this ability to infer the intended answer in a preferential looking experiment (Gagliardi et al., 2016; Perkins & Lidz, under review). But as learners' parsing abilities develop, they may be compelled to examine the rest of the sentence to determine the cause of the missing argument, eventually identifying that another expression in the sentence is satisfying the verb's transitivity requirement non-locally. This would allow them both to assign an appropriate parse to the sentence, and to begin to learn how various types of non-local dependencies are realized: e.g. identifying that *what* is a wh-word, which enters into a non-local relation with a predicate by virtue of a wh-dependency.

In Chapter 5, I investigate the computational feasibility of Gap-Driven Learning and ask what specific learning mechanisms it would require. I propose that this learning process follows three logically independent steps: (i) using verb argument structure knowledge to detect predicted but unexpectedly missing arguments of known verbs (gaps); (ii) identifying what surface forms are correlated with argument gaps; and (iii) inferring what types of syntactic dependencies are responsible for those correlations.

Input filtering provides a way for the first step of argument structure learning to be possible in principle, even before learners have gained any further knowledge about the features of non-basic clauses in their language. The second step of learning involves a type of informed distributional analysis: a learner tracks the surface morpho-syntactic properties of sentences that violate her expectations of verb tran-

sitivity, in order to identify clusters of properties that are correlated with those transitivity violations. This would allow the learner to identify the surface features that mark the realization of various types of non-basic clauses in the target language. For example, English learners may identify that sentence-initial functional expressions like *what* are correlated with questions and with argument gaps. Finally, the third step of learning requires inference about the underlying syntactic dependencies that are responsible for those surface feature distributions. For example, English learners must identify which grammatical properties give rise to correlations like those between *what*, questions, and argument gaps: these features mark the realization of one form of argument displacement in English, wh-dependency formation.

In this work, I provide a starting point for specifying this learning process. I instantiate the first two steps of learning computationally, and outline possibilities for the further processes involved in the third step of learning. And although my case study is English, this work bears on broader questions about language acquisition cross-linguistically. Given the highly variable ways in which different dependencies are realized across languages, what mechanisms help learners recognize those dependencies from their surface realizations? Are these mechanisms uniform at some level of analysis, or must they develop ad-hoc in response to the data learners are encountered with? For example, Gap-Driven Learning will only be helpful for wh-dependency acquisition in languages with overt wh-movement; for wh-in-situ languages, learning must proceed differently. At a very broad level, this work demonstrates how two components of learning must work together in tandem.

Learners must have the ability to identify regularities in their data via informed distributional analysis. But the generalizations they draw from those regularities must be guided by a prior over the types of dependencies that grammars make use of, and the consequences they have on the surface forms of sentences.

1.4 Overview

In summary, the work I present here examines the role of transitivity in verb learning as a window into the incremental nature of language acquisition in development. I posit that the acquisition of verb argument structure and transformations like wh-movement are tightly connected, and jointly inform our understanding of how clause structure representations develop in infancy. Children may be able to infer aspects of verb meanings on the basis of their syntactic distributions, but these inferences rely on the ability to recognize the transitivity of a clause, and may involve a sophisticated understanding of the relations between conceptual and linguistic structure. Languages vary in how particular syntactic transformations are realized, but detecting those properties depends on first identifying the language's argument structure profile.

The dissertation is structured as follows. In Chapter 2, I investigate whether verb-learning infants privilege the thematic content of clause arguments in their bootstrapping inferences, over and above the number of clause arguments. In Chapter 3, I ask when infants recognize transitivity in non-basic clauses, focusing on wh-dependencies. Together, these behavioral experiments show that infants privilege

the categories 'subject' and 'object' in their sentence representations when bootstrapping verb meanings and syntax, yet do not reliably recognize these categories when they are realized in non-argument positions.

In Chapters 4 and 5, I investigate the developmental trajectory of argument structure and non-basic syntax. I propose a learning process by which these two phenomena may be acquired in tandem: learners may initially treat non-basic clauses as noise for the purposes of learning verb transitivity, but once confident about those transitivity properties, they may use that knowledge to begin identifying the ways that argument displacement is realized in their language. I show that the first steps of this process are computationally feasible in a language like English. This lays a foundation for further investigation into the ways that learners of any language recover underlying grammatical dependencies from their variable surface realizations.

This case study points towards one way to incorporate development into a theory of grammar acquisition, speaking to two broad questions that such a theory must address. First, what is the nature of learners' earliest representations of their input? Do infants from the onset of syntax acquisition attempt to represent sentences in terms of abstract structure and dependencies over that structure? Second, how do learners use their initially immature input representations to draw further generalizations about their developing grammar, such that they can identify these structures and dependencies more veridically? By investigating the first steps of syntax acquisition in development, this work aims to provide a fuller picture of what portion of the input is useful to an individual child at any single point in

development, how the child perceives that portion of the input given her current grammatical knowledge, and what internal mechanisms enable the child to generalize beyond her input in inferring the grammar of her language.

Chapter 2: Bootstrapping from Transitivity

2.1 Background

Transitivity is robustly correlated with clause meaning cross-linguistically (Fisher et al., 2010; Gleitman, 1990; Hopper & Thompson, 1980; Lidz & Gleitman, 2004a; Naigles, 1990), making it a particularly useful cue for bootstrapping verb argument structure and clause structure properties. As discussed in Chapter 1, exploiting these correlations could allow learners to use a verb's distribution in transitive and intransitive clauses to draw inferences about its meaning and argument-taking properties: this is syntactic bootstrapping (Fisher et al., 2010; Gleitman, 1990; Landau & Gleitman, 1985; Lasnik, 1989). In this chapter, I examine the nature of these early bootstrapping inferences, and ask what transitivity representations they rely on.

A substantial literature shows that young learners are sensitive to the relation between transitivity and clause meaning. In a seminal preferential looking study, Naigles (1990) presented 25-month-olds with a novel verb in the context of two scenes: one intended to be viewed as a causative event of a duck pushing a bunny over, and one intended to be viewed as a non-causative event of a duck and a bunny each wheeling their arms independently. Infants who heard the novel verb

in a transitive clause (*The duck is gorping the bunny*) looked longer at the pushing scene then infants who heard an intransitive clause (*The duck and the bunny are gorping*). This shows that infants were sensitive to clause transitivity, inferring that *gorp* in a transitive frame was more likely to label the causative scene than the non-causative scene. This result has been borne out in extensive additional preferential looking tasks, with infants as young as 19 months reliably showing a preference for a causative scene when they hear a transitive clause (Arunachalam & Waxman, 2010; Arunachalam et al., 2013; Messenger et al., 2015; Yuan & Fisher, 2009; Yuan et al., 2012). 15-month-olds also show this same preference when tested with simplified visual stimuli (Jin & Fisher, 2014).

However, these results do not tell us how infants represent clause transitivity in order to draw these inferences, and what inferences they are drawing. In particular, they do not tell us whether infants who hear a transitive clause represent merely a set of two noun phrases, or whether they differentiate the subject and object of the clause in some manner. These different types of clause representations would enable different bootstrapping inferences. Here, I contrast two main families of proposed bootstrapping strategies. One account proposes a strategy based primarily on the number of arguments in a clause, which are expected to match the number of participants perceived in an event (Fisher et al., 2010; Lidz & Gleitman, 2004a; Naigles, 1990; Yuan et al., 2012). An alternative account proposes a strategy linking particular grammatical and thematic relations, e.g. transitive subject to agent and object to patient (Pinker, 1984; Williams, 2015; Lidz et al., 2017). These accounts implicate different representational resources for the learner at this stage

of development. The number-based bootstrapping strategy only requires infants to recognize and count the number of noun phrases in a clause, whereas the thematic linking strategy requires infants to differentiate those arguments in some way that would allow them to infer their thematic relations.

In this chapter, I introduce a new verb-learning task to differentiate these accounts, asking which strategy an infant will use when they would lead to different inferences about verb meaning. I show that infants at the age previously tested— between 19 and 22 months— rely on the thematic content of clause arguments above and beyond argument number. This provides evidence that learners at this developmental stage do not merely count the arguments of transitive and intransitive clauses, but privilege the grammatical relations of those arguments and the meanings they express.

2.1.1 Bootstrapping from Argument Number

On one account, infants at the onset of verb learning might represent a transitive clause as having two noun phrase arguments, but cannot yet reliably identify which is the subject and which is the object. If this is the case, learners need a way to map from the number of noun phrases they have identified to some properties of scenes in the world, as they perceive them. One influential hypothesis proposes that infants expect the noun phrases in a clause to be arguments, and expect that these will match one-to-one the number of participants in an event the clause describes, as they perceive it (Fisher et al., 2010; Lidz & Gleitman, 2004a; Naigles, 1990; Yuan

et al., 2012):

> We and others have proposed that syntactic bootstrapping originates in a powerful bias toward one-to-one mapping between nouns in sentences and participant-roles in events ... As a result of this bias, children treat the number of nouns in the sentence as a cue to its semantic predicate-argument structure (Fisher et al., 2010).

Or, stated another way:

> Every participant in an event as it is mentally represented shows up as a syntactic phrase in a sentence describing that event (Lidz & Gleitman, 2004a).

For purposes of the current discussion let us call this strategy "Participant-to-Argument Matching" (PAM) (Williams, 2015) (He 2015). This hypothesis accounts for previous experimental findings in the following way. Children who hear a transitive clause like *The duck is gorping the bunny*, and represent it as having two noun phrase arguments, should expect it to describe an event that they perceive as having two participants. If they perceive a pushing scene but not an arm-wheeling scene under a two-participant event concept, then, given these two choices, they should draw the inference that *gorping* most likely describes PUSHING rather than ARM-WHEELING.

PAM is an attractive hypothesis because it provides strong constraints on verb learning, yet is quite simple to implement. If children expect noun phrases to correspond to arguments, then they need only count the number of noun phrases in a sentence in order to infer the participant structure of the event described that

sentence.[1] A sentence with n noun-phrases-as-arguments will describe an event with n participants. Children do not need to access any finer-grained syntactic or semantic information: they do not need to parse the sentence, infer the thematic roles of the arguments, or even know what the nouns mean. Thus, PAM has the potential to be a powerful strategy for children at very early stages of syntactic development and word learning.

PAM's utility in the absence of finer-grained syntactic or semantic information relies on two assumptions. The first assumption is stable perception of both the number of arguments in sentences children hear and the number of participants in events they see. PAM might lead children astray if they mis-identify the number of arguments in a clause (Yuan et al., 2012; Gertner & Fisher, 2012), or if they perceive scenes under event concepts different from those intended by the speaker (Brandone, Addy, Pulverman, Golinkoff, & Hirsh-Pasek, 2006; Pozzan, Gleitman, & Trueswell, 2015), so the hope is that these kinds of mismatches will be rare in early learning.

Furthermore, if children can flexibly shift their representation of an event to a concept with different numbers of participants but the same entailments, then PAM will provide few constraints on which event out of many a speaker's sentence labels (Williams, 2015; Wellwood, He, Lidz, & Williams, 2015). For example, consider what might happen if a child could see a particular 'pushing' scene as either a

[1]While the literature commonly refers to this strategy as 'counting,' it may be more accurately understood as merely a process of identifying the NPs in a sentence and finding an event representation in which those NPs can be put into one-to-one correspondence. That is, infants do not actually need the ability to count; they only need the ability to hold clause arguments in memory in order to pair them with an event representation.

PUSHING with two participants, a pusher and a pushee, or as a PUSHING in which the pusher was still entailed, but not explicitly represented as a participant. That is, suppose a child could switch from representing this 'pushing' scene as PUSHING(x, y), 'a pushing of x by y,' to representing it as PUSHING(x), 'a pushing of x.' I mean these two representations to be mutually entailing, true in all of the same circumstances: they would both be made true by pushings, and differ only in whether something that is necessarily involved in pushings (the pusher) is represented explicitly. If it were possible to "down-shift" in this manner from a 2-place to a 1-place description of an event, while maintaining all of the same entailments of the event concept, then PAM's advice would be much less helpful. A transitive clause might describe the 2-participant representation of this PUSHING event, or any other event that can be suitably down-shifted to a 2-participant representation. An intransitive clause might describe the 1-participant representation of this PUSHING event, or any other event that can be suitably down-shifted to a 1-participant representation. In order for PAM to provide useful guidance about the mapping between verb meanings and events, children's event representations must be stable and similar to adults', and PAM must be an expectation that arguments will match participants in events as we readily perceive them (Williams, 2015; Wellwood et al., 2015).

The second assumption is that PAM is a strong, bi-directional matching hypothesis: arguments in a clause must match participants perceived in an event one-to-one. Consider a much weaker version of this hypothesis: perhaps children expect that clause arguments each name a participant, but do not need to match one-to-one. Under this alternative "Arguments Name Participants" strategy (ANP)

(Williams, 2015), a clause with n arguments should be able to label any event readily perceived with n or more participants. This strategy would only be constraining enough to guide verb learning under very specific circumstances. A learner who hears a new verb in a transitive clause could infer that it does not describe an event seen with only one participant. But a verb in an intransitive clause could describe an event seen with one participant, or two, or three: argument number would provide very little guidance in this case, and inferences about verb meaning would need to be drawn from another source of information. In order for PAM to provide strong enough constraints to guide verb learning in the general case, in the absence of other information, it must be formulated as an expectation of one-to-one matching between participants and arguments, for any number (Williams, 2015).

However, PAM faces some challenges on both theoretical and empirical grounds. Although participants and arguments might frequently align in a language like English, grammars in general do not require one-to-one participant-argument matching even in basic clauses (Williams, 2015). Even in the basic clauses of English, for example, it is not obvious that every perceived event participant is always realized as a clause argument. Imagine a scene in which a girl takes a toy truck from a boy. It is plausible that this scene might be viewed under a 3-participant event concept, one in which the girl, truck, and boy fill core participant roles. But it is also seems entirely acceptable for a transitive clause– *The girl took the truck*– to describe this 3-participant concept, even though only two participants are realized as arguments. But PAM would lead learners astray in this case. Using PAM, learners would erroneously conclude that *The girl took the truck* cannot describe a TAKING event

concept in which a victim or source is also represented as a participant.[2]

Moreover, once we turn to languages other than English, the trend towards one-to-one matching may be weaker. Consider the following example from St'át'imcets:

(1) Qámt kwskwimçxen
hit.with.projectile det.NAME
'Kwimçxen got beaned.' (Davis, 2010)

In St'át'imcets, it is possible to state that someone got hit with a thrown object using a simple intransitive clause, without realizing either the hitter or the thrown object as arguments; this clause is not passive and does not have syntactic null arguments (Davis, 1997; Davis & Demirdache, 2000; Davis, 2010). But if a particular BEANING event is perceived with a hitter, person hit, and thrown object as participants, PAM would erroneously tell learners that a sentence like (1) cannot describe this event. This would lead to incorrect inferences about the meaning of the verb in this sentence. As nearly every verb in this language can occur in a simple intransitive clause (Davis, 1997; Davis & Demirdache, 2000; Davis, 2010), PAM would provide very unhelpful advice for St'át'imcets learners (Williams, 2015).

Thus, because PAM does not generally characterize the relation between clause structure and meaning in adult grammars, its utility is only as an initial learning heuristic, which would need to be abandoned later in development. This raises a

[2]One might think that this problem could be solved in the following way. Perhaps a learner in this situation would assume that any clause arguments required by one-to-one matching, but not realized in the sentence, are syntactically present but silent. However, this solution is circular unless the learner has independent evidence to support this assumption: for instance, evidence that her language has syntactic null arguments, and evidence from the discourse structure to believe that one might be present in this particular sentence, independent of her event representations. The evidence cannot come from knowledge about what event concept the sentence is describing, because that is exactly the knowledge that the learner is trying to acquire.

difficult developmental question: how do children learn to abandon a heuristic based on counting arguments, and acquire a grammar that encodes structure rather than number?

Furthermore, while PAM can account well for the empirical data on transitive clauses, it faces challenges from children's behavior with other clause types. Naigles' (1990) seminal study found that infants who heard an intransitive clause preferred the scene intended to be viewed with one participant (ARM-WHEELING). But beyond this study, further work has found inconsistent behavior with intransitives. Infants who hear novel verbs in intransitive frames do not show a reliably above-chance preference for events intended to be viewed with one participant as opposed to two (Arunachalam & Waxman, 2010; Noble, Rowland, & Pine, 2011; Yuan et al., 2012). Because these results are not predicted under PAM, several methodological explanations have been proposed. Infants may not perceive the presented sentence as intransitive. To control the number of nouns across conditions, several studies used intransitive sentences with conjoined subjects (e.g. *The duck and the bunny are gorping*), which some have argued that infants might mistake for two separate clause arguments (Yuan et al., 2012; Gertner & Fisher, 2012). If so, PAM would guide infants towards the event intended to be viewed with two participants (henceforth "2-participant event"). Alternatively, it is possible that infants do not reliably perceive the presented scenes with the intended number of participants. A scene intended to be viewed as one person pushing another might also be viewed as two people playing (Arunachalam, Syrett, & Chen, 2016; Brandone et al., 2006; Pozzan et al., 2015). If so, then PAM would tell them that this scene could be described by

an intransitive clause.

But note that these results are equally compatible with the weaker number-based strategy, Arguments Name Participants. ANP says that an intransitive clause can describe either a 1-participant event or a 2-participant event, and all things otherwise equal, predicts no preference in matching the sentence to one event over the other. Thus, previous findings can be accounted for under PAM, but do not differentiate it from even this much weaker number-based alternative. The possibility that infants perceive a particular sentence or scene in a different way than experimenters intended makes it challenging to determine which way they relate those sentence and scene representations in bootstrapping.

2.1.2 Bootstrapping from Thematic Content

On another account, learners at this age might use information beyond the number of arguments in a clause, and instead draw inferences by linking the particular grammatical and thematic relations of those arguments. Let us call this family of inferences "Thematic Linking," following Pinker (1984). Suppose that an infant were not only able to identify that a transitive clause like *The duck is gorping the bunny* has two noun phrase arguments, but could also differentiate the subject and the object of that clause in some way. Suppose further that she were aware of the cross-linguistically robust trend for subjects of transitive clauses to name agents and objects to name patients of events (Baker, 1988; Dowty, 1991; Fillmore, 1968; Jackendoff, 1972). This would allow her to infer that this sentence describes an

event perceived with a duck as agent and a bunny as patient: more likely PUSHING than ARM-WHEELING in Naigles' (1990) task. Thus, Thematic Linking strategy predicts the same looking time preferences for transitive clauses as PAM, and is equally compatible with prior preferential looking findings (Arunachalam & Waxman, 2010; Arunachalam et al., 2013; Messenger et al., 2015; Naigles, 1990; Yuan & Fisher, 2009; Yuan et al., 2012).

Existing empirical evidence supports the hypothesis that infants at early stages of verb learning may use finer-grained information about the thematic relations of clause arguments in their bootstrapping inferences. Hirsh-Pasek and Golinkoff (1996) found that 17-month-olds who heard *Big Bird is washing Cookie Monster* preferred to look at a scene where Big Bird was the agent of washing and Cookie Monster the patient, over a scene where the two characters played the opposite roles. These infants differentiated the clause arguments of the known verb *wash* in some way that allowed them to infer the thematic relations of those arguments, and map the sentence to the appropriate event.

Further work shows that 21-month-olds can use the thematic relations of clause arguments in order to infer the meaning of a novel verb. In another preferential looking task, infants who heard *The duck is gorping the bunny* preferred a scene in which a duck pushed a bunny, over a scene in which the bunny pulled the duck (Gertner et al., 2006). Furthermore, infants preferred the duck-agent and bunny-patient event even for sentences like *He is gorping the bunny*: here, they could only rely on the referent of the object because the subject does not identify a unique referent in the discourse. These infants appeared able to link the argument in

subject position to the agent in an event, and the argument in object position to the patient. In this case, it would be unhelpful to attend only to the number of clause arguments. PAM predicts that these transitive clauses must describe 2-participant events, but does not by itself predict which 2-participant events. In order to draw appropriate inferences about the new transitive verbs in this task, children needed to use information not about the number of arguments in a sentence, but about the relationships between the grammatical relations (e.g. subject vs. object)[3] and thematic relations (e.g. agent vs. patient) of those arguments.

For intransitive verbs these relationships are more complicated: the subject of an intransitive clause can label either an agent (e.g. *John baked*) or a patient (e.g. *The bread rose*). These sub-classes of intransitives also display differences in meaning. Intransitives whose subject is an agent tend to label actions of that agent, whereas intransitives whose subject is a patient tend to label changes undergone by that patient (e.g. Fillmore, 1970; Levin & Hovav, 2005; Williams, 2015). Another line of work has asked whether children can draw these finer-grained inferences about verb meanings on the basis of the thematic role of the intransitive subject (Bunger & Lidz, 2004, 2008; Scott & Fisher, 2009). For example, Scott and Fisher (2009) familiarized 28-month-olds with a dialogue in which a novel verb alternated between transitive and intransitive uses. Infants either heard the intransitive with an animate subject (e.g. *Matt dacked the pillow. He dacked*) or an inanimate subject (e.g. *Matt dacked the pillow. The pillow dacked*). At test, infants heard the verb

[3]Note that these results do not tell us how infants differentiate these grammatical relations: whether by using particular structural positions in a clause, or by using their linear order. I return to this point below.

in a transitive frame in the context of two causative scenes: a "caused-motion" event in which a girl pushed a boy over, or a "contact-activity" event in which the girl dusted the boy with a feather duster. Infants who were familiarized with the animate-subject intransitive dialogue mapped this verb to the DUSTING event, whereas infants who were familiarized with the inanimate-subject dialogue mapped it to the PUSHING event. A solely number-based bootstrapping strategy would not have provided the information needed to perform this inference. According to PAM, a verb that alternates between a transitive and intransitive frame can label either a 2-participant or a 1-participant event, but PAM does not by itself predict which 2-participant event out of two a child should prefer. Children could only succeed on this task by using cues to the thematic relation of the intransitive subject: hearing a novel verb in a sentence about a likely agent informed them that the verb labelled an activity of that agent, and hearing a sentence about a likely patient informed them that the verb labelled a result undergone by that patient.

Animacy is a cue that infants are sensitive to from a very young age, making it particularly useful for inferring these thematic relations. Infants within their first year of life are sensitive to animate entities' behavioral characteristics (e.g. self-propelled motion) and physical characteristics such (e.g. eyes, faces, hands), and they expect animate entities to be able to engage in goal-directed action and serve as agents rather than recipients of change (see Rakison & Poulin-Dubois, 2001 for a review). Because agents, which are typically animate, are usually labelled by subjects of basic transitive clauses, subjects tend to be higher in animacy than objects cross-linguistically (e.g. Comrie, 1989). Becker (2015) argues that children

might expect subjects of all clauses to be animate, and treat inanimate subjects as a signal of displacement from a deep object position. Preschoolers appear able to use subject animacy as a cue to the structure of a sentence: 3-year-olds learn a novel *tough*-adjective more easily when it appears with only inanimate subjects (Becker, Estigarribia, & Gylfadottir, 2012; Becker, 2015). The younger children in Scott and Fisher (2009), as well as in other studies on intransitives (Bunger & Lidz, 2004, 2008), may have used animacy to infer whether an intransitive subject is a likely agent or patient, and thus constrain their inferences about what type of event the clause describes.

If children can exploit generalizations between clause positions and thematic roles when the number of arguments in a sentence is uninformative, we might wonder what information fed children's sensitivity to transitivity in previous tests of PAM (e.g Naigles, 1990; Fisher et al., 2010; Yuan et al., 2012). When 20-month-olds infer that a transitive sentence labels a 2-participant event rather than a 1-participant event, are they drawing this inference from the number of arguments in the transitive sentence, or from the roles of the arguments in that sentence? A strategy relying on argument role information appears to explain this behavior just as well as PAM. If children at this age know that subjects of transitives name agents and objects name patients, then they would infer that only the 2-participant event is compatible with the sentence, because only the 2-participant scene shows an event of an agent acting on a patient. Similarly, when infants show no preference for a 1- or 2-participant event when hearing a novel intransitive clause, this may be because they do not know whether the intransitive subject is intended to label an agent of an activity

or a patient undergoing a change. Animacy has not been a useful cue in most prior tests of PAM, where animate actors were used in all participant roles in order to match salience of the 1- and 2-participant scenes. For example, in Yuan et al. (2012), the subject of the intransitive *He is gorping* could be taken to refer to any of the animate, male actors on the screen: either the agent or patient of the 2-participant PUSHING event, or agent of the 1-participant ARM-WAVING event. This sentence would therefore be entirely ambiguous in context, predicting that infants would show no preference for either event.

Thus, it is possible that infants' inferences in prior tests of PAM may have been fed by the thematic content of clause arguments, without counting the number of arguments. Infants' early sensitivity to transitivity in verb learning might be explained without referencing argument number, appealing instead to the more specific relationships between grammatical and thematic relations that children in other studies appear able to use at the same age or several months later. And unlike PAM, there are generalizations about these relationships that are cross-linguistically robust (Baker, 1988; Dowty, 1991; Fillmore, 1968, 1970; Jackendoff, 1972; Levin & Hovav, 2005). To recap our discussion above, here is one way of summarizing these linking principles, adapted from Williams (2015):

(2) (a) The arguments of a clause name participants in its event.

(b) Subjects of basic transitive clauses tend to name agents and objects tend to name patients.

(c) A clause describing a change tends to realize the patient undergoing

change.

(d) A clause describing an action tends to realize the agent of that action.

If learners were aware of these linking principles, they would provide advice that is more specific but also more flexible than PAM. The principle in (2a) is a restatement of ANP, and only tells learners that they need to identify a participant role for each argument in a clause: arguments are not likely to be expletive. A learner using (2b) would infer that a transitive sentence must describe an event with both an agent and a patient, but with no commitment that these be the only participants that they represent. This would allow them to accept a transitive description of a 3-participant TAKING (e.g. *The girl took the truck*), provided that learners perceive that event with an agent named by the subject and a patient named by the object. A third perceived participant, such as the source or victim, need not be mentioned. And, a learner using (2c-d) would infer that an intransitive sentence will either describe a change or an action, depending on the thematic role of its subject: a clause whose sole argument is a patient more likely describes a change, and a clause whose sole argument is an agent more likely describes an action. This means that a bare intransitive could describe a 3-participant BEANING, if that event is seen as a change and the intransitive subject names the patient of that change. Thematic Linking therefore allows learners to generalize to more cases within and across languages. Because these principles are widely applicable to adult language, children would not need to abandon them over the course of development as has been suggested for PAM.

It is possible that a strategy like Thematic Linking could co-exist with PAM, and learners make use of either strategy when convenient (Fisher et al., 2010). It is also possible that a number-based bootstrapping strategy like PAM is a heuristic for learners at very early stages of syntactic development, and is eventually replaced by more sophisticated bootstrapping strategies once infants have acquired richer grammatical knowledge of their language (Fisher, 1996; Lidz & Gleitman, 2004a). That is, PAM might be a primary bias that guides young learners before they can bootstrap the clause representations that are needed to use a strategy like Thematic Linking. The linking principles in (2) are cross-linguistically robust because they are stated in terms of the abstract high-level grammatical categories 'subject' and 'object.' To exploit these principles, a learner would need to know something about how these categories are realized in their language.[4] Note that this strategy does not require that these categories be represented at the same level of richness as in the adult grammar, or be recognized accurately in all cases. A learner may still be able to employ these linking principles by using rougher and less accurate proxies for these grammatical relations, such as linear word order: for example, an English learner might infer that a pre-verbal noun phrase is a likely subject and a post-verbal noun phrase is a likely object. However, this means that learners would need to know the relevant cues that they can use to identify likely subjects or likely objects in their language. If they have not acquired this knowledge by the time they

[4]It should be noted that these linking principles hold over 'subject' and 'object' as loose descriptive categories, whether or not these exist as categorical primitives in the adult grammar. Exploiting these principles only requires a learner to access the high-level abstractions under which the principles are stated, whatever the specific syntactic relations are that variously exemplify them in the adult grammar.

are learning their first verbs, it might be that PAM is the only strategy available.

In summary, because prior results are compatible with both PAM and Thematic Linking, they do not tell us whether a number-based strategy is primary or co-exists with a thematic content-based strategy in early verb learning. To differentiate these alternatives, we would need to know whether infants privilege the numbers of arguments and participants in their bootstrapping inferences, and whether they do so in the youngest age range at which they have been argued to use PAM. Because PAM and Thematic Linking can account for the same behavior in prior tasks, our test case must be one in which these strategies would lead to different inferences about verb meanings. One way to do this is to move beyond the case of 2-argument sentences and 2-participant scenes, and test how infants behave with other numbers of arguments and participants.

In the current study, I differentiate these alternatives with 19- to 22-month-old infants, the youngest age that reliably succeeds in prior bootstrapping tasks. I ask whether infants at this age will allow a 2-argument description of a TAKING event that is perceived under a 3-participant concept. As previously discussed, transitive *take* is acceptable in adult English. But an infant using PAM should not come to this same conclusion. If she perceives a scene in which a girl takes a truck from a boy under a concept in which the girl, truck, and boy are all participants, then she would expect that scene to be described by a 3-argument clause. If she instead hears a novel verb in a 2-argument clause— *The girl pimmed the truck*— then PAM will tell her that this verb cannot describe the scene as she readily perceives it. It must instead describe a nearby 2-participant concept: perhaps it describes the girl's

moving or grabbing of the truck, or another sub-event in this scene where the boy is not a perceived participant. The learner should conclude that *pimmings* cannot be TAKINGS, but instead are more likely MOVINGS or GRABBINGS.

On the other hand, Thematic Linking predicts no difficulty taking *The girl pimmed the truck* as a description of a 3-participant TAKING concept. If a learner can identify *the girl* as a likely subject and *the truck* as a likely object, then she will infer that this sentence can describe any event in which she perceives a girl as agent and a truck as patient, regardless of participant number. If she views the stimulus scene as a TAKING with those participant roles, she can then infer that this clause describes that entire 3-participant event. It doesn't matter that the perceived third participant, the boy, isn't mentioned. Under this account, *pimmings* can be TAKINGS, and this may be the most likely conclusion for a learner to draw if she readily views this scene under that conceptual description. Thus, pitting a 2-argument sentence against a 3-participant event concept will allow us to tell whether infants privilege argument number above argument roles in their verb learning inferences at this age.

2.1.3 Diagnosing Event Representations

The current test case requires an independent diagnostic for how infants readily view the presented stimulus scene, in the absence of language. If we do not know the number of participants that infants readily perceive in a particular scene, then we cannot tell whether that conceptual representation matches or mismatches the

number of arguments that infants hear in a particular sentence. This is one reason why it is difficult to reason about infants' behavior with intransitive clauses in prior tests of PAM (Arunachalam et al., 2016; Brandone et al., 2006; Pozzan et al., 2015). Thus, identifying infants' bootstrapping strategies is only possible if we first fix the conceptual representations under which infants perceive the scenes in our task.

Any particular event concept entails many relations. To describe an event as a TAKING entails that it has an agent of taking, a patient that is taken, a victim or source, some manner of transfer, a duration of transfer, a particular location of taking, and so on. But surely not every relation entailed by a predicate corresponds to an argument of that predicate. The same could be said about our psychological representations of predicates. Not all entailed relations will be on par with each other psychologically, because any particular psychological representation will foreground some of those relations and background others (Williams, 2015). Thus, not all of the relations that are entailed when we describe something as a TAKING will be psychologically privileged to the same extent; only some will be explicit in the conceptual structure under which we represent that TAKING event. We might call the entailed relations that are psychologically privileged the participant relations (Williams, 2015). Our question is, for any given stimulus scene, which relations are privileged in the concept that infants view it under?

There are many possible ways of writing down the structure under which infants could view a particular TAKING event. Here are just some of these possibilities:

(3) (a) TAKING(e) & AGENT(e, x) & PATIENT(e, y) & SOURCE(e, z)

(b) TAKING(e) & AGENT(e, x) & PATIENT(e, y)

(c) TAKING(e) & AGENT(e, x)

Infants might view a particular TAKING event under a 3-place conceptual representation that privileges the agent, patient, and source as participants, or under a 2-place representation that privileges e.g. only the agent and patient, or under a 1-place representation that privileges only the agent— and there are many other options. These different possibilities for the structure of an event representation exist independently of the entailments of the event concept. Infants may represent a scene under a concept that entails a source, without necessarily representing the source as a participant that is explicit in their conceptual structure. So, in order to diagnose the number of participants that infants represent, we need to determine which relations they view as not only entailments, but also as entailments that are psychologically privileged.

Recent work has developed a new method for diagnosing the structure of infants' event representations (He, 2015; Knowlton, Perkins, Williams, & Lidz, 2018; Perkins, Knowlton, Williams, & Lidz, 2018; Wellwood et al., 2015), building off of a task introduced by Gordon (2003). In this study, 10-month-old infants were habituated to a silent scene that could be plausibly seen as a GIVING event: a girl gave a toy to a boy. At test, infants either saw another token of the same scene, or they saw a scene in which the motion of the two actors was constant but the toy was no longer present: the girl approached the boy with empty hands. In a control condition, a separate group of infants were habituated to a silent scene in

which the girl hugged the boy while holding the toy. At test, the same manipulation was performed. Infants either saw another token of the hugging-with-toy scene, or they saw the girl hug the boy in the same way without holding the toy. This same manipulation was found to affect infants' attention differently in the two conditions. Infants dishabituated (recovered attention) to the disappearance of the toy from the giving scene, but not from the hugging scene.

These results may be explained in the following way. If infants in the 'giving' condition represented the stimulus scene as a GIVING that entails a gift, then the change at test is a large conceptual change: there is no longer a toy filling that entailed relation. But if infants in the 'hugging' condition represented the stimulus scene as a HUGGING, the toy does not fill an entailed relation, and is merely incidental. Every HUGGING has an agent of hugging and a patient being hugged, but the agent need not be holding an object while hugging. The change at test is therefore a smaller conceptual change: the hugging-without-toy scene can still be described as a HUGGING with all of its entailed roles present. Infants' dishabituation in the 'giving' condition but not the 'hugging' condition might be understood as sensitivity to these different conceptual entailments. Because the same physical manipulation— removing a toy— led to different patterns of dishabituation for the two types of events, we might infer that infants considered this toy more central to one of these events than to the other. All else being equal, and controlling for other differences like physical salience, maybe this difference arose because infants represented the giving scene as a GIVING and the hugging scene as a HUGGING, with the toy filling an entailed relation in the one case and not in the other.

Thus, when the design is controlled in the right way, infants' patterns of attention to a change in a stimulus scene can provide a window into the conceptual representation under which they view that scene. However, the result from Gordon (2003) doesn't tell us whether infants in this study were responding merely to a change in conceptual entailments, or moreover to a change in participant structure. It is possible that infants viewed this scene under a 3-place representation in which the toy fills a participant role, along with the girl and the boy. But it is also possible that they viewed this scene under a concept that entails something given, without explicitly representing that relation as a participant. There are many other entailments of GIVINGS that might not be explicitly represented as participants, such as their locations, times, and manners. To tell whether the toy is filling a participant relation, we need to know whether the relation is not only entailed, but is also psychologically privileged.

Wellwood et al. (2015) and He (2015) adapted the habituation-switch task in order to address this question. In one experiment, infants were habituated to a silent scene in which a girl opened a box using a lever: we might see this as a JIMMYING event. At test, infants saw one of two types of change. In one condition, infants saw the girl now open the box with her hand; the lever was still present and visible, but no longer used as an instrument of opening. In a second condition, infants saw the girl still open the box using the lever, but from the right instead of from the left. All JIMMYINGS have a direction of opening, so this also represents a change to one of the entailed relations of the event concept. This change was larger perceptually, if measured by differences in pixels. However, infants dishabituated only when the

lever was no longer used as an instrument, and not to the perceptually more salient change in direction. This pattern was also observed when the lever was added as an instrument at test: infants dishabituated when the girl switched from opening the box with her hand to opening it with the lever, but not when she opened the box with her hand from the opposite direction. Infants' differential response to these two types of changes suggests that they viewed the lever as filling a more psychologically potent relation than the direction of opening— that the instrument relation was privileged in their event representation.

In Knowlton et al. (2018) and Perkins et al. (2018), we extended this paradigm to norm infants' representations of a TAKING scene. 9- to 12-month-old infants were habituated to several tokens of a silent video in which a girl picked up a toy truck and moved it towards herself, while a boy looked on. By hypothesis, this event is viewed as a 2-participant PICKING-UP. At test, infants in one condition saw the girl move the truck towards herself in a different manner: she slid the truck across the table instead of picking it up. This different motion changes an entailment of the event concept; perhaps this is now seen as a SLIDING rather than a PICKING-UP. However, despite this different motion, the participant structure should remain constant. Infants in the second condition saw the girl still pick up the truck, but with the boy now gripping it. By hypothesis, this should be seen as a 3-participant TAKING: the girl is now taking the truck from the boy. Infants dishabituated to both changes, but showed greater dishabituation when the boy was added as the possessor of the truck than when the manner of motion was changed. Despite the fact that the motion change is highly salient perceptually, infants cared more about

the change to the boy's role in the event. This suggests that infants viewed the TAKING scene under a concept with the boy filling a privileged relation, one that was more psychologically potent than the manner of motion.

This result gives us evidence for the conceptual structure under which infants view this particular TAKING scene, independent of language. Among the possibilities laid out in (3), it is now more likely that infants readily view this scene under a concept privileging the victim or source of taking. Under the assumption that infants also privilege the agent of taking and the patient being taken, this means that they view this scene under a 3-participant conceptual structure:

(4) TAKING(e) & AGENT(e, girl) & PATIENT(e, truck) & SOURCE(e, boy)

Additional work might take further steps to confirm this event representation. But given this initial evidence, we are now now in a position to ask what principles verb-learning infants use for mapping a sentence to this stimulus scene. Do infants expect that the number of arguments in a clause must match one-to-one the participants they perceive in this TAKING event (PAM), or do they deploy a more flexible strategy linking particular grammatical and participant relations (Thematic Linking)? In Study 1, I show that 19- to 22-month-old infants think that this 3-participant event concept can be described by a 2-argument sentence. In a control experiment (Study 2), I show that infants do not think that this event can be described by a particular 1-argument sentence. These results demonstrate that infants do not expect participants to match arguments in number, arguing against a strict number-based bootstrapping strategy. But they do draw different inferences

from transitive and intransitive clauses, demonstrating that they are sensitive to the meanings that can be expressed by these different clause types. Taken together, these results support Thematic Linking over PAM: infants at this age privilege the grammatical and thematic relations of arguments above argument number in their verb learning inferences.

2.2 Experiment 1: Transitive Frame

2.2.1 Method

Experiment 1 tested whether 19- to 22-month-old English learners allow a 2-argument clause to describe our TAKING scene perceived under a 3-participant concept. A novel verb learning task was adapted from the Verb Extension paradigm (Waxman, Lidz, Braun, & Lavin, 2009). Infants were familiarized to the TAKING stimulus scene: a girl takes a toy truck from the boy. This scene was described by a transitive clause containing a novel verb: *The girl pimmed the truck*. On the basis of these familiarization trials, infants should make an inference about what concept *pimming* describes. We can then test which inference they made by asking what else will count as an instance of *pimming* for them.

At test, infants were prompted to find *pimming* in the context of two candidate videos. One showed the girl still taking the truck from the boy (another token of the TAKING scene). The second showed the girl moving the truck towards herself in the same way, but without the boy present (a MOVING/GRABBING scene). By measuring infants' looking preference towards TAKING, we can determine whether they inferred

that *pimming* in a transitive frame could describe the entire 3-participant event concept under which they readily view that scene, or whether they inferred that it must describe a 2-participant sub-event involving only the girl and the truck.

Participants. Participants included 24 typically-developing infants (12 males) between the ages of 19;0 and 21;28 (mean = 20;8). Participants were recruited from the greater Washington, D.C. area with the criteria that they heard English during at least 80% of their waking hours. An additional 4 infants were tested but not included in the final sample due to inattentiveness (2), equipment malfunction (1), or less than 80% English exposure (1). Participants' total productive vocabulary was collected by parental report using the Words and Sentence MacArthur-Bates Communicative Development Inventory (MCDI) (Fenson et al., 1993). Mean total words produced were 160.5, with a median of 99.

Materials. Visual stimuli were created by filming two live actors, a girl and a boy, performing actions with inanimate objects. Different tokens of each event were filmed and edited in Adobe Premiere to create the trial structure in Table 2.1. Six tokens of each main event were presented: four during familiarization, one during the contrast phase, and one at test. Event tokens were edited to be 7.5 seconds in duration during the familiarization and contrast phase, and 5.5 seconds during the test phase. The different tokens of TAKING used in the critical trial were created to be almost identical to the TAKING scene normed in prior work (Knowlton et al., 2018; Perkins et al., 2018), with one difference: in the current study, the girl takes the truck from the boy by sliding it across the table. Adult piloting found this to be a more natural 'taking' motion than lifting the truck out of the boy's grip.

Phase		**Audio**	**Video**
Familiarization (30s) 4 scenes		*Look, the girl is gonna pim the truck! She just pimmed the truck!* *Wow, she's gonna pim it again! She just pimmed it again!* *Ooh, she's gonna pim the truck! She just pimmed the truck!* *Hey, she's gonna pim it again! She just pimmed it again!*	Girl takes truck from boy x 4
Contrast (15s) 2 scenes	Negative Positive	*Uh-oh, she's not gonna pim that. She didn't pim that.* *Yay, she's gonna pim the truck! She pimmed the truck!*	Girl pokes ring tower held by boy Girl takes truck from boy
Test (16.5s) Split-screen scenes looped 3 times	Baseline Response	*Now look, they're different!* *Find the one where she's pimming the truck. Where is she pimming the truck?*	Girl takes truck from boy / Girl moves truck without boy x 3

Table 2.1: Structure of Experimental Trial, Experiment 1

Audio stimuli were recorded by a female native speaker of American English using child-directed speech. Stimuli were edited in Adobe Audition and Praat, and synchronized with the video stimuli using Adobe Premiere. Because prior work finds mixed results in verb extension tasks depending on whether infants hear pronominal arguments or full lexical NPs (Arunachalam & Waxman, 2011, 2015). During familiarization, audio was timed to frame the main action in the video: a future-tense sentence (*The girl is going to pim the truck!*) ended as the action began, and a past-tense sentence (*The girl just pimmed the truck!*) began as soon as the action ended. At test, sentence onset was timed to coincide with the beginning of each looped video.

Procedure. Infants sat on a parent's lap or a high chair positioned 6 feet

away from a 51" widescreen television. Parents were instructed not to talk to their children or direct their attention, and they either looked away from the screen or wore a visor to block their view. Stimuli were presented using QuickTime. A camera located above the television was used to video-record the experiment at a capture rate of 29.97 frames per second. The camera's pan and zoom were controlled by an experimenter watching a video monitor in a separate room to ensure that an infant's face remained in the frame during the experiment.

Each experiment lasted 5.6 minutes. The structure of the experiment is as follows. Infants were first introduced to the two actors: the girl and the boy. Each actor appeared on a different side of a black screen for 7 seconds, waving and smiling. In order to familiarize infants with the descriptions for these actors, they were named by a full lexical NP and two pronouns (e.g. *Look, it's a girl! Do you see her? There she is!*). Next, the actors appeared for 15 seconds in split-screen. Infants were prompted to first find the girl, and then find the boy.

After actor introductions, infants saw 4 trials, each consisting of a familiarization phase, a contrast phase, and a test phase (Table 2.1), following Waxman et al. (2009). During the familiarization phase, infants saw four 7.5-second video tokens of a scene appearing on different sides of the screen. Screen side presentation was pseudo-randomized in two lists, and counterbalanced across participants. Each video was described by two sentences containing a verb in a particular syntactic frame (e.g. *Look, the girl is gonna pim the truck! She just pimmed the truck!*).

During the contrast phase, infants saw a new 7.5-second video labelled as a negative example of the verb, followed by another token of the familiarization scene

labelled as a positive example. This phase was included following Waxman et al. (2009), who found that it facilitated infants' recognition that the novel verb has a specific meaning. So as not to bias infants' attention towards any particular aspect of the familiarization scene, the negative contrast video involved the same human actors but was otherwise different across many dimensions: it introduced a new action, with a different manner of motion, being performed on a new object (e.g. the girl pokes a ring tower that the boy is holding, causing it to rock slightly). This scene was described as not involving the verb of interest, using negative, downcast intonation (e.g. *Uh-oh, she's not gonna pim that.*). The positive contrast video was another token of the familiarization scene, described by the same verb using positive, upbeat delivery (e.g. *Yay, she's gonna pim the truck!*). Each of the contrast videos was played on a different side of the screen, with screen side counterbalanced across participants.

During the test phase, two 5.5-second videos were presented concurrently, on loop, on different sides of the screen: another shortened video token of the familiarization scene (e.g. the girl takes the truck from the boy) and a new scene (e.g. the girl moves the truck, without the boy). The timing of the motions in the two videos was edited to be consistent. The screen side of the familiar vs. new video was counterbalanced across participants, and further counterbalancing controlled for whether it matched or mismatched the side on which the last contrast video had appeared. The two test videos were first accompanied by uninformative audio (*Now look, they're different!*). This allowed infants to examine both videos before the test sentence, and was intended to reveal any baseline preferences for one of the two

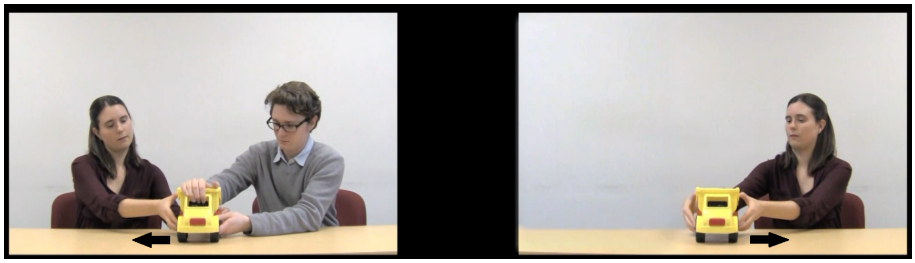

Figure 2.1: Image of Experimental Test Videos, TAKING (right) vs. MOVING (left)

videos that may have been driven by differences in salience. The videos then played on loop twice more, accompanied by two prompts to find the verb of interest (e.g. *Find the one where she's pimming the truck. Where is she pimming the truck?*). The videos then disappeared to a black screen, and a new trial began. Fig. 2.1 shows sample still images from the test videos in the experimental trial. In each video, the truck moves away from the center of the screen.

Of the 4 trials, the first 3 consisted of training trials with known verbs, and the last was the experimental trial with the novel verb *pim*. This type of single-trial verb learning design is common in other prior bootstrapping tasks (e.g. Yuan & Fisher, 2009; Scott & Fisher, 2009), and training with with known verbs has been found to facilitate infants' familiarity with the experimental procedure before the novel verb is introduced. So as not to bias infants to one particular argument structure during training, these three trials used a ditransitive, transitive, and intransitive frame. The verbs *give*, *shake*, and *open* were chosen because they ranked high in familiarity norming for 20-month-olds in the web-based WordBank database (Frank, Braginsky, Yurovsky, & Marchman, 2016). Each training trial had the same structure as the experimental trial in Table 2.1. A sample sentence and the events for these trials are provided in Table 2.2. Infants were assigned to one of two lists in order to

Sample Audio	**Familiarization**	**Contrast**	**Test**
The girl is gonna give the toy to the boy!	Girl gives stuffed owl to boy	Girl throws ball, boy watches	Girl gives owl to boy / Girl hugs boy, holding owl
The girl is gonna shake the bottle!	Girl shakes bottle of red liquid	Girl spins toy rattle	Girl shakes bottle / Girl taps lid of bottle
The box is gonna open!	Girl opens box	Girl lifts toy house	Girl opens box / Girl tilts box over

Table 2.2: Training Trials, Experiments 1 and 2

counterbalance the order of training trials across participants; in one list, the order of training trials was reversed. However, the experimental trial was always last in the experiment. To focus infants' attention, trials were interleaved with either a 4-second still image of a baby face with audio of a baby giggling, or a 14-second video of moving toys accompanied by music.

Predictions. The two bootstrapping strategies predict that infants will draw difference inferences about the meaning of the novel verb during the familiarization phase of the experimental trial, leading to different looking preferences at test. If infants at this age are primarily using a number-based bootstrapping strategy that requires one-to-one matching between clause arguments and perceived event participants (PAM), we make the following predictions. During the familiarization phase, infants who hear *pim* in a 2-argument clause— *The girl pimmed the truck*— will infer that it describes an event that they perceive with exactly 2 participants. But given prior evidence that infants perceive the TAKING scene under a 3-participant concept (Knowlton et al., 2018; Perkins et al., 2018), this means that there is a mismatch. This sentence cannot describe the entire TAKING event, as infants readily perceive

it. Instead, PAM will lead them to infer that the sentence describes a sub-event that they represent with only 2 participants, such as the girl's moving or grabbing of the truck. Thus, *pimmings* are not TAKINGS, they are MOVINGS or GRABBINGS. On the basis of this inference during familiarization, infants are predicted to accept both of the test videos as possible instances of *pimming*. Both the TAKING and the MOVING video show the girl moving the truck towards herself in the same way, so if *pimmings* are movings or grabbings, we predict no above-baseline preference for one test video over the other.

The alternative hypothesis is that infants are using a more flexible bootstrapping strategy linking particular grammatical and participant relations (Thematic Linking), e.g. transitive subject to agent and object to patient, with no expectation of one-to-one matching. This hypothesis makes the following predictions. During the familiarization phase, infants who hear *The girl pimmed the truck* will infer that it describes an event that they perceive with the girl as agent and the truck as patient. This sentence can therefore describe the 3-participant TAKING concept under which they view the familiarization scene, provided that they view the girl and the truck as filling those respective participant relations. Given that infants most readily perceive this scene as a TAKING, and the clause arguments can link appropriately to the participant relations that they represent in this event, we predict that they will infer that *pimmings* are TAKINGS; the sentence gives them no reason to conclude otherwise. On the basis of this inference during familiarization, infants are predicted to only accept a TAKING event as an instance of *pimming* at test. This means that they should show an above-baseline preference for the TAKING

video over the MOVING video.

2.2.2 Results

Data Preparation. The videotaped recordings of the test phase for each experimental trial were coded offline by an experimenter with the audio turned off. EyeCoder software was used (Fernald, Zangl, Portillo, & Marchman, 2008) in order to advance each muted video frame-by-frame and code whether infants were looking at the left or right side of the screen, or neither. Data were coded by two experimenters, with intercoder reliability established to be above 90% (Cohen's Kappa > 0.90).

Analysis. Frame-by-frame analysis calculated whether each infant was looking at the TAKING video, the MOVING video, or neither. These looking preferences were then averaged across participants to create a timecourse of the proportion of looks towards the TAKING video at each frame, out of looks towards either video. To determine whether looking patterns were conditioned on the test prompt, two window of analysis were selected within this timecourse. The Response window spans the 3 seconds after the offset of the first presentation of the novel verb *pim*, allowing us to measure infants' preferences after they were asked to find *pimming*. The Baseline window spans the 3 seconds prior to this point, allowing us to measure any baseline preferences before infants were asked to find *pimming*.

The looking timecourse is plotted in Fig. 2.2. The offset of the novel verb is marked by a vertical line. The shaded gray region represents the selected windows of

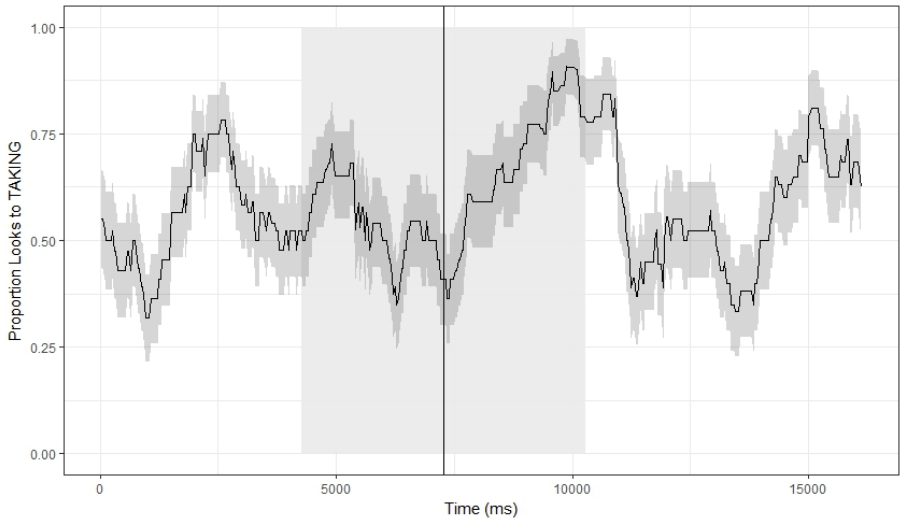

Figure 2.2: Timecourse of Proportion Looks to TAKING Video, Exp. 1

analysis. Visual inspection reveals no sustained preferences for either video during the Baseline window before the novel verb offset, but a strong and sustained preference for the TAKING video that emerges during the Response window immediately after the offset of the novel verb.

Looking times during the Baseline and Response windows were averaged for analysis. Infants' mean proportion of looking time towards the TAKING video during the Response window ($M = 0.71, SE = 0.06$) was found to be significantly higher than during the Baseline window ($M = 0.54, SE = 0.04$) in a two-tailed paired-sample t-test ($t(23) = 2.27, p < 0.03$). To examine whether looking times were affected by age or vocabulary, a simple linear regression was conducted to predict each infant's baseline-corrected looking preferences (mean Response looking time - mean Baseline looking time) by age in days and log-transformed total vocabulary, as reported on the MCDI. No significant relationship to age or vocabulary was found. Thus, infants on average showed a significantly above-baseline preference for the

TAKING video during the 3-second window after being asked to find *pimming*, and this preference was independent of age or vocabulary level.

While it is standard in the literature to analyze pre-selected windows conditioned on the linguistic stimulus, it is also known that this practice may obscure effects that emerge on an earlier or later timescale than originally expected (Delle Luche, Durrant, Poltrock, & Floccia, 2015). To control for this possibility, a second analysis was conducted over the entire trial to independently search for the time window(s) in which infants' proportion of looks to the TAKING video differed significantly from chance, defined here as 50%. The type of analysis used is called cluster-based permutation (Dautriche, Swingley, & Christophe, 2015; De Carvalho, Dautriche, & Christophe, 2016; De Carvalho, Dautriche, Lin, & Christophe, 2017; Maris & Oostenveld, 2007; Von Holzen & Mani, 2012). In this analysis, a one-sample t-test is first conducted at each timeframe in order to identify clusters of adjacent frames with t-values greater than a pre-determined threshold value (here, $t = 2.01$). The sum of the t-values within each cluster yield the size of the cluster. In order to control for multiple comparisons, the probability of observing a cluster of a particular size is calculated by comparing against a null distribution. This distribution is bootstrapped from the largest clusters observed in 1,000 random permutations of the data. Looking times during a time cluster are considered significantly different than chance ($p < 0.05$) if a cluster of that size is larger than 95% of the clusters in the bootstrapped null distribution.

The cluster-based permutation analysis was conducted over the length of the entire trial using the eyetrackingR package (Dink & Ferguson, 2015). A significant

time window was found between 8.90 and 10.98 seconds from trial onset (1.63 and 3.71 seconds from novel word offset) in which infants' proportion looking towards the TAKING video was significantly above chance ($p < 0.001$). No other significant time windows were found, including in the time before novel word offset. This is consistent with the results found in the standard analysis. Infants showed no baseline preference for either the TAKING or the MOVING video before the test prompt, but they showed a significant preference for TAKING within the approximately 3 seconds after being asked to find *pimming*.

2.2.3 Discussion

The results from Experiment 1 are inconsistent with the hypothesis that 19- to 22-month-olds expect the number of arguments in a clause to match one-to-one the number of participants they perceive in an event (PAM) (Fisher et al., 2010; Lidz & Gleitman, 2004a; Naigles, 1990; Yuan et al., 2012). Infants in this experiment did not seem affected by the mismatch between a 2-argument clause (*The girl pimmed the truck*) describing a TAKING event that they readily perceive with 3 participants (a girl took a toy truck from a boy). When asked to find *pimming* at test, infants showed above-baseline looking preferences for another token of the 3-participant TAKING event. They dispreferred a 2-participant MOVING event, in which the girl moved the truck towards herself in the same way, but without the boy present. It seems that this event was a less likely instance of *pimming* for these infants, given their inference about this verb's meaning on the basis of

familiarization. Instead, these infants appeared to consider another TAKING event to be a much better instance of *pimming*: they appeared to infer that the transitive clause described the entire 3-participant TAKING concept under which they readily viewed the familiarization scene.

This behavior is counter to the predictions of PAM. Infants using PAM should conclude that the transitive clause describes a 2-participant sub-event in the TAKING scene, like MOVING or GRABBING, in which event participants match clause arguments in number. This prediction was not confirmed, suggesting that infants were not relying on number-matching in this task.

These results are more consistent with the hypothesis that infants at this age are able to link particular grammatical and thematic relations in verb learning (Thematic Linking) (Pinker, 1984; Williams, 2015; Lidz et al., 2017). Under this hypothesis, infants are predicted to allow a transitive description of a 3-participant event concept, provided that the grammatical relations they represent in the clause can link to appropriate thematic relations in their event representation; arguments and participants need not match one-to-one. Infants' behavior in this task is consistent with the possibility that they used a strategy like Thematic Linking in their inference about the meaning of *pim*. When hearing *The girl pimmed the truck*, they may have linked the transitive subject to an agent relation and the object to a patient relation in their representation of the TAKING event. Because these relations align in the right way, this would enable them to infer that the transitive clause describes the entire 3-participant concept under which they readily viewed this event, with no need to shift to a 2-participant concept instead. Using this boot-

strapping strategy, infants would infer that *pimmings* can be TAKINGS, and most likely are. Thus, Thematic Linking can provide an explanation for why infants in this experiment preferred the TAKING scene at test.

However, these results are also consistent with alternative explanations. Perhaps infants in this task did draw inferences from argument number rather than thematic content, but used a weaker number-based bootstrapping strategy than PAM. These results would be consistent with the weakest such strategy that I discussed above: Arguments Name Participants (ANP). Infants may have merely expected every argument in the transitive clause to name a participant in their representation of the TAKING scene. With no expectation of one-to-one matching, this would allow them to infer that *pimmings* can be TAKINGS. Alternatively, an even weaker possible strategy would be one that does not involve syntactic bootstrapping at all. These results are consistent with the possibility that infants ignored the syntax of the familiarization sentence altogether, and relied entirely on their conceptual representation— the TAKING concept under which they readily viewed the familiarization scene— to draw inferences about the novel verb. That is, perhaps infants in this task expected that any novel verb presented in the context of this scene must describe TAKING, regardless of the verb's syntactic frame.

A control experiment is therefore needed to determine which alternative to PAM was guiding infants in this task: whether they are using a thematic content-based strategy like Thematic Linking, as compared to a weaker number-based strategy like ANP or no syntactic bootstrapping at all. We can differentiate these possibilities by testing how infants behave when they are familiarized to the TAKING

scene described by a different syntactic frame, one that Thematic Linking predicts will not be a good fit for a 3-participant TAKING event concept. One such case is an unaccusative intransitive frame: *The truck pimmed.* If infants can use the inanimacy of this intransitive subject to draw the inference that it is a likely patient, and furthermore know the types of meanings that intransitives with patient subjects are likely to express, this will lead them to draw different inferences about the likely meaning of *pim* than infants in Experiment 1. Recall the linking principles in (2c-d), repeated here as (5):

(5) (a) A clause describing a change tends to realize the patient undergoing change.

(b) A clause describing an action tends to realize the agent of that action.

A learner aware of these linking principles would expect that a clause whose sole argument is a patient is more likely to describe a change undergone by that patient, and less likely to describe some other action by an unmentioned agent. By this reasoning, *The truck pimmed* in the context of the familiarization scene is more likely to describe some aspect of the truck's motion, and less likely to describe the girl's taking of the truck from the boy. In other words, because we do not naturally see TAKINGS as changes spontaneously undergone by the thing being taken, verbs describing TAKINGS are unlikely to distribute in unaccusative frames. Thus, if infants can use a strategy like Thematic Linking to constrain their inferences about the meaning that this type of intransitive clause can express, they should infer that intransitive *pimmings* are not TAKINGS but are more likely MOVINGS in the

context of this scene.

By contrast, the two weaker alternative hypotheses predict that infants will still infer that intransitive *pimmings* are TAKINGS. If infants are bootstrapping primarily from argument number using the weaker ANP strategy, then they should infer that *The truck pimmed* can describe any event in which they perceive a truck as a participant. Given that they perceive the truck as a participant in the TAKING event, this clause should be able to describe the entire 3-participant TAKING concept under which they view this event. Thus, ANP predicts that infants will infer that intransitive *pimmings*, like transitive *pimmings*, are most likely TAKINGS. The no-syntax strategy makes the same prediction. If infants are ignoring the syntax of the clause altogether, and relying only on their conceptual representation of this scene as a TAKING, then they should also infer that intransitive *pimmings* are TAKINGS. Thus, we can differentiate Thematic Linking from these two weaker alternative accounts by determining whether infants who hear *pim* in an intransitive frame continue to infer that it describes a 3-participant TAKING event.

2.3 Experiment 2: Intransitive Frame

Experiment 2 tested a second sample of 19- to 22-month-old English learners to determine whether infants will allow a 1-argument clause to describe our TAKING scene perceived under a 3-participant concept. Using the same design as in Experiment 1, infants were familiarized to the TAKING scene now described by an intransitive clause: *The truck pimmed*. At test, infants were prompted to find *pim-*

ming in the context of the same two candidate videos, TAKING and MOVING. I ask whether infants continue to prefer the TAKING video after intransitive familiarization, or whether they now think that the MOVING video is a possible instance of *pimming*. This will illuminate whether infants in this task are using the syntactic context of the novel verb, and are moreover sensitive to the different meanings that can be expressed by intransitive vs. transitive frames.

Participants. Participants were 24 typically-developing infants (13 males) between the ages of 19;1 and 21;25 (mean = 20;3). Participants were recruited from the greater Washington, D.C. area and heard English during at least 80% of their waking hours. An additional 4 infants were tested but not included in the final sample due to inattentiveness (3) or less than 80% English exposure (1). Participants' mean total productive vocabulary was 107 words, with a median of 68.5, from parental report on the MCDI (Fenson et al., 1993). Although the mean total vocabulary in this sample appears qualitatively lower than the mean of the sample in Experiment 1, this difference was not found to be statistically significant (Welch's $t(39.11) = 1.31, p < 0.20$).

Materials and Procedure. The same procedure and task design was used as in Experiment 1, but with different audio stimuli during the critical trial. Infants in Experiment 2 heard the novel verb *pim* in an intransitive frame during the experimental trial (see Table 2.3). Audio stimuli were recorded by a female native speaker of American English using child-directed speech, and were edited in Adobe Audition and Praat and combined with the video stimuli in Adobe Premiere. The experimental trial followed the same structure as in Experiment 1, using identical

visual stimuli. This trial was preceded by the same character introductions and 3 known-verb training trials as in Experiment 1.

Phase		**Audio**	**Video**
Familiarization (30s) 4 scenes		*Look, the truck is gonna pim! The truck just pimmed!* *Wow, it's gonna pim again! It just pimmed again!* *Ooh, the truck is gonna pim! The truck just pimmed!* *Hey, the truck is gonna pim again! It just pimmed again!*	Girl takes truck from boy x 4
Contrast (15s) 2 scenes	Negative Positive	*Uh-oh, that's not gonna pim. That didn't pim.* *Yay, the truck is gonna pim! The truck pimmed!*	Girl pokes ring tower held by boy Girl takes truck from boy
Test (16.5s) Split-screen scenes looped 3 times	Baseline Response	*Now look, they're different!* *Find the one where the truck is pimming. Where is the truck pimming?*	Girl takes truck from boy / Girl moves truck without boy x 3

Table 2.3: Structure of Experimental Trial, Experiment 2

Predictions. If infants at this age are using a bootstrapping strategy like Thematic Linking, and can draw inferences from the thematic content of an intransitive subject, then we make the following predictions. During the familiarization phase, infants who hear *pim* in an intransitive clause with an inanimate subject— *The truck pimmed*— should infer that the intransitive subject is likely to be a patient, and that the clause is likely to describe some change undergone by that patient. But given that infants readily see the familiarization scene as a taking of a truck by a girl from a boy, and not primarily as a change undergone by the truck, this means that the sentence is unlikely to describe the entire 3-participant TAKING

event as they readily perceive it. Instead, it is more likely to describe the motion of the truck, a sub-event in this scene. Intransitive *pimmings* are therefore not likely to be TAKINGS, but more likely MOVINGS. On the basis of this inference during familiarization, infants are predicted to accept both the TAKING and the MOVING video as possible instances of *pimming* at test, because both show the truck moving in the same way. Thematic Linking thus predicts no above-baseline preference for one video over the other.

Under the two alternative hypotheses that infants are using a less sophisticated verb-learning strategy, we make different predictions. One alternative hypothesis is that infants expect that every argument in a clause should name a participant they perceive in an event (Arguments Name Participants, ANP), with no one-to-one matching or further inferences about the thematic content of those arguments. Under ANP, infants who hear *The truck pimmed* during familiarization should infer that this sentence describes any event in which they perceive a truck as a participant. This sentence can therefore describe the 3-participant TAKING concept under which they view the familiarization scene. Given that this is how they readily view the scene, we predict that they will infer that intransitive *pimmings* are TAKINGS, and will show an above-baseline preference for the TAKING video at test.

The same prediction is made by the even weaker alternative hypothesis that infants ignore the familiarization syntax altogether, and infer verb meaning from their conceptual representations alone. Under this no-syntax hypothesis, infants will infer that any sentence paired with the familiarization scene describes the 3-participant TAKING concept under which they readily view this scene. This should

also lead them to conclude that intransitive *pimmings* are TAKINGS, and to prefer the TAKING video at test. Thus, this control experiment does not differentiate between the two alternative accounts for infants' behavior in Experiment 1, but will tell apart a strategy based on thematic content from these two alternatives.

2.3.1 Results

Data Preparation. The videotaped recordings of the test phase for each experimental trial were coded frame-by-frame in the same way as in Experiment 1. Data were again coded by two experimenters, with intercoder reliablity established to be above 90% (Cohen's Kappa > 0.90).

Analysis. Infants' average proportions of looks towards the TAKING video at each frame, out of looks to either video, were calculated in the same way as in Experiment 1. The looking timecourse is plotted in Fig 2.3. The shaded gray region marks the same two windows of analysis as in Experiment 1: the Baseline window spanning the 3 seconds prior to the offset of the novel word (marked by a vertical line) and the Response window spanning the 3 seconds after novel word offset. Visual inspection reveals no sustained preferences for either video during either of these test windows. There appears to be a short-lived increase in looks to the TAKING video that begins in last second of the Response window; however, this apparent uptick is delayed and much more modest than was seen in Experiment 1, where infants oriented strongly towards TAKING immediately after novel word offset.

Looking times during the Baseline and Response windows were averaged for

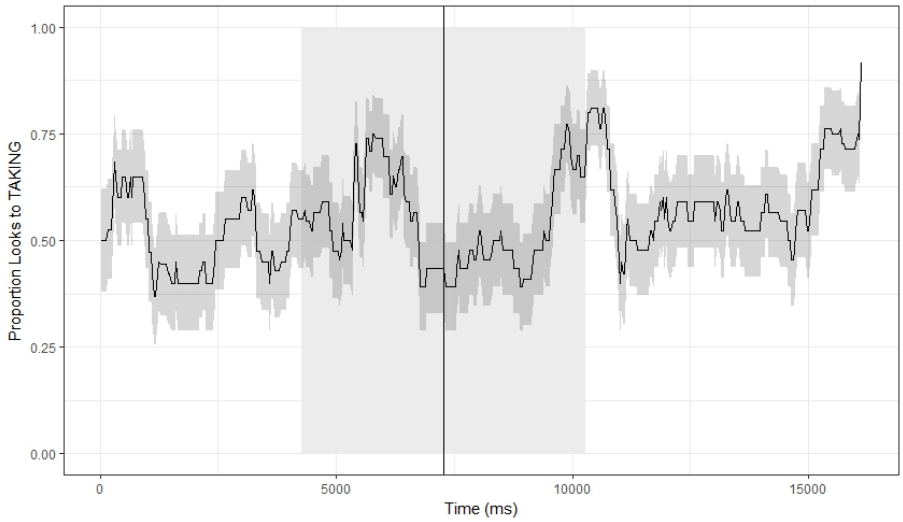

Figure 2.3: Timecourse of Proportion Looks to TAKING Video, Exp. 2

analysis. Infants' mean proportion of looking time towards the TAKING video during the Response window ($M = 0.60, SE = 0.06$) and during the the Baseline window ($M = 0.53, SE = 0.06$) were not significantly different from each other ($t(23) = 0.93, p < 0.36$). Additionally, the infants' differences in looking preferences from Baseline to Response were not found to be significantly predicted by their age in days or their log-transformed total vocabulary. Thus, infants on average did not show a significantly above-baseline preference for the TAKING video in the 3-second window after being asked find *pimming*, regardless of age or vocabulary.

One might be concerned that these predetermined 3-second test windows could obscure effects that emerge on a different timescale than predicted; in particular, these windows may not capture the short increase in looking towards TAKING that visually appears to emerge about 2 seconds after novel word offset in the timecourse in Fig. 2.3. To control for this possibility, a cluster-based permutation analysis was again conducted over the entire trial length to identify the time windows in

which proportion looks to TAKING were significantly different from chance (50%). This analysis was conducted in the same manner as described for Experiment 1. No significant time windows were found, either before or after novel word offset. The apparent short increase in looks towards TAKING at the end of the Response window did not reach the significance threshold ($p < 0.21$). Thus, infants' looking preferences appeared to fluctuate throughout the trial, but did not significantly differ from chance in any time window before or after they were asked to find *pimming*.

2.3.2 Discussion

The results from Experiment 2 are consistent with the hypothesis that infants in our task are drawing inferences from the thematic content of clause arguments (Thematic Linking) (Pinker, 1984; Williams, 2015; Lidz et al., 2017). When familiarized with an intransitive clause (*The truck pimmed*) paired with our TAKING scene, infants did not appear to infer that this clause described the entire 3-participant concept under which they view that event. When asked to find *pimming* at test, infants showed no above-baseline looking preferences for another token of the 3-participant TAKING event. Instead, they looked to the same extent at this event and at a 2-participant MOVING event. It seems that these two events were equally likely instances of *pimming* for them. This behavior is consistent with a strategy like Thematic Linking. If infants can use the likely thematic relation of the intransitive subject to constrain their inferences about what type of event the clause describes, then they will infer that *The truck pimmed* more likely describes some aspect of the

truck's motion than the girl's taking of the truck from the boy. The truck's motion is the same in both the TAKING and MOVING scenes, so infants who have inferred that *pimmings* are MOVINGS are predicted to show no preference for either of the two test videos. The current results confirm this prediction.

Taken together with the results of Experiment 1, this control experiment also helps differentiate a bootstrapping strategy based on thematic content from a weaker alternative. If infants in this task drew inferences about verb meaning solely based on their conceptual representation of the familiarization scene as a 3-participant TAKING, without using the syntactic context of the novel verb, then they would infer that any novel verb presented in the context of this scene describes TAKING. This account predicts that infants will always show an above-baseline preference for the TAKING scene at test, regardless of whether they heard transitive or intransitive familiarization. However, infants only showed this preference for transitive *pimming* in Experiment 1, and not for intransitive *pimming* in Experiment 2. This demonstrates that infants attended to the syntactic context of the novel verb in our task.

These results are also inconsistent with the possibility that infants used a weaker number-based strategy to bootstrap from that syntactic context. If infants merely expected that each argument in a clause should name a participant in the clause's event (ANP), then they would allow either a 2-argument or a 1-argument clause as a description of an event seen with 3 participants, provided each clause argument names one of those participants. This account thus predicts the same behavior by infants who hear either *The girl pimmed the truck* or *The truck pimmed*

as a description of an event viewed as a TAKING with both the girl and the truck as participants: infants should infer that *pimmings* are TAKINGS in both cases, and show above-baseline preferences for the TAKING scene at test. This prediction was not confirmed by the results of Experiment 2. This shows that it did not suffice for these infants that the arguments of a transitive or intransitive clause named some of the participants in their TAKING event representation. Instead, it mattered which types of arguments were present: infants drew different inferences about clause meaning when *the truck* was a transitive object, compared to when it was the sole argument of an intransitive.

Thus, Experiments 1 and 2 together show that infants use the syntactic context of the novel verb in our task, and did so in a way that was not predicted by a strictly number-based bootstrapping account: infants took a transitive clause but not an intransitive clause to describe an event seen with 3 participants. These results are inconsistent with the hypothesis that infants expect clause arguments to match one-to-one the perceived event participants, and are also inconsistent with the weaker hypothesis that this matching is not one-to-one, but that it suffices for each clause argument to name a perceived event participant. These results are predicted by Thematic Linking, under which infants may have used information about the likely thematic relations of the arguments in a clause to draw fine-grained inferences about the types of events that the clause was likely to describe: whether it primarily described a change undergone by a patient, or an action of an agent.

However, I note that infants' lack of preference in Experiment 2 leaves open other possibilities for characterizing their sensitivity to the differences between tran-

sitive and intransitive clauses. For example, perhaps infants at this age merely know that intransitives cannot describe a TAKING event seen with 3 participants, without knowing what other types of events they can describe. If infants in our task identified that intransitive *pimmings* were not likely to be TAKINGS, but did not know what other type of events they could be, this confusion might also result in chance behavior at test. This possibility may be less likely given prior evidence that slightly older infants can draw fine-grained inferences about the meanings of verbs in different intransitive frames (Scott & Fisher, 2009; Bunger & Lidz, 2004, 2008), but further investigation would be needed to show that infants are able to perform these types of inferences at the age currently tested, and indeed do so in our task. Nonetheless, the current experiments show that infants' bootstrapping strategies at this age go beyond an expectation of one-to-one correspondence between clause arguments and event participants, or even merely an expectation that clause arguments name event participants. Instead, infants in our task appeared to use some form of more sophisticated information about the different meanings that transitive and intransitive clauses can express.

2.4 General Discussion

In this chapter, I investigate the way that infants at early stages of grammar learning use clause transitivity to draw inferences about verb meanings. Prior work has proposed that young infants are guided by a bias towards one-to-one matching between the number of arguments they hear in a clause and the number of partici-

pants they represent in an event (PAM) (Fisher et al., 2010; Lidz & Gleitman, 2004a; Naigles, 1990, 1996). This strategy would provide strong constraints on verb learning from only the number of arguments in a clause, without requiring learners to differentiate the grammatical and thematic relations of those arguments. However, previous results taken in support of number-based bootstrapping (Arunachalam & Waxman, 2010; Arunachalam et al., 2013; Messenger et al., 2015; Naigles, 1990; Yuan & Fisher, 2009; Yuan et al., 2012) are also compatible with a finer-grained bootstrapping strategy. Perhaps infants even at early stages of verb learning can differentiate subjects and objects of clauses in some manner, and are sensitive to the cross-linguistically robust links between those grammatical relations and the thematic relations they express (Thematic Linking) (Pinker, 1984; Williams, 2015; Lidz et al., 2017). Here, I aim to address the following open question: when bootstrapping from argument number vs. thematic content would lead to different inferences about verb meaning, which strategy will infants fall back on?

The two experiments presented above show that verb-learning infants at the age previously tested— between 19 and 22 months— do not fall back on one-to-one number matching. Infants in Experiment 1 did not behave as if they expected clause arguments to match event participants in number: they inferred that a novel verb in a transitive clause described an event that they viewed under a 3-participant concept. Moreover, infants in Experiment 2 drew different inferences when that novel verb appeared in an intransitive frame. These infants inferred that the verb was unlikely to describe that same 3-participant event concept, even though the subject of the intransitive clause (*The truck pimmed*) named the same participant

as the object of the transitive clause in Experiment 1 (*The girl pimmed the truck*). Thus, infants' bootstrapping in this task did not appear to be driven primarily by the numbers of clause arguments and perceived event participants. Instead, infants appeared to differentiate the arguments of transitive and intransitive clauses in a manner that allowed them to draw finer-grained inferences about the types of meanings that those clauses could express.

These results can be explained by appealing to the same sensitivities to thematic content that have been attested in prior experimental findings (Gertner et al., 2006; Bunger & Lidz, 2004, 2008; Scott & Fisher, 2009). If infants expected that subjects of transitive clauses name agents and objects name patients, as shown in other work with 21-month-olds (Gertner et al., 2006), this would allow them to link the grammatical relations of the arguments in the transitive clause to the participant relations that they perceived in the TAKING event. But if they inferred that the subject of the intransitive clause is likely to be a patient, and knew that a clause whose sole argument is a patient is likely to describe a change undergone by that patient, then they would conclude that this clause is an unlikely description of an event seen as a TAKING. This is not the only possible way that infants may have arrived at a different meaning for the intransitive clause in the current study, but it is consistent with prior evidence that 22- to 28-month-old infants can perform this type of inference (Bunger & Lidz, 2004, 2008; Scott & Fisher, 2009). Thus, the current results are consistent with other findings that infants at this age or a little later can bootstrap verb meanings from finer-grained information about the grammatical and thematic relations of clause arguments.

Moreover, this work contributes a novel finding: infants at this age rely on this finer-grained information above and beyond the information gained solely from the number of clause arguments. Both PAM and the more flexible Thematic Linking strategy were in principle available to infants in this task, but would lead to different inferences about verb meaning. By showing that infants did not prefer to use PAM, this work argues against the claim that a bias towards one-to-one matching between participants and arguments is the primary or privileged bootstrapping strategy at this age (Fisher, 1996; Fisher et al., 2010; Lidz & Gleitman, 2004a). Because infants in this task instead drew more sophisticated inferences from the types of arguments in a clause, it may be that prior verb learning results from the same age range were driven by a similarly sophisticated bootstrapping strategy, rather than PAM (Arunachalam & Waxman, 2010; Arunachalam et al., 2013; Messenger et al., 2015; Naigles, 1990; Yuan & Fisher, 2009; Yuan et al., 2012). It is still possible that PAM is a primary guide for verb learning in even earlier stages of development; further investigation with infants younger than 19 months is needed to determine whether this is the case. However, at least by the age at which infants most reliably succeed at novel verb-learning tasks, it appears that they do not adhere tightly to PAM in their verb-learning inferences, and instead use information about argument relations in a more flexible way than number-matching would allow.

Further investigation is needed to determine the exact information about argument relations that infants were using in this task, particularly in the case of intransitive clauses. In order to more precisely characterize infants' inference about intransitive *pimming*, it would be necessary to differentiate whether they only infer

that *pimmings* are unlikely to be TAKINGS, or whether they also infer that this verb describes some aspect of the truck's motion by virtue of having *the truck* as subject. Future work may address this issue by testing whether infants will consider other events with the same motion as possible *pimmings*, but reject events in which the truck is affected in a different way. This will allow for a more precise understanding of how infants at this age map between their representations of events and the argument relations in an intransitive vs. transitive clause.

More broadly, the results in this chapter illuminate the sophisticated ways that infants relate linguistic and conceptual representations in early grammar learning. To bootstrap verb meanings in the current experiment, infants needed to map between two structures: the structure under which they perceived the sentences in our task, and the structure under which they viewed the stimulus scene. By using a stimulus scene whose conceptual representation had been normed in prior work (Knowlton et al., 2018; Perkins et al., 2018), we were able able to fix the conceptual side of the equation in order to diagnose this mapping process. This step has not explicitly been taken in previous work, in which uncertainty over infants' conceptual representations has left open questions about which bootstrapping strategy was being used (Arunachalam et al., 2016; Brandone et al., 2006; Pozzan et al., 2015). Having first diagnosed the conceptual structure under which infants readily view the scene in our task, we are able to show the circumstances under which infants map a sentence to this readily available conceptual representation, and the circumstances under which they shift to a different concept. Infants were able to link the argument relations in a transitive clause in to appropriate participant relations in this event

representation, and preferred to do so even despite the mismatch between argument and participant number. But infants did not allow an intransitive description of this event concept, and instead shifted away from their initial scene representation. This provides further evidence for the independent contribution of both conceptual and linguistic structure in bootstrapping. Infants attempt to relate a sentence to the conceptual representations under which they readily perceive the world around them, but simultaneously take the linguistic form of the sentence as evidence for which conceptual representation is being tokened (Gleitman, 1990).

Finally, this work has implications for how infants represent the arguments in a clause when drawing inferences about verb meaning. If infants' bootstrapping inferences at this age are primarily driven by the grammatical and thematic relations of clause arguments rather than argument number, this implies that those relations are somehow privileged in their clause representations. That is, infants at this age do not merely track the number of noun phrase arguments they hear; instead, they differentiate those arguments in some fashion in order to draw inferences about their likely thematic relations, and thus draw further inferences about clause meaning. Infants in this task drew different inferences about transitive subjects and objects than they did about intransitive subjects, and these inferences were strikingly similar to those that adults would draw on the basis of the robust links between grammatical and thematic relations that exist cross-linguistically (Williams, 2015). This would be explained if infants' clause representations privilege categories like 'subject' and 'object'— the high-level abstractions under which these linking principles are stated.

Thus, it is possible that infants even at early stages of verb learning may be at-

tempting to recover these syntactic categories in their input, with argument number serving only as a proxy for doing so. The current work does not tell us how richly these categories might be represented: whether they are represented *qua* subjects and objects, in an adult-like hierarchical clause structure, or whether they are represented in a rougher and less accurate way. It is also an open question how infants learn to recognize these categories in their language, a question I will continue to return to in the following chapters. But by pointing towards clause representations in which subjects and objects are asymmetrically differentiated in some manner, the current work bring us one step closer to understanding the representational resources that infants bring to the task of bootstrapping into the grammar of their language.

Chapter 3: Representing Transitivity in Non-Basic Clauses

3.1 Background

Chapter 2 focused on the basic resources that infants bring to the task of representing arguments in a clause, in order to draw inferences about the meanings and argument structure of new verbs. I provided new evidence that by the age of 20 months, infants are bootstrapping from clause representations that differentiate the grammatical and thematic relations of clause arguments, and privilege these relations over and above the number of noun phrase arguments in sentences they hear.

But this does not mean that infants always succeed at identifying these relations when they are present. Infants do not only hear basic, active, declarative sentences like those in (1) and (2); they also have to contend with so-called "non-basic" clauses like the wh-questions in (1), in which arguments have been displaced from their canonical positions.

(1) John ate a sandwich. Amy fixed her bicycle.

(2) John ate. (*Amy fixed.)

(3) What did John eat? What did Amy fix?

The subject and object of a clause might be harder to recognize in sentences like (1), in which an argument of the verb is not pronounced in an argument position. Adult English speakers know that *what* acts as the verb's object by virtue of a particular type of non-local dependency. But in order to understand that these sentences have objects, learners need to identify that this dependency is present. If a child doesn't yet know the form of wh-dependencies in her language, she might think that these sentences are intransitive. This would be very misleading for verb learning: she might think that *fix* can freely occur in an intransitive clause, like *eat*, with consequences for what she thinks it means.

The case of wh-questions serves to illustrate a much broader problem for bootstrapping in early grammar learning. Bootstrapping allows young learners to draw inferences about grammar and meaning by relating syntactic representations of subjects and objects in sentences to conceptual representations of events. But these inference depend on learners recognizing subjects and objects when they are present, and may fail if other linguistic properties interfere with learners' abilities to recognize those core clause arguments (Gleitman, 1990; Lidz & Gleitman, 2004a, 2004b; Pinker, 1984, 1989). This problem was first noted by Pinker in his earliest work on semantic bootstrapping, following Keenan (1976):

> One must place an important proviso, however, on the use of semantic information to infer the presence of syntactic symbols, especially grammatical relations. Keenan argues that the semantic properties of subjecthood hold only in what he calls "basic sentences": roughly, those that are simple, active, affirmative, declarative, pragmatically neutral, and minimally presuppo-

sitional. In nonbasic sentences, these properties may not hold. In English passives, for example, agents can be oblique objects and patients subjects, and in stylistically varied or contextually dependent sentences the agent can be found in nonsubject positions (e.g., *eats a lot of pizza, that guy*). Thus one must have the child not draw conclusions about grammatical relations from nonbasic sentences (Pinker, 1984).

In summary, the meaning-distribution relations that learners rely on for bootstrapping may be disrupted in sentences with argument displacement. Examples include not only wh-questions, but also relative clauses (5) and passives (6):

(4) What did Amy fix?

(5) I like the bicycle that Amy fixed.

(6) The bicycle was fixed (by Amy).

In each of these examples, an expression acting as the verb's object is being realized in a non-canonical object position, rather than after the verb. Adult English speakers know that these are all instances of particular dependencies that hold non-locally between a predicate and an argument, and they know that these dependencies have particular locality properties: some of them, like wh-objects and relative clauses, contain the same types of dependencies (wh- or A-bar dependencies) and others, like passives, are fundamentally different (A-movement).

Children, however, must learn the shapes that these various dependencies take in their language. Until they have learned this, non-basic clause types will provide

misleading data for bootstrapping, whether syntactic or semantic. A child who does not know that *what* and *the bicycle* are displaced objects in the sentences above will relate these sentences improperly to her perceptions of events in the world, resulting in faulty inferences about grammar and/or meaning. A semantic bootstrapper who takes these sentences to be descriptions of causative events might conclude that these sentences are evidence for base object-initial word order in English, rather than evidence for the transformations that actually produced this non-canonical word order. Or if these phrases are not even recognized as arguments, these sentences might be erroneously taken as evidence that *fix* can take an implicit object, or that English has syntactic null objects. Conversely, a syntactic bootstrapper who fails to perceive the transformations in these sentences might draw erroneous inferences about verb properties: she might conclude that *fix* does not take a direct object and therefore "fixings" are not likely to be causative events, or she might conclude that *fix* belongs to a class of verbs that can alternate between transitive and intransitive uses, like *eat* or *rise*.

In summary, non-basic clauses introduce a chicken-and-egg problem for early grammar acquisition. If learners had a way of identifying the structure of these sentences, and in particular to recognize transitivity when it is present, then they could avoid drawing faulty inferences about clause structure and verb argument structure on the basis of these data. On the other hand, if learners knew the argument structure of verbs in these sentences— that a verb like *fix* requires an object, and does not allow object-drop— then they could use that information to identify when that object has been displaced from its canonical position. Which comes first, non-basic

clause acquisition or argument structure learning? And how do learners find their way out of this chicken-and-egg problem? In this chapter, I focus on the first of these questions, the 'which' question. Using a novel method to probe the wh-dependency representations of English-learning infants, I show that the answer appears to be argument structure: infants' ability to detect local verb transitivity violations is developmentally prior to their ability to recognize a displaced argument in a wh-object question. This new evidence for the developmental trajectory of non-basic clause acquisition lays the empirical groundwork for examining the second learning question, the 'how' question, in Chapters 4 and 5.

3.1.1 Acquiring Wh-Dependencies

My case study is wh-dependencies, which are among the most common non-basic clause types in child-directed speech. English-learning children hear a large number of wh-questions even before their second birthday (around 15% of their total input), the majority of which contain non-canonical word orders (Newport et al., 1977; Stromswold, 1995). These dependencies are also present in relative clauses, which are rarer in the speech to young children but share similar structural properties (Chomsky, 1977). Both cases involve a particular type of non-local dependency between a fronted argument and the thematic position where the argument is interpreted. The length of this dependency can hold across arbitrarily long distances, as in (7), but cannot cross certain structures that are islands for dependency formation (8) (Chomsky, 1977; Ross, 1967).

(7) (a) What did Jake believe that Susan claimed that John ate?

(b) I made the sandwich that Jake believed that Susan claimed that John ate.

(8) (a) *What did Jake believe Susan's claim that John ate?

(b) *I made the sandwich that Jake believed Susan's claim that John ate.

Unlike in English, in some languages wh-dependencies do not on the surface appear to involve displacement. In "wh-in-situ" languages like Chinese, Japanese, and Korean, wh-phrases are pronounced in their thematic position, although on many accounts they still take scope in a higher clausal position by undergoing covert movement that happens to be inaudible (e.g. Aoun et al., 1981; Huang, 1982)[1]:

(9) Hufei mai-le shenme (Mandarin Chinese; Cheng, 2003)

Hufei buy-PERF what

'What did Hufei buy?'

Given the variety of forms that wh-dependencies can take, there are several problems that children must solve in order to be able to recognize them in their input. Children need to learn whether their language has overt or covert wh-movement, which surface forms in their language signal that movement has taken place, and (for languages with overt movement) how to identify the thematic position from which movement occurred. In English, surface signals for wh-movement include

[1]Some have also argued for non-movement accounts of wh-in-situ, such as binding by a covert operator (Reinhart, 1998), or for different wh-in-situ representations across different languages (Cole & Hermon, 1994). See Cheng (2003) for an overview.

wh-forms (e.g. *what*), subject-auxiliary inversion, do-support, and relativizers (e.g. *that*). Adult speakers make use of these signals efficiently in sentence processing to identify a displaced argument (a "filler") and predict upcoming gaps, or thematic positions where it could be interpreted ("active gap filling") (Aoshima, Phillips, & Weinberg, 2004; Crain & Fodor, 1985; Frazier & d'Arcais, 1989; Frazier & Clifton, 1989; Sussman & Sedivy, 2003; Traxler & Pickering, 1996). But because these signals are language-specific, children must learn them. And in order to arrive at the correct interpretation of wh-dependencies, children furthermore must identify the particular relation that holds between the moved argument and a non-local predicate— to identify the thematic position where the moved argument originated. This is not a trivial task, as wh-movement gaps are phonologically null in languages like English.

3.1.2 Hypothesis: Gap-driven Learning

On one proposal, children begin to identify wh-dependencies by detecting when a phrase stands in relation to a verb that is locally missing a predicted argument (Gagliardi et al., 2016). In other words, although mature parsing of these dependencies is filler-driven, the acquisition of these dependencies may be gap-driven. For example, a child who knows that *fix* requires a direct object might detect that it is missing after the verb in a sentence like *What did Amy fix* __ ? She may then be driven to examine the sentence for cues to what happened to this unexpectedly missing argument, and start learning the signals of the wh-dependency that is re-

sponsible: for example, that *what* is an argument wh-word and that do-support can occur in wh-object questions. The detection of a direct object gap will also allow the child to correctly interpret this particular wh-dependency, interpreting *what* as questioning some unknown patient of fixing by relating it to the position where *fix* assigns that thematic role.

If correct, this proposal has implications for verb learning. Learners need to know which verbs require direct objects, in order to notice when those arguments are needed and missing. In other words, learners need to detect that sentences like *What did Amy fix* __ ? contain direct object gaps, rather than intransitive uses of these verbs— but in order to do so, they must know which verbs are transitive. This account therefore posits that learners use verb argument structure knowledge to drive wh-dependency acquisition, rather than the other way around.

This proposal stands in contrast to another logical alternative: that wh-dependency acquisition is actually filler-driven. Under this alternative, the first step for the learner is to cluster the wh-words in her language into an equivalence class, by tracking function words that appear at clause boundaries in questions. The second step is to label this cluster *as* the set of wh-words in her language, and to identify that these words stand in particular non-local relationships with a predicate. Having done so, the learner might then use the presence of a wh-word to facilitate argument structure acquisition: an argument wh-word signals an upcoming argument gap, and therefore enables the learner to differentiate direct object gaps from intransitive uses of verbs. This proposal therefore posits that verb transitivity acquisition occurs after learners identify at least some of wh-dependencies

in their language.

This alternative account may be feasible, but is not straightfoward. Despite their orthography, wh-words do not have a signature morphology. The forms [hu] 'who' and [wʌt] 'what' do not have a single phoneme in common. Distributional analysis may be more informative for identifying these words: Mintz, Newport, and Bever (2002) found that an algorithm that clustered words based on their immediately preceding and following sentence environments in child-directed speech was able to cluster the set of English wh-words when tested on the Nina corpus (Suppes, 1974), although it did not appear to identify such a cluster for the Peter corpus (Bloom, 1970) in CHILDES (MacWhinney, 2000)[2]. But even if learners can identify a cluster of sentence-initial question words in the language, it is not trivial to identify these *as* wh-words. Many languages have question particles that aren't wh-words, but can appear at sentence boundaries in both wh- and polar questions. An example is the particle *la* in Tz'utujil Mayan (10). A Tz'utujil learner needs a way to tell that *la* is a question particle and not a wh-word, and conversely an English learner needs a way to tell that *what* is a wh-word and not a question particle.

(10) La xwari ja ch'uuch'? (Tz'utujil Mayan; Dayley, 1981

Q slept the baby

'Did the baby sleep?'

Furthermore, supposing a learner can identify the set of wh-words in her language, it is less clear how she would identify the particular non-local relationships

[2]Note that the goal of this computational model was not to identify closed-class categories like wh-words, but rather to use closed-class items to help identify lexical categories like nouns and verbs.

that they participate in without using verb transitivity information. Because adjunct wh-words (like *when* and *where*) do not predict upcoming argument gaps, a learner cannot use the mere presence of a wh-word to infer that a verb is occurring with a displaced argument, rather than intransitively. Instead, the learner would need to rely on the semantics of particular wh-words to determine which of them are questioning unknown arguments of predicates, and which are questioning times, locations, manners, and reasons. This introduces a new puzzle, which is how a child identifies the semantics of these words— particularly, how a child determines which are the argument wh-words, without first knowing whether they stand in relation to an argument gap.

The gap-driven and filler-driven learning hypotheses make different empirical predictions. Under gap-driven learning, because wh-dependency acquisition depends on learning verb transitivity, it should come developmentally later. Under filler-driven learning, the reverse is true: wh-dependency acquisition facilitates verb transitivity learning, so it should come at the same time or developmentally earlier.

3.1.3 Prior Experimental Results

Previous preferential looking studies suggest that infants' abilities to identify wh-dependencies in English develops in tandem with basic argument structure knowledge. English-learning infants as young as 15 and 16 months old show sensitivity to verb transitivity: Jin and Fisher (2014) found that 15-month-olds are able to draw inferences about the meaning of a novel verb on the basis of hearing it

in a transitive frame, and Lidz et al. (2017) found that high-vocabulary 16-month-olds predicted an upcoming direct object for a known transitive verb during online sentence processing. However, wh-dependency knowledge at this same age seems somewhat fragile. One early preferential looking study found that 15-month-olds were able to comprehend subject but not object wh-questions (Seidl et al., 2003). In two additional studies that found apparent success with object questions at this age (Gagliardi et al., 2016; Perkins & Lidz, under review), the authors argued that 15-month-olds' performance was not due to an adult-like representation of the wh-dependencies in these sentences, but rather to a "gap-driven" interpretation heuristic based on verb knowledge. Infants who knew that *bump* is transitive may have noticed that a predicted argument was unexpectedly missing in a question like *Which dog did the cat bump* __ ? They may then have inferred that the missing argument was the intended answer in the experimental task, and searched the display for the animal that got bumped. Identifying the answer would thus be possible without representing the wh-phrase *which dog* as that missing argument.

In support of this account, Gagliardi et al. (2016) found that 15-month-olds were not merely able to identify the right answer for wh-object questions, but also performed better than their 20-month-old peers on relative clauses with the same transitive verbs (e.g. *Find the dog that the cat bumped* __). In a similar task, Perkins and Lidz (under review) found that 15-month-olds' performance was predicted by vocabulary, a likely correlate of verb knowledge. These experiments do not demonstrate directly when children represent the full structure of wh-dependencies, but several results suggest that this may occur around the age of 20 months. Infants at

this age are beginning to produce wh-questions in their own speech (Rowland, Pine, Lieven, & Theakston, 2003; Stromswold, 1995) and reliably comprehend them regardless of vocabulary level (Gagliardi et al., 2016; Seidl et al., 2003). Their poorer performance on relative clauses may be explained by difficulty processing these dependencies in sentences where the cues to argument displacement are less apparent (Gagliardi et al., 2016)[3].

In summary, the current experimental evidence suggests the following developmental trajectory. Basic verb transitivity knowledge may develop jointly with infants' ability to recognize wh-dependencies between the ages of 15 and 20 months, and may emerge before infants can reliably recognize a fronted phrase as an argument in a wh-question. If correct, this account supports the hypothesis that the acquisition of these dependencies, and other non-basic clause types, is gap-driven. Infants' ability to detect a predicted but unexpectedly missing argument of a verb may not only enable them to infer the right answer in an experimental task at 15 months, but may also drive their identification of wh-dependencies over the next several months. As they attempt to integrate more of the linguistic material in these sentence into a complete parse, they may identify that a displaced argument stands in relation to that gap, and infer the type of dependency responsible.

However, the evidence from these previous studies is still highly indirect. These

[3]For example, relative clauses lack subject-auxiliary inversion and do-support, and the relativizer *that* is homophonous with other words in the language (such as demonstrative *that*). These weaker cues may make it challenging for learners to encode the filler or retrieve it in memory during online sentence processing, resulting in comprehension difficulty. This account predicts that 20-month-olds should improve on relative clauses if they contain stronger cues to displacement, and Gagliardi et al. (2016) confirmed this prediction: 20-month-olds were successful at responding to wh-relatives like *Find the dog who the cat bumped* __ , where the wh-word *who* more strongly signals the presence of a wh-dependency.

data demonstrate whether infants can identify the correct image as an answer to a question, but do not diagnose how infants represent the structure of that question in order to arrive at their interpretation. Therefore, these results cannot directly demonstrate that younger infants' representations of wh-dependencies are immature. Doing requires more direct evidence of infants' syntactic representations: in particular, evidence that 15-month-olds recognize when an argument of a known transitive verb is missing, but do not yet represent the fronted wh-phrase as an argument of the verb.

Here, I introduce a listening-time task that probes infants' syntactic representations in more directly. My test case is whether infants distinguish auditorily presented filled-gap wh-object questions (e.g. **Which dog should the cat bump him*?) from questions with gaps (*Which dog should the cat bump* __ ?). I contrast these sentences with simple declaratives with and without direct objects (*The cat should bump him* and **The cat should bump* __). Sentences are presented in the absence of referential context: we no longer measure whether infants can identify an event that matches the sentence, but merely whether infants listen longer to a grammatical or to an ungrammatical sentence. This allows us to determine whether infants represent a fronted wh-phrase as an argument in a wh-question, or whether they only notice when a verb is locally missing an argument.

Under the gap-driven hypothesis, if 15-month-olds know the argument structure requirements of these transitive verbs and do not yet represent the wh-phrase as an object of the verb, they should process a wh-question the same way they would process a simple transitive clause with no wh-phrase. In both cases, the absence of

a direct object should be more surprising than an overt direct object. However, if older infants represent the wh-phrase as an argument, they should process a filled-gap question differently from a simple transitive clause: they should notice that the filled-gap question has too many arguments.[4] In two experiments, I test and confirm these predictions. I find that 15-month-olds respond similarly to overt direct objects and object gaps in both declaratives and wh-questions, whereas 18-month-olds differentiate between these sentence types. This provides the first direct evidence that infants' wh-dependency representations develop in during the second year of life. Moreover, this development is consistent with the gap-driven learning hypothesis: infants show sensitivity to local argument structure violations prior to the point at which they represent a wh-phrase as an argument.

3.2 Experiment 1: 14- and 15-month-olds

3.2.1 Method

Experiment 1 tested a sample of 14- and 15-month-old English learners using a listening time task based on the Sequential Listening Preference Procedure (Maye, Werker, & Gerken, 2002; Shi, Werker, & Cutler, 2006). In this task, infants hear blocks of sentences accompanied by an abstract, unrelated video. Each trial is infant-controlled: both the video and audio stop after infants look away for more than 2 seconds. This allows us to use looking time towards the visual stimuli as a

[4]Some languages allow pronominal elements, called resumptive pronouns, in the thematic position of a wh-question (see McCloskey, 2017 for an overview). English is not one of these languages: in general, the base position of wh-movement in English must be phonologically null. I set aside for now the interesting question of how a learner might identify this property of her language.

Sentence Type	**Overt Object**	**Object Gap**
Declarative	*Wow, a tiger! The lion should hug him.*	**Wow, a tiger! The lion should hug __ .*
Wh-question	**Wow, which tiger should the lion hug him?*	*Wow, which tiger should the lion hug __ ?*

Table 3.1: Sample Test Sentences, Experiments 1 and 2

measure of infants' attention to the auditory stimuli.

Infants heard sentences with familiar transitive verbs in alternating trials, with and without overt direct objects following the verb. Infants in one condition heard these verbs in wh-object questions, and infants in the other condition heard these verbs in basic declarative clauses. Table 3.1 presents the fully crossed design: sentence type is manipulated between subjects, and object vs. object gap is manipulated within subjects. The dependent measure is looking time for each trial type. This allows us to determine whether infants differentiate between trials with and without direct objects, and whether they show different listening preferences for declaratives and wh-questions.

Participants. Participants in the current sample included 63 typically-developing infants (35 males), 32 between the ages of 14;0 and 14;29, and 31 between the ages of 15;0 and 15;29. The target sample size is $n = 64$, with data collection ongoing. In the current sample, the mean age of the 14-month-olds was 14;14, and the mean age of the 15-month-olds was 15;13. Participants were recruited from the greater Washington, D.C. area with the criteria that they heard English during at least 80% of their waking hours. An additional 28 infants were tested but not included in

the sample due to failing to complete the full familiarization phase as described below (3), failing to complete the full test phase (8), fussiness or inattentiveness (13), parental interference during the study (2), age below 14;0 on day of test (1), or coder error (1). Participants' total productive vocabulary was collected by parental report using the Words and Sentences MacArthur-Bates Communicative Development Inventory (MCDI) (Fenson et al., 1993). Mean total words produced by the 14-month-olds were 14.9, and mean total words produced by the 15-month-olds were 31.7.

Materials. Each 25-second trial presented six sentences, one for each of six familiar, transitive verbs (*kiss*, *hug*, *tickle*, *bump*, *hit*, and *cover*). Verbs were presented in a pseudo-random order with different animals as NP arguments. These particular verbs were selected because they are highly transitive and ranked high in familiarity norming for 16-month-old infants, the youngest age represented in the web-based WordBank database (Frank et al., 2016). The animals used as NP arguments were likewise selected for familiarity in the same age group. Pronominal objects (*him* and *her*, in equal numbers) were used in order to control for the number of full lexical NPs across sentence types.

All sentences were recorded by a female native speaker of American English using child-directed speech. In order to maintain natural-sounding prosody, all ungrammatical conditions were created by splicing together two grammatical sentences. Ungrammatical wh-questions were created by splicing a grammatical wh-question with a causative sentence (*Which tiger should the lion hug + I made the lion hug him = Which tiger should the lion hug him*). Ungrammatical declaratives

were created by splicing a grammatical declarative with an embedded question (*The lion should hug him + I know who the lion should hug = The lion should hug*). The modal *should* was used in order to avoid differences in verbal morphology across sentence types.

Audio stimuli were edited using Adobe Audition and Praat, and concatenated with variable 750-1000 ms of silence between sentences. Audio for each trial was combined with one of two videos of animated, slowly rotating shapes in Adobe Premiere. One of these videos was used during the familiarization phase of the experiment, and one during the test phase. A silent video of a butterfly on a leaf was separately edited for use as an attention-getter stimulus.

Procedure. Infants sat on a parent's lap or a high chair positioned 6 feet away from a 51" widescreen television. Parents listened to music played over noise-cancelling headphones, and were instructed not to talk to their children or direct their attention. Stimuli were played using the Habit program (Cohen, Atkinson, & Chaput, 2004). A camera located above the television was used to video-record the experiment. The camera feed was connected to a video monitor in a separate room to allow an experimenter in a separate room to live-code infants' eye fixations. The experimenter was not able to hear the audio for the experiment, and therefore was blind to the particular trial type that the infant was hearing.

Each experiment began by displaying the attention-getter stimulus. Once the infant fixated on the attention-getter, the experimenter initiated the first trial. During each trial, the experimenter pressed a key on the computer to record when the infant was looking at the screen, and released the key as soon as the infant looked

away. A trial ended after the full 25-second duration of the audio stimulus, or after the computer program registered that the infant looked away from the screen for more than 2 seconds continuously. At the end of a trial, the attention-getter stimulus was displayed. The next trial was initiated as soon as the infant re-oriented back towards the screen.

The experiment had two phases. During the familiarization phase, infants were familiarized to at least 72 seconds of the six test verbs in basic transitive clauses with direct objects. Because 14- and 15-month-olds may have variable prior experience with the experimental verbs, this phase was intended to aid their lexical processing by facilitating retrieval of these lexical items and their argument structure from memory. The 72-second familiarization duration was informed by prior work that used dialogues of similar length to familiarize infants with novel verbs (Arunachalam et al., 2013; Scott & Fisher, 2009; Yuan & Fisher, 2009), and pilot testing confirmed that this duration of training would not result in substantial loss of attention before the test phase began. Familiarization sentences had the same structure as the overt object sentences presented in the declarative test condition (e.g. *Wow, a giraffe! The bird should hug him.*), but did not include any of the same sentences presented at test. In order to ensure that infants in the declarative condition were exposed to sufficiently novel stimuli at test, different videos of rotating shapes were used for the familiarization and test phases of the experiment. However, the differences in stimuli between familiarization and test were still smaller for the declarative condition than for the wh-question condition, a point I return to in the discussion section.

Four 25-second familiarization trials were prepared, each containing six sen-

tence with the test verbs presented in a pseudo-random order. The order of familiarization trials was randomized across participants. The familiarization phase ended after the trial during which an infant reached the 72-second threshold of looking time. If this threshold was not reached during the first presentation of the familiarization trials, they repeated in a random order, for up to 12 trials total. Infants who did not reach 72 seconds of looking time over the course of 12 familiarization trials were excluded from analysis.

After the familiarization phase, the experiment proceeded to the test phase. At test, infants heard 12 trials of declarative sentences or wh-object questions, alternating between trials with overt objects and trials with object gaps. Infants in each condition were randomly assigned to one of four lists, counterbalancing two factors across participants. Grammaticality of the first test trial was counterbalanced by presenting half of participants with a grammatical first test trial and half with an ungrammatical test trial in their condition, and test trial order was counterbalanced by reversing the order of trial presentation for half of the participants. If an infant became fussy over the course of the test phase, the experiment was stopped and the infant was excluded from analysis.

Predictions. If 14- and 15-month-old infants are learning the transitivity of frequent verbs in their input but do not yet represent wh-phrases as arguments in wh-questions, this makes two predictions. The first is that infants at this age should be able to detect when the transitivity requirements of a verb are locally violated: they should differentiate sentences with overt direct objects from sentences with object gaps. The second is that their behavior should be the same for both declaratives

and wh-questions. In both cases, an infant should expect a known transitive verb to have a following direct object, and should be surprised by an object gap. This should result in similar looking time preferences for either overt objects or object gaps in each condition. Note that in this method we cannot predict the direction of infants' preferences in advance. It is possible that infants will prefer to listen to sentence types they find surprising or unfamiliar; conversely, they may prefer to listen to sentence types they find unsurprising or familiar. Many interacting factors have been found to affect when infants display a novelty or a familiarity preference in a listening-time task (see DePaolis, Keren-Portnoy, & Vihman, 2016 and Houston-Price & Nakai, 2004 for reviews). Because age is known to be one of these factors, I examine infants' preferences as a function of their age in months.

Under the alternative hypothesis that infants at this age do represent wh-phrases as arguments, this makes a different prediction. Infants should still be surprised to hear an object gap following a transitive verb in a declarative sentence, but they should be unsurprised in a wh-object question, because they have recognized that the wh-phrase is acting as that object. Conversely, infants should be unsurprised to hear an overt object in a declarative, but they should be surprised in a wh-object question, because they should notice that the question has too many arguments. Thus, this alternative hypothesis predicts that infants should show opposite patterns of preference across the two conditions: whichever direction infants show a preference for overt objects vs. object gap trials in declarative sentences, their preferences should be flipped for wh-questions.

3.2.2 Results

Total looking time during overt object and object gap trials was calculated for each infant. Fig. 3.1 plots the mean total looking times across infants for each trial type, as a function of condition and age in months. A visual scan of these plots reveals that infants in the 14-month-old and 15-month-old age groups appear to show opposite patterns of preference for object gap vs. overt object trials. 14-month-olds as a whole appear to listen longer to overt object trials, whereas 15-month-olds as whole appear to listen longer to object gap trials. However, these preferences did not seem to differ by condition: neither age group appears to show different patterns of preference for declarative sentences vs. wh-questions.

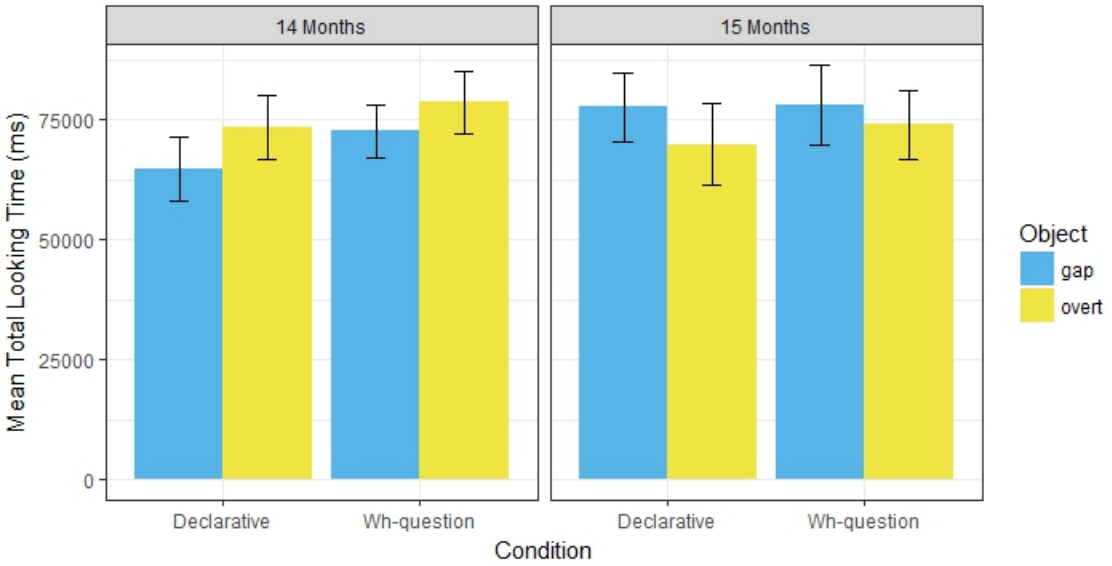

Figure 3.1: Looking Time by Trial Type, Condition, and Age, Exp. 1

To normalize for individual differences in participants' overall looking times during the experiment, a preference score was calculated for each participant following Shi et al. (2006): total looking time during object gap trials divided by total

looking time during overt object trials, minus 1. This yields a percent advantage for object gap vs. overt object trials for each participant. These individual preference scores are plotted in Fig. 3.2 as a function of condition and age in months, revealing a pattern consistent with that in Fig. 3.1. Although there is variability in each group, 14-month-olds in both conditions tend towards a slight preference for overt object trials, and 15-month-olds in both conditions tend towards a slight preference for object gap trials. This visual trend was confirmed by a 2x2 ANOVA (condition x age in months), with mean preference for object gap trials as the dependent measure. This analysis revealed a significant main effect of age in months ($F(1, 59) = 4.05, p < 0.05$), but no main effects or interactions by condition. Infants' preferences for overt objects vs. object gaps changed between the ages of 14 and 15 months, but these preferences were consistent for both declaratives and wh-questions.

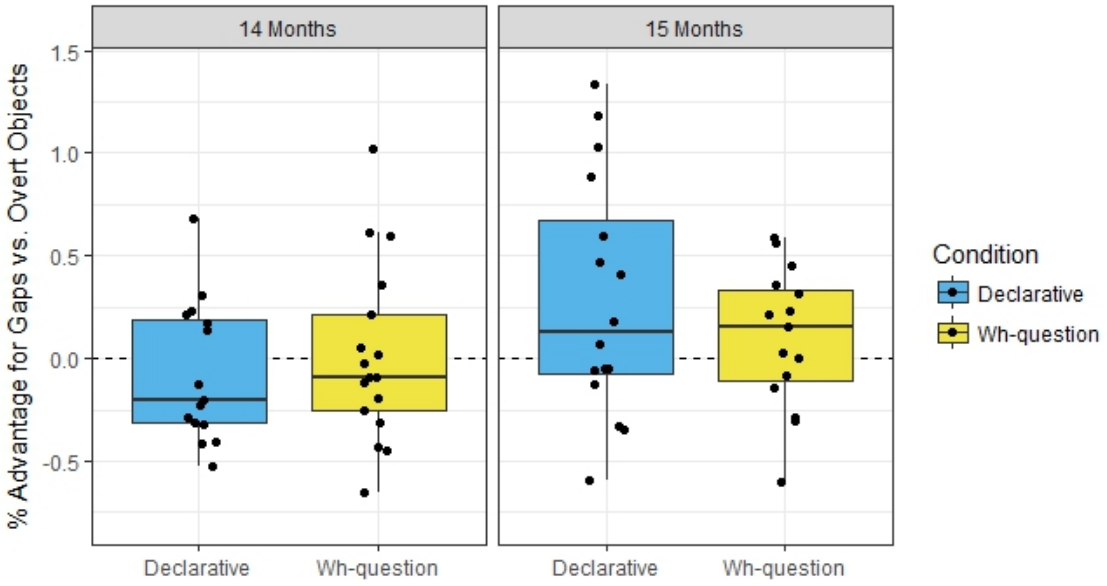

Figure 3.2: Preference for Object Gap Trials by Condition and Age, Exp. 1

Because no significant effects of condition were found, looking times for both conditions were analyzed together in follow-up analyses. 14-month-olds' overall looking preferences were not found to significantly differ from zero ($t(31) = -0.38, p < 0.70$): the apparent numerical preference for overt objects in this age group did not reach statistical significance. However, 15-month-olds as a group showed a significant preference for object gap trials ($t(30) = 2.26, p < 0.03$). These results show that infants at 15 months differentiated between sentences with and without direct objects following the transitive test verbs, and listened longer when these sentences did not have direct objects.

3.2.3 Discussion

The results from Experiment 1 are consistent with the hypothesis that infants by the age of 15 months show sensitivity to verb transitivity, but do not yet represent wh-phrases as arguments in wh-questions. The 15-month-olds in the sample listened longer to sentences in which transitive verbs did not have following direct objects, compared to sentences where direct objects were present after the verb. The ability to differentiate between these sentence types appears to develop between the ages of 14 and 15 months. Unlike the 15-month-olds, the 14-month-olds in the sample did not differentiate between overt object and object gap sentences. These results are consistent with previous findings that basic argument structure knowledge— in particular, the ability to encode verb transitivity properties, and use these properties to predict an upcoming direct object — emerges around the age of 15-16 months

(Jin & Fisher, 2014; Lidz et al., 2017).

Moreover, the 15-month-olds showed the same pattern of preference for both basic declarative sentences and wh-object questions. For both sentence types, they listened longer when the sentence had an object gap than when it had an overt object after the verb. This is despite the fact that these sentence types have opposite patterns of grammaticality. Object gaps are ungrammatical in declaratives with transitive verbs but grammatical in wh-object questions; overt objects are grammatical in declaratives with transitive verbs but ungrammatical in wh-object questions. These differences in grammaticality arise because wh-object questions have a fronted argument acting as the verb's object, and that argument should not be pronounced in canonical object position after the verb. By responding in the same way to overt objects vs. object gaps in both sentence types, the 15-month-olds did not appear to detect this non-local predicate-argument dependency in wh-questions. Instead, they appeared to process these sentences in the same way as they processed basic declarative clauses: by listening longer to object-gap sentences in both conditions, they appeared to notice only when the verb's requirement for an object was locally violated. This supports the hypothesis from previous literature that 15-month-olds respond to local verb transitivity violations in wh-questions, but do not yet represent fronted wh-phrases as arguments of the verb (Gagliardi et al., 2016; Perkins & Lidz, under review).

15-month-olds' preferences for sentences with object gaps could be interpreted as a type of novelty preference. After familiarization to declarative sentences with overt objects following transitive verbs, 15-month-olds then preferred to listen to

sentences in which those objects were missing. I note that these results do not differentiate between two alternative explanations for 15-month-olds' behavior. On the one hand, infants at this age may have developed knowledge of the transitivity requirements of these verbs prior to entering the lab, and their responses at test reflected that prior knowledge rather than any new information gained during familiarization. On this account, the familiarization phase may have only helped infants deploy their prior knowledge of verb transitivity more efficiently during sentence processing, by facilitating lexical access. On the other hand, these infants may have had no prior knowledge of the transitivity of these verbs prior to our experiment, but were able to efficiently encode this information over the course of the familiarization phase. The first account is consistent with earlier work showing transitivity knowledge around this age (Lidz et al., 2017), but further work is needed to differentiate these accounts by testing infants without a familiarization phase.

In summary, the results of Experiment 1 are consistent with the gap-driven learning hypothesis. These results point towards a developmental stage in which infants show sensitivity to verb argument structure, and respond to local argument structure violations, but do not yet represent wh-phrases as arguments. However, these results do not show that this task can detect when infants do represent wh-phrases as arguments. By hypothesis, this ability develops between the ages of 15 and 20 months. In Experiment 2, I test this hypothesis by examining the behavior of 17- and 18-month-old infants in the middle of this developmental window. If wh-dependency representations indeed develop over this age range, and our measure is sensitive to this development, then we should find that older infants are able to

differentiate between declarative sentences and wh-questions.

3.3 Experiment 2: 17- and 18-month-olds

3.3.1 Method

Experiment 2 tested a sample of 17- and 18-month-old English learners using the same listening time task as in Experiment 1. This age range was chosen in order to determine the first point in development at which infants begin representing wh-dependencies in wh-object questions. Because children begin producing their first argument wh-questions around the age of 20 months (Rowland et al., 2003; Stromswold, 1995) and reliably show comprehension of these questions at the same age (Gagliardi et al., 2016; Seidl et al., 2003), it is likely that knowledge of these dependencies may be emerging a couple of months earlier.

Participants. Participants in the current sample were 58 typically-developing infants (32 males), 29 between the ages of 17;0 and 17;29 and 29 between the ages of 18;0 and 18;29. Data collection is ongoing with a target $n = 64$. In the current sample, mean participant age of the 17-month-olds was 17;13 and mean age of the 18-month-olds was 18;11. Participants were recruited from the greater Washington, D.C. area and heard English during at least 80% of their waking hours. An additional 25 infants were tested but not included in the sample due to failing to complete the full familiarization phase (5), failing to complete the full test phase (9), fussiness or inattentiveness (5), parental interference during the study (3), or less than 80% English exposure (3). Mean total productive vocabulary of the 17-month-olds was

43.0 words and mean total productive vocabulary of the 18-month-olds was 90.5 words, as reported by parents on the the Words and Sentences MacArthur-Bates Communicative Development Inventory (MCDI) (Fenson et al., 1993).

Materials and Procedure. The materials and procedure for Experiment 2 were identical to those in Experiment 1. Just as in Experiment 1, infants were tested in two between-subjects conditions differing by sentence type (declaratives and wh-questions), with trial type (overt object vs. object gap) manipulated within-subjects.

Predictions. Under the hypothesis that infants in the 17- to 18-month age range are beginning to represent wh-phrases as arguments in wh-questions, infants should show different patterns of responses for wh-questions as compared to basic declarative clauses. Infants should still differentiate between overt objects and object gaps in declaratives, finding object gaps more surprising than overt objects following transitive verbs. We still cannot predict the direction of infants' listening time preferences: infants at this age may prefer to listen to sentences they find surprising, or to sentences they find unsurprising. But whichever direction their preference goes, this preference should be reversed for wh-questions. If infants at this age know that the wh-phrase is acting as the verb's object, they should find overt objects more surprising than object gaps in these questions, because the overt-object questions have too many arguments. This predicts that an analysis of infants' looking times will find an interaction of trial type by condition.

Under the alternative hypothesis that 17- to 18-month-olds do not yet represent wh-phrases as arguments, they should show the same pattern of preference for

both wh-questions and declarative clauses, like the 15-month-olds in Experiment 1. This predicts that we will again find no effect of condition at this age.

3.3.2 Results

Fig. 3.3 plots infants' mean total looking times during overt object and object gap trials, as a function of condition and age in months. Visually, these plots reveal that 17 and 18-month-olds behaved differently on our task. 17-month-olds do not appear to differentiate between overt object and object gap trials regardless of sentence type, whereas 18-month-olds appear to show a preference for overt objects in declarative sentences and a preference for object gaps in wh-questions.

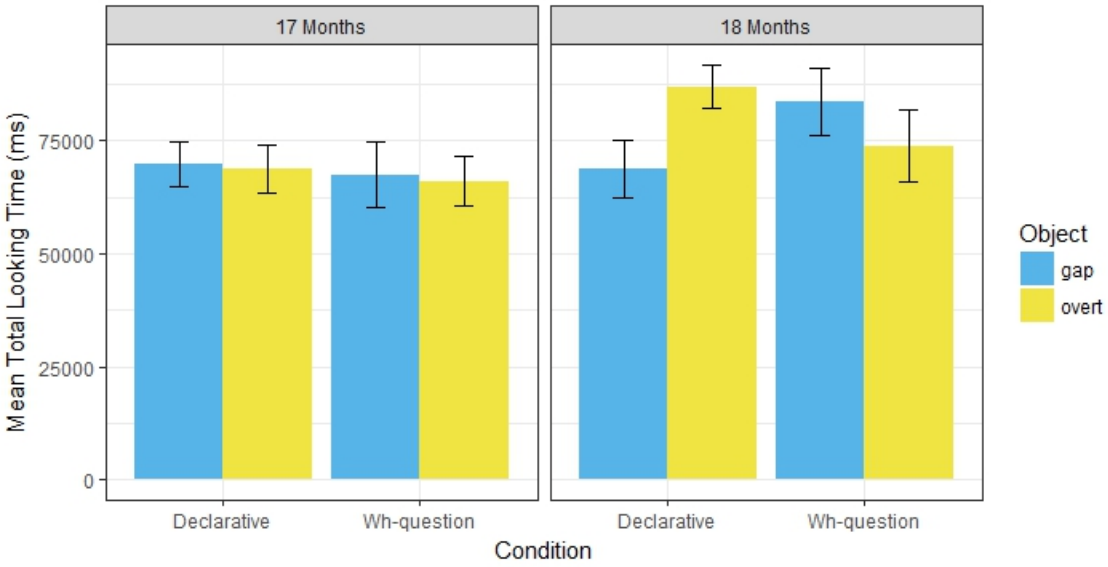

Figure 3.3: Looking Time by Trial Type, Condition, and Age, Exp. 2

Just as in Experiment 1, a preference score was calculated for each participant by dividing total looking time during object gap trials by total looking time during overt object trials and subtracting 1. Individual preference scores for Experiment

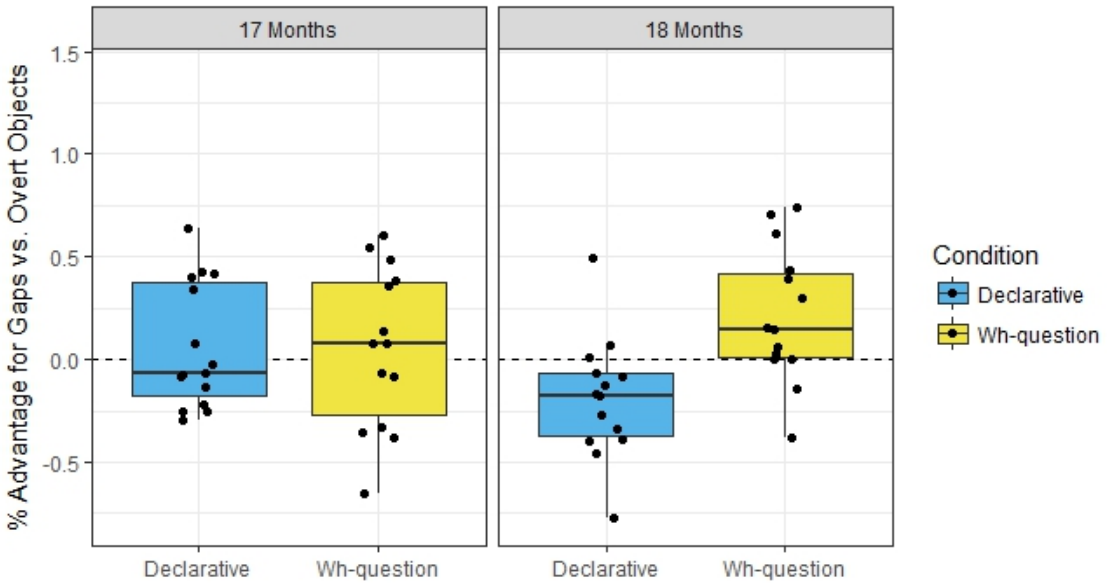

Figure 3.4: Preference for Object Gap Trials by Condition and Age, Exp. 2

2 are plotted in Fig. 3.4 as a function of condition and age in months. A 2x2 ANOVA (condition x age in months) found a significant main effect of condition ($F(1, 54) = 5.15, p < 0.03$) and a significant interaction of condition by age in months ($F(1, 54) = 5.28, p < 0.03$). The main effect of condition appears to be carried by 18-month-olds' behavior, as supported by the interaction with age. Thus, follow-up analyses examined each age group separately.

Follow-up analyses in each age group found that 18-month-olds showed significantly different looking preferences in the two conditions (Welch's $t(27.0) = 3.45, p < 0.002$), but 17-month-olds did not (Welch's $t(24.6) = 0.02, p < 0.98$). 18-month-olds' preference scores in the declarative condition were significantly lower than zero ($t(13) = -2.42, p < 0.03$), indicating a significant preference for overt object trials. Conversely, 18-month-olds' preference scores in the wh-question condition were significantly greater than zero ($t(14) = 2.47, p < 0.03$), indicating a significant

preference for object gap trials. These results show that 18-month-olds not only differentiated between overt object and object gap trials, but also differentiated between declaratives and wh-questions: they had opposite trial type preferences for these different sentence types.

3.3.3 Discussion

The results from Experiment 2 are consistent with the hypothesis that infants by the age of 18 months not only show sensitivity to verb transitivity, but also begin to represent wh-phrases as arguments in wh-questions. The 18-month-olds in the sample showed different listening preferences for direct objects in basic declarative clauses compared to wh-object questions. They listened longer to declarative sentences in which transitive verbs had following direct objects, compared to when those verbs were missing direct objects. Conversely, they listened longer to wh-object questions in which transitive verbs did not have following direct objects, compared to when an object was present after the verb. Their listening time therefore patterns together with the differences in grammaticality between these sentence types: 18-month-olds listened longer to grammatical than to ungrammatical sentences.

These results indicate that the 18-month-olds had acquired more mature knowledge about the grammar of wh-questions in English, compared to the 15-month-olds in Experiment 1. Specifically, the 18-month-olds no longer responded in the same way when a transitive verb is locally missing its object in a wh-object question, compared to in a declarative clause. Instead, they acted as if they knew that an

object gap is grammatical in a wh-object question, and a locally present overt object is ungrammatical. This suggests that these infants had learned that the fronted wh-phrase is acting as that object, satisfying the verb's transitivity requirement non-locally.

This ability appears to develop within the age range I tested. Only the 18-month-olds in the sample appeared to differentiate between declaratives and wh-questions; the 17-month-olds in the sample did not show this same ability. Indeed, I did not find that the 17-month-olds, in aggregate, discriminated between object gap and overt object trials for either sentence type. It is possible that this age range captures an intermediate window of development, in which some infants have acquired more mature knowledge of these sentence types and some are still in the process of acquiring that knowledge. 17-month-olds show larger variability than 18-month-olds in their preference scores in Fig. 3.4, with few infants showing a preference of zero; instead, these preferences seem to span a wide range above and below chance. If these data were due to a heterogenous population at this age, aggregating across preference scores would lead to the appearance of at-chance behavior. However, because sentence type was manipulated between-subjects, it is not possible to tell in the current design whether some of these infants would show the same patterns of preference for both sentence types, and others would differentiate between them. Future work testing sentence type within-subjects would help address this issue and illuminate the development occurring between 15 and 18 months.

Interestingly, where the 15-month-olds in Experiment 1 showed a novelty preference for object gaps in declarative clauses, the 18-month-olds showed a familiarity

preference: they listened longer to overt objects in declarative clauses. It is not immediately apparent why this difference arose, but I note that it is not very well-understood when infants will display a novelty or a familiarity preference in a looking or listening-time task (DePaolis et al., 2016; Houston-Price & Nakai, 2004). Stimulus complexity, age of participant, and type and length of familiarization phase are all factors known to affect these preference types, but do so in complex interacting ways (e.g. Colombo & Bundy, 1983; Hunter & Ames, 1988; Wagner & Sakovits, 1986). In the current task, there are many possibilities for why these two age groups may have shown different preference types. For instance, 18-month-olds may have processed not only the wh-questions but also the declarative sentences in a qualitatively different manner than the 15-month-olds, or may have responded differently to the type and amount of familiarization that was provided. Further investigation without a familiarization phase may help eliminate one of these potential factors. Nonetheless, what is important in this design is not what type of preference infants display, but whether they show different patterns of preference for declaratives and wh-questions. The finding that 15-month-olds prefer object gaps to the same extent in both sentence types, whereas 18-month-olds show different patterns of preference depending on sentence type, suggests that grammatical development has occurred between these two ages.

3.4 General Discussion

This chapter investigates a chicken-and-egg problem in early grammar acquisition. Learners rely on argument relations like 'subject' and 'object' to bootstrap into the core argument structure and clause structure properties of their language. But it is not trivial to identify these relations in non-basic clauses, in which transformations have applied to displace arguments from their canonical positions. If learners could identify argument displacement in these sentences, then they could avoid being misled when drawing bootstrapping inferences about verbs and clause structure. Or, if infants could identify the basic argument structure of some frequent verbs in their language, then they might be able to detect when arguments have been displaced. Here, I aim to answer the chicken-and-egg question: which comes first, argument structure or non-basic clause syntax?

The experiments in this chapter provide new empirical evidence that argument structure acquisition comes first. 15-month-olds were sensitive to the argument structure requirements of common transitive verbs: they discriminated sentences in which the verb's object appeared in canonical object position from sentences in which it was missing in that position. However, they did not respond differently to sentences where that object appeared in a non-canonical position as a fronted wh-phrase. Only the 18-month-olds showed this ability. They differentiated between object gaps in wh-object questions, in which the verb's requirement for a direct object is being satisfied non-locally, and object gaps in declarative clauses, in which the verb does not have a direct object at all. Thus, 18-month-olds' different listening

preferences for these sentence types demonstrated that they had acquired knowledge of both the local and non-local predicate-argument dependencies in these sentences.

This work contributes two novel findings. First, it provides new evidence that infants' wh-question representations develop over the second year of life, and identifies the first age at which argument displacement is represented. Infants appear to represent the non-local predicate-argument dependencies in wh-questions about two months before they begin to produce these sentences in their own speech (Rowland et al., 2003; Stromswold, 1995). Second, it provides new evidence that this ability developmentally follows argument structure acquisition, a hypothesis that had only indirect support from prior work (Gagliardi et al., 2016; Perkins & Lidz, under review). At 15 months, infants detect local verb argument structure violations, but not until 18 months do they appear to represent fronted wh-phrases as arguments.

However, as these results come from a new experimental paradigm, additional converging evidence is needed to show that they reflect the developmental change I propose. These results are consistent with prior work showing the development of argument structure knowledge around the age of 15 months (Jin & Fisher, 2014; Lidz et al., 2017), and the development of wh-dependency comprehension between 15 and 20 months (Gagliardi et al., 2016; Perkins & Lidz, under review; Seidl et al., 2003). Further work will aim to reproduce the current findings with a modified experimental task in order to better understand the differences observed across this age range. In particular, the familiarization phase in the current task may have introduced complications for interpreting the results. This phase may have been processed differently by the two age groups that were tested.

Moreover, because the wh-question test stimuli were less similar to the familiarization stimuli than were the declaratives, it is possible that these differences between familiarization and test may have influenced what preference type— novelty or familiarity— was displayed across conditions. These differences across conditions were minimized as much as possible by introducing a new visual stimulus at test, and by controlling the number of noun phrases in both sentence types. This helps justify the inference that preferences in the same direction reflect the same underlying processing of the two types of sentences, and preferences in opposite directions reflect different types of processing. However, I cannot rule out the possibility that these preference directions were also affected by the change from familiarization to test in the current design. Further investigation with no familiarization phase, or with a better control for complexity between the familiarization and test phases, will address these issues.

These findings help illuminate how syntactic representations develop and interact with infants' comprehension abilities. The 15-month-olds in our study are able to detect argument gaps— unexpectedly missing arguments of transitive verbs— but do not yet appear to represent the wh-dependency between the fronted wh-phrase and the verb in a wh-question. Nonetheless, 15-month-olds in prior comprehension studies act as if they can comprehend these questions (Gagliardi et al., 2016; Perkins & Lidz, under review), indicating that they are bootstrapping some aspect of sentence meaning from a partial or non-adultlike sentence representation.

The current findings do not tell us exactly how 15-month-olds represent the wh-questions in our task. However, both these findings and prior results can be

explained if infants represent at least the verb with one locally identified argument and an argument gap. In the tasks in Gagliardi et al. (2016) and Perkins and Lidz (under review), an infant who heard *Which dog did the cat bump?* may have been able to identify the local subject for the verb *bump*, and may have further expected an object for this verb even though it was not present in its canonical position. This partial parse, together with an understanding of how grammatical relations link to thematic relations, may have allowed infants to infer some aspects of the sentence meaning: there was an event of bumping by a cat, and this also likely involved something that was bumped, even though that relation was realized by an overt post-verbal object. In the experimental context, an infant could guess that she was being directed to find the image or video that best matched this partial sentence representation. If only two choices existed— a dog who was a bumper, and a dog who got bumped— the infant might have guessed that she was being asked to find the animal that bore the relation of 'thing bumped' in the event of bumping by the cat.

That is, infants in prior preferential looking tasks may have intuited that the experimental task was to locate the object or event that could link in an appropriate way to their incomplete syntactic representation, and fill in the gaps. Further work is needed to support this hypothesis, and to specify the nature of that syntactic representation. In particular, do 15-month-olds ignore the fronted wh-phrase in these questions, or do they parse it in some fashion but fail to integrate it into the rest of their sentence representation, or do they integrate it into a representation that is not fully adult-like? This speaks to the broader question of how infants at very

early stages of syntactic development might bootstrap aspects of sentence meaning from partial syntactic representations, in combination with top-down information about the discourse context.

Furthermore, the finding that argument structure knowledge is developmentally prior to wh-dependency representations is consistent with the hypothesis that non-basic clause acquisition is gap-driven. By hypothesis, 15-month-olds' ability to detect argument gaps drives their search for non-local predicate-argument dependencies in these sentences, allowing them to eventually discover the signals of wh-movement and other forms of argument movement in their language. However, this developmental trajectory introduces an apparent paradox. Learners acquire verb transitivity before they can recognize arguments in non-canonical positions, arguably because identifying the structure of non-basic clauses depends on knowing which verbs are transitive. But how do learners accurately identify which verbs are transitive in order to parse non-basic clauses, if those sentences themselves provide misleading data for learning verb transitivity? What enables learners to avoid being misled by non-basic clauses in their input, at the stage at which they do not recognize that they are non-basic?

Thus, the empirical results in this chapter help answer the question of which comes first, argument structure or non-basic clause syntax, but they do not show how learners find their way out of this chicken-and-egg problem. In Chapters 4 and 5, I propose a computational account for how this learning succeeds. In Chapter 4, I focus on the first step of learning, and show that it is in principle possible for learners to accurately identify argument structure even before they can recognize

non-basic clauses in their input. In Chapter 5, I model how learners might use that argument structure knowledge to identify the surface forms of non-basic clause types in their language. These models provide an account of the learning mechanisms that learners could recruit in order to incrementally build on their prior linguistic knowledge, and draw the right grammatical generalizations from partial and immature representations of their input.

Chapter 4: Filtering Input for Transitivity Acquisition

4.1 Background

The experimental findings in Chapter 3 point towards the following developmental trajectory for verb argument structure and non-basic clause acquisition. Infants as young as 15 months show emerging knowledge of the basic argument-taking properties of verbs in their language, such as transitivity. These first steps of verb learning appear to take place before infants recognize displaced arguments in common non-basic clause types, such as the wh-question in (1).

(1) What did Amy fix?

This developmental trajectory is consistent with the hypothesis that verb argument structure knowledge bootstraps the acquisition of non-basic clause syntax. It may be the case that learning to identify argument movement in non-basic clauses depends on knowing when verbs require particular arguments: learning that *what* is the object of *fix* in (1) may depend on knowing that *fix* takes a direct object. Here, I computationally investigate the specific learning mechanisms that this hypothesis would require. To begin, I address how the first steps of learning attested empirically are even possible. How can learners begin acquiring the argument-taking properties

of verbs when they do not yet recognize displaced arguments in non-basic clauses? How do they learn the transitivity of a verb like *fix* when their input contains sentences like (1), which they cannot yet recognize as transitive?

I propose that learners need a mechanism to avoid being misled by non-basic clauses in their input, at the developmental stage when they are bootstrapping into the basic argument structure and clause structure properties of their language. In this chapter, I present a computational model that learns to filter its data in order to infer verb transitivity, effectively ignoring misleading data from non-basic clauses without knowing what non-basic clauses look like. The model instantiates a learner that considers the possibility that its data contains some amount of noise, because it does not have the grammatical knowledge to accurately represent all of the sentences it hears. The learner infers what portion of its data is signal and what portion is noise in order to identify the transitivity of verbs in its input. This mechanism, in principle, provides a way for learners to cope with their own immature representations of their input, and to draw the right grammatical generalizations even when those representations are incomplete and inaccurate.

4.1.1 Filtering

I follow a solution that has been assumed by both syntactic and semantic bootstrapping theories: learners have a way to bootstrap into the core argument structure properties of their language primarily using data from basic clauses, in which core arguments are easier to identify and correspondence relations between

syntax and meaning will hold more reliably (Gleitman, 1990; Lidz & Gleitman, 2004a, 2004b; Pinker, 1984, 1989). Under this approach, first proposed by Pinker in his earliest work on semantic bootstrapping (Pinker, 1984), children somehow "filter out" non-basic clauses from their bootstrapping data. That is, they avoid learning about verb meanings, argument structure, and clause structure from sentences with argument displacement, because these sentences obscure the meaning-distribution relations that bootstrapping relies on (Gleitman, 1990; Lidz & Gleitman, 2004a, 2004b; Pinker, 1984, 1989).

This approach has implicitly assumed that learners know which sentences to filter out, but the mechanism by which they identify these sentences has not yet been established. Pinker (1984) proposes two options. Either parents might do the filtering and avoid producing these sentences in their children's presence, or children might internally filter these sentences themselves. As noted in the previous chapter, parental filtering does not seem to occur: wh-questions are prevalent in the input to 1-year-old infants (Cameron-Faulkner, Lieven, & Tomasello, 2003; Newport et al., 1977; Stromswold, 1995). The second logical solution is for children to filter out non-basic clauses themselves. This approach implicitly assumes that children know which sentences to filter out. But this solution risks being circular. Learners need to filter non-basic clauses in order to learn argument structure, but argument structure knowledge comes developmentally before infants can identify the structure of non-basic clauses, and may even be needed for that learning to take place. How can learners identify non-basic clauses in order to filter them, if they do not know yet what argument displacement looks like in their language?

Pinker (1984, 1989) argues that this circularity can be avoided if children can use non-syntactic cues to flag certain utterances as likely to contain non-basic clauses, without recognizing the structure of those clauses. These cues might include "special intonation, extra marking of the verb, presuppositions set up by the preceding discourse or the context, nonlinguistic signals of the interrogative or negative illocutionary force of an utterance" (Pinker, 1984). The challenge with this solution is identifying how learners know which cues to use. Attempting to define the criteria by which children should filter their input creates its own learning problem: this introduces a new set of categories which the learner must know to track, and which in many cases may be far from transparent (Gleitman, 1990).

One might imagine the following solution to the circularity problem: perhaps learners acquire non-basic clause syntax and verb argument structure by attempting to learn both of these phenomena at the same time. This simultaneous learning hypothesis may be possible, but I choose to model a different hypothesis instead, one that allows learning to take place in incremental steps over development. This solution does not require learners to know the criteria for identifying non-basic clauses in order to learn verbs. One way of thinking about this is that it provides learners with a way of arriving at a fruitful starting point from which a subsequent joint learning process might proceed—an initial wedge into the system.

Here, I present computational solution for how a learner might, in principle, filter its input to infer verb transitivity, without knowing what types of sentences it should filter out. This solution proposes that young learners implicitly assume that they will not accurately parse everything they hear, and expect that their data

will contain a certain amount of noise: erroneous parses that they shouldn't trust for the purposes of verb learning. Children might be able to learn the right way to filter erroneous parses out of their input in order to solve a particular learning problem— in this case, jointly inferring verb transitivity along with how much of their data to trust in making that inference. Crucially, this solution doesn't require learners to know where those errors came from, thereby sidestepping the problem of which cues learners should track for identifying non-basic clauses. Under this approach, children might filter non-basic clauses from the data they use for verb learning without knowing that they are non-basic clauses. This will allow them to use relations between syntax and meaning to bootstrap into the target grammatical system, even though they do not yet know when those relations are masked by other grammatical properties of the language.

4.1.2 Computational Models of Verb Learning

I adopt a Bayesian framework, in which a learner observes a data pattern and infers the probability of some properties of the system that may have generated that data. This framework conveniently allows us to specify the alternative systems (verb transitivity properties vs. erroneous parses) that the learner considers for the verb distributions it observes.

The model follows previous Bayesian approaches to argument structure acquisition (Alishahi & Stevenson, 2008; Barak, Fazly, & Stevenson, 2014; Parisien & Stevenson, 2010; Perfors, Tenenbaum, & Wonnacott, 2010), but considers a differ-

ent problem than the one explored in that literature. The goal of the learners in Alishahi and Stevenson (2008) and Perfors et al. (2010) is to identify which verb classes exist in the language, and how verbs in those classes generalize across syntactic frames. The acquisition phenomenon being modelled— the over-generalization of verbs across argument structures that they do not actually participate in— is a stage of verb learning in preschool-aged children who are older than the infant bootstrappers discussed in the current work. This behavior is the output of at least three logically independent steps of learning: (1) perceiving how verbs distribute in particular syntactic frames; (2) performing an initial classification of verbs according to their argument-taking properties, e.g., as one-, two-, or three-place predicates; and finally (3) identifying how productively verbs in a class can generalize across different types of argument structures, e.g., from the prepositional dative to the double-object dative. The primary focus of prior models is the third step of learning, but I am concerned with the earlier processes involved in the first two steps. In particular, I ask how learners are able to establish a veridical percept of verbs' syntactic distributions, when they may not have the linguistic knowledge to reliably identify syntactic arguments in non-basic clauses.

This question has not yet been answered in previous models of argument structure acquisition, in which a learner's ability to veridically represent its input has been largely assumed. Alishahi and Stevenson (2008) acknowledge that this assumption is most likely unrealistic, and simulate noise in their learner's syntactic representations by randomly removing some of the distributional features that it learns from. Yet the "noise" faced by a learner in real life is not random. As the authors note,

"A more accurate approach must be based on careful study of the types of noise that can be observed in child-directed data, and their relative frequency" (Alishahi & Stevenson, 2008). This invites us to consider the ways in which learners might mis-perceive the data in their input, and how a learner can avoid being misled by that data when identifying a verb's basic syntactic distribution in the language.

Other previous computational models have investigated how learners might bootstrap between syntactic and conceptual representations, using conceptual structure to identify the core grammatical rules and word order properties of the language, and then using syntactic structure to infer the meaning of words and utterances (Abend, Kwiatkowski, Smith, Goldwater, & Steedman, 2017; Kwiatkowski, Goldwater, Zettlemoyer, & Steedman, 2012; Maurits, Perfors, & Navarro, 2009). While shedding light on how semantic and syntactic bootstrapping might proceed in tandem, these models still presuppose the step of learning that concerns this chapter: how a learner gains access to either type of representation, syntactic or conceptual, that it uses in these bootstrapping inferences. I focus on the learner's syntactic percept, but in doing so, do not deny that learners might simultaneously make use of both syntactic and conceptual information. My goal is simply to ask how far a learner can get in identifying a verb's argument-taking properties on the basis of its syntactic distributions, when those distributions may not be accurately perceived.

In the experiments below, I test the computational feasibility of this proposed solution: whether a learner could, in principle, jointly infer verb transitivity along with the parameters for filtering errorful sentence representations from the data it uses for learning. In Simulation 1, I demonstrate that a learner can accomplish this

joint inference on the basis of the syntactic distributions of frequent English action verbs in child-directed speech. The learner performs this inference using only rates of overt direct objects after verbs, and does not condition on any other utterance features, such as wh-words, prosody, or extra-linguistic discourse context; it succeeds even though it cannot distinguish object wh-questions from basic intransitive clauses. In Simulation 2, I ask how much the learner's performance in Simulation 1 depended on its *a priori* assumption that transitive, intransitive, and alternating verbs are equally likely. I show that the learner performs no better when it assumes these categories will occur in the proportions in which they actually do occur in child-directed English. However, it does not differentiate transitivity categories as well when it is extremely biased towards the alternating class, showing that the deterministic categories must be weighted sufficiently in the model's hypothesis space in order to be identified in its input. Thus, I provide a proof of concept that a child may be able to filter non-basic clauses from her input in order to correctly identify verbs' argument structure properties, without knowing in advance which clauses are non-basic. This inference requires prior knowledge about what types of transitivity properties a verb might display, but but does not require specific knowledge about the frequency of those transitivity categories in the learner's target language.

4.2 Model

I present a Bayesian model that learns how to filter its input in order to infer verb transitivity. The learner performs this inference only on the basis of observing

how verbs distribute with and without direct objects, and does not use any other syntactic or non-syntactic cues to identify its filter. Instead, the learner assumes that some of its parses are not trustworthy sources of information for learning its language, because it does not have enough linguistic knowledge to accurately parse every sentence in its input. The learner infers the right way to filter erroneous parses out of the data it uses for verb learning, without knowing why those parses were erroneous.

In this section, I first specify the generative model, which encodes the learner's assumptions about how its direct object observations are generated. Then, I specify how the learner jointly infers verb transitivity along with the parameters for filtering its input, given its data. In the following sections, I present simulations demonstrating that this joint inference is successful when tested on child-directed speech. I do not claim that the Bayesian inference performed by our model represents the exact algorithms performed by child learners. Although there is substantial literature on young children's statistical inference capabilities (Gómez & Gerken, 2000), this model is intended only as a proof of concept that such joint inference is possible. However, although this model may not provide a realistic implementation of the inference process that children use, it provides a more realistic account than previous models of the steps of learning involved in bootstrapping: specifically, how learners establish a veridical percept of verbs' syntactic distributions, in order to enable further bootstrapping inferences.

4.2.1 Generative Model

A generative model represents a learner's assumptions about the processes that generated its observed data. In our case, the observed data are counts of direct objects with particular verbs, as the learner represents them; specifically, the learner tracks how frequently it represents an overt direct object or no overt direct object following the verb. It assumes that there are two reasons why it might observe direct objects or no direct objects. On one hand, the transitivity of the verb determines whether it always, never, or sometimes takes a direct object. This means that the rate of direct objects following the verb gives the learner evidence for inferring whether the verb is transitive, intransitive, or alternating. But on the other hand, the learner might also mis-perceive whether a direct object is present, because it lacks the grammatical knowledge to identify the full structure of some sentences in its input. If this is the case, some of the observed data points might not reflect the true transitivity of the verb and should be filtered from the data that the learner uses to infer transitivity. Thus, there is some probability of error in the learner's direct object observations, and our learner infers two parameters for filtering this error: how frequently mis-parses of sentences occur, and whether the learner is more likely to miss a direct object that is underlyingly present or mistake another constituent for a direct object.

Figure 4.1 provides the graphical model for the learner. The model's observations of direct objects or no direct objects are formalized as the Bernoulli random variable X. Each $X^{(v)}$ represents an observation from a sentence containing verb

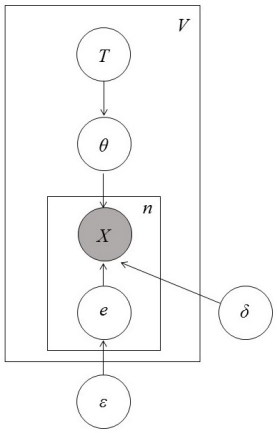

Figure 4.1: Graphical Model

v in the model's input, with a value of 1 if the sentence contains a direct object and 0 if it does not. These observations of direct objects can be generated by two processes: the transitivity of verb v, represented by the variables T and θ in the upper half of the model, or an erroneous parse of the sentence, represented by the variables e, ϵ, and δ in the lower half of the model. I will describe each of these processes in turn.

In the upper half of the model, each $X^{(v)}$ is conditioned on the parameter $\theta^{(v)}$, a continuous random variable defined for values from 0 to 1 inclusive. This parameter controls how frequently a verb v will be used with a direct object: the learner assumes that for every observation $X^{(v)}$, a biased coin is flipped to determine whether the sentence contains a direct object, with probability $\theta^{(v)}$, or does not, with probability $1-\theta^{(v)}$. The parameter $\theta^{(v)}$ is conditioned on the variable $T^{(v)}$, which represents the transitivity of verb v. T is a discrete random variable that can take on three values, corresponding to transitive, intransitive, and alternating verbs. Each of these values determines a different distribution over θ. For the transitive

category of T, θ always equals 1: the verb should always occur with a direct object. For the intransitive category, θ always equals 0: the verb should never occur with a direct object. For the alternating category, θ takes a value between 0 and 1 inclusive. The prior probability distribution over θ in this case is a uniform $Beta(1, 1)$ distribution. The learner begins with the simplifying assumption that all three values of T have equal prior probability— that is, the learner assumes that any verb in the language is equally likely *a priori* to be transitive, intransitive, or alternating. In later simulations, I explore the model's behavior when this assumption is changed.

In the lower half of the model, each X is conditioned on a Bernoulli random variable e, which represents the input filter. If $e_i^{(v)} = 0$, the observation in $X_i^{(v)}$ was generated by $\theta^{(v)}$ and $T^{(v)}$, and accurately reflects the transitivity of verb v. But if $e_i^{(v)} = 1$, the observation in $X_i^{(v)}$ was generated by an erroneous parse (henceforth an "error"), meaning the learner did not have adequate grammatical knowledge to parse the sentence correctly. This observation was not generated by $\theta^{(v)}$ and $T^{(v)}$, and may not accurately reflect the transitivity of verb v, so it should be ignored for the purpose of inferring $T^{(v)}$. Each $e^{(v)}$ is conditioned on the variable ϵ, which represents the probability of an erroneous parse occurring for any sentence in the input. The model learns a single parameter value for ϵ across all verbs.

The second parameter of the input filter is δ, which represents the probability of observing a direct object when an observation was generated in error. Thus, whether a sentence contains a direct object or no direct object depends on one of two biased coins. If $e_i^{(v)} = 0$ and the observation accurately reflects the verb's transitivity properties, then one biased coin is flipped and the sentence contains a

direct object with probability $\theta^{(v)}$. If $e_i^{(v)} = 1$ and the observation was generated in error, then a different biased coin is flipped and the sentence contains a direct object with probability δ. Like ϵ, δ is a shared parameter across all verbs. Both ϵ and δ are assumed to have a uniform $Beta(1, 1)$ prior distribution.

4.2.2 Joint Inference

The learner uses Gibbs sampling (Geman & Geman, 1984) to jointly infer the transitivity of each verb (T) and the two parameters of the input filter (ϵ and δ). In this form of sampling, we start with randomly-initialized values for ϵ and δ, and use those values to calculate the posterior probability of each transitivity category T for each verb, given the observed data and those filter parameters. We sample values for T from this posterior probability distribution. Then, we use the sampled transitivity categories to sample new values for ϵ and δ from estimates of their posterior probability distributions. This cycle is repeated over many iterations until the model converges to a stable distribution over T, ϵ and δ, which represents the optimal joint probability solution for these three variables. See Appendix A for details of the sampling procedure.

4.3 Simulation 1

In Simulation 1, I ask whether inferring the parameters of an input filter will allow a learner to accurately identify the transitivity categories of verbs in the speech that children hear. I tested the joint inference model on a dataset containing dis-

Corpus	#Children	Ages	#Words	#Sentences
Brown: Adam, Eve, Sarah (Brown, 1973)	3	1;6-5;1	391,848	87,473
Soderstrom (Soderstrom et al., (2008))	2	0;6-1;0	90,608	24,130
Suppes (Suppes, 1974)	1	1;11-3;11	197,620	35,904
Valian (Valian, 1991)	21	1;9-2;8	123,112	25,551

Table 4.1: Corpora of Child-Directed Speech

tributions of the 50 most frequent transitive, intransitive, and alternating verbs in corpora of child-directed speech. In order to determine whether this inference is successful, I compare the model's performance to an oracle model that already knows appropriate parameters for filtering its input, and baseline models with inappropriate filter parameters.

4.3.1 Data

A dataset was prepared from four corpora selected from the CHILDES Treebank (Pearl & Sprouse, 2013). This resource provides parse trees for several corpora of child-directed speech on CHILDES (MacWhinney, 2000), generated by the Charniak or Stanford parser and hand-checked by undergraduates. The selected corpora contain 803,188 words of child-directed speech, heard by 27 children between the ages of 6 months and 5 years. See Table 4.1 for corpus details.

The dataset was created by extracting sentences with the 50 most frequent action verbs in these corpora that could be characterized as transitive, intransitive, or alternating. Verbs with other argument-taking properties were excluded, such as obligatorily ditransitive verbs or those that frequently take clausal or verbal complements: mental state verbs (e.g. *want*), aspectual verbs (e.g. *start*), modals (e.g.

should), auxiliaries (e.g. *have*), and light verbs (e.g. *take*).[1] The selected 50 verbs were then sorted into transitive, intransitive, and alternating categories according to the English verb classes described in Levin (1993), supplemented by native speaker intuitions for verbs not represented in that work. These classes provide a target for learning meant to align with adult speaker intuitions, independent of the corpus data that the model learns from. The transitive and intransitive categories are conservative; any verb that could occur in a transitivity alternation was classified as alternating, regardless of the frequency or type of alternation. So, verbs like *jump* are considered alternating even though they occur infrequently in their possible transitive uses (e.g. *jump the horses over the fence*). These target categories thus set a very high bar for our model to reach.

An automated search was then conducted over the Treebank trees for the total occurrences of each verb in the corpora, in all inflections, and the total occurrences with overt direct objects following the verb (right NP sisters of V). These direct object counts included basic transitive clauses, but not wh-object questions or any other sentences with object gaps. Thus, we assume a learner who uses the canonical word order properties of English to identify direct objects when they occur after verbs, but does not yet know how to identify arguments in non-canonical positions. Table 4.2 lists the complete dataset provided to the learner: counts of the selected 50 verbs, along with their counts of overt post-verbal direct objects. For legibility I also report the percentages of direct objects with each verb, although this model

[1] Verbs with these other argument-taking properties were excluded for the sake of simplicity, as I am only modelling the acquisition of transitivity. Modelling the learning of other argument-taking properties would require expanding the learner's hypothesis space to include many more argument structure categories and alternations, a complex problem I leave for future work.

Verb	Total	# DO	% DO
Transitive Verbs			
feed	220	205	0.93
fix	337	305	0.91
bring	605	541	0.89
throw	312	275	0.88
hit	214	187	0.87
buy	358	299	0.84
catch	185	141	0.76
hold	579	406	0.70
wear	477	287	0.60
Alternating Verbs			
pick	331	299	0.90
drop	169	149	0.88
lose	185	160	0.86
close	166	141	0.85
touch	183	153	0.84
leave	356	297	0.83
wash	195	161	0.83
pull	331	268	0.81
push	352	274	0.78
open	342	265	0.77
cut	263	198	0.75
bite	191	140	0.73
turn	485	350	0.72
build	299	215	0.72
knock	160	115	0.72
read	509	350	0.69

Verb	Total	# DO	% DO
Alternating Verbs, cont.			
break	550	347	0.63
drink	366	221	0.60
eat	1318	777	0.59
sing	306	161	0.53
blow	255	132	0.52
draw	375	193	0.51
move	238	112	0.47
ride	281	114	0.41
hang	151	53	0.35
stick	192	57	0.29
write	583	155	0.27
fit	227	49	0.22
play	1568	308	0.19
stand	294	21	0.07
run	228	13	0.06
walk	253	11	0.04
jump	197	8	0.04
swim	180	7	0.04
sit	859	11	0.01
Intransitive Verbs			
wait	383	57	0.15
work	256	11	0.04
cry	275	8	0.03
sleep	451	13	0.03
stay	308	4	0.01
fall	605	3	0.00

Table 4.2: Dataset: Uses with Overt Direct Objects (DO) of 50 Verbs

learns from raw counts rather than percentages.

4.3.2 Results

Verb Transitivity Inference. The joint inference model infers a probability distribution over transitivity categories for each verb in its dataset. These distributions are displayed in Figure 4.2. Black bars represent the posterior probability assigned to the transitive category, dark gray bars represent the probability assigned to the intransitive category, and light gray bars represent the probability assigned to the alternating category. The target categories for each verb are shown below the horizontal axis.

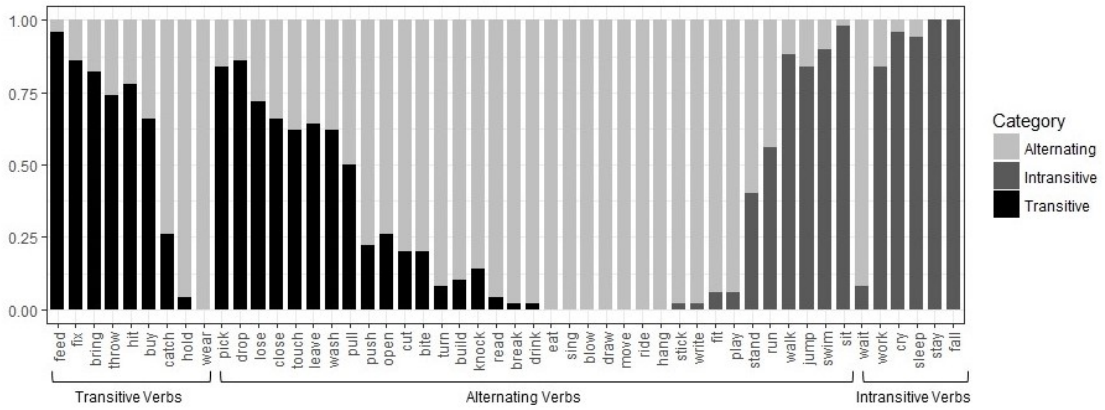

Figure 4.2: Posterior Distributions over Verb Categories (T), Joint Inference Model

Accuracy was calculated by determining which transitivity category was assigned highest probability to each verb by our model, and comparing these category assignments to the target categories for each verb. The percentage of verbs categorized correctly by the model is reported in Table 4.3. Overall, the model infers the correct transitivity properties for 2/3 of the verbs in our dataset. This is substantially better than chance performance: a model that randomly assigned categories to verbs would achieve 33% accuracy, because there are three possible options for each verb. The joint inference model performs significantly better on each verb class, and nearly twice as well overall.

The model achieves highest accuracy in categorizing the intransitive verbs: for all but one of these verbs, the model assigns highest probability to the intransitive

Model	**% Transitive**	**% Intransitive**	**% Alternating**	**% Total Verbs**
Joint Inference	0.67	0.83	0.63	0.66
Oracle	0.78	0.83	0.51	0.60
No-Filter Baseline	0.00	0.00	1.00	0.70
Chance	0.33	0.33	0.33	0.33

Table 4.3: Percentages of Verbs Categorized Correctly, Simulation 1

category. The exception is the verb *wait*, which the model assigns highest probability under the alternating category. This is due to prevalent uses of *wait* with temporal adjuncts, as in *wait a minute*, that were indistinguishable from NP direct objects in the CHILDES Treebank parse trees. Thus, a learner who cannot differentiate these adjuncts from direct objects would infer that *wait* is an alternating rather than intransitive verb.[2]

The model assigns 6 out of the 9 transitive verbs highest probability under the transitive category. Three transitive verbs are assigned highest probability under the alternating rather than the transitive category: *catch, hold,* and *wear*. This is likely because these verbs display different behavior than the other transitive verbs in the corpus. The verb *hold* occurs frequently in verb-particle constructions (e.g. *hold on*), which might be treated differently than simple verbs by learners. The verbs *catch* and *wear* appear to occur at much higher rates than other transitive verbs in non-basic clauses: *catch* occurs frequently in passives (e.g. *get caught*), and *wear* occurs frequently in wh-object questions (e.g. *what are you wearing?*). I leave for future work the question of whether children likewise mis-classify these verbs, or whether they can accommodate their different distributional behavior by using more sophisticated information than our modeled learner.

The model assigns highest probability for most of the alternating verbs to the alternating verb category. There are 13 exceptions. The verbs *pick, drop, lose, close, touch, leave*, and *wash* are assigned highest probability under the transitive category

[2]We can tell that *a minute* is not a direct object by observing that it can occur non-adjacent to the verb, which is not generally the case for direct objects in English: compare **Fix here the car* vs. *Wait here a minute*. Furthermore, direct objects can generally be passivized, but *a minute* resists passivization: *The car was fixed* vs. **A minute was waited*.

because they infrequently occur in their possible intransitive uses in child-directed speech. The verb *pull* is assigned equal probability under the transitive and alternating categories for the same reason. (This verb was not considered to be correctly assigned to the alternating category in the accuracy calculation.) The verbs *run, swim, walk, jump,* and *sit* are assigned highest probability under the intransitive category because these verbs very infrequently occur in their possible transitive uses.[3] Thus, the model over-regularizes the alternating verbs that alternate infrequently, preferring the transitive and intransitive verb categories.

Filter Parameter Inference. Recall that the model identifies verb transitivity categories by jointly inferring parameters for filtering its input. These parameters are ϵ, which represents the frequency of erroneous parses, and δ, which represents whether those errors are likely to cause direct objects to go missing, or to spuriously appear. Figure 4.3 displays the posterior probability distributions inferred by the model for ϵ and δ. In order to evaluate the model's inference of these parameters, we can estimate their true value in our dataset. The proportion of transitive verbs with missing overt post-verbal direct objects in the dataset gives us an estimate of $(1-\delta)\times\epsilon$, and the proportion of intransitive verbs with spurious direct objects (e.g. *wait a minute*) gives us an estimate of $\delta \times \epsilon$. Solving these two equations, we find that $\delta = 0.18$ and $\epsilon = 0.24$. The posterior probability distribution over δ inferred by the model has a mean of 0.25, and the probability distribution over ϵ has a mean of

[3]Note that four out of these five verbs are manner of motion verbs (*run, swim, walk, jump*), and their transitive uses do not typically involve agent-patient relations (e.g., *walk a mile*, *swim the channel*, *jump the turnstile*). Even when a causative meaning may be used, as in the case of *jump the horse*, this implies less direct causation than a typical alternating verb such as *break* or *open*. So, even though the conservative target categories treated these verbs as alternating, in some ways they behave more typically like intransitives.

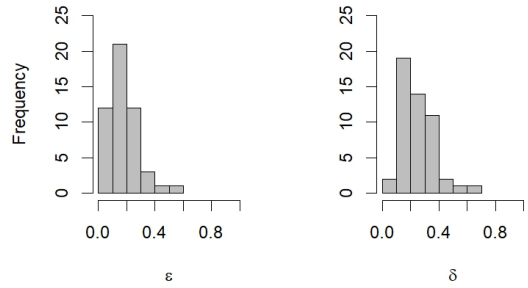

Figure 4.3: Posterior Distributions over ϵ and δ, Simulation 1

0.19. The model thus slightly over-estimates the value of δ and under-estimates the value of ϵ, but it infers values for these parameters that are close to the true values in the corpus.

4.3.3 Model Comparisons

Oracle Model. The primary contribution of the joint inference model is demonstrating that a learner can filter its input without knowing anything in advance about what needs to be filtered out. Therefore, it makes sense to compare this model against an "oracle" that knows a lot about what needs to be filtered out. I instantiated an oracle model in which δ is fixed to 0.18 and ϵ to 0.24 in order to reflect their true values in the dataset, as estimated in the previous section. This oracle model thus knows the parameters for the input filter in advance: it knows how frequently erroneous parses are likely to occur, and how they will behave. By comparing the joint inference model to this oracle, we can determine whether our model's performance is impaired by having to learn these parameters.

The posterior probability distributions over verb categories inferred by the

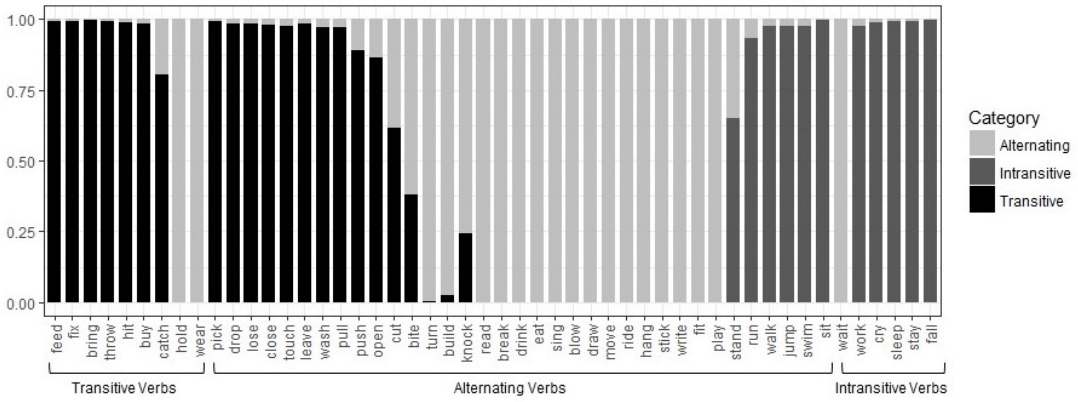

Figure 4.4: Posterior Distributions over Verb Categories (T), Oracle Model

oracle model are displayed in Figure 4.4. The posterior probabilities inferred by the oracle are less graded than those inferred by our joint inference model; this is unsurprising, as the oracle considers only one value each for δ and ϵ instead of sampling over multiple values. But when considering which transitivity category is assigned highest probability to each verb, the two models classify most of these verbs in the same way. The joint inference model classifies intransitive verbs identically to the oracle model, and performs almost as well with transitive verbs: the oracle succeeds in identifying one more transitive verb, *catch*, as transitive. Our model performs better than the oracle in categorizing alternating verbs: the oracle has an even higher tendency to over-regularize the verbs that alternate infrequently. Inferring the parameters of the input filter thus results in comparable, and maybe slightly better, accuracy in categorizing verbs than knowing these parameters in advance.

Random Filter Parameters. If the values of the filter parameters aren't important, then it wouldn't be remarkable that our joint inference model performs

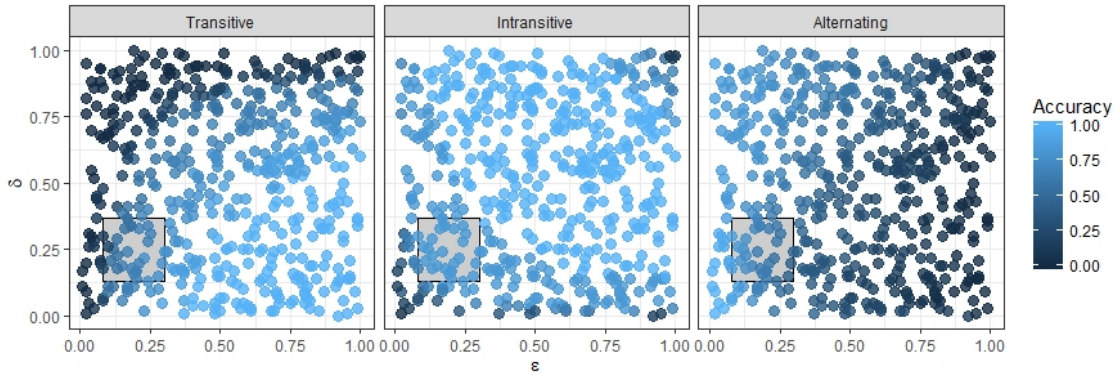

Figure 4.5: % Verbs Categorized Correctly by Varying Values of ϵ and δ

comparably to the oracle model. To test whether the filter parameters actually matter, I ran 500 model simulations in which ϵ and δ were fixed to randomly-sampled values. Fig. 4.5 displays the model's resulting accuracy in inferring transitivity categories given each set of filter parameters, with ϵ along the x-axis and δ along the y-axis. Lighter colors denote higher percentages of verbs categorized correctly. The gray rectangle marks the range of filter parameter values that were considered highest probability by our joint inference model— specifically, these are the values within one standard deviation of the mean in the posterior probability distributions that our model inferred.

A visual scan of these plots shows that it is not trivial to infer filter parameters that will result in high accuracy across all three transitivity categories. Higher values of ϵ yield higher accuracy on categorizing transitive and intransitive verbs, but lower accuracy for alternating verbs. This is because the learner assumes there is more error in its transitive and intransitive verb observations, and lowers the threshold for assigning verbs to those categories. The learner thus assigns more verbs in its dataset to the transitive and intransitive categories rather than the

alternating category. On the other hand, higher values of δ yield lower accuracy for transitive verbs, but higher accuracy for intransitive verbs. With higher values of δ, the learner assumes that more of its errorful sentence observations contain mistaken direct objects, rather than missing direct objects. The learner therefore expects more error in its intransitive verb observations because there should be more intransitive verbs appearing with spurious direct objects. This lowers the threshold for assigning verbs to the intransitive class, resulting in higher accuracy for intransitives. Conversely, the learner expects less error in its transitive verb observations because there should be fewer transitive verbs appearing with missing direct objects. This raises the threshold for assigning verbs to the transitive class, resulting in lower accuracy for transitives.

Thus, successfully categorizing verbs in all three transitivity classes requires inferring filter parameters that fall within a somewhat narrow range. Our model performs comparably to the best-case oracle model not merely because it infers an input filter, but because it infers the best parameters for such a filter given our dataset. Note that our model is not actually optimizing for the accuracy values plotted in the graph in Fig. 4.5, because it is not trained on the target classifications for verb transitivity. Instead, the model is optimizing for probability: it is searching for the best joint-probability solution for verb transitivity categories and filter parameters to explain the distributions in its data. The fact that our model performs well with respect to our target verb classifications means that the parameter values that have high probability under our model also result in good accuracy across all three verb classes.

No-Filter Baseline. Our model accurately categorizes verbs across transitivity categories by inferring appropriate parameters for a filter on its input, and the model comparisons above show that the values of these filter parameters are important. Models with grossly inappropriate filter parameters might have better accuracy on some verb classes, but do not perform as well across all three transitivity categories. A special case of these models would be those where ϵ equals exactly zero, representing zero probability of parsing errors: this produces models that do not have an input filter at all. Comparing against a no-filter baseline tells us how much having a filter matters in identifying verb transitivity.

As values of exactly zero were never randomly sampled in the simulations reported in Fig. 4.5, I conducted an additional simulation setting ϵ to zero. The value of δ in this case does not matter, because it is never used. Because every verb in our dataset occurs some but not all of the time with overt post-verbal direct objects, and this no-filter model assumes there are no parsing errors to filter out, it assigns every verb to the alternating category. It thus categorizes 100% of the alternating verbs correctly, achieving 70% overall accuracy because alternating verbs make up 70% of our dataset. However, this accuracy comes at the cost of failing to categorize any verbs as transitive or intransitive. The joint inference model performs substantially better in this regard, categorizing the majority of transitive and intransitive verbs correctly. This demonstrates that an input filter is important for differentiating alternating from non-alternating verbs.

Threshold Comparisons. By inferring how frequently parsing errors occur in its sentence observations and the behavior of those errors, our model is essentially

inferring where to put thresholds for classifying verbs as transitive or intransitive based on rates of observed direct objects. Another way of evaluating our model's performance is to compare it against a simple threshold model, which classifies verbs as transitive if their percentage occurrence with overt direct objects falls above a certain threshold, and as intransitive if their percentage occurrence with overt direct objects falls below a certain threshold. There are several differences between this type of threshold model and our model. Instead of setting hard thresholds that delineate each of these categories, our model uses soft thresholds that take into account how much data it has available for any particular verb. And the primary advance in our model is that these soft thresholds are learned: the model does not need to know the true distributions of transitive and intransitive verbs in advance. If our model performs comparably to a model that knows the best thresholds for classifying its data, this will give us another indication that it is learning successfully.

To create these comparisons, I hand-fit the thresholds for classifying verbs by percentage overt direct objects to maximize accuracy on the model's dataset. Table 4.4 reports the accuracy of the best-performing threshold models, compared to our joint inference model. The thresholds that yielded the best performance overall were 87% and 4%: this model classifies verbs as transitive if they occur with direct objects above 87% of the time, and verbs as intransitive if they occur with direct objects less than 4% of the time. This model was able to achieve 80% accuracy overall. However, its performance on classifying transitive and intransitive verbs was lower than for our joint inference model. The second threshold comparison thus aimed to maximize overall accuracy without performing lower than our joint

Model	**% Transitive**	**% Intransitive**	**% Alternating**	**% Total**
Joint Inference	0.67	0.83	0.63	0.66
87% & 4% Thresholds	0.56	0.66	0.89	0.80
83% & 5% Thresholds	0.67	0.83	0.71	0.72
76% & 15% Thresholds	0.78	1.00	0.54	0.64

Table 4.4: Percentages of Verbs Categorized Correctly by Threshold Models

inference model on these two verb classes. Thresholds of 83% and 5% allowed the model to achieve 72% overall accuracy, while achieving the same accuracy as our joint inference model on transitive and intransitive verbs. Finally, the third threshold comparison attempted to maximize overall accuracy while achieving *higher* accuracy than our joint inference model on transitive and intransitive verbs. The best thresholds for this model were 76% and 15%. This threshold model's higher performance on transitive and intransitive verbs led to lower accuracy on alternating verbs, and it only achieved 64% accuracy overall.

Although our joint model is not explicitly learning thresholds, we can use the filter model parameters that our model inferred to estimate the soft thresholds it is effectively using. Because ϵ is the inferred rate of error and δ is the inferred proportion of error that has direct objects, $\epsilon \times (1 - \delta)$ gives an estimate of the rate of missing direct objects for transitive verbs. Therefore, $1 - \epsilon \times (1 - \delta)$ can be interpreted as a threshold of direct object rates above which verbs are more likely classified as transitive. Conversely, $\epsilon \times \delta$ estimates the rate of spurious direct objects for intransitive verbs, and thus provides an estimate for a threshold below which verbs are more likely classified as intransitive. When we estimate thresholds based on the means of the distributions over ϵ and δ that our model inferred (0.19 and

0.25), we obtain estimated thresholds of 85% and 5%. These are very close to the thresholds that yielded the best performance in our threshold models.

In summary, these comparisons show that it is possible for a simple threshold model to achieve higher overall accuracy than our joint inference model, if it is allowed to use thresholds that are hand-fit to maximize performance on this dataset. However, it is not trivial to find hard thresholds that will ensure high performance across all three verb classes. In particular, the best-performing threshold models may have exceeded the overall accuracy of our joint inference model, but they never exceeded our model's accuracy on both transitive and intransitive verbs without reducing overall accuracy. This shows us that the soft thresholds that our model is essentially learning are appropriate to its dataset: our model performs just as well as the best-performing threshold models on identifying these deterministic verb categories. And this is true even though our model is not optimizing for accuracy. Unlike the threshold models, our model does not have access to the target classifications for verb transitivity in its dataset, and cannot use those classifications to identify its thresholds. Instead, our model learns where to put these soft thresholds by finding the best joint probability solution for verb transitivity categories and the parameters for error in its dataset.

4.3.4 Discussion

Our model accurately categorizes 2/3 of the most frequent transitive, intransitive, and alternating verbs in child-directed speech on the basis of their distribu-

tions with and without direct objects, by learning to filter out sentences that were likely mis-parsed. This enables the learner to avoid drawing faulty inferences about verb transitivity from non-basic clause types that may be mistaken for intransitive clauses. Our model performs comparably to an oracle model that knows in advance the best parameters for a filter given its dataset, and better than many models with inappropriate filter parameters. It performs substantially better in categorizing transitive and intransitive verbs than a baseline model that lacks an input filter altogether, and performs twice as well overall as would be expected by chance. It also performs just as well on categorizing transitive and intransitive verbs as the best-performing threshold models, which categorize verbs using thresholds of direct object rates that are hand-fit to the dataset. These results demonstrate that an input filter both matters for verb transitivity learning, and can be learned.

The model makes two types of mistakes in inferring verb categories. First, it is unable to correctly categorize some transitive and intransitive verbs that behave differently than other verbs in their category, such as *catch, hold, wear*, and *wait*. Further investigation is necessary to determine whether these verbs pose difficulties for child learners as well. A second type of mistake is over-regularizing alternating verbs that alternate infrequently: the model prefers to assign these verbs to the transitive and intransitive categories. This is an example of a learner preferring a more deterministic analysis for probabilistic input, a tendency also found in child learners in artificial language studies (Hudson Kam & Newport, 2009). The error-filtering mechanism I present here could thus potentially provide a way to model other forms of over-regularization in learning.

There are three factors that contribute to our model's ability to regularize its input. First, our learner only needs to infer two parameters for its input filter: it makes the simple assumption that there is a single value for ϵ and δ shared across all verbs, rather than having to infer separate values for these parameters on a verb-by-verb basis. This allows the learner to use distributions of direct objects across verbs to inform its estimates of how much error is present in its sentence representations, and what that error looks like. If instead the learner expected a different ϵ and δ for each verb, it would be difficult for the learner to tell whether a particular rate of direct objects observed for a verb is due to a particular rate of transitivity alternation (θ) or due to a particular type of error that occurs only with that verb.

Intuitively, the expectation of a single shared value for these filter parameters corresponds to the expectation that the noise process generating the error in the learner's sentence representations reflects some properties that are independent of the particular verbs in those sentences. Arguably, this expectation is not only a helpful simplification, but also a realistic one. While our learner has no commitment to what this noise process is, in reality it reflects the contribution of a variety of grammatical operations that the learner has mis-parsed. These operations are due to independent properties of the grammar, and apply to entire classes of verbs, not on a verb-by-verb basis. A more sophisticated learner might identify that there are several noise processes at work, corresponding to these different grammatical properties, and use distributions of direct objects across verbs along with other surface features of these sentences to infer a different ϵ and δ for each of these properties.

Additionally, the learner's inference of its input filter is successful because it encounters a wide variety of verb behavior in its data. Some verbs appear more deterministic than others: they alternate less frequently, instead show a stronger preference for solely transitive or intransitive frames. Just as we used the true transitive and intransitive verbs in the dataset to arrive at our estimates of the true values for ϵ and δ, our learner can anchor its estimates of these parameters by using the distributions of direct objects with the more deterministic verbs it observes—those that it thinks are more likely to be transitive or intransitive. If instead all verbs alternated at exactly the same rate, the learner would have difficulty knowing whether all verbs have exactly the same transitivity properties, or whether there is additional error present. This raises the question of whether all languages have enough variety in verb distributions to enable successful learning by this filtering mechanism. Answering this question would require testing this model with cross-linguistic corpora of child-directed speech.

Finally, our learner's ability to successfully regularize depends on having deterministic categories in its hypothesis space: it expects that some verbs will only occur in transitive or intransitive frames, and makes the simplifying assumption that these verbs are equally likely *a priori* as verbs that can alternate. However, we might ask how realistic it is for a learner to have this assumption, as in reality these categories will occur in different proportions in the target language. Will a learner perform just as well if it expects transitive, intransitive, and alternating verbs to occur with different frequency? We can answer this question by examining the model's performance when it has different prior beliefs about the probability of

these verb classes. If there is no difference in performance, then it suffices to merely have transitive or intransitive categories in the learner's hypothesis space, regardless of how they are weighted. But if there is a difference in performance, this would show that the model's prior beliefs about the relative probabilities of transitivity classes matter for its ability to identify these classes in its input.

4.4 Simulation 2

In Simulation 2, I ask whether our model will still accurately identify the transitivity categories of verbs in child-directed speech if it does not expect transitive, intransitive, and alternating verbs to be equally likely *a priori*. Instead of setting a uniform prior over transitivity categories ($P(T^{(v)})$ in Equation 1), the model's prior is now biased in favor of alternating verbs. In Simulation 2a, the model's prior was set to match the actual frequencies of verb transitivity categories in its input: a prior probability of 0.70 was set for alternating verbs, 0.18 for transitive verbs, and 0.12 for intransitive verbs, to match the proportion of the target verb categories in our dataset. This allows us to determine whether our learner's verb transitivity inference is affected if it expects to find verb categories in the same proportions as they will actually occur in its input. In Simulation 2b, the model's prior was skewed even more heavily in favor of the alternating category: a prior probability of 0.90 was set for alternating verbs and 0.05 each for transitive and intransitive verbs. By giving the alternating category substantially greater prior probability than the two deterministic verb categories, we can determine whether simply having transitive

and intransitive categories in the learner's hypothesis space, in any proportion, is sufficient for identifying them in its input.

4.4.1 Data

I tested the skewed-prior models on the same dataset of transitive, intransitive, and alternating verbs in child-directed speech that was prepared for Simulation 1.

4.4.2 Results

Verb Transitivity Inference. Fig. 4.6 displays the posterior probability distribution over transitivity categories that our model inferred for each verb in Simulation 2a, when it expected 70% alternating verbs. Fig. 4.7 displays the distribution over transitivity categories inferred in Simulation 2b, when the model expected 90% alternating verbs. Table 4.5 reports the proportion of verbs categorized correctly in each transitivity category, compared to our original joint inference model in Simulation 1.

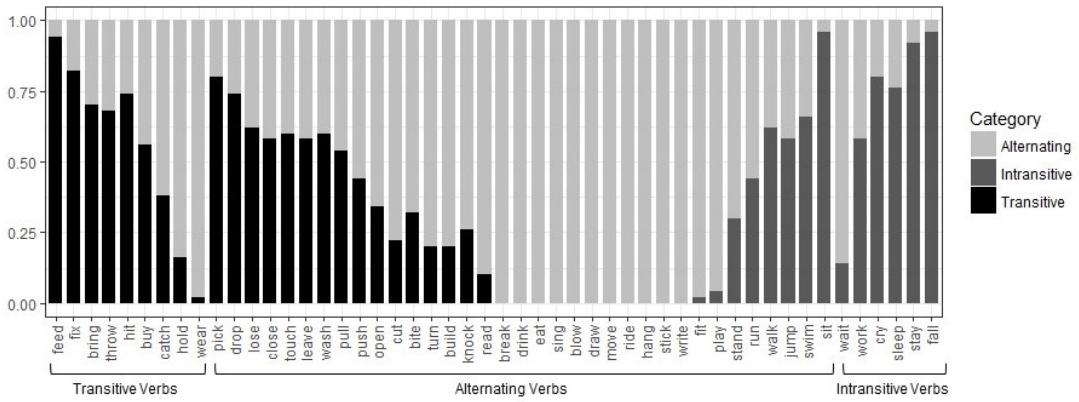

Figure 4.6: Posterior Distributions over Verb Categories (T), Simulation 2a

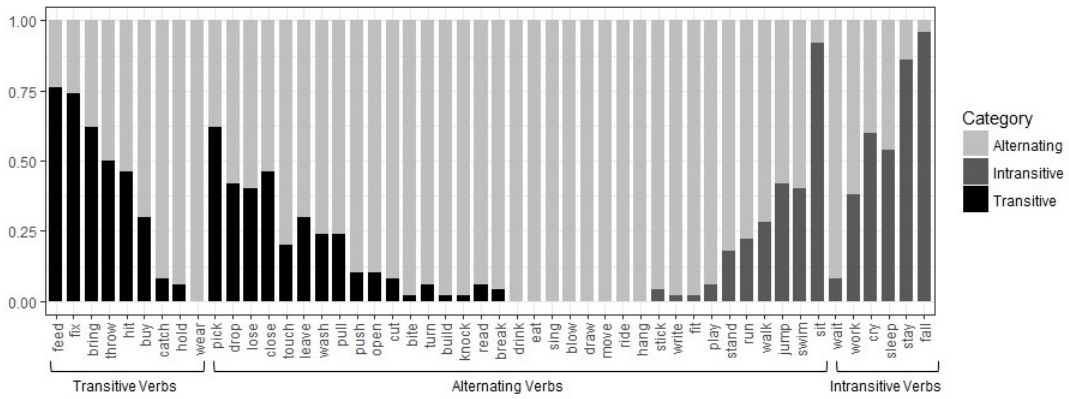

Figure 4.7: Posterior Distributions over Verb Categories (T), Simulation 2b

In Simulation 2a, the inferred distribution over transitivity categories is very similar to the distribution inferred by our original model in Simulation 1. This model assigns highest probability under the transitive category to the same 6 out of 9 transitive verbs as our original model, and it assigns highest probability under the intransitive category to the same 5 out of 6 intransitive verbs. The model also assigns highest probability under the alternating category to 23 alternating verbs, and considers the remaining 12 to be either transitive or intransitive, over-regularizing at nearly the same rate as our original model. Thus, skewing the model's prior to expect alternating verbs 70% of the time resulted in very little difference in verb categorization accuracy compared to our original model.

In Simulation 2b, when the model's prior was skewed to expect alternating verbs 90% of the time, the model inferred a different distribution over transitivity

Model	**% Transitive**	**% Intransitive**	**% Alternating**	**% Total Verbs**
Simulation 1	0.67	0.83	0.63	0.66
Simulation 2a	0.67	0.83	0.66	0.68
Simulation 2b	0.33	0.67	0.94	0.80

Table 4.5: Percentages of Verbs Categorized Correctly, Simulations 1 and 2

categories. There are two general trends to observe in these data. First, even though this learner was heavily biased against the transitive and intransitive categories, there are still several verbs that it assigns high probability under these categories. To some extent, the model was able to overcome its biased prior and identify some deterministic verbs in its input.

On the other hand, there are fewer verbs that this model assigns highest probability under the transitive and intransitive categories, and more verbs that it assigns highest probability under the alternating category. This results in higher accuracy for alternating verbs: this model only over-regularizes one of these verbs (*pick*) as transitive, and one of these verbs (*sit*) as intransitive. Because alternating verbs are most frequent in the model's data, the model's higher accuracy on alternating verbs leads to higher total accuracy as well. But the model achieves lower accuracy for the transitive and intransitive categories. The model assigns highest probability to the transitive category for only 3 of the 9 transitive verbs, and it assigns highest probability to the intransitive category for only 4 of the 6 intransitive verbs. Of the target transitive verbs, the model now considers *throw, hit*, and *buy* to be alternating, along with *catch, hold*, and *wear*. Of the intransitive verbs, the model now considers *work* to be alternating along with *wait*. The model still performs better than chance in categorizing intransitive and alternating verbs, but it is no different from chance in categorizing transitive verbs.

In summary, the model performed comparably to our original model when its prior was skewed to expect transitive, intransitive, and alternating categories in the same proportions as they actually occur in the input. However, when the model

was biased more strongly towards the alternating category, it identified transitive and intransitive verbs at a much lower rate. The model's rate of regularization was not affected by its bias against deterministic categories in Simulation 2a, but was affected by its stronger bias in Simulation 2b.

Filter Parameter Inference. Figs. 4.8 and 4.9 display the posterior probability distributions over ϵ and δ inferred by the skewed-prior models. Although the shapes of these distributions are different, they are centered around similar values as those inferred by our original model in Simulation 1. The mean of the distribution over ϵ is 0.22 in Simulation 2a and 0.19 in Simulation 2b, compared to 0.19 for our original model. The mean of the distribution over δ is 0.23 in Simulation 2a and 0.21 in Simulation 2b, compared to 0.25 for our original model. Just as for our original model, these values are close to the estimated true values of $\epsilon = 0.24$ and $\delta = 0.18$ in the model's dataset, as calculated for Simulation 1.

Thus, changing the learner's prior beliefs about how transitivity categories distribute in its input did not substantially affect its inference about the parameters

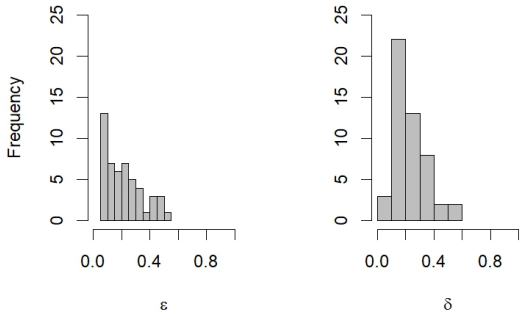

Figure 4.8: Posterior Distributions over ϵ and δ, Simulation 2a

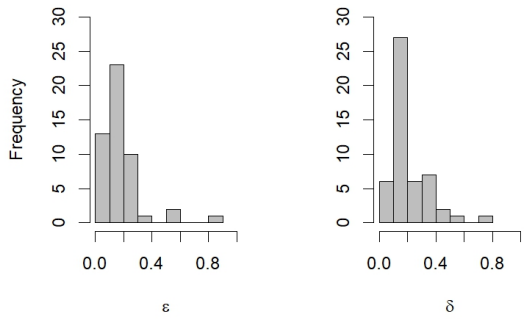

Figure 4.9: Posterior Distributions over ϵ and δ, Simulation 2b

of its input filter: it still inferred appropriate values for the frequency and behavior of error in its data. This might be because the learner is anchoring that inference on the distributions of the verbs that it considers to be transitive and intransitive with the highest probability. Because both models in Simulation 2 did identify some transitive and intransitive verbs, and those verbs are a subset of the verbs that our original model categorized as transitive and intransitive with highest probability, it is not so surprising that all three models found similar parameters for their input filters. Moreover, inferring these parameters is what allowed the model in Simulation 2b to still categorize some verbs as transitive and intransitive, despite its strong bias against those categories. Without a filter, the model would perform identically to the no-filter baseline in Simulation 1, and categorize all verbs as alternating.

4.4.3 Model Comparisons

Random Prior Parameters. Different results for the model's verb transitivity inference were found depending on how much its prior was biased against

transitive and intransitive verbs. This raises the question: under what circumstances does the model's prior substantially affect its ability to identify verb transitivity, and under what circumstances does it not matter? That is, how much bias against deterministic verb categories can our learner accommodate and still accurately identify those categories in its input?

To answer this question, I ran 500 model simulations in which the model's prior probabilities over transitive, intransitive, and alternating categories were fixed to randomly sampled values that summed to 1. Because the models in Simulations 1 and 2 inferred similar values for ϵ and δ, for ease of computation I set these filter parameters to the mean values of $\epsilon = 0.20$ and $\delta = 0.23$ that were inferred in those previous simulations. Fig. 4.10 plots the learner's accuracy in categorizing transitive, intransitive, and alternating verbs as its prior becomes more skewed towards the alternating category. The x-axis displays varying values of the model's prior on alternating verbs, and the y-axis displays the average percentage of verbs in each class categorized correctly at each of those values. A curve of best fit is plotted using a running LOESS regression (local nonparametric regression; Cleveland & Devlin, 1988).

This plot shows that the learner's accuracy in verb categorization remains steady across a large range of prior parameter values. When its prior probability on alternating verbs is less than approximately 0.75, the learner's performance is fairly consistent: it correctly categorizes on average 6/9 transitive verbs, 5/6 intransitive verbs, and 22/35 alternating verbs. Performance only begins to vary when its prior probability on alternating verbs is pushed above 0.75. Above this value, its accuracy

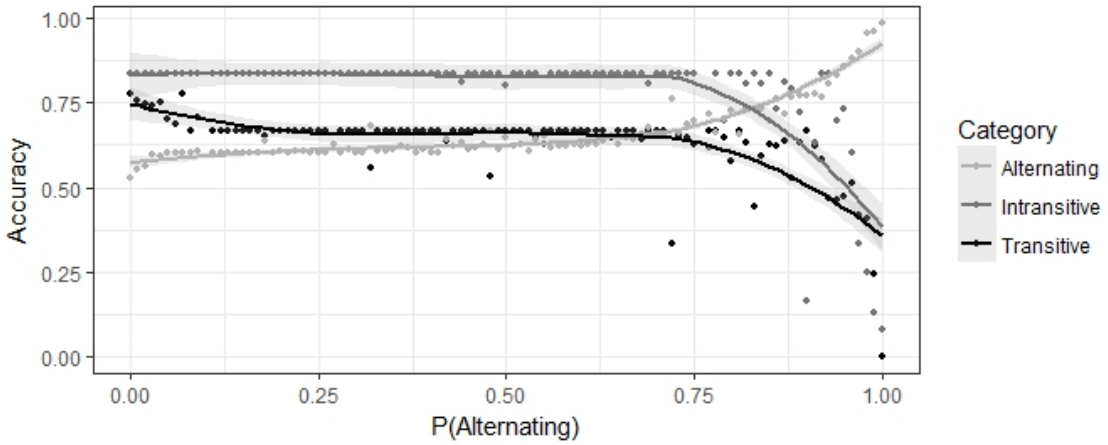

Figure 4.10: Accuracy by Prior Probability on Alternating Verbs

on categorizing transitive and intransitive verbs declines and its accuracy on alternating verbs increases, as it categorizes fewer verbs as transitive and intransitive. Thus, it appears that there is a large range of bias towards or against deterministic verb categories that our learner can accommodate without affecting its ability to identify those verbs in its input. It only begins to lose that ability when its bias against deterministic categories becomes extreme.

4.4.4 Discussion

While Simulation 1 shows that an appropriate input filter is important for learning verb transitivity, Simulation 2 shows that learning is also affected to some extent by the learner's prior beliefs about the relative frequency of transitivity categories in its input. Skewing the model's prior to expect verb transitivity categories in the same proportions that it would actually encounter in its input did not affect its performance; its accuracy in categorizing transitive, intransitive, and alternating verbs was nearly identical to our original model. However, skewing the model's prior

more extremely in favor of alternating verbs resulted in different performance. With a heavy bias against deterministic categories, the model over-regularized alternating verbs much less, leading to higher accuracy on that verb class and higher accuracy overall. But the model was also less successful at identifying the target transitive and intransitive verbs, and did not perform above chance levels in categorizing transitive verbs.

This behavior reveals two properties of the learner. First, it did not matter whether the learner expected transitive, intransitive and alternating verbs to be equally likely *a priori*, or whether it expected them to occur in the same proportions as they actually do occur in child-directed English. In fact, it appears that our model's performance would be very similar across a large range of prior parameters. It is desirable that the learner can succeed at identifying verb transitivity without prior expectations that match the proportions of transitivity categories in the input— this will allow the learners to be somewhat flexible in learning different target languages, even if transitivity categories distribute differently in those languages compared to English or compared to the learners' own priors. However, there is a point where the learner's prior does exert an influence on its verb categorization. When it was extremely biased to expect alternating verbs, our learner was not able to successfully categorize transitive verbs. This means that merely having deterministic categories in the learner's hypothesis space, in any proportion, does not suffice for accurately identifying those categories in the learner's input. A learner must give those categories sufficient prior weight in order to find them.

Second, even a learner strongly biased in favor of alternating verbs was able to

infer appropriate parameters for filtering sentences that were likely mis-parsed. This allowed it to identify at least some of the transitive and intransitive verbs in its input, and to avoid drawing the mistaken inference that all verbs are alternating. This filtering was less effective for a learner with an extreme bias against the transitive and intransitive categories: its bias hampered its ability to detect the signals of these deterministic categories in the data that it let through its filter. However, the fact that the learner inferred appropriate filter parameters even in this case points towards a promising direction for future research. A more sophisticated learner might incrementally update its prior over transitivity categories given more evidence about their distribution in its input, inferring the parameters of that distribution in a hierarchical model. In this case, the learner's correct initial estimates of its input filter parameters could be very helpful in identifying the right distribution over transitivity categories. Thus, even if a learner's prior beliefs about transitivity are grossly inaccurate, inferring an input filter might allow it to appropriately adjust those beliefs as it learns more about its language.

4.5 General Discussion

This chapter provides a computational solution for a chicken-and-egg problem that arises in early grammar acquisition. English learners are sensitive to verb transitivity at very young ages: infants as young as 15-16 months use verbs' distributions in transitive and intransitive clauses to draw inferences about their argument-taking properties and meanings (Jin & Fisher, 2014; Lidz et al., 2017). The experiments in

Chapter 3 show that verb transitivity knowledge emerges before infants are able to identify displaced arguments in common non-basic clause types, such as wh-object questions; by hypothesis, this may be because transitivity knowledge facilitates non-basic clause structure acquisition. But accurately identifying verb transitivity is difficult if infants cannot recognize when arguments of the clause have been displaced from their canonical positions, and therefore cannot reliably recognize clause transitivity when it is present. In this chapter I have followed a proposal that children need to filter non-basic clauses out of the data they use for verb learning (Lidz & Gleitman, 2004a, 2004b; Pinker, 1984, 1989), but filtering non-basic clauses by identifying them *as* non-basic only re-introduces our paradox. Identifying the structure of non-basic clauses may depend on knowing some of the core argument structure properties of the language, and yet learners need to filter non-basic clauses in order to bootstrap their learning of those very properties.

The model proposed here offers a new solution to this paradox, which does not require the learner to detect any direct or indirect signals to non-basicness (Pinker, 1984, 1989; Gleitman, 1990). This model instantiates a learner that considers the possibility that it occasionally parses sentences erroneously. The learner infers how to filter out errors from the data it uses for verb learning, without knowing where those errors came from. It observes only verbs' distributions with and without direct objects, and does not track any additional syntactic or non-syntactic cues that might correlate with non-basicness—to this learner, a wh-object question is indistinguishable from an intransitive clause. Nonetheless, the model successfully infers appropriate parameters for filtering its input in order to identify the transitivity of

the majority of frequent verbs in child-directed speech. This demonstrates that it is in principle possible for a learner to filter non-basic clauses for verb learning, without knowing which clauses are non-basic and without needing to infer what the features of non-basic clauses are. This provides an account for how the first attested steps of verb argument structure learning in infancy can take place even as non-basic clause acquisition is still developing.

More broadly, by introducing a mechanism for a learner to filter erroneous parses of its input, this model helps answer what has remained an open question in bootstrapping and verb learning: how learners manage to avoid drawing faulty inferences about grammar and meaning, at stages of development when they lack the linguistic knowledge to arrive at veridical syntactic representations of sentences they hear. This ability has been traditionally assumed by theories of both syntactic and semantic bootstrapping (Lidz & Gleitman, 2004a, 2004b; Gleitman, 1990; Pinker, 1984, 1989), and has been presupposed by previous computational models of verb learning (Alishahi & Stevenson, 2008; Barak et al., 2014; Parisien & Stevenson, 2010; Perfors et al., 2010; see also Siskind, 1996). These previous models assume that learners can veridically represent the arguments in a clause, and use those syntactic percepts to identify verbs' core argument-taking properties and their ability to productively generalize across different argument structure alternations. This model addresses the question of how this process begins. I propose that a learner equipped with a filtering mechanism can still identify a verb's basic argument structure, even before that learner can reliably identify all of the arguments in sentences she hears.

This model diverges from previous computational models of bootstrapping

(Abend et al., 2017; Kwiatkowski et al., 2012; Maurits et al., 2009; Siskind, 1996) by learning from a very limited type of data. The learner identifies verb transitivity only by using rates of overt direct objects, and does not have access to any additional syntactic or non-syntactic features of the sentences or the discourse environment. By limiting the learner's data in this way, I do not imply that real-life learning proceeds only from this type of distributional information. On the contrary, it is likely that children make simultaneous use of a much fuller set of information in inferring a grammar, including conceptual representations of the extra-linguistic contexts of the sentences they hear. But by investigating how much can be learned solely from verbs' syntactic distributions, I am testing the viability of the proposal that infants can use syntactic information to draw helpful generalizations even if they do not know which event in the world a particular sentence describes (Gleitman, 1990).

This issue has not been fully examined in prior bootstrapping models, which assume that learners begin by accessing the exact meaning (or set of possible exact meanings) of a sentence, represented under a structure that is homomorphic with the syntactic structure (Abend et al., 2017; Kwiatkowski et al., 2012; Maurits et al., 2009; Siskind, 1996). Given access to this full conceptual representation, or instead to the full syntactic representation of a sentence, these models show that it is simple to learn how to convert from one representation to the other. This is because the learner's meaning representation is in a form that encodes all and only the predicate-argument relations in the syntactic representation of the sentence, and there is an assumption built into the learner that those two representations will mirror each other. The bootstrapping task thus reduces to the problem of identifying which

lexical items express which predicates and arguments in the learner's conceptual structure. Given this information, the learner can infer the syntactic representation of a sentence by reading off of its structured conceptual representation, and vice versa.

But bootstrapping is not so simple if learners only have access to approximations of these representations, or if conceptual structures encode more relations than those expressed in the sentence's argument structure. Even if children can perceive events and event relations in the world in the same way as adults do— and the work discussed in Chapter 2 provides some evidence for how we might be able to tell that this is the case (He, 2015; Knowlton et al., 2018; Perkins et al., 2018; Wellwood et al., 2015)— it is still not straightforward to identify which event relations a sentence expresses solely from its context of use (Gleitman, 1990). And when we consider the wide range of syntactic relations that might be instantiated in a particular sentence, including the various non-local dependencies found in non-basic clauses, it seems even less straightforward for the child's non-linguistic perception of the world to yield a meaning in a form that is homomorphic with the syntax of that sentence. Here, we ask whether learning can still succeed in cases where a child might not have access to conceptual and syntactic representations that mirror each other in their structure. If either of these representations is approximate or incomplete, then children must use whatever partial information might be useful in one domain— syntax or meaning— as probabilistic evidence for drawing inferences about the other domain. I show how learners might accommodate error in their syntactic percepts, such that those percepts are still useful as evidence for drawing

further generalizations about their language.

Crucially, an input filtering mechanism like the one I propose can flexibly adapt over the course of a learner's development. As a learner gains more knowledge of the grammar of her language, her syntactic percepts will change: she will be learning from more complete and more accurate parses of the sentences she hears. This means that the error in her syntactic percepts will also reduce over time, and she will not need to filter as much of her data for learning. In the current case, our model is learning from data that reflects the parses of an immature learner at a particular stage of development: one who cannot identify objects when they are realized in non-canonical positions, and who mistakes certain NP adjuncts for arguments. These data do not veridically reflect the distributions of verbs with direct objects in the actual input to the learner. Thus, the learner is not inferring filter parameters to fit its actual input— instead, the learner is inferring filter parameters to fit its erroneous representations of that input (cf. Gagliardi, 2012; Gagliardi & Lidz, 2014). A more mature learner who has learned to identify argument displacement in English will have access to a different dataset, one that has a lower rate of error. This more mature learner would identify different parameters for filtering its data in order to learn more about its grammar.

This model is merely a starting point: we have looked at only a subset of English action verbs. But having presented a proof of concept that our filtering solution is possible, we can ask how far it could generalize. It remains to be seen how well filtering would work if we consider a less idealized version of the learning problem: could this mechanism still succeed if a learner is not only acquiring the transitivity

properties of verbs in her language, but also many other complement-taking properties at the same time? Moreover, future work aims to test whether this model could be extended to languages with freer word order or rampant argument-drop. These linguistic properties may make it difficult for learners to identify clause transitivity even in simple, active, declarative clauses. For example, the relatively free word order of Japanese compared to English means that word order is less helpful for identifying subjects and objects in a clause, and learners must use language-specific case morphology instead; furthermore, the ability of Japanese speakers to freely drop the subject and/or object of a clause if it is salient in the discourse means that a learner must use discourse cues to recognize when silent arguments are present. For these reasons, languages like Japanese are potentially problematic for syntactic bootstrapping strategies that rely on learners accurately identifying transitive verbs (Lee & Naigles, 2005, 2008), but see Fisher, Jin, and Scott (in press) and Suzuki and Kobayashi (2017) for evidence that learners do nonetheless succeed. If our model can learn appropriate parameters for filtering out the relatively higher rate of potentially misleading data in languages like Japanese, this may help clarify how syntactic bootstrapping is possible in these languages.

More broadly, we might ask whether this filtering mechanism could generalize beyond verb transitivity learning, to other cases in language acquisition where learners must ignore misleading data in order to draw correct inferences about their language. For example, prior work has proposed that some form of input filtering is helpful in identifying vowel categories (Adriaans & Swingley, 2012), and in drawing the right generalization about the constraints on the antecedent of anaphoric

one in English (Pearl & Lidz, 2009). Filtering may also provide a mechanism for understanding why young learners tend more strongly than adults to regularize probabilistic input in artificial language studies (Hudson Kam & Newport, 2009), or how a learner can acquire the correct generalizations about his first language from noisy input by second-language speakers (Singleton & Newport, 2004). When our learner expects that error might be masking regularities in its data, filtering allows it to identify those regularities, and even to over-regularize in some cases. A combination of determinism in children's hypothesis spaces, along with the expectation of error in their input representations, may help explain when children draw deterministic generalizations about their language and how they draw the right ones.

Thus, this type of filtering mechanism might have broad utility for learners attempting to generalize from immature representations of their input. But more narrowly, it serves an important role in accounting for how infants begin to identify both the local and the non-local predicate-argument relations in their language. This model provides a proof-of-concept that infants can use the local relations they are more reliably able to perceive— subjects and objects in their canonical argument positions— to begin to identify the argument-taking properties of verbs in their language, even if they cannot always recognize those arguments in non-local positions. This provides a way out of the chicken-and-egg problem introduced by non-basic clauses in early grammar learning. It is not in principle necessary for infants to identify non-basic clauses in order to learn basic verb argument structure properties; instead, infants may be able to acquire this knowledge first, and then use it to identify non-basic clauses in their language. In the next chapter, I now turn to

the question of which mechanisms enable this next step of learning: identifying the forms that argument movement can take in the target language.

Chapter 5: Learning the Surface Forms of Argument Movement

5.1 Background

The filtering mechanism introduced in the previous chapter provides an in-principle solution for learners to identify the basic syntactic distributions of verbs in their language, despite the misleading data introduced by non-basic clauses in their input. Learners may then use those distributions to bootstrap into other grammatical properties of their language, including the movement dependencies that are present in those non-basic clauses. I have been considering a "gap-driven" hypothesis for how this bootstrapping takes place in languages with overt movement. Infants as young as 15 months may use their knowledge of verb argument structure to notice when predicted arguments of verbs are unexpectedly missing in their canonical positions— for example, noticing that predicted objects of *fix* are missing after the verb in (1) and (2).

(1) What did Amy fix?

(2) I like the bicycle that Amy fixed.

As their parsing abilities develop, infants may then examine the rest of the sentence to determine the cause of the missing argument, eventually identifying that

another expression is acting as that argument non-locally. This will allow them to learn the underlying syntactic dependencies in these sentences, and the forms that these dependencies take in their language: i.e. that (1) and (2) both contain wh-dependencies, which are marked in English by various surface signals, such as *what*, *that*, *do*-support, and subject-auxiliary inversion.

In Chapter 3, I provided empirical evidence consistent with the Gap-Driven Learning hypothesis: the ability to detect local verb transitivity violations is developmentally prior to the ability to recognize fronted arguments in wh-questions. In Chapter 4, I showed that it is computationally feasible for a learner equipped with a filtering mechanism to accurately identify verb transitivity even before being able to parse non-basic clauses in its input, providing proof-of-concept that the first step of learning under this hypothesis is possible. In this chapter, I investigate the remainder of this learning process. I propose that this follows three logically independent steps:

(i) using knowledge of verb argument structure to detect argument gaps: predicted arguments that are unexpectedly missing in their local positions;

(ii) identifying what surface forms are correlated with these argument gaps; and

(iii) inferring what types of syntactic dependencies are responsible for those correlations.

I present a computational model that instantiates the first two steps of learning under this hypothesis. The learner builds off of the model in Chapter 4, using

the knowledge of verb transitivity properties that the previous model identified by filtering its input. It tracks the surface morphosyntactic properties of sentences that violate its expectations of verb transitivity, in order to identify clusters of properties that are correlated with those transitivity violations. These properties include the forms that characterize movement dependencies in English. Furthermore, I show that prior verb transitivity knowledge is necessary for this distributional learning to be successful: assuming that learners must build incrementally on prior knowledge, learners would do better to acquire verb transitivity knowledge before identifying the clusters of morphosyntactic properties that characterize non-basic clauses. This type of syntactically-informed distributional learning provides a mechanism that, in principle, allows learners to identify the forms that mark the realization of various types of argument movement in the target language.

5.1.1 Syntactically-Informed Distributional Learning

The learning mechanism proposed in this chapter is a form of syntactically-informed distributional learning. Under this account, children track the statistical distributions of morphosyntactic features present in the surface forms of sentences, and apply their knowledge of verb argument structure in order to draw inferences about the more abstract properties underlying those distributions. This distributional learning takes the form of categorization: learners jointly infer 'categories' of sentences according to their surface feature distributions, and which sentence categories contain locally missing arguments of verbs (gaps). This allows learners to

generalize across sentences that share similar surface features, and to identify which of those shared features signal movement dependencies in the target language.

This distributional learning mechanism follows prior computational work that has proposed similar mechanisms for the acquisition of phonetic categories in infancy, and for category learning domain-generally (e.g. Anderson & Matessa, 1990; Feldman, Myers, White, Griffiths, & Morgan, 2013; Maye et al., 2002; McMurray, Aslin, & Toscano, 2009; Sanborn, Griffiths, & Navarro, 2010). Similar to these previous models, the current account envisions the learning task as requiring two simultaneous inferences: discovering the underlying system of categories that give rise to distributions of surface features that a learner observes, and identifying which observations belong to which category. However, it departs from previous literature by envisioning this categorization process as merely a means to an end. Whereas the phonetic learning literature operates under the assumption that there is a set of phonetic categories to be acquired, here I do not assume that adult grammars represent 'categories' of sentences in any meaningful sense. Instead, the categories inferred by this learner are an intermediate step of learning: they enable further inference about the underlying properties of sentences that share similar features. For example, a learner who has identified that one sentence in a category contains an argument gap may then infer that this property holds of other sentences in the category as well, and may then posit that one of the shared features of these sentences— such as a wh-word— is instantiating the movement that generated those gaps. Thus, the formation of categories allows the learner to draw generalizations across sentences that are formally similar.

The success of this learning process hinges on which features of sentences learners are able to use for distributional learning at the relevant point in development, and whether those feature representations provide enough signal for a learner to accurately infer which sentences contain movement. Under most phonetic learning accounts, learners draw inferences about the sound system of their language by primarily using relevant acoustic features of sounds. Likewise, I assume that learners draw inferences about the syntactic operations in their language by primarily using relevant morphosyntactic features of sentences. I furthermore assume that inferences about argument movement should be primarily driven by information relevant to the predicate-argument structure of a sentence: morphosyntactic features pertaining to subjects, objects, and verbs. But these features must be restricted to those that would be identifiable by an infant at the developmental stage in question. Here, I briefly review the empirical evidence for the relevant morphosyntactic features that infants can represent prior to 18 months.

Local Subjects and Objects. The prior experimental findings discussed in Chapters 2 and 3 provide evidence that English-learning infants before 18 months can recognize subjects and objects of verbs in their canonical sentence positions. In the right experimental settings, infants at this age can use these local predicate-argument relations to draw inferences about sentence meanings (Jin & Fisher, 2014; Hirsh-Pasek & Golinkoff, 1996; Seidl et al., 2003; Gagliardi et al., 2016; Perkins & Lidz, under review). Moreover, Lidz et al. (2017) showed that 16-month-olds can differentiate direct objects from post-verbal prepositional objects, drawing different inferences about the meanings of nouns when they occur after a verb but follow

prepositions like *with* and *on*. These results suggest that learners at this age are attempting to identify subjects and objects as arguments in particular syntactic positions, rather than merely as nominal elements before or after verbs, although linear position may be a useful proxy for recovering the syntactic positions of these arguments.

Auxiliaries. Additional work shows that infants before 18 months can track functional elements with respect to subjects, verbs, and objects. Geffen and Mintz (2015) found that infants as young as 12 months were able to use the presence of subject-auxiliary inversion to distinguish between auditorily-presented declaratives and polar questions, even when all prosodic information was removed. This indicates that infants at this age are able to notice when a subject of a clause is sentence-initial, versus when it is preceded by other lexical items, such as auxiliaries. Other work finds that 14- to 18-month-olds are sensitive to the presence of auxiliaries in their non-moved positions, and can use them to categorize an upcoming novel word as a verb (Hicks, Maye, & Lidz, 2007; He & Lidz, 2017). 15- to 18-month-olds are furthermore sensitive to the dependencies that can hold between auxiliaries and particular verbal forms, such as the dependency between *is* and *-ing* in English (Santelmann & Jusczyk, 1998). These results suggest that infants at this age attend to auxiliaries, notice when they are displaced before subjects, and are aware that they select for verbs.

Verbal Morphology. Infants at this age also show awareness of the various affixes that verbs can take. 15-month-olds recognize that the verbal suffix *-ing* is a bound morpheme of English, differentiating it from pseudo-suffixes on novel verbs

(Mintz, 2013). 15- to 19-month-olds moreover show emerging sensitivity to the licensing environments for this suffix, as well as for the past-tense suffix *-ed* and for the 3rd-person agreement marker *-s* (Figueroa & Gerken, 2019; Omaki, Orita, & Lidz, in prep; Santelmann & Jusczyk, 1998; Soderstrom, Wexler, & Jusczyk, 2002; Soderstrom, White, Conwell, & Morgan, 2007). Similar knowledge of verbal affixes has been found for other languages for infants in this age range, including French, Dutch, and German (Höhle, Schmitz, Santelmann, & Weissenborn, 2006; Nazzi, Barrire, Goyet, Kresh, & Legendre, 2011; Van Heugten & Shi, 2010). This indicates that infants at this point in development are beginning to identify the patterns of verbal morphology in their language, and may be taking steps to identify the tense, aspect, and agreement dependencies that these patterns instantiate.[1]

Other Functional Categories. In addition to auxiliaries and verbal affixes, a handful of other functional categories may be represented by infants younger than 18 months. Experimental findings show that 14- to 18-month-olds are aware of some basic syntactic properties of determiners (Hicks et al., 2007; Höhle, Weissenborn, Kiefer, Schulz, & Schmitz, 2004; Shi & Melanon, 2010; Cauvet et al., 2014; He & Lidz, 2017), pronouns (Cauvet et al., 2014), prepositions (Lidz et al., 2017), and negation (De Carvalho, Crimon, Barrault, Trueswell, & Christophe, under review). Infants in these studies used these functional categories to aid recognition of co-occurring known words, and to draw inferences about the grammatical properties

[1] I note that these results do not yet tell us how infants at this age represent these morphological dependencies: whether as a dependency between two strings, or as a structural dependency between two abstract grammatical categories, such as a subject or auxiliary and a verb. For the purposes of the simulations below, I make the more conservative assumption that infants are able to identify the surface forms that verbal morphology can take, but have not yet identified the function of these morphemes.

and meanings of co-occurring novel words. But other functional categories may not yet be identified by this age. The results in Chapter 3 suggest that infants prior to 18 months do not yet know which words are wh-words in English, and we have no evidence at this age for knowledge of complementizers, quantifiers, focus particles, or conjunctions (except for *and*, which may be acquired around this age or a little older (Arunachalam et al., 2013)). However, infants may still be able to recognize these lexical items as functional based on their acoustic and phonological properties, as infants even within the first months of life are able to differentiate function and content words (Monaghan, Chater, & Christiansen, 2005; Shi, Morgan, & Allopenna, 1998; Shi, Werker, & Morgan, 1999).

In summary, the current empirical evidence indicates that infants prior to 18 months, acquiring the languages that have been studied to date, can represent many basic morphosyntactic features of sentences they hear. These include local subjects and objects of verbs, auxiliaries in canonical and moved positions, verbal morphology, and a few other functional categories, such as determiners, pronouns, prepositions, and negation. There are many open empirical questions regarding the richness of these representations. For example, it is not known whether infants at this age are aware that determiners not only form an equivalence class of lexical items before nouns, but also have particular syntactic and semantic properties. And it may be that infants do not yet recognize all forms of an irregular verb or auxiliary (e.g. *be*, *is*, *was*, *were*, *am*, *are*) as the same verb or auxiliary (see Tincoff, Santelmann, & Jusczyk, 2000). But even if these representations are not as rich as those of adults, these findings show that infants at this age have important scaf-

folding in place for further syntactic learning. The proposed distributional learning mechanism need not operate over strings; instead, it can can operate over sentence representations that encode syntactic properties, albeit partial and rudimentary. In other words, this account proposes that learners leverage their syntactic knowledge in order to constrain the distributional learning they perform for drawing inferences about syntax.

In focusing on the morphosyntactic properties of sentences that learners may use for syntactic learning, I do not make the claim that this is the sole information at a learner's disposal. Indeed, it is quite plausible that infants may represent and use information from other non-syntactic domains, such as the prosodic contour of an utterance (Christophe, Millotte, Bernal, & Lidz, 2008) and the communicative intent of the speaker (see e.g. Csibra, 2010; Meltzoff, 1995; Woodward, 2009 on infants' abilities to read and reason about the intentions of others). In particular, interrogative force is likely to be an important signal for a learner attempting to identify particular non-basic clause types, such as wh-questions. Although there is little existing empirical evidence for when infants can perceive that they are being asked a question, 15-month-olds' ability to identify the right answer to wh-questions in prior preferential looking tasks (Seidl et al., 2003; Gagliardi et al., 2016; Perkins & Lidz, under review) provides suggestive support that they may be aware that these sentences are being used to ask questions, even before they are aware that they contain wh-dependencies.

For the purposes of the experiments below, I make the assumption that 'question' is a pragmatic feature that infants at this age are able to identify, at least

for some of the sentences they hear (cf. Carruthers, 2018 on 'questioning attitudes' as a basic component of human minds). However, much more work is needed to determine how and when infants successfully identify interrogative force, and what other pragmatic and prosodic information they may be able to use to inform their syntactic inferences. I return to this point in the Discussion.

Finally, the current model departs from previous Bayesian models of clause structure acquisition (Maurits et al., 2009; Kwiatkowski et al., 2012; Abend et al., 2017) by focusing on learning primarily from formal distributional information, and not conceptual information. As discussed in Chapter 4, these prior models assume that learners bootstrap grammatical properties of their language using precise information about the meaning of utterances they hear. In taking a different approach, I am not making the claim that learners do not have access to conceptual information in their bootstrapping inferences; indeed, it is possible that a guess about the intended meaning of an utterance can work in concert with inferences on the basis of formal syntactic distributions. By focusing primarily on those distributional inferences, I merely ask how far that information could take a learner. That is, what distributional information is present in the learner's input, that could allow the learner to identify the syntactic properties responsible for those distributions? This will help understand how learning may be possible even in cases when the precise meanings of utterances are difficult to identify (Gleitman, 1990).

In the experiments below, I test the computational feasibility of the proposed distributional learning mechanism. I demonstrate that a learner can jointly categorize sentences according to their surface feature similarities, and infer which sentence

‘categories’ contain argument gaps. This allows the learner to identify forms that characterize movement in English. The learner does this only on the basis of the relevant syntactic information available to an infant at the developmental stage under investigation: morphosyntactic features pertaining to subjects, verbs, and objects in their canonical positions, together with known verb transitivity properties. Moreover, I show that the learner does better if it uses its verb transitivity knowledge to guide distributional learning, rather than categorizing sentences only using their surface features, without knowing which verbs require objects. These simulations demonstrate that a learner can use distributional analysis to identify forms that are characteristic of movement in English, and that doing so incrementally requires building on prior verb transitivity knowledge, consistent with the empirically attested developmental trajectory. This provides a proof of concept that Gap-Driven Learning may be a feasible mechanism for the incremental learning of syntactic dependencies.

5.2 Model

I present a Bayesian model that performs joint inference, categorizing sentences based on their surface features and inferring which sentence ‘categories’ contain locally missing arguments of verbs (gaps). To do this, the learner uses known verb transitivity properties, as learned by the model in Chapter 4. When sentences in a category violate the learner’s expectation that a verb has an object, it posits that those sentences, and therefore the category, contain argument gaps. Learning such

categories allows for further inference about which surface features signal different types of movement dependencies.

In this section, I first specify the generative model, encoding the learner's assumptions about how its observations of sentence features are generated. Then, I specify how the learner jointly infers sentence categories and argument gaps, given its data and its knowledge of verb transitivity. In the following sections, I present simulations demonstrating that this joint inference allows the learner to successfully identify features that characterize movement dependencies in English, when tested on child-directed speech.

5.2.1 Generative Model

The data that our learner observes consists of overt post-verbal direct objects and other morphosyntactic features of sentences containing known verbs. These are the 50 most frequent transitive, intransitive, and alternating action verbs whose transitivity properties were learned by the model in Chapter 4. Similar to the previous model, the current learner assumes that there are two reasons why it might observe direct objects or no direct objects following a verb. On the one hand, the known transitivity of that verb determines whether it should always, never, or sometimes occur with a direct object. On the other hand, the sentence might belong to a category of sentences with shared grammatical properties that produce argument gaps.

The learner assumes that there is a set of underlying sentence categories that

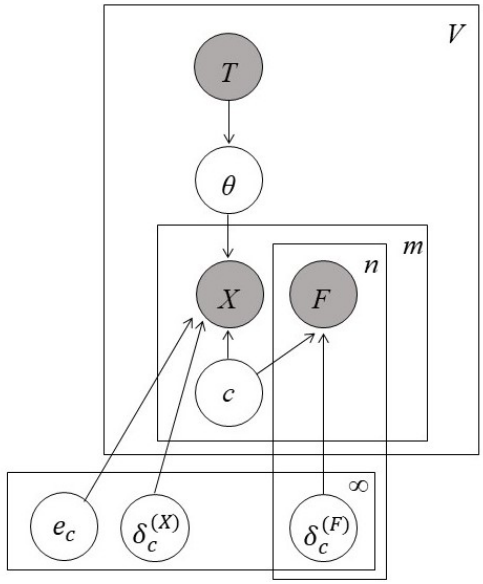

Figure 5.1: Graphical Model

give rise to the distributions of morphosyntactic features that it observes. Moreover, there is a parameter for each category that governs whether it produces object gaps: if it does, then observations of direct objects in that category may no longer reflect the known transitivity properties of these verbs, but may instead be due to other grammatical properties of that sentence category. Using the distributions of direct objects and sentence features in its observed data, the learner infers what categories of sentences are present, and which of those categories produce object gaps. This allows the learner to identify specific morphosyntactic features that are correlated with argument gaps for different types of sentences in its data.

Figure 5.1 provides the graphical model for the learner. Just as in the previous model, observations of direct objects or no direct objects are formalized as the Bernoulli random variable X, which takes a value of 1 if a sentence contains a direct object, and 0 if it does not. The model's observations of other relevant

Type	**Features**
Subject	Subject of known verb is overt in canonical subject position; sentence-initial; preceded by an auxiliary; preceded by another noun
Verb	Known verb is first verb in sentence; followed by a preposition or particle; has -*ed*, -*en*, -*ing*, -*s*, or irregular morphology
Tense & Auxiliaries	Verb is preceded by *to*, *be*, *have*, *get*, or *do*
Other	Question; unknown function word sentence-initially, sentence-medially before verb, sentence-medially after verb, or sentence-finally

Table 5.1: Morphosyntactic Features (F)

morphosyntactic features of a sentence are represented by the vector of Bernoulli random variables $\vec{F}$, which take a value of 1 if a sentence contains that feature, and 0 if not. These features encode the relevant properties of subjects, verbs, and auxiliaries that are observable to an infant prior to 18 months, as supported by the prior empirical findings discussed above. Table 5.1 lists the features that are input to our current learner. Wh-words are coded as "unknown function words," a hyper-category that includes all functional elements assumed to be unknown at this age: wh-words, complementizers, quantifiers, focus particles, and conjunctions other than *and*. This assumes a learner that can identify these lexical items as functional due to their phonological properties, but has not yet identified their grammatical categories. The observed features in $\vec{F}$ also include the pragmatic feature 'question,' which encodes whether an utterance has interrogative force. (In the model's dataset, interrogative force was identified by the presence of a question mark in the transcription; this does not distinguish constituent questions from polar questions.)

The upper part of the model has the same structure as for the previous model

in Chapter 4. Each observation $X^{(v)}$ of a direct object for a particular verb is conditioned on the parameter $\theta^{(v)}$, a continuous random variable that controls the probability that verb v will be used with a direct object. $\theta^{(v)}$ is conditioned on the variable $T^{(v)}$, a discrete random variable that can take on three values corresponding to transitive, intransitive, or alternating verbs. Unlike the previous learner, the current learner knows values of T for each verb: it uses the transitivity categories that were inferred by the previous learner in Chapter 4. This means that the learner knows some of the values of θ as well. For a verb in the transitive category of T, θ equals 1, because the verb should always take a direct object. For an intransitive verb, θ equals 0, because the verb should never occur with a direct object. For the alternating category, θ takes a value between 0 and 1 inclusive. The prior probability over θ in this case is a $Beta(\alpha, \beta)$ distribution, where the parameters α and β are counts of direct objects and no direct objects for verb v in sentence categories without argument gaps, excluding the current category.

In the lower part of the model, each $X^{(v)}$ and $F^{(v)}$ is conditioned on the discrete random variable c, defined for all positive integers, which represents the category that the sentence belongs to. Each value of c represents a different sentence category, assumed to reflect a particular set of underlying grammatical properties that give rise to the distributions of direct objects and other features in a sentence. For example, one value of c might represent English wh-object questions, which gives high probability to features like sentence-initial function words, subject-auxiliary inversion, and the auxiliary *do*, as well as object gaps. Another value of c might represent English polar questions, which also gives high probability to features like

subject-auxiliary inversion and the auxiliary *do*, but does not produce sentence-initial function words or object gaps.

For each sentence, a coin is flipped to determine which category of c that sentence belongs to. The prior probability over sentence categories is a Dirichlet process (Ferguson, 1973), with parameter α. In this process, a particular category c has prior probability proportional to the number of sentence observations already assigned to that category. This process also reserves a small non-zero probability for new categories, allowing the model to flexibly converge on the number of sentence categories that best explains the distribution of features and direct objects in its data. See Appendix B for details.

Each direct object observation $X^{(v)}$ is also conditioned on two random variables, e_c and $\delta_c^{(X)}$, representing parameters of the sentence category that the observation belongs to. The Bernoulli random variable e_c represents whether a given category c produces argument gaps. If $e_c = 0$, then the category does not produce argument gaps, and all observations of a direct object in $X^{(v)}$ were generated by the transitivity properties of verb v: $T^{(v)}$ and $\theta^{(v)}$. But if $e_c = 1$, then the category does produce argument gaps, and observations of direct objects $X^{(v)}$ were generated by a particular grammatical property of category c. In this case, these observations may not conform to the transitivity properties of verb v. The prior probability that $e_c = 1$ is assumed to be 0.19, which is the mean value of ϵ learned by the model in Chapter 4. This represents the probability of a transitivity violation inferred by that model for sentences containing these same verbs.[2]

[2]The probability of observing a transitivity violation for a particular sentence, ϵ in the previous

The random variable $\delta_c^{(X)}$ represents the probability of observing a direct object in a category with argument gaps. Thus, whether a sentence contains a direct object or no direct object depends on one of two biased coins. If $e_c = 0$ and the observation was generated by the verb's transitivity properties, then one biased coin is flipped and the sentence contains a direct object with probability $\theta^{(v)}$. But if $e_c = 1$ and the observation was generated by grammatical properties of category c, then a different biased coin is flipped and the sentence contains a direct object with probability $\delta_c^{(X)}$. The parameter $\delta_c^{(X)}$ is assumed to have a uniform $Beta(1, 1)$ prior distribution. This means that it is equally likely *a priori* for a sentence category to create argument gaps as it is for a sentence category to "add" an extra argument that isn't licensed by the verb. This represents a simplifying assumption for the current learner, but one that may complicate the learner's goal of identifying argument gaps rather than transitivity violations more broadly. We will return to this point when discussing the simulations below.

Each feature observation $F^{(v)}$ is conditioned on the random variables in $\vec{\delta}_c^{(F)}$, which represent the probabilities of observing those morphosyntactic features in a particular sentence category. Each sentence in category c contains feature F_1 with probability $\delta_c^{(F_1)}$, feature F_2 with probability $\delta_c^{(F_2)}$, and so on. Each $\delta_c^{(F)}$ is also assumed to have a uniform $Beta(1, 1)$ prior distribution.

model, is not necessarily equivalent to the probability of a transitivity violation in a sentence category, e_c in the current model. These two parameters will only be equivalent if sentences are equally distributed among sentence categories. Although this assumption may not be borne out, it is adopted here as a simplifying assumption of the learner's prior, which can be overridden as the learner updates its hypotheses upon seeing data.

5.2.2 Joint Inference

The learner uses Gibbs sampling (Geman & Geman, 1984) to jointly infer the category of each observed sentence (c) and whether or not each category contains argument gaps (e). We first initialize values of c for each sentence, starting with three initial categories: one with argument gaps and two without. These values are sampled for each sentence from the posterior probability distribution over an error in that sentence observation, as inferred by the learner in Chapter 4. This posterior probability is calculated given the transitivity of the verb ($T^{(v)}$), whether the observation contains a direct object ($X^{(v)}$), and the other parsing error parameters inferred by the previous learner.

After initialization, we use the observed data in X and F, the known verb transitivity properties T, the argument-gap property for each category e, and the other sentence category assignments c to calculate a posterior probability distribution over new category assignments for a given sentence. We re-sample new values of c for each sentence sequentially from this posterior probability distribution. Then, we use the re-sampled category values, along with X, F, and T, to sample new values for the argument-gap property of each category e_c from its posterior probability distribution. This cycle is repeated over many iterations until the model converges to a stable distribution over c and e. See Appendix B for details of the sampling procedure.

5.3 Simulations

In the simulations below, I ask whether a learner can identify the surface signals of movement in English by jointly categorizing sentences according to their surface feature distributions, and inferring which sentence categories have argument gaps. The learner was tested on a dataset of child-directed speech containing the same transitive, intransitive, and alternating verbs learned by the model in Chapter 4. In order to evaluate the model's performance, I compare it to two baseline models: one that infers argument gaps but does not categorize sentences based on their feature distributions, and one that categorizes sentences without jointly inferring argument gaps. I then analyze the specific features of the model's categories in order determine whether these categories are informative for the next step of learning: identifying the underlying syntactic dependencies that are responsible for these feature distributions. Results of this analysis find that the model's categories contain both signal and noise for this step of learning. The learner successfully identifies the forms that characterize movement dependencies in English, but also identifies many forms that are accidentally correlated with those dependencies.

5.3.1 Data

The dataset for this learner was prepared from the 18,503 CHILDES Treebank sentences (Pearl & Sprouse, 2013) containing the same 50 transitive, intransitive, and alternative verbs as in Chapter 4. Each sentence was coded for whether an overt direct object or no direct object followed the verb of interest, using the same

procedure as in Chapter 4. In addition, each sentence was coded for the presence of the morphosyntactic features in Table 5.1. Coding was conducted using automated scripts to search over the Treebank trees. To verify the accuracy of these scripts, a random sample of 500 sentences from the dataset was also coded by hand. Percentage agreement between the hand-coding and automated coding was well above 90% for 20 out of 21 sentence features. The exception was the feature 'verb followed by a preposition or particle,' which had slightly lower percentage agreement (89%) due to inconsistent part-of-speech tagging of prepositions and particles in the Treebank.

In order to evaluate our model's performance, the sentences in the dataset were also coded for their underlying clause types, listed in Table 5.2. These included three clause types with movement: wh-questions, passives, and relative clauses. A given clause might be coded as multiple types, e.g. as both a question and a passive. For sentences with multiple clauses, coding was conducted for the clause containing the verb of interest. For example, if the verb of interest was contained in an embedded clause within a matrix wh-question (e.g. *What do you want to* ***eat****?*), the clause was coded as an embedded clause and also a wh-question if it contained the base position of movement.

Clause type coding was conducted by another set of automated scripts. Accuracy was again evaluated by comparing against hand-coding for a random sample of 500 sentences. Percentage agreement between the hand-coding and automated coding was above 90% for 7 of 9 clause types. The two exceptions were embedded clauses (84% agreement) and basic intransitive clauses (89% agreement), due to parsing inconsistencies in the Treebank parse trees for some embedded clauses. Ad-

Clause Type	# Clauses	Description
Basic transitive	2855 (15%)	Matrix, finite, declarative clause with overt direct object following known verb
Basic intransitive	2704 (15%)	Matrix, finite, declarative clause without overt direct object following known verb
Wh-question	2336 (13%)	Clause has canonical syntactic form of a wh-question, with wh-element in a dependency with the known verb
Polar question	3641 (20%)	Clause has canonical syntactic form of a polar question
Other question	1922 (10%)	Clause was transcribed with a question mark, but does not have canonical syntactic form of a wh-question or polar question: includes tag, fragment, and echo questions, and rising intonation declaratives
Passive	268 (1%)	Known verb has been passivized, excluding forms that are clearly adjectival
Relative clause	298 (2%)	Known verb is in a full or reduced relative clause
Other embedded clause	4905 (27%)	Known verb is in a finite or non-finite embedded, non-relative clause
Imperative	2176 (12%)	Clause has canonical syntactic form of an imperative

Table 5.2: Distribution of Clause Types in Dataset

ditional hand-coding was conducted for wh-questions and relative clauses in order to annotate the gap site in these sentences, which could not be reliably identified automatically for the entire dataset.

5.3.2 Results

Sentence Category Inference. The joint inference model inferred 35 total sentence categories, 15 containing argument gaps and 20 without argument gaps. For each of these categories, Figure 5.2 reports the total sentences in the category and the proportion of the category made up of each underlying clause type, collapsing across both types of basic clauses. For example, 0.83 of the 560 sentences in the model's Category 1 are basic transitives and intransitives, 0.08 are polar questions, and so on. Note that these proportions do not necessarily sum to 1 because a single

	Model's Categories																
Clause Type	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
Basic transitive/intransitive	0.83	0.82	0.76	0.72	0.68	0.61	0.61	0.57	0.48	0.47	0.00	0.00	0.01	0.01	0.04	0.05	0.00
Wh-question	0.00	0.00	0.01	0.01	0.00	0.03	0.10	0.00	0.15	0.01	0.96	0.93	0.92	0.87	0.85	0.68	0.67
Polar question	0.08	0.07	0.03	0.14	0.16	0.11	0.06	0.04	0.16	0.12	0.04	0.04	0.02	0.11	0.04	0.25	0.30
Other question	0.08	0.10	0.10	0.12	0.16	0.08	0.11	0.07	0.14	0.12	0.00	0.03	0.05	0.00	0.06	0.03	0.03
Passive	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Relative clause	0.00	0.00	0.01	0.00	0.00	0.03	0.16	0.01	0.13	0.01	0.00	0.01	0.01	0.00	0.06	0.00	0.01
Other embedded clause	0.00	0.01	0.09	0.02	0.00	0.19	0.03	0.29	0.06	0.34	0.03	0.97	0.01	0.03	0.03	0.08	0.03
Imperative	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.05	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00
Total Sentences	560	1821	285	810	1168	154	148	652	212	418	448	307	171	517	106	110	91

	Model's Categories, Continued																	
Clause Type	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**
Basic transitive/intransitive	0.01	0.00	0.09	0.03	0.20	0.04	0.08	0.21	0.26	0.23	0.01	0.03	0.02	0.16	0.36	0.23	0.11	0.29
Wh-question	0.21	0.25	0.07	0.02	0.03	0.10	0.05	0.00	0.03	0.21	0.08	0.02	0.06	0.02	0.03	0.03	0.00	0.04
Polar question	0.77	0.75	0.69	0.14	0.13	0.21	0.23	0.15	0.15	0.14	0.45	0.09	0.16	0.17	0.14	0.15	0.01	0.04
Other question	0.01	0.00	0.11	0.64	0.42	0.14	0.14	0.03	0.12	0.18	0.27	0.06	0.11	0.15	0.10	0.09	0.03	0.12
Passive	0.00	0.00	0.00	0.00	0.00	0.90	0.82	0.65	0.59	0.03	0.00	0.00	0.07	0.01	0.00	0.02	0.00	0.01
Relative clause	0.00	0.00	0.02	0.00	0.02	0.03	0.04	0.03	0.02	0.33	0.00	0.01	0.05	0.12	0.02	0.25	0.00	0.01
Other embedded clause	0.01	0.02	0.07	0.00	0.01	0.00	0.00	0.06	0.03	0.05	0.98	0.93	0.87	0.67	0.54	0.45	0.00	0.03
Imperative	0.00	0.00	0.00	0.22	0.23	0.00	0.00	0.00	0.00	0.04	0.00	0.02	0.00	0.00	0.00	0.00	0.85	0.49
Total Sentences	1616	352	54	140	509	72	73	34	116	78	1520	1311	471	206	1272	265	2110	326

■ Argument-Gap Categories
■ Non-Argument Gap Categories

Figure 5.2: Proportions of Clause Types in Sentence Categories (c) Inferred by Joint Inference Model

clause might be of multiple types. Categories that the model inferred to have argument gaps are represented with gray numbers, and categories without argument gaps are in black.

Overall, the model's categories have high purity when compared to the underlying clause types that are being used for evaluation. For example, we find categories of 96% wh-questions (11), 90% passives (23), and 98% embedded clauses (28). We can calculate the overall purity of the model's categories by adding together the total number of sentences that belong to the predominant clause type in each category, and dividing by the total number of sentences in the dataset (Manning, Raghavan,

& Schütze, 2008). When compared against a gold standard, this measure has a minimum value of 0 for poor clustering and a maximum value of 1 for perfect clustering. Our model's overall cluster purity is 0.76. This tells us that the features used by the model were informative about the underlying clause types in the corpus, and the model was able to use these features to track the target syntactic distinctions in these sentences.

Of the model's 15 argument-gap categories, 9 were primarily comprised of clause types that were coded as having movement. Categories 11, 12, 14, 16, and 17 were predominantly wh-questions, and Categories 23-26 were predominantly passives. Wh-questions were also the primary clause type in two categories (13 and 15) that were not identified as having argument gaps. Relative clauses made up a small plurality of Category 27, which was also not inferred to have argument gaps.

For purposes of illustration, Table 5.3 provides a sample sentence for each of the model's argument-gap categories, and the three other categories whose main clause type had movement. Appendix C provides a full description of these categories. Focusing here on the categories whose predominant main clause type contained movement, there are several distinctions the model appears to be making. The model's categories of passives (Categories 23-26) are differentiated between *get*-passives (Category 23) and *be*-passives (Categories 24-26). These categories of *be*-passives are further split by whether the verb is embedded and has overt -*en* vs. -*ed* or irregular morphology.

The model's wh-question categories (11-17) are differentiated along several other dimensions. Long wh-questions (Category 12) are differentiated from short

Category	Main Clause Type	Sample Sentence
3	Basic	*I'll close.*
9	Basic	*What's that you're writing?*
11	Wh-question	*What are you bringing?*
12	Wh-question	*What would you like to read?*
13	Wh-question	*What fell down?*
14	Wh-question	*What does he eat?*
15	Wh-question	*Who's playing with the balloons?*
16	Wh-question	*Very good and what did you build?*
17	Wh-question	*And what is he wearing?*
21	Other question	*Open what?*
23	Passive	*It got lost.*
24	Passive	*That can't be eaten.*
25	Passive	*I don't think that car is broken.*
26	Passive	*It can't be fixed.*
27	Relative clause	*You're the doggie who ran away?*
30	Embedded	*They're very easy to lose.*
33	Embedded	*Let's see if we have any clean pants that Thomas can wear.*
35	Imperative	*Throw away.*

Table 5.3: Sample Sentences from Model's Argument-Gap and Other Movement Categories

wh-questions. Progressive wh-questions (Categories 11, 15, and 17) are differentiated from non-progressive wh-questions. Wh-questions with interjections or other expressions before the wh-word (Categories 16-17) are differentiated from those in which the wh-word is sentence-initial. Finally, subject wh-questions (Categories 13 and 15) are differentiated from object questions. The progressive/non-progressive distinction, driven by the robust correlation between *be* and *-ing*, is one that the model attended to in many of its other categories as well. Furthermore, because the model only infers argument gaps on the basis of missing direct objects and not subjects, the subject question categories were not inferred to have argument gaps. This was also the case for the model's category that had a small plurality of relative clauses (Category 27), of which most were subject relatives.

Argument Gap Accuracy. The model's accuracy was evaluated by compar-

ing the sentences that it assigned to argument-gap categories against the sentences that were coded as belonging to an underlying clause type with movement. These clause types included wh-questions, relative clauses, and passives. They did not include other rarer cases of movement such as *tough*-movement and movement out of purposive clauses. Other cases of movement that were not coded separately in the gold standard included clefting, pseudo-clefting, topicalization, comparative movement, and raising. Thus, the model's accuracy score reflects when it identified argument gaps for only the most frequent types of movement in its data, but not cases when it identified argument gaps for less frequent movement types. This means that the model is being evaluated against an incomplete gold standard, but one that provides an estimate for the most frequent movement types in the corpus.

The accuracy of the joint inference model is summarized in Table 5.4 using three metrics (Perry, Allen, & Berry, 1955). Precision measures the proportion of sentences in the model's argument-gap categories that contained movement according to our gold standard— that is, the proportion of these categories made up of wh-questions, relative clauses, or passives. Recall measures the proportion of wh-questions, relative clauses, and passives in the corpus overall that were identified as belonging to one of the model's argument-gap categories. These metrics are not always aligned: it would be possible to achieve perfect recall by identifying all sentences as having movement, but this would result in very poor precision. The F1 score, the harmonic mean of precision and recall, reflects the model's overall accuracy by taking into account both of these metrics. The model achieved an F1 score of 0.56, with 0.51 precision and 0.62 recall. This recall measure indicates that it

Model	Precision	Recall	F1 Score
Joint Inference	0.51	0.62	0.56
No-Category Baseline	0.25	0.37	0.30
Distributional Baseline	0.15	1.00	0.27
Chance	0.15	0.50	0.23

Table 5.4: Accuracy on Identifying Sentences with Movement

identified 62% of sentences with movement in its data. While not perfect, this is substantially above chance performance. A learner that randomly identified sentences as having movement by flipping a fair coin would achieve an F1 score of only 0.23.

The model achieves this performance despite several factors that limit its accuracy. First, as previously discussed, the model does not receive credit for identifying cases of movement other than wh-questions, passives, and relative clauses. Second, the model only infers movement from sentences with object gaps, as its argument-gap inference is driven by missing direct objects for verbs that it knows require objects. This means that the current evaluation is partially measuring how well the model was able to generalize in two ways: from object gaps with transitive verbs to object gaps with freely alternating verbs; and from object movement to other types of movement in sentences that look formally similar.

To evaluate these two types of generalization, Table 5.5 reports the percentages of sentences with subject, object, prepositional object, or adjunct movement that were identified as having gaps by the model, broken down by the known transitivity classes of verbs in these sentences. The model achieves high accuracy on identifying sentences with object movement: it correctly identifies that 85% of these

Movement Type	# Total Sentences	% Identified
Object	**1391**	**0.85**
Transitives	296	0.81
Intransitives	14	0.93
Alternators	1081	0.86
Prepositional Object	**246**	**0.60**
Transitives	19	0.37
Intransitives	53	0.72
Alternators	174	0.60
Adjunct	**774**	**0.51**
Transitives	125	0.28
Intransitives	258	0.66
Alternators	391	0.48
Subject	**419**	**0.14**
Transitives	64	0.14
Intransitives	120	0.11
Alternators	235	0.16

Table 5.5: % Sentences Identified as Having Gaps by Joint Inference Model, by Movement and Verb Type

sentences have gaps, where chance performance would be only 50%. Moreover, the model achieved this degree of accuracy even though the majority of sentences with object movement occurred with verbs that it believed to be alternating, rather than obligatorily transitive. Of the 1391 sentences coded as having object movement in the corpus, only 296 contained known transitive verbs, compared to 1081 containing known alternating verbs.[3] The model's overall high accuracy across verb classes tells us that it was able to generalize in one way: it used the presence of object gaps with known transitive verbs to identify the forms that object movement takes in its data, even with verbs that do not obligatorily require objects.

On the other hand, the model showed weaker performance on identifying other types of movement. It achieved slightly above-chance accuracy on identifying prepositional object movement; in this case, its accuracy for intransitive and alternating

[3]The few cases of object movement with intransitive verbs were uses of the verb in a rare or ungrammatical transitive frame (e.g. *What did you run?*).

verbs was higher than for transitive verbs. This is likely because prepositional object wh-questions and relative clauses are less likely to contain a direct object after intransitive and alternating verbs. The model appeared to recognize some of the surface distributional similarities between direct and prepositional object movement, but was still looking for gaps in direct object position because it did not encode the transitivity of specific prepositions. For adjunct movement, the model's overall performance was no better than chance. Its performance on subject movement was worse than chance: it inferred that most subject wh-questions and relative clauses did not contain gaps.

This tells us that the model did not tend to generalize across different types of movement. It did not use the presence of object gaps to infer that there were also gaps in subject or adjunct wh-questions and relative clauses, on the basis of the formal similarities of these movement types, and it only drew this inference for some cases of prepositional object movement. To some extent this lack of generalization is desirable, because a learner would be incorrect to infer that direct object gaps are present in these sentences. Indeed, a learner should not posit argument gaps at all when there is adjunct movement. However, this learner misses the opportunity to learn about the shared formal features that mark all types of wh-questions and relative clauses in the language. A more sophisticated learner might use the absence of other required arguments, such as subjects and prepositional objects, to condition its inference about argument gaps and identify movement of different types.

5.3.3 Model Comparisons

Our model achieves above-chance performance on identifying sentences with movement, specifically those with object movement, by jointly inferring two properties: how sentences should be categorized together according to their surface feature distributions, and which sentence categories violate the model's expectations about verb transitivity. To evaluate how important this joint inference is, we can compare our model to baseline learners that only perform one step of inference at a time.

No-Category Baseline. If it didn't matter that our learner categorized sentences according to their surface features, then a learner should do just as well at identifying movement on a sentence-by-sentence basis, by noting when objects are unexpectedly missing for known transitive verbs. To test whether the model's categorization process matters, I compare our model against a baseline learner that uses only the presence or absence of direct objects in individual sentences, together with the known transitivity properties of verbs in these sentences, to infer which sentences likely contain argument gaps. This baseline learner has the architecture of the model in Chapter 4, but we now fix the transitivity properties of each verb (T) and the parameters of the input filter (ϵ and δ) to the values inferred by that previous model. We can sample individual error values (e) for each sentence in the corpus from the posterior probability distribution over errors, given the observed direct objects and known model parameters. This uses the same sampling equations that were already derived to initialize the current learner; see Section 1 of Appendix B for details of this sampling process. Sentences that are sampled as containing errors

	# Total Sentences	% Identified
Overall	**1391**	**0.55**
Transitives	296	0.76
Intransitives	14	0.36
Alternators	1081	0.50

Table 5.6: % Object Movement Sentences Identified by No-Category Baseline

are those that violate the learner's expectations of verb transitivity.

To determine how well this 'No-Category' baseline model identified movement, I compared the sentences that it inferred to be errorful against the actual cases of movement that were coded in the gold standard. The precision, recall, and F1 score are reported in Table 5.4. The model achieved slightly better than chance accuracy overall, but scored much lower than the joint inference model on all three metrics. If we examine the baseline model's identification of object movement only (Table 5.6), we find that it only identified 55% of the sentences with object movement in the corpus, barely above chance performance. This is not entirely surprising, as the baseline model's only source of reliable information for object gaps comes from the small percentage of verbs that it knows to be obligatorily transitive. Indeed, it achieved high accuracy (76%) on identifying object movement with these verbs. But for the much larger percentage of verbs that are alternating, it must guess at random which sentences contain gaps, because it uses no other features in the sentences to inform this inference. This results in overall worse performance than the joint inference model.

These results tell us that the categories inferred by the joint inference model were useful. The ability to categorize sentences using a wide range of surface mor-

phosyntactic features, and to generalize across sentences in a category, results in substantially better performance than inferring movement on a sentence-by-sentence basis from direct object observations alone.

Distributional Baseline. Our model tests the hypothesis that it is useful for prior verb transitivity knowledge to guide the identification of movement: specifically, that it is helpful for a learner to use verb transitivity knowledge in the process of identifying the surface forms that signal movement dependencies. It therefore makes sense to compare our model against a learner that performs distributional analysis over the surface features of sentences without using verb knowledge. This distributional baseline learner categorizes sentences only using their surface features, without knowing which verbs require direct objects. It treats direct object observations identically to other surface features: for this learner, all direct objects are generated by the grammatical properties of a sentence category, not by the properties of verbs in the sentences. Once these categories are formed, the model then applies knowledge of verb transitivity in order to infer which sentence categories contain argument gaps. If this baseline distributional learner can perform just as well as our joint inference model in identifying sentences with movement, this will tell us that it is not necessary for verb knowledge to guide distributional learning—that verb knowledge can instead be acquired and applied after distributional analysis over sentence forms has already taken place.

The first stage of this baseline learner has part of the architecture of the joint inference model in Fig 5.1, but omits the variables T, θ, and e. When the variables T and θ are omitted, the learner now assumes that all direct object observations X

are generated by $\delta^{(X)}$, grammatical properties of the sentence category, rather than by any properties of the verbs in these sentences. When the variable e is omitted, the learner is no longer inferring which sentence categories contain argument gaps at this stage of learning; it is only inferring which sentence categories are present and which sentences belong to which categories. The learner samples category values for each sentence in the corpus from the posterior probability distribution over c given X and F, integrating over $\delta^{(X)}$ and $\delta^{(F)}$. This inference uses the subset of the sampling equations in Appendix B that condition only on $\delta^{(X)}$ and $\delta^{(F)}$.

The second stage of this learner then applies verb knowledge to identify whether its already-formed sentence categories do or do not have argument gaps. The second stage learner now has the architecture of the full model in Fig. 5.1, but sentence categories are fixed to the values of c inferred by the first stage learner. The learner samples argument-gap properties for each known sentence category from the posterior probability distribution over e given X, F, c, and T, again using the sampling equations in Appendix B.

The distributional baseline learner inferred 36 sentence categories. The proportions of underlying clause types in these categories are reported in Fig. 5.3. On the surface, these categories appear similar to those inferred by the joint inference model. Their purity is similarly high when compared against the clause types in the gold standard: the baseline model's overall cluster purity is 0.77, compared to 0.76 for the joint inference model. This shows again that the morphosyntactic features being tracked by both learners are informative for differentiating the different underlying clause types in the corpus.

	Baseline Model's Categories																	
Clause Type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Basic transitive/intransitive	0.82	0.82	0.77	0.73	0.70	0.68	0.66	0.64	0.55	0.48	0.29	0.00	0.00	0.03	0.02	0.03	0.00	0.18
Wh-question	0.00	0.00	0.00	0.00	0.01	0.04	0.00	0.00	0.10	0.00	0.10	0.98	0.97	0.85	0.84	0.84	0.81	0.30
Polar question	0.07	0.07	0.10	0.14	0.14	0.06	0.06	0.18	0.13	0.05	0.08	0.01	0.03	0.04	0.02	0.04	0.10	0.19
Other question	0.10	0.10	0.12	0.10	0.12	0.14	0.12	0.18	0.09	0.04	0.14	0.00	0.00	0.08	0.13	0.08	0.07	0.14
Passive	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Relative clause	0.00	0.00	0.00	0.00	0.00	0.09	0.01	0.00	0.08	0.02	0.24	0.00	0.00	0.03	0.01	0.11	0.01	0.07
Other embedded clause	0.01	0.02	0.01	0.03	0.02	0.02	0.15	0.00	0.14	0.45	0.15	0.03	0.03	0.08	0.04	0.01	0.98	0.30
Imperative	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Total Sentences	1673	917	271	226	725	314	268	893	379	781	143	660	516	80	120	118	371	149

	Baseline Model's Categories, Continued																	
Clause Type	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Basic transitive/intransitive	0.02	0.03	0.01	0.01	0.00	0.01	0.04	0.10	0.29	0.23	0.01	0.02	0.13	0.27	0.35	0.04	0.09	0.19
Wh-question	0.03	0.11	0.19	0.24	0.21	0.03	0.11	0.05	0.00	0.00	0.05	0.06	0.02	0.04	0.06	0.00	0.00	0.00
Polar question	0.94	0.79	0.77	0.74	0.71	0.20	0.19	0.25	0.01	0.15	0.34	0.23	0.17	0.17	0.17	0.02	0.01	0.01
Other question	0.01	0.08	0.02	0.01	0.00	0.74	0.15	0.11	0.03	0.08	0.19	0.18	0.13	0.14	0.11	0.01	0.01	0.01
Passive	0.00	0.00	0.00	0.00	0.64	0.00	0.90	0.84	0.66	0.59	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00
Relative clause	0.00	0.03	0.00	0.00	0.00	0.01	0.03	0.05	0.01	0.00	0.01	0.03	0.17	0.06	0.00	0.00	0.00	0.00
Other embedded clause	0.01	0.03	0.00	0.01	0.07	0.01	0.03	0.00	0.01	0.05	0.98	0.93	0.66	0.58	0.42	0.00	0.01	0.14
Imperative	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.93	0.88	0.65
Total Sentences	916	38	385	609	14	447	73	73	77	39	1785	1108	257	925	215	393	274	2271

■ Argument-Gap Categories
■ Non-Argument Gap Categories

Figure 5.3: Proportions of Clause Types in Sentence Categories (c) Inferred by Baseline Distributional Learner

Similarly to the joint inference model, the baseline model identified 11 categories whose predominant clause type contained movement. These include 7 categories of wh-questions (Categories 12-18) and 4 categories of passives (Categories 25-28). The baseline model did not identify any category predominantly made up of relative clauses. The model's wh-question categories appear to mark the same dimensions as were marked by the joint inference model. Long wh-questions (Category 17) are differentiated from short wh-questions; progressive wh-questions (Categories 13 and 16) are differentiated from non-progressive questions; and subject questions (Categories 14-16) are differentiated from object questions. The small plurality of

wh-questions in Category 18 contain questions with an interjection or other expression before the wh-word. Similarly, the model differentiates between *get*-passives (Category 25), and *be*-passives (Categories 26-28). *Be*-passives are further distinguished by whether the verb is embedded or has overt -*en* vs. irregular morphology.

These similarities in distributions between the baseline and joint inference models show that a learner does not need verb knowledge to identify clusters of different wh-questions and passives according to their distinctive surface forms. Subject and object wh-questions differ enough in their surface feature distributions, as do questions with and without progressive morphology and passives with *get* vs. *be*, that a learner will differentiate between the surface forms of these sentences regardless of whether it knows which verbs require objects.

However, despite these surface similarities, the baseline model's categories are different from those of the joint inference model in an important way. When the baseline learner applied verb transitivity knowledge in its second stage of learning, it inferred that all of its already-formed categories contained argument gaps, as indicated by gray numbers in Fig. 5.3. That is, it was unable to differentiate sentences with argument gaps from sentences without; the categories the model identified via distributional analysis alone did not enable it to generalize effectively in its second stage of learning. By positing argument gaps for every sentence in the corpus, the model achieved perfect recall on identifying sentences with movement (Table 5.4). However, its precision was at chance, leading to an F1 score that was barely above chance performance. The joint inference learner achieved substantially higher overall accuracy. This shows that it is important for verb transitivity knowledge to

guide distributional learning when inferring categories of sentences with and without movement. While the distributions of these morphosyntactic surface features convey a great deal of information about the distinctions among different clause types, learning which of these distinctions signal movement and which do not requires the use of verb knowledge during distributional analysis.

5.3.4 Interim Summary

In summary, our model identifies 62% of sentences with movement in child-directed speech, and 85% of sentences with object movement, by tracking the surface morphosyntactic features of sentences that violate its expectations of verb transitivity. The model jointly infers how to categorize sentences according to their surface feature distributions, and which of these sentence categories contain argument gaps— missing objects of known verbs. This allows the learner to generalize across sentences that share the same form and posit object gaps even for verbs that it does not know to be transitive, resulting in overall accuracy twice as high as would be expected by chance. It performs substantially better than a learner that relies only on known verb transitivity knowledge and does not categorize sentences on the basis of their surface feature distribution. This shows that the model's categorization process is helpful. It also out-performs a baseline distributional learner that categorizes sentences using their surface features alone, without knowing which verbs require objects, and only afterwards attempts to use verb knowledge to infer which categories had argument gaps. The baseline learner inferred categories with

similar distributions as our model, but was unable to differentiate categories with and without argument gaps, showing that verb knowledge is an important guide for learning movement.

5.3.5 Features of Argument-Gap Categories

Under our hypothesis, the sentence categories inferred by the joint inference model are an intermediate step of learning. Jointly inferring how to categorize sentences according to their surface features, and which sentence categories contain argument gaps, helps a learner identify the particular forms that characterize different types of movement dependencies in the target language. In the evaluations above, I showed that our model's categories enable it to perform well above chance in identifying sentences with movement in its dataset, and substantially better than two baselines that do not perform this joint inference. Here, I ask whether the model's categories will be useful for the next step of learning: identifying which specific surface features are the footprints of movement.

Distinctive Features of Argument-Gap Categories. We can answer this question by assessing which surface features are most distinctive in the categories that the model inferred to have argument gaps. If these include the characteristic forms of English movement dependencies, then the model's sentence categories contain helpful information for identifying the ways that movement can be realized in English. There are many ways that feature distinctiveness may be measured. One way is by calculating the odds ratio of each surface feature in the model's

argument-gap categories. This measure divides the odds of observing a feature in a given category by the odds of observing that feature outside of that category; an odds ratio significantly greater than 1 indicates that a feature is more likely to be observed within than outside of the category. For the current purposes, significance was calculated using a Fisher's exact test with a Bonferroni correction for multiple comparisons.

Note that by choosing this particular analysis, I do not make the claim that learners perform a Fisher's exact test on their data. There are many possible ways to identify which features are suspiciously more likely in a given category, all of which would likely yield similar results. The statistical analysis here merely provides an estimate of the information that could, in principle, be available to a learner who is sensitive to it.

Table 5.7 reports the features with odds ratios significantly greater than 1 for each of the model's argument-gap categories. Appendix D reports the odds ratios, confidence intervals, and p-values for these features. The features reported in this table include the characteristic forms of the non-basic clause types present in each of the model's categories. Categories 9, 30 and 33, which together contain 65% of the relative clauses in the corpus, have greater odds of including sentence-medial function words and nouns before the subject of the known verbs. These are the forms that complementizers and heads of relative clauses take in English. Categories 11, 12, 14, 16, and 17 have greater odds of being questions and including subject-auxiliary inversion, *do*, and function words sentence-initially or medially before the verb, which are all distinctive forms of wh-questions in English. Category 21 has

Category	**Main Clause Type**	**Features**
3	Basic	Subject is overt, sentence-initial or preceded by a noun; verb is first in sentence, has -*ed* or -*s*; function word after verb sentence-medially or finally
9	Basic	Subject is overt, preceded by a noun; verb has -*ing*, preceded by *be*; sentence-medial function word before verb
11	Wh-question	Subject is overt, preceded by an aux; verb is first in sentence, has -*ing*, preceded by *be*; sentence-initial function word; question
12	Wh-question	Verb is preceded by *to*; sentence-initial function word; question
14	Wh-question	Subject is overt, preceded by an aux; verb is first in sentence, preceded by *do*; sentence-initial function word; question
16	Wh-question	Subject is overt, preceded by an aux or a noun; verb is first in sentence, preceded by *do*; function word sentence-medially before or after verb; question
17	Wh-question	Subject is overt, preceded by an aux or a noun; verb is first in sentence, has -*ing*, preceded by *be*; function word sentence-medially before verb; question
21	Other question	Verb is first in sentence; sentence-final function word, question
23	Passive	Verb has -*en* or irregular form, preceded by *get*
24	Passive	Subject is overt, sentence-initial; verb is first in sentence, has -*en*, preceded by *be* or *have*
25	Passive	Subject is overt, preceded by a noun; verb has -*en* or irregular form, preceded by *be* or *have*
26	Passive	Subject is sentence-initial; verb is first in sentence, has irregular form, preceded by *be* or *have*
30	Embedded	Verb preceded by *to* or *get*; sentence-medial function word before verb
33	Embedded	Subject is overt, preceded by a noun; verb has -*s* or irregular form; sentence-medial function word before verb
35	Imperative	Verb is first in sentence

Table 5.7: Features with Significantly Higher Odds in Argument-Gap Categories

greater odds of including a sentence-final function word, as occurs in English echo-questions. Categories 23-26 have greater odds of including *get*, *be*, and -*en* verbal morphology, which are the distinctive forms of English passives. Several of these passive categories also have greater odds of including irregular verbal morphology, as many verbs lack an overt -*en* morpheme when passivized (e.g. *get caught*).

On the other hand, the features in this table also include forms that are irrelevant to the movement dependencies in these categories. These include many positional characteristics of subjects and verbs, but also some specific morphemes: *be* and -*ing* in Categories 9, 11, and 17; *to* in Categories 12 and 30; -*ed*, -*s*, and irregular verbal morphology in Categories 3 and 33; and *have* in Categories 24-26. Some of these features mark the realization of non-movement dependencies. For example, *to* introduces an embedded nonfinite clause, *be* and -*ing* mark the progressive aspect, and the presence of *have* together with -*en* in Categories 24-26 marks the perfect aspect. The specific verbal morphology in Categories 3 and 33 is entirely accidental.

Thus, the model's categories contain both signal and noise for learning which surface features are the footprints of particular types of movement. The learner correctly identified the forms that characterize the most frequent types of movement in English, but it also identified some irrelevant features that are accidentally correlated with these forms. This invites the question of how a learner could effectively use this information for further steps of learning— how a learner could separate signal from noise by explaining some correlations as movement, and others as different dependencies. For example, the learner needs to recognize that certain

sentence-initial function words are the moved wh-words in a wh-question, and *do*-support and subject-auxiliary inversion are reflexes of English question formation, but *be* and -*ing* mark an aspectual dependency that is orthogonal to wh-movement or question formation, even though these forms are statistically correlated with certain wh-questions.

Distinctive Features of Non-Argument Gap Categories. One might imagine an additional source of distributional information for solving this problem. If a surface feature is not realizing a movement dependency, but is realizing another syntactic dependency that is orthogonal to movement, then it should also be present in some sentences without movement. Under this reasoning, our hypothesized learner might not only identify which features are distinctive in sentence categories that have argument gaps, but might also determine which of those features are distinctive in other non-argument gap categories. This might be a signal that such features are not present due to movement.

To test the viability of this solution, we can calculate the odds ratios for the features of the model's categories without argument gaps, restricting our attention to those features that were identified as distinctive in argument-gap categories (Table 5.7). Table 5.8 lists the non-argument gap categories in which each of these features have odds ratios significantly higher than 1. This table shows that all of the features identified as distinctive in the model's argument-gap categories were also identified as distinctive in one or more of the model's non-argument gap categories. The few exceptions include *get* and -*en* verbal morphology, which are only found in the model's categories of passives and are indeed the features that indicate movement in

Feature	**Non-Argument Gap Categories**
Overt subject	1, 2, 4, 5, 6, 10, 18, 19, 20, 31, 32
Sentence-initial subject	1, 2, 4, 5, 10
Subject after aux	18, 19, 20
Subject after noun	6, 31, 32
Verb is first in sentence	1, 2, 4, 5, 10, 13, 15, 18, 19, 22, 34
Verb has -*ed*	
Verb has -*en*	
Verb has -*ing*	5, 6, 7, 8, 15, 19, 20, 22
Verb has -*s*	2, 13, 27, 32
Verb is irregular	4, 13, 27, 31
Verb preceded by *to*	28, 29
Verb preceded by *be*	5, 6, 7, 15, 19
Verb preceded by *have*	
Verb preceded by *get*	
Verb preceded by *do*	1, 18, 34
Sentence-initial function word	10, 13, 15, 18, 19
Sentence-medial function word before verb	6, 7, 27, 31, 32
Sentence-medial function word after verb	2, 10
Sentence-final function word	4, 22
Question	13, 15, 18, 19, 20, 22, 28

Table 5.8: Distributions of Distinctive Features from Table 5.7 in Non-Argument Gap Categories

these categories. However, *have* and -*ed* verbal morphology are also only distinctive features in argument-gap categories, despite the fact that their presence is irrelevant to movement.

Thus, the distinctiveness of a feature in an argument-gap category vs. a non-argument gap category is not by itself a strong signal for determining whether that feature instantiates movement. If a learner were to use this information to infer what type of dependency is instantiated by a particular feature, in some cases the learner would draw a correct inference, but in other cases it would not. A learner who notices that *get* and -*en* are only distinctive features in argument-gap categories

would correctly infer that these features are marking a movement dependency. But such a learner would also incorrectly infer that *have* and *-ed* are signals of movement.

Likewise, a learner who notices that *be* and *-ing* co-occur several non-argument gap categories would be correct to infer that this correlation is not due to movement. And a learner who notices that subject-auxiliary inversion and *do* can co-occur with questionhood in Categories 18, 19, and 20 (polar questions) would also be correct to infer that these features have more to do with interrogatives than with argument movement, since these categories do not have argument gaps. However, such a learner would also draw the same incorrect inference about wh-words, because sentence-initial function words are distinctive in many of many of the non-argument gap categories inferred by the model. A more sophisticated learner might track the distributions of individual unknown function words, rather than collapsing across all unknown function words as a group. But even this learner would be misled by the presence of wh-words in Categories 13, 15, 18, and 19, which contain large numbers of subject and adjunct wh-questions; because the learner does not track missing subjects in its argument-gap inference, it did not identify that these questions had argument gaps. Such a learner might infer that wh-words are only marking interrogative force, and not movement.

In summary, the distributional learning mechanism instantiated by the current learner can successfully identify a superset of the features that characterize different types of movement in English. However, it cannot by itself solve the problem of separating signal from noise in the features it has identified, in order to determine which features are instantiating different movement dependencies, and which are

instantiating other syntactic dependencies or are entirely accidental. It would seem logically possible to compare the distinctive features of sentences with and without argument gaps, in order to isolate the features that are responsible for those argument gaps. But this solution, at least as currently implemented, appears to fall short for two reasons. First, some features that mark non-movement dependencies, such as the perfect aspectual morphology *have* and *-en*, are so highly correlated with movement dependencies in the learner's data that they only occur in argument-gap categories. Second, because the learner imperfectly identifies sentences with movement, failing to recognize cases of non-object movement in e.g. wh-subject and adjunct questions, some features occur frequently in non-argument gap categories even though they are marking movement dependencies.

It is possible that a more sophisticated distributional learning mechanism might be able to overcome these obstacles. Further investigation is needed to determine whether the signal-to-noise ratio in the model's categories improves if it tracks individual function words, or infers argument gaps using not only missing direct objects but also other required but missing arguments (subjects and prepositional objects). This might allow the learner to better identify non-object movement, potentially making its inference about categories with argument gaps more precise. But some accidental correlations may be pernicious to disentangle even for such a learner. Thus, it is also possible that distributional learning, even when guided by prior verb argument structure knowledge, can only take a learner so far. While this process may allow a learner to make progress in identifying which sentences have movement, additional non-distributional information may be needed in order

to identify which movement or non-movement dependencies underlie the surface forms in these sentences. I consider some possible options in the Discussion that follows.

5.4 Discussion

This chapter provides a novel mechanism for the incremental learning of syntactic dependencies. By hypothesis, learners use prior verb argument structure knowledge to guide distributional learning over the surface morphosyntactic forms that they can represent at the current stage of development. Doing so allows them to identify forms that are characteristic of argument movement in the target language, enabling further inference about which specific syntactic dependencies underlie these surface forms. The current model instantiates the first steps of learning proposed under this account:

(i) using knowledge of verb argument structure to detect argument gaps;

(ii) identifying what surface forms are correlated with these argument gaps; and

(iii) inferring what types of syntactic dependencies are responsible for those correlations.

The experiments in this chapter demonstrate the computational feasibility of steps (i) and (ii). The model jointly categorizes sentences according to similarities in their surface forms, and infers which of these sentence categories contain argument gaps. It does so by combining prior knowledge of verb transitivity, as learned by the

model in Chapter 4, with new observations of the relevant morphosyntactic features pertaining to subjects, verbs, and objects that are available to a child prior to 18 months. On the basis of this partial and rudimentary syntactic information, the model accurately identifies a large majority of object movement in child-directed speech.

This process requires building on prior verb transitivity knowledge. A baseline model that first performs distributional analysis over the surface forms of sentences, without knowing which verbs require objects, is unable to later use verb knowledge to determine which clusters of surface properties correlate with argument gaps and which do not. This shows that it is helpful for verb transitivity knowledge to serve as an early guide for distributional learning. Acquiring and applying verb knowledge later— after distributional learning over these surface forms is already underway— leads the learner to miss generalizations about how these surface distributions correspond to argument movement. Thus, a learner who builds incrementally on prior knowledge would do better to identify the basic argument-taking properties of some frequent verbs before attempting to learn the forms that movement can take in a language like English.

This finding is consistent with the empirically attested developmental acquisition for non-basic clause syntax in English. As shown in Chapter 3 and suggested by prior work (Lidz et al., 2017; Jin & Fisher, 2014; Gagliardi et al., 2016; Perkins & Lidz, under review), English-learning infants show sensitivity to verb transitivity by the age of 15 months, but do not yet represent fronted objects in wh-questions as objects until several months later, around the age of 18 months. By showing

that verb transitivity knowledge is needed to guide the current model's learning mechanism, I provide an account for this developmental trajectory. This account proposes that the observed developmental timecourse is not accidental: verb transitivity knowledge is developmentally prior to the acquisition of wh-movement in English because it is needed to facilitate that acquisition.

Importantly, the learning mechanism in this chapter is incremental. The learner does not jointly infer verb transitivity properties, surface feature distributions, and gaps of movement all at once. Rather, in order to model development, I assume that the learner must build on prior syntactic knowledge one step at a time. The joint inference that the learner performs is limited to two specific clause-level properties— how features and argument gaps distribute in sentences— with verb properties having already been acquired. The baseline distributional learner is similarly incremental, but applies its syntactic knowledge in a different order, with verb knowledge applied after rather than before distributional analysis.

However, this means that the findings reported in this chapter do not tell us what would happen if learning does not take place incrementally. It is possible that a non-incremental learner could out-perform this baseline, and perform just as well as our model, by jointly inferring all of these phenomena in tandem. Such a non-incremental approach would be more similar to that taken by prior computational approaches to clause structure acquisition. In the models in Maurits et al. (2009), Kwiatkowski et al. (2012), and Abend et al. (2017), learners jointly infer the language's word order properties together with the argument-taking properties and meanings of verbs and other lexical items. But by modeling the acquisition

of lexical properties separately from the acquisition of clause-level properties, the current approach allows us to assess the independent contribution of each of these learning processes, and how they interact in a particular order. This brings us closer to understanding how learning unfolds over time in a child's development. A model that jointly infers as much as possible about its grammar at once may succeed in different ways than the current model, but would be at a farther remove from learning in development.

Prior models of clause structure acquisition (Maurits et al., 2009; Kwiatkowski et al., 2012; Abend et al., 2017) differ from the current model in another way, by assuming that learners bootstrap grammatical properties of their language using detailed conceptual information. As discussed in Chapter 4, these prior learners have access to meaning representations that specify all and only the predicate-argument relations in a sentence, in a structural format that mirrors the syntax of the sentence. This makes it simple to convert between conceptual to syntactic structure in order to infer both word meanings and clause structure properties.

In taking a different approach, I do not deny that learners perceive the world under particular conceptual descriptions (cf. Chapter 2), and may be able to infer aspects of a speaker's message using non-linguistic contextual cues. The question is how reliably detailed and accurate we can assume these non-linguistic conceptual representations to be. Other work has shown that it is not so easy even for adults to identify what specific meaning is conveyed by a particular linguistic utterance, without access to the form of that utterance (Gillette, Gleitman, Gleitman, & Lederer, 1999; Gleitman, 1990). Thus, the goal of this model is to ask how much

syntactic structure learners might be able to identify by using their representations of the formal features of sentences, even when they do not have access to sentence meaning represented under exactly the right conceptual structure. This bears on the broader question of how learning proceeds even if the information that learners use— linguistic or conceptual— is represented under a less complete or less accurate structural format than prior work has assumed.

In focusing on formal distributional learning, the current work is closer to the approach assumed in Stabler's (1998) learnability analysis of grammars with movement. Stabler identified that movement operations introduce challenges for a learner attempting to draw inferences from string distributions. Even with constraints on the forms that grammars can take, when movement is permitted, many different grammars can generate the same set of strings. Stabler considers two solutions to this problem. First, he shows that it helps to use structured representations rather than strings. A learner can succeed with access to representations of the syntactic "skeleton" of a sentence, which encodes all of the hierarchical branching structure but leaves the nodes unlabelled. These syntactic skeletons include phonologically null terminals (what we have been calling gaps) that make it possible to identify where movement has occurred. Second, Stabler shows that even using string representations only, a learner with a bias towards smaller, simpler grammars will prefer grammars with movement in order to account for the distributions of lexical items across sentences. This is because a learner can use movement to avoid positing different lexical entries for a lexical item that occurs in different structural positions across sentences.

The approach I posit here combines both of these learning strategies: distributional learning across sentences, and learning based on (limited) structural evidence, rather than merely strings. Stabler expresses concern that the skeletal syntactic representations he proposes make very unrealistic assumptions about the data available to learners at the onset of grammar acquisition. However, this concern may be mitigated if we consider grammar learning as taking place incrementally, rather than all at once. Because the current account assumes that some verb argument structure acquisition can take place independently, the proposed learner can infer some of the most important features of Stabler's structural representations: argument gaps, which allow for inference about where movement has occurred. Our learner uses the partial syntactic information that a young infant has access to, resulting in much less detailed representations than the syntactic skeletons proposed by Stabler. Yet the simulations above show that it is possible for a distributional learning mechanism, operating over these partial representations, to identify the surface signals of movement dependencies in English. This learner does not yet perform the final step of inference to identify the underlying dependencies in these sentences. But by identifying when particular surface forms are correlated with empty argument positions, the learner brings us one step closer to a solution for Stabler's problem.

5.4.1 Which Syntactic Dependencies?

A major challenge remains in the current account: how to perform the last step of learning, and infer which underlying syntactic dependencies are responsible

for the correlations between surface forms and argument gaps that the learner has identified. The results of the simulations in this chapter show that distributional analysis alone cannot solve this problem. Our learner may have identified most cases of object movement in its data, but does not know which features are instantiating that movement, or even that movement *per se* is responsible for the argument gaps it has inferred. Its distributional learning mechanism has yielded both signal and noise: the distinctive features of its argument-gap categories include forms that characterize movement dependencies in English, such as sentence-initial function words and *-en* verbal morphology, but also include forms that are instantiating other non-movement dependencies, such as *have*, *be*, and *-ing*. Comparing the distinctive features of argument-gap and non-argument gap categories does not resolve this issue. Moreover, the learner does not know that some forms of its argument-gap categories are instantiating wh-movement, others are instantiating A-movement, and others are not instantiating movement at all, but are marking argument-drop in special pragmatic contexts. How, then, does the learner identify whether a particular formal property is a footprint of movement, and what type of movement?

The problem becomes even thornier when we consider whether this learning mechanism could generalize to languages beyond English. There are several properties of English that make gap-driven learning possible in principle. First, English's relatively fixed word order gives learners a reason to expect that subjects and objects will appear in particular positions relative to the verb. As previously discussed, prior experimental work indicates that English learners are aware of these word order properties at an early age— as young as 15-17 months (Gagliardi et al., 2016;

Perkins & Lidz, under review; Lidz et al., 2017; Hirsh-Pasek & Golinkoff, 1996). This makes it possible for a learner to detect gaps when arguments are missing in their expected positions. But in many languages, word order will not be a reliable cue to argument relations. It is possible that other information, such as case morphology, will aid learners in these cases; see Fisher et al. (in press) and Suzuki and Kobayashi (2017) for evidence that Korean- and Japanese-learning 2-year-olds are sensitive to this information in verb learning. The proposed learning mechanism hinges on the assumption that learners at this stage of development have already acquired some knowledge of how their language marks local predicate-argument relations, but whether this is the case cross-linguistically is an open empirical question.

Moreover, English also differs from languages like Korean and Japanese in not allowing syntactic null arguments. The omitted objects for transitive verbs that occasionally occur in the current dataset are licensed only under very specific discourse contexts, if they are grammatically licensed at all. In the general case, English omitted objects do not appear to have the properties of syntactic null arguments in other languages, and many analysis have treated them as implicit arguments represented outside of the syntax (Bresnan, 1978; Fillmore, 1986; Fodor & Frazier, 1980; Koenig & Mauner, 1999; Rizzi, 1986). This means that English learners need to somehow rule out a syntactic null argument analysis of the argument gaps they observe, whereas Korean and Japanese learners need to rule this in. Furthermore, as discussed in Chapter 4, it is possible that languages with free and prevalent argument-drop may make it even more difficult for a learner to acquire the verb transitivity knowledge needed to detect argument gaps when they are present.

It remains to be seen whether the filtering mechanism proposed in Chapter 4 can withstand the additional noise that syntactic argument-drop introduces, in order to identify the argument-taking properties of verbs in languages with this property.

Finally, English primarily fronts wh-phrases in wh-questions, but fronting is optional in languages like French, and is very rare in primarily wh-in-situ languages like Mandarin. Yet French and Mandarin learners still need to identify wh-dependencies even when there has been no overt movement (Aoun et al., 1981; Huang, 1982); see Gotowski and Becker (2016) and Gotowski (2017) for evidence that French learners by the age of 3 have acquired the ability to produce both fronted and in-situ wh-questions. The current hypothesis does not account for how a learner could identify a movement dependency when no argument gap is observable.

Thus, it is possible that the proposed learning mechanism will only be useful for learners facing certain kinds of data, and other mechanisms will be needed for a learner whose data looks vastly different. In Chapter 6, I discuss the broader theoretical implications for proposing a learning strategy that needs to be sensitive to the shape of the data a learner is exposed to. But in the remainder of this chapter, I would like to consider possibilities for how learning succeeds even in English under the current hypothesis: how learners even in this case can move beyond the noisy signal that is yielded by distributional analysis, and infer the syntactic dependencies that give rise to the surface form distributions they have identified. I propose that this inference requires simultaneous use of these formal distributions, along with additional information about the likely dependencies in a given sentence and the ways that those dependencies might be realized. In particular, prosody and

pragmatics may be two sources of information available to a learner at this stage of development, at least at a certain grain size. Together with a prior over the types of dependencies that grammars make available, these additional sources of information may help constrain the learner's inference about when movement has occurred in a given sentence, and what is moving.

Prosodic Information. Infants are sensitive to patterns of prosodic prominence and prosodic breaks from their first weeks of life (e.g. Christophe, Dupoux, Bertoncini, & Mehler, 1994; Christophe, Mehler, & Sebastin-Galls, 2001; Christophe, Gout, Peperkamp, & Morgan, 2003; Gerken, Jusczyk, & Mandel, 1994; Jusczyk et al., 1992; Nazzi, Bertoncini, & Mehler, 1998). While languages vary in whether or how they use prosodic cues to particular syntactic transformations, prosody may serve as a more cross-linguistically robust signal for constituency structure, because prosodic breaks tend to fall at the edges of syntactic phrases (see Büring, 2013 for a review). For this reason, a large literature proposes that infants can use prosody to bootstrap some basic syntactic properties of their language: using prosodic breaks to help identify some of the constituent boundaries in an utterance, and using additional information, such as stress patterns, to draw further inferences about the syntactic properties of those constituents (Christophe et al., 2008; De Carvalho et al., 2016; Gleitman, Gleitman, Landau, & Wanner, 1988; Gout, Christophe, & Morgan, 2004; Gutman, Dautriche, Crabb, & Christophe, 2015; Morgan, 1986; Morgan & Demuth, 1996; Morgan, Meier, & Newport, 1987; Morgan & Newport, 1981; Wanner & Gleitman, 1982). Recent experimental work shows that infants as young as 18 months can use the presence of prosodic breaks to infer constituent boundaries

and disambiguate the syntactic parse of a string-ambiguous utterance (De Carvalho, He, Lidz, & Christophe, 2019).

Prosody might therefore be a valuable source of information to help learners identify some syntactic properties of the non-basic clauses that they cannot yet parse completely. In particular, learners may be able to use prosodic information to identify some of the constituency structure of the sentential material that they do not yet know how to integrate into a complete parse. For example, if a fronted wh-phrase receives focal stress or is eventually followed by a prosodic break before the subject, a learner may use that information to infer that the fronted wh-phrase is in a different syntactic constituent from the subject and verbal complex. A learner's initial parse of a wh-question, as in (3), might then gain some additional structural detail, as in (4).[4]

(3) What is the dog [$_{\text{VP}}$ eating ___] ?

(4) [$_{\text{XP}}$ What is] [[$_{\text{NP}}$ the dog] [$_{\text{VP}}$ eating ___]]?

Recall that from the perspective of the learner we are considering, the fronted wh-word is represented as an unknown function word. A learner who has a prior over the various grammatical categories that are possible for functional elements, and how those grammatical categories distribute, would gain information from knowing that this unknown function word is in a different constituent from the subject.

[4]Note that it is also possible that a learner could choose a different bracketing here, with the auxiliary in the same constituent as the subject. Given that infants at this age seem to treat fronted auxiliaries differently from determiners within a subject NP (Geffen & Mintz, 2015), this may be less likely at this stage of development. For the purposes of the current discussion I will set aside this possibility for now.

For example, *what* is now much less likely to be a morpheme that marks case or agreement within the subject NP. Instead, its category must be that of a functional element that can occur in a separate phrase on the left edge of a clause.

However, this still does not necessitate that *what* is an argument wh-word. Even though a learner can identify that this sentence contains an argument gap, there are other formal properties of the sentence that might be responsible for the missing argument. Our model identified that this sentence belongs to an argument-gap category with several distinctive features in addition to the unknown function word: the auxiliary has moved before the subject, it is a form of *be*, and the verb has -*ing* morphology. Here is a possible non-English analysis of this sentence that accounts for these features. Perhaps *what* is a particle that occurs in various types of questions, like *la* in Tz'utujil Mayan or *est-ce que* in French; auxiliary fronting also occurs in questions; and the co-occurence of *be* and -*ing* verbal morphology together mark object demotion, as would be found in an antipassive (see Polinsky, 2017). Under this analysis, this sentence would be parsed as a polar question meaning 'Did the dog eat (the thing previously discussed)?'

One could imagine even more possibilities, many of which would be attested cross-linguistically and therefore need to be available hypotheses for a learner. Thus, prosodic information may help constrain the structural analysis that learners assign to the distributional patterns they observe. But it still leaves many options open, even when combined with prior knowledge about the types of parses that grammars make available.

Pragmatic Information. Another useful source of information may come

from pragmatics. Our model tracks whether or not a sentence has interrogative form, and this is a feature that it identified as distinctive in several of its argument-gap categories. As the discussion above acknowledged, it is still an empirical question whether infants at this stage of development can reliably identify the force of an utterance. This would need to be possible independently of knowing the canonical syntactic forms that assertions, questions, and commands take in the language, as the learner is attempting to bootstrap those very syntactic properties. But if learners could use their understanding of the discourse context and other non-syntactic cues to intuit the intended speech act, then this might provide top-down information to constrain the parse that they are attempting to generate.

Here is one example of how this information might be applied in the case of wh-questions. Suppose a learner has managed to use prosodic, syntactic, and distributional information to identify a partial parse of a wh-object question that looks something like (4), repeated here as (5). The learner has further identified that *what*, *be*, *-ing*, and auxiliary inversion are suspiciously present in argument-gap sentences that take this shape.

(5) $[_{\text{XP}}$ What is$]$ $[[_{\text{NP}}$ the dog$]$ $[_{\text{VP}}$ eating ___$]]$?

Suppose further that in this context, the learner is able to identify not only that the sentence has interrogative form but that it is a particular type of interrogative: a content question. Setting aside the question of how the learner would intuit this speech act, if she could manage to do so, it would constrain the possible meanings

that this sentence might express. That is, suppose that the learner knows that the logical form for content questions has a two-part format in which an operator is taking scope over a proposition with an unvalued variable; informally, this might follow the template 'What is the value of x such that *P(x)* is true?' (e.g. Karttunen, 1977). The learner's task would then be to align this template to a possible syntactic parse of the sentence, by picking one of the possibilities that she has generated bottom-up.

In this case, the two-part template for the logical form aligns well with the syntactic constituent structure in (5). There is a constituent containing a subject NP and a VP with an argument gap, which looks like a complete sentence with an empty slot for a variable. The leftmost constituent containing *what* could then be inferred to be the operator. This provides much stronger constraints on the grammatical category of *what*: wh-words are operators binding empty variable slots, whereas question particles are not. A learner who chooses to parse this sentence as a wh-question, with *what* as a wh-word, would arrive at an interpretation that accords with the understood speech act. A learner who chooses to parse this sentence as a polar question, with *what* as a question particle, would not. And having selected the correct parse in this case, the learner would then be able to generalize to other sentences that have the same form— identifying that these sentences all contain wh-movement, and that *what* and other sentence-initial functional elements realize the moved wh-element.

Thus, if learners have access to the right sort of discourse information, it is possible that they could bootstrap simultaneously from syntax and pragmatics. Un-

der this hypothesis, syntactically-informed distributional analysis gets the process started. Syntactic knowledge allows learners to parse part of a sentence, identify when argument gaps are present, and generate a space of possibilities for the structures that could explain both those gaps and the distributions of other formal features. The ability to intuit the intended speech act, at least at the grain size of "question" or "content question," gives learners access to a template for a logical form. Aligning these two structures involves picking a parse that could fit that template, thereby making a hypothesis about which forms are instantiating movement vs. other dependencies. This hypothesis will be supported if it allows the learner to generate interpretations that are sensible in context and match the perceived communicative intent of the speaker.

Although similar in its use of simultaneous top-down and bottom-up information, this proposal differs from the semantic bootstrapping accounts proposed in prior computational work (Maurits et al., 2009; Kwiatkowski et al., 2012; Abend et al., 2017). The current account does not assume that learners have access to a complete meaning representation for a sentence, specifying all predicate-argument relations. Instead, pragmatic information provides only a template for a possible logical form, with details like predicate-argument relations unspecified. Inference is required to align this format for a meaning representation with the learner's partial syntactic representation. And this inference is only possible if the learner has the syntactic knowledge to represent the sentence under a partial structure that could align in the right way.

Further investigation is needed to determine whether this bootstrapping strat-

egy could be feasible. I have sketched out a possible story for how pragmatic information could inform one type of argument movement— wh-movement in wh-questions— but it remains to be shown that this account could generalize to other types of movement dependencies. And even if this strategy is possible in principle, much more empirical work is needed to determine whether infants can access the pragmatic information it requires, by tracking the speech acts of their parents and other speakers around them. One major challenge that this strategy faces is indirectness. The direct speech act, which is closely tied to the syntactic form of the sentence, may be difficult to disentangle from the indirect speech acts that a speaker is performing in using that sentence (Hacquard & Lidz, 2018). For example, a parent might use the sentence *What are you eating?* in order to communicate that a child should spit out the object that she is ingesting. Or, a parent might use the sentence *Where are your shoes?* in order to communicate that a child should put on shoes and leave the house. If a learner were to try to use these utterances as data for bootstrapping, she would need to identify that the direct speech acts are still content questions, even though the sentences are being used to issue indirect commands. For this form of bootstrapping to be viable, an important question for future work is whether infants can guess the direct speech act of an utterance reliably enough for learning to succeed.

5.4.2 Conclusion

In conclusion, this chapter contributes two main findings. First, I show that it is in principle possible for a learner to build incrementally on prior verb argument structure knowledge in order to identify the surface forms that characterize movement in English. The learner does so by performing syntactically-informed distributional analysis over the relevant formal properties of sentences that can be represented at this developmental stage. This is consistent with the empirically observed developmental trajectory for verb transitivity and non-basic clause structure acquisition, and supports the Gap-Driven Learning hypothesis for the learning mechanisms that underlie that development. More broadly, this shows how prior grammatical knowledge can guide statistical learning, allowing learners to incrementally make more precise generalizations about their grammar from incomplete representations of their input.

Second, I show that formal distributional analysis has its limits. The learning mechanism as currently implemented succeeds at identifying various surface forms that might instantiate movement in the target language, but stops there. The learner's distributional analysis does not allow it to identify what syntactic dependencies— movement or non-movement— underlie these surface forms. I propose that this step of inference requires syntactic knowledge working in concert with cross-domain information, potentially from prosody and pragmatics. But this information can only be applied when a learner has a formal syntactic representation for which it could be useful. More work is needed to show whether this proposal is log-

ically and empirically feasible. But if it is, it would point towards a way to combine two sources of knowledge— knowledge of the linguistic structures that grammars make available, and knowledge of how those structures relate to the communicative goals of speakers— in order to characterize the learning mechanisms for grammar acquisition in development.

Chapter 6: Conclusion

This dissertation examines the acquisition of argument structure as a window into the role of development in grammar learning. The way that children represent the data for language acquisition depends on the grammatical knowledge they have at any given point in development. Children use their immature grammatical knowledge, together with other non-linguistic conceptual, pragmatic, and cognitive abilities, to parse and interpret their input. But until children have fully acquired the target grammar, these input representations will be incomplete and potentially inaccurate. Our learning theory must take into account how learning can operate over input representations that change over the course of development. What allows learners to acquire new knowledge from partial and noisy representations of their data, one step at a time, and still converge on the right grammar?

The case study in this dissertation points towards one way in which a theory of language acquisition can account for the role of development. I ask how we can enrich the traditional model that previous linguistic theories have assumed, in which learning is idealized as observing a corpus of primary linguistic data, applying general cognitive abilities and knowledge of the properties that grammars can have, and inferring the grammatical properties of the target language (Chomsky, 1965;

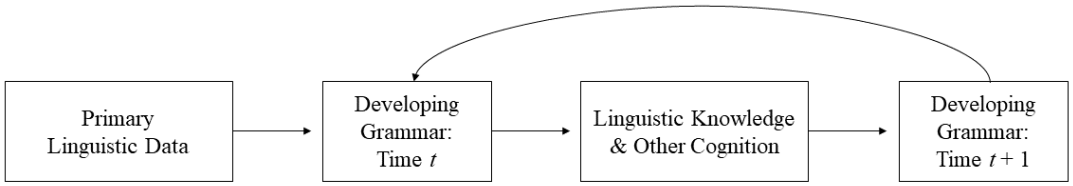

Figure 6.1: Developmental Model of Language Acquisition

Wexler & Culicover, 1980); (see also Yang, 2002; Perfors et al., 2006). I propose a move towards a developmental model that contains the same ingredients, but combines them one step at a time (Fig. 6.1). In this model, learners must use their developing grammar at the current point in time to represent their primary linguistic data in whatever format this immature grammar makes available. They then apply their cognitive abilities and knowledge of the properties that grammars can have in order to gain new knowledge of their language. This process then iterates: their updated developing grammar is now used to represent their input in a richer way, allowing them to draw new generalizations about their grammar and incrementally move closer to the target state.

This developmental model requires specifying in a more fine-grained way the resources that children bring with them to the language learning task. I consider two types of resources. The first is representational: learners need resources for representing their input in useful ways, even early in development. For example, a representation that encodes some amount of linguistic structure, even if that structure is incomplete or inaccurate, will be more useful for a learner acquiring syntax than a representation that only encodes string properties (e.g. Stabler, 1998). The second resource includes mechanisms for learning from these input representations

even if they are incomplete or inaccurate. Learners need tools for identifying the underlying grammatical dependencies in their immature input representations— identifying signal in data that will contain a great deal of noise early in the learning process.

In the experiments in the previous chapters, I examine these two types of resources in the domain of argument structure. With respect to the first representational question, I ask what resources infants use to represent their input for argument structure acquisition: whether they differentiate the core arguments of a clause by their grammatical relations, and when they recognize that clause arguments are present or missing in their canonical positions. With respect to the second learning question, I ask what resources infants use to learn about verb meanings, argument structure, and argument movement on the basis of these input representations. I investigate how learners could use a combination of domain-specific linguistic knowledge and domain-general cognition— statistical learning combined with conceptual and pragmatic abilities— in order to draw inferences about verb meanings and their argument-taking properties, and eventually identify not only local predicate-argument dependencies but also non-local ones. In the discussion that follows, I summarize the main findings in the dissertation and consider their broader implications.

6.1 Summary of Findings

I begin in Chapter 2 by asking how infants at early stages of verb learning represent and use clause transitivity to draw inferences about verb meanings. I contrast two families of hypotheses in the literature. One hypothesis proposes that infants primarily rely on the number of arguments in a clause and expect these to match one-to-one the number of participants that they view in an event (Fisher et al., 2010; Lidz & Gleitman, 2004a; Naigles, 1990, 1990). An alternative hypothesis proposes that infants primarily rely on the grammatical relations of clause arguments and link these in principled ways to the thematic relations that they view in an event (Pinker, 1984; Williams, 2015; Lidz et al., 2017). In two experiments, I pit these two bootstrapping strategies against each other, and ask which one infants will use when the two strategies would lead to different inferences about verb meanings. I show that infants at the age previously tested (19 to 22 months) do not primarily rely on a number-matching strategy. Infants allow a transitive description of an event that they view under a 3-participant concept, showing that they do not expect arguments to match participants in number. But they do not allow an unaccusative intransitive description of this same event concept, showing that they differentiate the arguments of transitive and intransitive clauses and are sensitive to the different meanings that those clause types can express. This result suggests that infants from very early in grammatical development privilege the grammatical and thematic relations of clause arguments above argument number in the representations that they use for bootstrapping verb meanings, and invites further questions into how richly these

relations are represented.

In Chapter 3, I turn to the question of how infants recognize these clause arguments not only when they are in local dependencies with the verb, but also when those dependencies are non-local. I argue that so-called "non-basic" clauses, in which arguments have been moved from their canonical positions, introduce a chicken-and-egg problem in early grammar acquisition. Learners use the subjects and objects of clauses to bootstrap verb meanings and the core argument structure properties of their language. If they could identify when these arguments have been moved, then they could use those moved arguments to draw accurate bootstrapping inferences. Or, if they could identify the argument-taking properties of some frequent verbs in their language, then they could use this information to identify when argument movement has taken place, and learn the ways that various movement dependencies are realized in their language. Which comes first, argument structure or non-basic syntax?

The two experiments in Chapter 3 find that the answer to this chicken-and-egg question is argument structure. I show that 15-month-old English learners are sensitive to the transitivity properties of some frequent verbs, differentiating sentences in which a direct object was present after the verb from sentences in which it was missing in that position. However, only 18-month-olds show awareness that a fronted wh-phrase acts as that locally missing object in a wh-object question. 18-month-olds responded differently to object gaps in wh-object questions compared to object gaps in simple declarative clauses, showing that they recognized both the local and non-local dependencies in these sentences. This result provides evidence

that English learners acquire the ability to represent non-local predicate-argument dependencies in wh-questions in their second year of life, about 2 months before they begin producing them in their own speech (Rowland et al., 2003; Stromswold, 1995). Furthermore, this ability comes developmentally after infants have acquired knowledge of verb transitivity, supporting the hypothesis in Gagliardi et al. (2016) and Perkins and Lidz (under review).

In Chapters 4 and 5, I ask which learning mechanisms might enable this development. I propose that the developmental trajectory observed in Chapter 3 is not accidental, but follows from a mechanism that allows infants to use knowledge of the argument-taking properties of some frequent verbs to bootstrap their acquisition of argument movement. I consider a "Gap-Driven" Learning hypothesis that involves the following steps of learning:

(i) using knowledge of verb argument structure to detect argument gaps: predicted arguments that are unexpectedly missing in their local positions;

(ii) identifying what surface forms are correlated with these argument gaps; and

(iii) inferring what types of syntactic dependencies are responsible for those correlations.

Chapter 4 investigates how the first step of learning required under this hypothesis is even in principle possible. How might a learner identify verbs' argument-taking properties at the developmental stage before she can recognize displaced arguments in non-basic clauses? I follow a solution proposed in prior literature

(Pinker, 1984, 1989; Lidz & Gleitman, 2004b): learners need to "filter" their input in such a way as to learn verb argument structure and clause structure properties only from basic clauses, and ignore the potentially misleading data contributed by non-basic clauses that they cannot yet parse completely. This solution risks being circular if learners need to identify particular sentences as non-basic in order to filter them. However, I show that this circularity can be avoided if learners do not filter non-basic clauses *per se*, but instead merely filter their data in a way that allows them to separate signal from noise.

The model in Chapter 4 implicitly assumes that some of its representations of its data will be erroneous, and it infers how to filter its data in order to learn verb transitivity, without knowing ahead of time where that error comes from or how much error is present. Instead, it infers the right way to filter by using very specific distributions in its data: it tracks observations of post-verbal direct objects within and across verbs, and uses no other information in a sentence. The model is nonetheless able to effectively ignore the non-basic clauses in its data, and accurately categorize the majority of frequent transitive, intransitive, and alternating verbs in a corpus of child-directed speech. And it does so without knowing which sentences are non-basic clauses, or even what the features of non-basic clauses are. This mechanism provides a flexible and potentially powerful way for learners to avoid being misled by the messy data that comes from immature parses of their input: learners may be able to use a form of filtering to maximize the signal in their data, and ignore the noise.

This model provides proof of concept that a learner could, in principle, learn

the argument-taking properties of frequent verbs before being able to identify non-basic clauses. In Chapter 5, I turn to the question of how this verb argument structure knowledge could be used to bootstrap the acquisition of non-basic syntax. I instantiate a model that builds off the learner in Chapter 4 in order to implement steps (i) and and (ii) in the Gap-Driven Learning hypothesis. The model in Chapter 5 jointly clusters together sentences according to similarities in their surface forms, and infers which of those sentence clusters contain argument gaps. It does this by conducting distributional analysis over the relevant morphosyntactic features of sentences that are observable to a child prior to 18 months, guided by the verb transitivity knowledge acquired by the model in Chapter 4. On the basis of these immature sentence representations and partial syntactic knowledge, the model accurately identifies the majority of object movement in a corpus of child-directed speech, and identifies many of the surface forms that characterize movement in English.

However, the mechanism modelled in Chapter 5 stops short of the final step of learning under the Gap-Driven Learning hypothesis: inferring which underlying syntactic dependencies are responsible for the observed correlations between argument gaps and particular surface forms in these sentences. Many spurious correlations exist in the learner's data, and the distributional learning mechanism cannot by itself determine which of these correlations are due to movement vs. other non-movement dependencies. This shows where the limits of distributional learning lie. In order to perform the final step of inference to identify particular movement dependencies in the language, I propose that learners need to make use of not only the formal dis-

tributions they observe, but also additional information— potentially from prosody and pragmatics— about the likely dependencies in a given sentence and the ways those dependencies might be realized. This invites further investigation into how the learning mechanisms for grammar acquisition might be fed by informed statistical learning working in concert with knowledge of the syntactic dependencies that grammars make use of, and knowledge of how those dependencies relate to speakers' goals in discourse.

6.2 Future Directions and Broader Implications

In summary, this work shows one case in which we can characterize the role of development in grammar acquisition by probing more deeply into the resources that learners bring to their learning task. The finding that infants privilege the grammatical and thematic content of clause arguments in verb learning has implications for what resources infants use to represent their input for syntax acquisition. Infants may be biased to look for asymmetrical relations between subjects and objects in a clause, enabling more sophisticated inferences about the types of meanings that a clause can express. The finding that English learners identify the local argument-taking properties of verbs before identifying non-local predicate-argument relations has implications for the mechanisms that they use to learn from immature representations of their input. I show that it is in principle possible for learners to acquire knowledge of verb argument structure by being strategic about how they learn from messy data; learners may then identify non-local dependencies in the language by

using this argument structure knowledge to guide formal distributional learning. This case study thus illuminates how learners may incrementally apply their current linguistic knowledge to their partial and changing input representations in order to infer a grammar.

This dissertation models only one narrow corner of grammatical development: the acquisition of argument structure and argument movement in English. We might now ask how far this model could generalize. The hypothesis I pursue requires that learners identify the core argument structure profile of their language: that they have some way to recognize subjects and objects in their canonical clause positions. It is still an open empirical question how richly these grammatical relations are represented, whether *qua* subjects and objects or in some other format. But given that these relations need to be differentiated in some manner, this raises the question of how learners know what information to use. The relatively fixed word order of English makes the linear position of a noun phrase relative to the verb a strong cue to its grammatical relation. But learners of languages with freer word order or free argument-drop will need other ways to identify when arguments are present even in local dependencies with the verb, and which arguments these are. It is an open empirical question whether other sources of information, such as case morphology, are available and used by learners of such languages at the relevant stage of development (e.g. Fisher et al., in press; Suzuki & Kobayashi, 2017).

Furthermore, the mechanisms proposed above will be unhelpful for identifying movement dependencies in which no argument gap is observable. Yet learners of wh-in-situ languages still need to recognize that a wh-dependency is present even

when the wh-element has not overtly moved to the clause position where it takes scope (Aoun et al., 1981; Huang, 1982). Identifying the underlying covert movement dependencies in these languages would seem to require a different learning process, one in which argument structure knowledge might be used in a different way. It is possible that learners of these languages can more readily recognize when an in-situ wh-element bears a particular grammatical relation, but would need to use other formal, prosodic, or pragmatic information to recognize that this element is in a non-local dependency with a higher node in the clause, corresponding to the scope of the interrogative. How this type of learning proceeds is another topic for future work.

This cross-linguistic comparison raises the possibility that the mechanisms specified at the current grain size may only be helpful for learners facing certain kinds of data, and other mechanisms may be needed for learners whose data have different properties. This raises an important theoretical question. How does language learning converge, and look so remarkably uniform across languages, if the underlying learning mechanisms need to be sensitive to the shape of the data a learner is exposed to— and the shape of the data varies widely from language to language? Or, put another way, what possible resources could allow a learner to identify the right strategy for learning incrementally from the data she has available, when some strategies might work better than others for her language?

I suggest that the analysis of English argument structure acquisition proposed in this dissertation is merely one instance of a more general learning mechanism that might be flexibly tailored to fit the evidence provided by a learner's data. Infants

may approach the task of syntax acquisition with a bias to attend first to the core predicate-argument structure of a clause. This structural representation is crucial for constructing an interpretation of an utterance beyond the single word level; however incomplete this representation may be, identifying some of the core argument relations in a clause will allow at least a partial understanding of who did what to whom. An English learner may identify that word order provides a decent initial signal for identifying these argument relations, and then use other information— formal, prosodic, and pragmatic— to identify the types of non-local dependencies that may disrupt the expected canonical word order. A Japanese learner may identify that case morphology is a better signal for these argument relations, and use that information together with other formal, prosodic, and pragmatic cues to identify both the local and non-local dependencies that are present in a sentence.

Considered at this grain size, these two learning strategies are not so different from each other. A learner's initial knowledge of the types of predicate-argument dependencies that grammars make available, and how these dependencies relate to sentence meanings and their uses in discourse, guides the learner to the relevant evidence that her language provides for detecting these dependencies. This may allow her to arrive at a learning strategy that is in part guided by general principles of grammar, and in part constructed ad-hoc on the basis of her data. Thus, taking an incremental approach to language acquisition requires enriching our understanding of the abstract prior knowledge that learners bring with them to the learning task, and how that abstract knowledge can be applied to enable learning from the specific data a learner has available, as she represents it at any given point in development.

We may also wonder how the mechanisms presented in this work will generalize in other ways. The mechanisms I propose may help a learner out of a particular chicken-and-egg problem: acquiring verb argument structure and non-basic clause syntax in tandem in a language like English. But learners must also acquire many other syntactic properties of their language. How well would these mechanisms generalize if the chicken-and-egg problem becomes multi-dimensional— when learners also need to identify whether their language has e.g. argument-drop, verb raising, and subject-auxiliary inversion? The toy model in this dissertation provides a starting place for understanding one corner of grammar acquisition, but it remains to be seen how it might scale up in order to understand how other aspects of grammar develop incrementally over time.

More broadly, this work illuminates the complex interplay of linguistic, cognitive, and conceptual development in the very first steps of acquiring a grammar. Infants may be able to infer aspects of verb meanings by mapping between the structure of a sentence and a conceptual representation of the world around them. But these inferences rely on the particular representations that infants have available in any given instance, and may involve a sophisticated understanding of the principled relations between linguistic and conceptual structure. Domain-general statistical learning may help a learner extract signal from noise in her own input representations, and identify formal regularities that are present. But statistical learning may need to be guided by prior grammatical knowledge in order for learners to draw the right generalizations to explain those regularities. Understanding a speaker's communicative intent in using a particular sentence may help constrain the structure

and interpretation that a learner assigns to that sentence. But this pragmatic information is not by itself constraining enough to provide a complete parse; a learner must also have available a partial syntactic representation for which this top-down information could be useful. Thus, studying grammatical development provides a window into how linguistic knowledge interacts with the rest of cognition in order to enable learning from input.

Finally, this work has novel implications not only for theories of language acquisition, but also for learning in general. This work proposes novel mechanisms by which a language learner might draw the right generalizations from partial and immature input representations, offering a new perspective on the use of data in learning. Under this approach, learners infer the regularities underlying a particular phenomenon in their input by jointly inferring what data to use in order to best identify those regularities. Learning succeeds through a combination of specific hypotheses that guide the learner to the relevant evidence in her data, and the assumption of a noisy relationship between her data and the hypotheses she is evaluating. The flexibility in this approach invites further investigation into how broadly these mechanisms might generalize beyond language learning. Understanding when learners choose to learn from their input, and when they choose not to learn, may help illuminate how learning succeeds in many other domains in which learners must generalize from incomplete or unreliable representations of data.

Appendix A: Details of Gibbs Sampling for Verb Transitivity Learner

The learner in Chapter 4 uses Gibbs sampling (Geman & Geman, 1984) to jointly infer T, ϵ, and δ, integrating over θ and summing over e, with Metropolis-Hastings (Hastings, 1970) proposals for ϵ and δ.

To begin, values of ϵ and δ are randomly initialized, and values of T for each verb are then sampled given values for those input filter parameters. From observations of a verb with and without direct objects, the model determines which value of T was most likely to have generated those observations. For $k^{(v)}$ direct objects in $n^{(v)}$ sentences containing verb v, we can use Bayes' Rule to compute the posterior probability of each value for $T^{(v)}$,

$$p(T^{(v)}|k^{(v)}, \epsilon, \delta) = \frac{pk^{(v)}|T^{(v)}, \epsilon, \delta)p(T^{(v)})}{\sum_{T'^{(v)}} p(k^{(v)}|T'^{(v)}, \epsilon, \delta)p(T'^{(v)})} \tag{1}$$

Bayes' Rule tells us that the posterior probability of a particular value of T given $k^{(v)}$ and the other model parameters is proportional to the likelihood, the probability of $k^{(v)}$ given that value of T and those parameters, and the prior, the probability of T before seeing any data. T is assumed to be independent of ϵ and δ. Simulation 1 uses a uniform prior over T, which is adjusted to reflect different

biases about the proportions of transitivity categories in Simulation 2.

To calculate the likelihood, we must sum over e. This sum is intractable, but because all of the values of e for the same verb and the same direct object status are exchangeable, we can make the computation more tractable by simply considering how *many* errors were generated for sentences with and without direct objects for a particular verb. We can divide the $k^{(v)}$ observed direct objects for a verb into $k_1^{(v)}$ direct objects that were observed accurately and $k_0^{(v)}$ direct objects that were observed in error. The total $n^{(v)}$ observations for verb v are likewise divided into $n_1^{(v)}$ accurate observations and $n_0^{(v)}$ errorful observations. We then calculate the likelihood by marginalizing over $n_1^{(v)}$ and $k_1^{(v)}$, again assuming independence among T, ϵ, and δ,

$$p(k^{(v)}|T^{(v)},\epsilon,\delta) = \sum_{n_1^{(v)}=0}^{n^{(v)}} \left[\sum_{k_1^{(v)}=0}^{k^{(v)}} p(k^{(v)}|k_1^{(v)},n_1^{(v)},\delta)p(k_1^{(v)}|n_1^{(v)},T^{(v)}) \right] p(n_1^{(v)}|\epsilon) \qquad (2)$$

The first term in the inner sum is equivalent to $p(k_0^{(v)}|n_0^{(v)},\delta)$, assuming we know $n^{(v)}$, the total number of observations for a particular verb. This is the probability of observing $k_0^{(v)}$ errorful direct objects out of $n_0^{(v)}$ errorful observations, which follows a binomial distribution with parameter δ,

$$p(k^{(v)}|k_1^{(v)},n_1^{(v)},\delta) = p(k_0^{(v)}|n_0^{(v)},\delta) = \begin{cases} \binom{n_0^{(v)}}{k_0^{(v)}} \delta^{k_0^{(v)}} (1-\delta)^{n_0^{(v)}-k_0^{(v)}} & \text{if } k_0^{(v)} \leq n_0^{(v)} \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

The second term in the inner sum in (2) is the probability of observing $k_1^{(v)}$

accurate direct objects out of $n_1^{(v)}$ accurate observations, which follows a binomial distribution with parameter $\theta^{(v)}$,

$$p(k_1^{(v)}|n_1^{(v)}, T^{(v)}) = \begin{cases} \binom{n_1^{(v)}}{k_1^{(v)}} (\theta^{(v)})^{k_1^{(v)}} (1-\theta^{(v)})^{n_1^{(v)}-k_1^{(v)}} & \text{if } k_1^{(v)} \leq n_1^{(v)} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Recall that $\theta^{(v)} = 1$ for the transitive category of T, and $\theta^{(v)} = 0$ for the intransitive category of T. For the alternating verb category, $\theta^{(v)}$ is unknown, so we integrate over all possible values of $\theta^{(v)}$ to obtain $\frac{1}{n_1^{(v)}+1}$.

The last term in (2) is the probability of observing $n_1^{(v)}$ accurate observations out of the total $n^{(v)}$ observations for verb v, which follows a binomial distribution with parameter $1-\epsilon$,

$$p(n_1^{(v)}|\epsilon) = \binom{n^{(v)}}{n_1^{(v)}} (1-\epsilon)^{n_1^{(v)}} (\epsilon)^{n^{(v)}-n_1^{(v)}} \tag{5}$$

After sampling values for T for each verb in the dataset, we then sample values for ϵ and δ. If T denotes the set of values $T^{(1)}, T^{(2)}, ..., T^{(V)}$, and k denotes the full set of observations of direct objects $k^{(1)}, k^{(2)}, ..., k^{(V)}$ for all V verbs in the input, we can define functions proportional to the posterior distributions on ϵ and δ, $f(\epsilon) \propto p(\epsilon|T, k, \delta)$ and $g(\delta) \propto p(\delta|T, k, \epsilon)$, as

$$f(\epsilon) = p(k|T, \epsilon, \delta)p(\epsilon) \tag{6}$$

$$g(\delta) = p(k|T, \epsilon, \delta)p(\delta) \tag{7}$$

where the likelihood $p(k|T, \epsilon, \delta)$ is the product over all verbs v of $p(k^{(v)}|T^{(v)}, \epsilon, \delta)$, as calculated in (2).

Within the Gibbs sampler, ϵ is resampled using 10 iterations of a Metropolis-Hastings algorithm. After randomly initializing ϵ, at each iteration, a new value ϵ' is proposed, sampled from the proposal distribution $Q(\epsilon'|\epsilon) = N(\epsilon, 0.25)$. Because the proposal distribution is symmetric, this new value is accepted with probability

$$A = min\left(\frac{f(\epsilon')}{f(\epsilon)}, 1\right) \tag{8}$$

If the new value ϵ' has higher probability given T, k and δ under equation (6), it is accepted. If it has lower probability under equation (6), it is accepted at a rate corresponding to the ratio of its probability and the probability of the old value of ϵ. After sampling ϵ, δ is resampled with 10 iterations of Metropolis-Hastings. The proposal and acceptance functions are analogous to those for ϵ.

Multiple chains from different starting points were run to test convergence of T, ϵ, and δ. The simulations reported here used 1,000 iterations of Gibbs sampling. Every tenth value from the last 500 iterations were taken as samples from the posterior distribution over T, ϵ, and δ.

Appendix B: Details of Gibbs Sampling for Argument Gap Learner

The learner in Chapter 5 uses Gibbs sampling (Geman & Geman, 1984) to jointly infer c and e, integrating over θ, $\delta^{(X)}$, and $\delta^{(F)}$. The learner is initialized using the input filter parameters and transitivity categories inferred by the model in Chapter 4.

B.1 Initialization

To begin, values of c for each sentence are initialized to one of three initial sentence categories: one category with argument gaps and two without. These initial categories are sampled from the posterior probability distribution that a given sentence contains an error— the variable e in the model in Chapter 4— given the values of T, ϵ, and δ inferred by that model. If a sentence is sampled as containing an error under that model, it is initialized to the argument-gap category; if not, it is randomly initialized to one of the two non-argument gap categories. We can use Bayes' Rule to compute the posterior predictive probability for the error value e_i of a particular sentence observation X_i, given all other error values $\mathbf{e}$, other sentence

observations $\mathbf{X}$, and other model parameters:

$$p(e_i|X_i^{(v)}, T^{(v)}, \epsilon, \delta, \mathbf{e}, \mathbf{X}) = \frac{p(X_i^{(v)}|e_i, T^{(v)}, \epsilon, \delta, \mathbf{e}, \mathbf{X})p(e_i|\epsilon)}{\sum_{e_i'} p(X_i^{(v)}|e_i', T^{(v)}, \epsilon, \delta, \mathbf{e}, \mathbf{X})p(e_i'|\epsilon)} \tag{1}$$

The posterior predictive probability of a particular value of e_i given $X_i^{(v)}$, all other observations, and other model parameters, is proportional to the likelihood, the probability of $X_i^{(v)}$ given that value of e_i and the other observations and parameters, and the prior, the probability of e_i before seeing any data. The prior probability of e_i only depends on ϵ: e_i takes a value of 1 with probability ϵ, and 0 with probability $1 - \epsilon$.

For the errorful value $e_i = 1$, the likelihood term $p(X_i^{(v)}|e_i = 1, T^{(v)}, \epsilon, \delta, \mathbf{e}, \mathbf{X})$ depends only on the inferred value for δ, the probability that an errorful observation contains a direct object. $X_i^{(v)}$ takes a value of 1 with probability δ, and 0 with probability $1 - \delta$.

For the non-errorful value $e_1 = 0$, the likelihood term depends on the probability that verb v occurs with a direct object, given by $\theta^{(v)}$ for the verb's inferred transitivity category $T^{(v)}$. If the verb is inferred to be transitive or intransitive, $\theta^{(v)}$ is known. In these cases, $X_i^{(v)}$ takes a value of 1 with probability θ, and 0 with probability $1-\theta$. If the verb is inferred to be alternating, we must integrate over all possible values of $\theta^{(v)}$, conditioning on other non-errorful observations of this verb $\mathbf{X}_1^{(v)}$,

$$p(X_i^{(v)}|e_i = 0, T^{(v)}, \mathbf{X}_1^{(v)}) = \int p(X_i^{(v)}|\theta^{(v)})p(\theta^{(v)}|T^{(v)}, \mathbf{X}_1^{(v)})d\theta^{(v)} \tag{2}$$

The first term inside the integral is equal to $\theta^{(v)}$ if $X_i^{(v)} = 1$, or $1 - \theta^{(v)}$ if $X_i^{(v)} = 0$. We can use Bayes' Rule to compute the second term inside the integral, the probability of $\theta^{(v)}$ given all other non-errorful observations $\mathbf{X}_1^{(v)}$,

$$p(\theta^{(v)}|T^{(v)}, \mathbf{X}_1^{(v)}) = \frac{p(\mathbf{X}_1^{(v)}|\theta^{(v)}, T^{(v)})p(\theta^{(v)}|T^{(v)})}{\int p(\mathbf{X}_1^{(v)}|\theta^{(v)}, T^{(v)})p(\theta^{(v)}|T^{(v)})d\theta^{(v)}} \tag{3}$$

The prior probability $p(\theta^{(v)}|T^{(v)})$ is assumed to follow a uniform $Beta(1, 1)$ distribution. We can re-write $\mathbf{X}_1^{(v)}$ as $k_1^{(v)}$ direct object observations for this verb out of $n_1^{(v)}$ total non-errorful observations. The likelihood term, $p(\mathbf{X}_1^{(v)}|\theta^{(v)}, T^{(v)})$, is the probability of observing $k_1^{(v)}$ direct objects in $n_1^{(v)}$ total observations. This follows a binomial distribution with parameter $\theta^{(v)}$,

$$p(\mathbf{X}_1^{(v)}|\theta^{(v)}, T^{(v)}) = p(k_1^{(v)}|n_1^{(v)}, \theta^{(v)}) = \binom{n_1^{(v)}}{k_1^{(v)}} {\theta^{(v)}}^{k_1^{(v)}} (1 - \theta^{(v)})^{n_1^{(v)} - k_1^{(v)}} \tag{4}$$

Solving the integrals in equations (2) and (3), we calculate that $X_i^{(v)}$ takes a value of 1 with probability $\frac{k_1^{(v)}+1}{n_1^{(v)}+2}$, and 0 with probability $\frac{n_1^{(v)}-k_1^{(v)}+1}{n_1^{(v)}+2}$.

Values of $e = 1$ and $e = 0$ were randomly initialized for every sentence in the dataset. These values were then re-sampled for each sentence sequentially from the posterior predictive probability distribution defined in Equation 1, using the values for T and the means of the distribution over ϵ and δ (0.19 and 0.25 respectively) that were inferred by the model in Chapter 4, Simulation 1. This process was repeated over 1,000 iterations of Gibbs sampling. The final sample was used to initialize the simulations in Chapter 5 in the following way: all sentences sampled as being errorful

were initialized to an argument-gap sentence category, and all sentences sampled as being non-errorful were randomly divided among two sentence categories without argument gaps.

B.2 Sampling c

After initializing c, new values of c for each sentence are re-sampled sequentially. From observations of direct objects and other features in a sentence, and across other sentences in the model's data, the model determines which previously seen or new value of c was most likely to have generated those observations. For direct object observation $X_i^{(v)}$ and other feature observations $\vec{F}_i^{(v)}$ in sentence i, together with all other direct object observations $\mathbf{X}$, feature observations $\vec{\mathbf{F}}$, and sentence category assignments $\mathbf{c}$ for other sentences in the dataset, we can use Bayes' Rule to compute the posterior predictive probability of each value for c,

$$p(c_i|X_i^{(v)}, \vec{F}_i^{(v)}, T^{(v)}, e_c, \mathbf{X}, \vec{\mathbf{F}}, \mathbf{c}) = \frac{p(X_i^{(v)}, \vec{F}_i^{(v)}|c_i, e_c, T^{(v)}, \mathbf{X}, \vec{\mathbf{F}}, \mathbf{c})p(c_i|\mathbf{c})}{\sum_{c_i'} p(X_i^{(v)}, \vec{F}_i^{(v)}|c_i', e_c, T^{(v)}, \mathbf{X}, \vec{\mathbf{F}}, \mathbf{c})p(c_i'|\mathbf{c})} \quad (5)$$

The posterior predictive probability of a particular value of c given the observed data, known transitivity categories, and other sentence category values is proportional to the likelihood, the probability of $X_i^{(v)}$ and $\vec{F}_i^{(v)}$ given that value of c, other observed data and category values, and the prior probability of c. We assume that c is independent of all other model parameters. The prior probability of c is a Dirichlet process (Ferguson, 1973) with parameter α. In this process, each

category value c_i has prior probability proportional to the number of sentence observations already assigned to that category, n_{c_i}. This process also reserves a small non-zero probability for new categories of c, with half of this probability reserved for new argument-gap categories and half reserved for new categories without argument gaps. This probability is determined by the parameter α, which we set equal to 1. For n total observations of sentences across all categories, we define the prior on c,

$$p(c_i|\mathbf{c}) = \begin{cases} \dfrac{n_{c_i}}{n+\alpha} & \text{for previously seen values of } c \\ \dfrac{0.5\alpha}{n+\alpha} & \text{for new values where } e_c = 1 \\ \dfrac{0.5\alpha}{n+\alpha} & \text{for new values where } e_c = 0 \end{cases} \tag{6}$$

Assuming independence between X and F, we calculate the likelihood as the product of the probabilities of observing $X_i^{(v)}$ and $\vec{F}_i^{(v)}$, given the other observations and model parameters,

$$p(X_i^{(v)}, \vec{F}_i^{(v)}|c_i, e_c, T^{(v)}, \mathbf{X}, \vec{\mathbf{F}}, \mathbf{c}) = p(X_i^{(v)}|c_i, e_c, T^{(v)}, \mathbf{X}, \mathbf{c})p(\vec{F}_i^{(v)}|c_i, e_c, \vec{\mathbf{F}}, \mathbf{c}) \tag{7}$$

The first term in this likelihood function is calculated differently depending on the value of e_c for the current category c_i. If c_i is an argument-gap category, then direct objects are generated by the grammatical property of that category $\delta_{c_i}^{(X)}$. We calculate the probability of a direct object by integrating over all possible values of

$\delta_{c_i}^{(X)}$, conditioning on other observations of sentences in this category,

$$p(X_i^{(v)}|c_i, e_i = 1, T^{(v)}, \mathbf{X}, \mathbf{c}) = \int p(X_i^{(v)}|\delta_{c_i}^{(X)}) p(\delta_{c_i}^{(X)}|c_i, \mathbf{X}) d\delta_{c_i}^{(X)} \tag{8}$$

This equation takes the same form as in equation (2). Let n_{c_i} be the total observations in category c_i and k_{c_i} be the total direct object observations in this category. Following equations analogous to (2)-(4), we calculate that $X_i^{(v)}$ takes a value of 1 with probability $\frac{k_{c_i}+1}{n_{c_i}+2}$, and 0 with probability $\frac{n_{c_i}-k_{c_i}+1}{n_{c_i}+2}$.

If c_i is not an argument-gap category, then direct objects in this category are generated by the transitivity properties of each verb. The first term in the likelihood function in (7) thus depends on the known transitivity category $T^{(v)}$. If verb v is transitive or intransitive, then θ is known, and $X_i^{(v)}$ takes a value of 1 with probability θ, and 0 with probability $1 - \theta$. If verb v is alternating, we again integrate over all possible values of $\theta^{(v)}$, conditioning on observations of this verb in other categories without argument gaps. This is the same integral as in equation (2). Here, let $n_1^{(v)}$ be the total observations for verb v in categories where $e_c = 0$, and $k_1^{(v)}$ be the total direct object observations for verb v in these categories. Following equations (2)-(4), we calculate that $X_i^{(v)}$ takes a value of 1 with probability $\frac{k_1^{(v)}+1}{n_1^{(v)}+2}$, and 0 with probability $\frac{n_1^{(v)}-k_1^{(v)}+1}{n_1^{(v)}+2}$.

The second term in (7) is the probability of the other observed features occurring in the given category. Assuming independence among features, this is equivalent

to the product over the probabilities of observing each feature in this category,

$$p(\vec{F_i}^{(v)}|c_i, e_c, \vec{\mathbf{F}}, \mathbf{c}) = \prod_{F_i'^{(v)}} p(F_i'^{(v)}|c_i, e_c, \mathbf{F}, \mathbf{c}) \tag{9}$$

The probability of observing a particular feature F in a category c_i is given by $\delta_{c_i}^{(F)}$ for that feature and that category. We integrate over all possible values of $\delta_{c_i}^{(F)}$, conditioning on other observations of feature F. Let n_{c_i} be the total observations in category c_i and $k_{c_i}^F$ be the total observations of feature F in this category. Again following equations analogous to (2)-(4), we calculate that $F_i^{(v)}$ takes a value of 1 with probability $\frac{k_{c_i}^F+1}{n_{c_i}+2}$, and 0 with probability $\frac{n_{c_i}-k_{c_i}^F+1}{n_{c_i}+2}$.

B.3 Sampling e

After sampling values for c for each sentence in the dataset, we then sample new values of e for each category. Bayes' Rule allows us to calculate the posterior probability of each value of e_c for a category c given all of the direct object observations in that category $\mathbf{X}_c$ and known verb transitivity properties T,

$$p(e_c|c, \mathbf{X}_c, T) = \frac{p(\mathbf{X}_c|e_c, c, T)p(e_c)}{\sum_{e_c'} p(\mathbf{X}_c|e_c', c, T)p(e_c')} \tag{10}$$

We assume that e_c is independent of T and c, and that the prior probability $p(e_c) = 1$ is set to 0.19, the mean value of ϵ inferred by the model in Chapter 4. In other words, the learner assumes that the prior probability of a transitivity-violating category is equivalent to the probability that any single sentence contains

a transitivity violation, as inferred by the previous learner. This will only be the case if sentences are equally distributed among categories, a simplifying assumption of the learner's prior that may be overridden if not supported by the data.

The likelihood term, $p(\mathbf{X}_c|e_c, c, T)$, is the probability of seeing particular observations of direct objects for verbs in this category. If $e_c = 1$ and c_i is an argument-gap category, this probability is determined by $\delta_{c_i}^{(X)}$. We calculate the joint probability of the direct object observations for each verb in that category given $\delta_{c_i}^{(X)}$, integrating across all possible values of $\delta_{c_i}^{(X)}$,

$$p(\mathbf{X}_c|e_c = 1, c, T) = \int \prod_{v'} \Big(p(\mathbf{X}_c^{(v')}|\delta_{c_i}^{(X)})\Big) p(\delta_{c_i}^{(X)}|c_i) d\delta_{c_i}^{(X)} \tag{11}$$

The first term inside the integral is the product across all verbs of probability of the direct observations for that verb $\mathbf{X}_c^{(v)}$ in the category, given $\delta_{c_i}^{(X)}$. For each verb v, we can re-write $\mathbf{X}_c^{(v)}$ as k_c^v direct object observations out of n_c^v total observations for that verb in the category. The probability $p(\mathbf{X}_c^{(v)}|\delta_{c_i}^{(X)})$ follows a binomial distribution with parameter $\delta_{c_i}^{(X)}$,

$$p(\mathbf{X}_c^{(v)}|\delta_{c_i}^{(X)}) = p(k_c^{(v)}|n_c^{(v)}, \delta_{c_i}^{(X)}) = \binom{n_c^{(v)}}{k_c^{(v)}} \delta_{c_i}^{(X)^{k_c^{(v)}}} (1 - \delta_{c_i}^{(X)})^{n_c^{(v)} - k_c^{(v)}} \tag{12}$$

We assume that the prior probability $p(\delta_{c_i}^{(X)}|c_i)$ follows a uniform $Beta(1, 1)$ distribution. Let n_c be the total observations in a particular category and k_c be the total direct object observations in that category. Solving the integral in equation

(11), we find that

$$p(\mathbf{X}_c|e_c = 1, c, T) = \frac{\Gamma(k_c+1)\Gamma(n_c-k_c+1)}{\Gamma(n_c+2)}\Big(\prod_{v'}\frac{\Gamma(n_c^{(v')}+1)}{\Gamma(k_c^{(v')}+1)\Gamma(n_c^{(v')}-k_c^{(v')}+1)}\Big) \quad (13)$$

If $e_c = 0$ and c_i is not an argument-gap category, the likelihood term in equation (10) is determined by the known transitivity $T^{(v)}$ of each verb in the category. The probability of the particular direct object observations $\mathbf{X}_c$ in the category is the joint probability of seeing those direct object observations for each verb, given the transitivity of that verb,

$$p(\mathbf{X}_c|e_c = 0, c, T) = \prod_{v'}\Big(p(\mathbf{X}_c^{(v')}|T^{(v')}\Big) \quad (14)$$

We can again re-write $\mathbf{X}_c^{(v)}$ as k_c^v direct object observations out of n_c^v total observations for a given verb in a given category. The probability of observing k_c^v direct objects out of n_c^v total observations of a verb follows a binomial distribution with parameter $\theta^{(v)}$,

$$p(\mathbf{X}_c|e_c = 0, c, T) = p(k_c^{(v)}|n_c^{(v)}, T^{(v)}) = \binom{n_c^{(v)}}{k_c^{(v)}}(\theta^{(v)})^{k_c^{(v)}}(1-\theta^{(v)})^{n_c^{(v)}-k_c^{(v)}} \quad (15)$$

Recall that $\theta^{(v)} = 1$ for transitive verbs and $\theta^{(v)} = 0$ for intransitive verbs. For

alternating verbs, we must integrate across all possible values of $\theta^{(v)}$,

$$p(k_c^{(v)}|n_c^{(v)}, T^{(v)}) = \int p(k_c^{(v)}|n_c^{(v)}, \theta^{(v)}) p(\theta^{(v)}|T^{(v)}) d\theta^{(v)} \tag{16}$$

We assume that $p(\theta^{(v)}|T^{(v)})$ follows a $Beta(\alpha, \beta)$ distribution, where the parameters α and β are counts of direct object observations and no direct object observations for verb v in other categories without argument gaps. Solving the integral in equation (16), we find that

$$p(k_c^{(v)}|n_c^{(v)}, T^{(v)}) = \\ \Big(\frac{\Gamma(n_c^{(v)}+1)}{\Gamma(k_c^{(v)}+1)\Gamma(n_c^{(v)}-k_c^{(v)}+1}\Big)\Big(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\Big)\Big(\frac{\Gamma(k_c^{(v)}+\alpha)\Gamma(n_c^{(v)}-k_c^{(v)}+\beta)}{\Gamma(n_c^{(v)}+\alpha+\beta)}\Big) \tag{17}$$

B.4 Sampling with Annealing

Multiple chains from different starting points were run to test convergence of c and e. The simulations reported here used 5,000 total iterations of Gibbs sampling. To aid in the model's search process, simulated annealing was used during the first 1,000 iterations (e.g. Goldwater Griffiths 2007). In this process, we raise the posterior probabilities of c and e to the power of an annealing constant defined as $1/t$, where t is the current 'temperature.' Then, we slowly lower the temperature (reduce t) until the annealing constant reaches 1. While the temperature is warm, the posterior probability distributions are flattened so the learner is able to explore more of its hypothesis space. After 1,000 iterations of Gibbs sampling with annealing,

another 4,000 iterations were run without annealing. The final interation was taken as a sample from the posterior distribution over c and e.

Appendix C: Description of Argument-Gap and Other Movement Categories

The learner in Chapter 5 identified 15 categories that it inferred to have argument gaps, and 3 additional non-argument gap categories whose predominant clause type contained movement. I provide a description of these categories below.

Categories 3 & 9 are categories whose predominant clause type is basic, but with properties that led the model to identify transitivity violations. Setting aside Category 3 for the moment, the clauses in Category 9 tended to start with conjunctions or other unknown function words, and were grouped together with a non-negligible number of relative clauses and wh-questions (example provided in Table 5.3). For this reason, the model inferred that Category 9 had argument gaps.

Category 9 is also characterized by the presence of progressive morphology. This is a distinction that was also made in several of the model's other categories. For example, the model differentiated progressive wh-questions (Categories 11, 15, and 17) from non-progressive wh-questions (Categories 12, 13, 14, and 16). It also differentiated progressive relative and other embedded clauses (Category 33) from non-progressive relative clauses (Category 27). It seems that correlation between the features *be* and *-ing* was sufficiently robust in the corpus that the model considered

it important to explain in its categorization of several sentence types.

The progressive/non-progressive distinction is one of several dimensions that the model differentiated in its categories of wh-questions (***Categories 11-17***). The model differentiated long wh-questions with movement out of an embedded clause, in Category 12, from short wh-questions in the remaining categories. It also differentiated wh-questions that started with a wh-forms (Categories 11-15) from wh-questions with a conjunction, interjection, or other expression before the wh-word (Categories 16-17). Finally, it differentiated subject wh-questions, in Categories 13 and 15, from object questions in the remaining categories. Subject questions appear to have sufficiently different formal properties, such as lack of subject-auxiliary inversion and *do*-support, that the model categorized them separately from other wh-questions. Because the model only attended to missing objects and not subjects in its argument-gap inference, the subject question categories were not inferred to have argument gaps. Adjunct questions did not make up the majority of any single wh-question category, but were distributed among the model's two object question categories and Categories 18-19, which are categories of predominantly polar questions.

Category 21 contains many questions that fell into the 'other' category—i.e., those that did not have the canonical syntactic form of an English wh-question or a polar question. Because of the prevalence of fragment and echo questions in this category, many transitive verbs in these sentences were missing nominal direct objects, which led the model to infer that this sentence category also contains argument gaps.

Categories 23-26 are categories of passives. The model differentiated between *get*-passives, in Category 23, and *be*-passives, in Categories 24-26. Category 25 contains primarily embedded passives. The two categories of matrix *be*-passives differ by their verbal morphology: overt -*en* morphology is consistently present in Category 24, whereas the passivized verbs in Category 26 have -*ed* or an irregular form rather than an an overt -*en* morpheme. Some of the passives in these categories may be adjectival passives, as the learner does not have a way to identify these.

Categories 27, 30 & 33 contain various embedded clauses. Category 27 is a fairly mixed category whose most common clause type, by a small plurality, is relative clauses. These are mostly subject relatives, and because they do not have object gaps, the model inferred that this was not an argument-gap category. Most of the object relative clauses in the corpus appear in argument-gap Category 33, where they were categorized with other embedded clauses. The relevant distributional difference for the learner between subject and object relatives appeared to be the presence or absence of an overt subject for the verb of interest, in canonical subject position before the verb. Category 30 is predominantly made up of nonfinite embedded clauses. Many of these sentences contain embedded passives or rarer types of movement that were not coded as a separate underlying clause type, such as *tough*-movement (e.g. *They're very easy to lose*) and movement out of nonfinite purposive clauses (e.g. *It's not a ball to throw*). For this reason, this category was also inferred to have argument gaps. It appears that the model was sensitive to some less frequent cases of object movement that were not included in the target coding scheme.

Finally, ***Category 35*** is a category whose primary clause type is imperatives, but also includes many basic clauses. This category, along with Category 3, shows the interesting property of not actually having much movement. Instead, these two categories appear to have a large proportion of fragments and object-drop that might be ungrammatical, or licensed only under special discourse contexts (e.g. *I'll open*). These categories also have a small proportion of known *intransitive* verbs occurring with direct objects (8% of Category 3 and 3% of Category 35). Some of these extra direct objects occur in grammatical but infrequent uses of a verb that is normally intransitive (e.g. *I'll sit you up*). Others are temporal adjuncts that the learner could not tell apart from direct objects (e.g. *We went swimming nearly every day*). Due to the presence of both missing direct objects for transitive verbs and extra direct objects for intransitive verbs, the learner inferred that the direct objects in these categories likely came from the grammatical property of the category rather than the transitivity of the verbs in these sentences.

This means that the term 'argument-gap category' is a misnomer in some cases; more accurately, the model inferred that some categories contained transitivity violations, which in some cases produced argument gaps and in some cases produced unexpected extra objects. Because these extra objects are rare, I will continue to treat these categories as primarily argument-gap categories in the analyses that follow, with the caveat that in some cases these contain different types of transitivity violations. A more sophisticated learner might draw difference inferences on the basis of these different types of violations, a possibility that I leave for future work.

Appendix D: Odds Ratios for Features of Argument Gap Categories

The learner in Chapter 5 identified 15 sentence categories that it inferred to have argument gaps. To evaluate which surface morphosyntactic features were the most distinctive of these categories, I calculated the odds ratios for each feature in each of these categories. The odds of observing a feature in a particular category were divided by the odds of observing the feature outside of that category. An odds ratio greater than 1 indicates that a feature has higher-than-usual odds inside a category; an odds ratio less than 1 indicates that a feature has lower-than-usual odds inside a category. An odds ratio of infinity can occur if a feature is always present inside a category, and an odds ratio of 0 can occur if a feature is never present.

A Fisher's exact test was conducted to determine which odds ratios were significantly different from 1. A Bonferroni correction was applied to correct for multiple comparisons: because 21 odds ratios were conducted for each category, the critical value for each comparison was established by setting α equal to $0.05/21 = 0.002$. Tables D.1-D.15 list the odds ratios (OR), along with their 95% confidence intervals (CI) and p-values, for features whose odds ratios were significantly greater than 1 in each of the model's argument-gap categories.

Feature	**OR**	**CI**	***p*-value**
Overt subject	47.01	(18.15, 173.90)	< 0.001
Sentence-initial subject	9.37	(7.12, 12.47)	< 0.001
Noun before subject	1.77	(1.29, 2.40)	< 0.001
Verb is first in sentence	10.22	(5.75, 20.03)	< 0.001
Verb has -*ed*	2.68	(1.62, 4.19)	< 0.001
Verb has -*s*	3.10	(1.89, 4.86)	< 0.001
Sentence-medial function word after verb	4.23	(2.73, 6.35)	< 0.001
Sentence-final function word	2.69	(1.50, 4.51)	< 0.001

Table D.1: Distinctive Features of Category 3 (Basic)

Feature	**OR**	**CI**	***p*-value**
Overt subject	139.89	(24.90, 5391.99)	< 0.001
Noun before subject	11.59	(8.70, 15.48)	< 0.001
Verb has -*ing*	∞	(289.35, ∞)	< 0.001
Verb preceded by *be*	36.94	(25.31, 55.49)	< 0.001
Sentence-medial function word before verb	6.49	(4.85, 8.66)	< 0.001

Table D.2: Distinctive Features of Category 9 (Basic)

Feature	**OR**	**CI**	***p*-value**
Overt subject	∞	(82.40, ∞)	< 0.001
Aux before subject	2401.00	$(0.00, > 10,000)$	< 0.001
Verb is first in sentence	203.69	(36.40, 7672.94)	< 0.001
Verb has -*ing*	∞	(671.79, ∞)	< 0.001
Verb preceded by *be*	3370.99	$(614.42, < 10,000)$	< 0.001
Sentence-initial function word	208.56	(124.72, 375.14)	< 0.001
Question	329.80	(91.31, 2830.29)	< 0.001

Table D.3: Distinctive Features of Category 11 (Wh-Questions)

Feature	**OR**	**CI**	***p*-value**
Verb preceded by *to*	186.15	(89.26, 475.39)	< 0.001
Sentence-initial function word	1053.39	(283.62, 8192.00)	< 0.001
Question	∞	(120.52, ∞)	< 0.001

Table D.4: Distinctive Features of Category 12 (Wh-Questions)

Feature	**OR**	**CI**	***p*-value**
Overt subject	∞	(95.48, ∞)	< 0.001
Aux before subject	474.41	(216.39, 1339.56)	< 0.001
Verb is first in sentence	21.05	(11.65, 42.47)	< 0.001
Verb preceded by *do*	31.85	(25.42, 40.19)	< 0.001
Sentence-initial function word	61.33	(46.13, 83.04)	< 0.001
Question	13.59	(10.14, 18.56)	< 0.001

Table D.5: Distinctive Features of Category 14 (Wh-Questions)

Feature	**OR**	**CI**	***p*-value**
Overt subject	∞	(19.32, ∞)	< 0.001
Aux before subject	101.98	(42.31, 320.98)	< 0.001
Noun before subject	7.38	(4.97, 10.95)	< 0.001
Verb is first in sentence	1.87	(1.15, 3.18)	0.009
Verb preceded by *do*	17.19	(11.18, 27.02)	< 0.001
Sentence-medial function word before verb	47.85	(28.89, 83.45)	< 0.001
Sentence-medial function word after verb	6.50	(3.54, 11.23)	< 0.001
Question	7.25	(4.34, 12.79)	< 0.001

Table D.6: Distinctive Features of Category 16 (Wh-Questions)

Feature	**OR**	**CI**	***p*-value**
Overt subject	∞	(15.90, ∞)	< 0.001
Aux before subject	83.03	(34.20, 262.06)	< 0.001
Noun before subject	2.86	(1.72, 4.60)	< 0.001
Verb is first in sentence	13.00	(4.30, 64.35)	< 0.001
Verb has -*ing*	216.25	(58.25, 1800.69)	< 0.001
Verb preceded by *be*	142.79	(53.65, 543.38)	< 0.001
Sentence-medial function word before verb	29.14	(17.88, 49.19)	< 0.001
Question	∞	(34.34, ∞)	< 0.001

Table D.7: Distinctive Features of Category 17 (Wh-Questions)

Feature	**OR**	**CI**	***p*-value**
Verb is first in sentence	61.91	(10.92, 2432.58)	< 0.001
Sentence-final function word	37.34	(25.80, 53.74)	< 0.001
Question	5.68	(3.73, 8.95)	< 0.001

Table D.8: Distinctive Features of Category 21 (Other Questions)

Feature	**OR**	**CI**	***p*-value**
Verb has -*en*	43.75	(22.31, 81.23)	< 0.001
Verb has irregular form	18.00	(10.97, 29.78)	< 0.001
Verb preceded by *get*	∞	(8402.05, ∞)	< 0.001

Table D.9: Distinctive Features of Category 23 (Passives)

Feature	**OR**	**CI**	***p*-value**
Overt subject	3.02	(1.64, 6.01)	< 0.001
Sentence-initial subject	8.56	(4.99, 15.31)	< 0.001
Verb is first in sentence	31.89	(5.54, 1269.77)	< 0.001
Verb has -*en*	15443.69	(3062.43, $> 10,000$)	< 0.001
Verb preceded by *be*	23.07	(13.05, 43.15)	< 0.001
Verb preceded by *have*	761.15	(393.03, 1600.00)	< 0.001

Table D.10: Distinctive Features of Category 24 (Passives)

Feature	**OR**	**CI**	***p*-value**
Overt subject	6.76	(2.11, 34.55)	< 0.001
Noun before subject	29.22	(12.40, 79.44)	< 0.001
Verb has -*en*	191.83	(89.79, 412.45)	< 0.001
Verb has irregular form	5.54	(2.36, 12.04)	< 0.001
Verb preceded by *be*	10.32	(4.92, 22.46)	< 0.001
Verb preceded by *have*	4541.88	(664.29, $> 10,000$)	< 0.001

Table D.11: Distinctive Features of Category 25 (Passives)

Feature	**OR**	**CI**	***p*-value**
Sentence-initial subject	5.32	(3.59, 7.98)	< 0.001
Verb is first in sentence	12.43	(4.72, 46.44)	< 0.001
Verb has irregular form	23.68	(15.96, 35.54)	< 0.001
Verb preceded by *be*	8.57	(5.83, 12.67)	< 0.001
Verb preceded by *have*	214.41	(139.30, 333.63)	< 0.001

Table D.12: Distinctive Features of Category 26 (Passives)

Feature	**OR**	**CI**	***p*-value**
Verb preceded by *to*	121.74	(75.20, 210.26)	< 0.001
Verb preceded by *get*	5.19	(2.38, 10.17)	< 0.001
Sentence-medial function word before verb	4.91	(3.99, 6.01)	< 0.001

Table D.13: Distinctive Features of Category 30 (Embedded Clauses)

Feature	**OR**	**CI**	***p*-value**
Overt subject	∞	$(47.65, \infty)$	< 0.001
Noun before subject	93.00	(59.80, 151.43)	< 0.001
Verb has *-s*	3.36	(2.05, 5.27)	< 0.001
Verb has irregular form	4.95	(3.69, 6.59)	< 0.001
Sentence-medial function word before verb	17.17	(13.24, 22.37)	< 0.001

Table D.14: Distinctive Features of Category 33 (Embedded Clauses)

Feature	**OR**	**CI**	***p*-value**
Verb is first in sentence	2.94	(2.12, 4.16)	< 0.001

Table D.15: Distinctive Features of Category 35 (Imperatives)

References

Abend, O., Kwiatkowski, T., Smith, N. J., Goldwater, S., & Steedman, M. (2017). Bootstrapping language acquisition. *Cognition*, *164*, 116–143.

Adriaans, F., & Swingley, D. (2012). Distributional learning of vowel categories is supported by prosody in infant-directed speech. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 34).

Alishahi, A., & Stevenson, S. (2008). A computational model of early argument structure acquisition. *Cognitive science*, *32*(5), 789–834.

Anderson, J. R., & Matessa, M. (1990). A rational analysis of categorization. In *Machine Learning Proceedings 1990* (pp. 76–84). Elsevier.

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of memory and language*, *51*(1), 23–54.

Aoun, J., Hornstein, N., & Sportiche, D. (1981). Some aspects of wide scope quantification. *Journal of Linguistic Research*, *1*(3), 69–95.

Arunachalam, S., Escovar, E., Hansen, M. A., & Waxman, S. R. (2013). Out of sight, but not out of mind: 21-month-olds use syntactic information to learn verbs even in the absence of a corresponding event. *Language and cognitive processes*, *28*(4), 417–425.

Arunachalam, S., Syrett, K., & Chen, Y. (2016). Lexical disambiguation in verb learning: evidence from the conjoined-subject intransitive frame in English and Mandarin Chinese. *Frontiers in psychology*, *7*, 138.

Arunachalam, S., & Waxman, S. R. (2010). Meaning from syntax: Evidence from 2-year-olds. *Cognition*, *114*(3), 442–446.

Arunachalam, S., & Waxman, S. R. (2011). Grammatical form and semantic context in verb learning. *Language Learning and Development*, *7*(1), 169–184.

Arunachalam, S., & Waxman, S. R. (2015). Let's see a boy and a balloon: Argument labels and syntactic frame in verb learning. *Language Acquisition*, *22*(2), 117–131.

Baker, M. C. (1988). *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.

Barak, L., Fazly, A., & Stevenson, S. (2014). Learning verb classes in an incremental model. In *Proceedings of the Fifth Workshop on Cognitive Modeling and Computational Linguistics* (pp. 37–45).

Becker, M. (2015). Animacy and the acquisition of tough adjectives. *Language Acquisition*, *22*(1), 68–103.

Becker, M., Estigarribia, B., & Gylfadottir, D. (2012). Tough-adjectives are easy to learn. *Supplemental Proceedings of BUCLD*, *36*, 1–12.

Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: MIT Press.

Brandone, A., Addy, D. A., Pulverman, R., Golinkoff, R. M., & Hirsh-Pasek, K. (2006). One-for-one and two-for-two: anticipating parallel structure between events and language. In *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 36–47). Cascadilla Press Somerville, MA.

Bresnan, J. (1978). A realistic transformational grammar. In M. Halle & G. A. Miller (Eds.), *Linguistic Theory and Psychological Reality* (pp. 1–59). Cambridge, MA: MIT Press.

Brown, R. (1973). *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.

Bunger, A., & Lidz, J. (2004). Syntactic bootstrapping and the internal structure of causative events. In *Proceedings of the 28th Annual Boston University Conference on Language Development* (p. 74). Cascadilla Press.

Bunger, A., & Lidz, J. (2008). Thematic relations as a cue to verb class: 2-year-olds distinguish unaccusatives from unergatives. *University of Pennsylvania Working Papers in Linguistics*, *14*(1), 4.

Büring, D. (2013). Syntax, information structure and prosody. In M. den Dikken (Ed.), *The Cambridge Handbook of Generative Syntax* (pp. 860–895). Cambridge University Press.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, *27*(6), 843–873.

Carruthers, P. (2018). Basic questions. *Mind & Language*, *33*(2), 130–147.

Cauvet, E., Limissuri, R., Millotte, S., Skoruppa, K., Cabrol, D., & Christophe, A. (2014). Function words constrain on-line recognition of verbs and nouns in French 18-month-olds. *Language Learning and Development*, *10*(1), 1–18.

Cheng, L. L.-S. (2003). Wh-in-situ. *Glot International*, *7*(4), 103–109.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1977). On Wh-Movement. In P. Culicover, A. Akmajian, & Wasow (Eds.), *Formal Syntax* (pp. 71–132). New York: Academic Press.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994, March). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *The Journal of the Acoustical Society of America*, *95*(3), 1570–1580.

Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of phonetics*, *31*(3-4), 585–598.

Christophe, A., Mehler, J., & Sebastin-Galls, N. (2001, July). Perception of Prosodic Boundary Correlates by Newborn Infants. *Infancy*, *2*(3), 385–394.

Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008, March). Bootstrapping Lexical and Syntactic Acquisition. *Language and Speech*, *51*(1-2), 61–75.

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American statistical association*, *83*(403), 596–610.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2004). *Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0).*

Cole, P., & Hermon, G. (1994). Is there LF wh-movement? *Linguistic inquiry*, *25*(2), 239–262.

Colombo, J., & Bundy, R. S. (1983). Infant response to auditory familiarity and novelty. *Infant Behavior and Development*, *6*(2-3), 305–311.

Comrie, B. (1989). *Language universals and linguistic typology: Syntax and morphology*. University of Chicago press.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers. In D. Dowty, D. Kartunnen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 94–128). Cambridge University Press.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*(2), 141–168.

Dautriche, I., Swingley, D., & Christophe, A. (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition*, *143*, 77–86.

Davis, H. (1997). Deep unaccusativuty and zero syntax in St'át'imcets. *Anuario del Seminario de Filologa Vasca "Julio de Urquijo"*, 55–96.

Davis, H. (2010). *A Teaching Grammar of Upper St'át'imcets*. ms., University of British Columbia and The Upper St'át'imc Language Culture and Education Society.

Davis, H., & Demirdache, H. (2000). On lexical verb meanings: Evidence from Salish. In C. Tenny & J. Pustejovsky (Eds.), *Events as grammatical objects* (pp. 97–142). CSLI Publications.

Dayley, J. (1981). *Tzutujil Grammar* (Dissertation). University of California, Berkeley.

De Carvalho, A., Crimon, C., Barrault, A., Trueswell, J., & Christophe, A. (under review). "Look! It is not a bamoule!" 18- and 24-month-olds can use negative sentences to constrain their interpretation of novel word meanings.

De Carvalho, A., Dautriche, I., & Christophe, A. (2016). Preschoolers use phrasal prosody online to constrain syntactic analysis. *Developmental science*, *19*(2), 235–250.

De Carvalho, A., Dautriche, I., Lin, I., & Christophe, A. (2017). Phrasal prosody constrains syntactic analysis in toddlers. *Cognition*, *163*, 67–79.

De Carvalho, A., He, A. X., Lidz, J., & Christophe, A. (2019). Prosody and function words cue the acquisition of word meanings in 18-month-old infants. *Psychological science*, *30*(3), 319–332.

Delle Luche, C., Durrant, S., Poltrock, S., & Floccia, C. (2015). A methodological investigation of the intermodal preferential looking paradigm: methods of analyses, picture selection and data rejection criteria. *Infant Behavior and Development*, *40*, 151–172.

DePaolis, R. A., Keren-Portnoy, T., & Vihman, M. (2016). Making sense of infant familiarity and novelty responses to words at lexical onset. *Frontiers in psychology*, *7*, 715.

Dink, J. W., & Ferguson, B. (2015). *eyetrackingR: An R library for eye-tracking data analysis.* Retrieved from `http://www.eyetrackingr.com`

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 547–619.

Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., & Morgan, J. L. (2013, June). Word-level information influences phonetic learning in adults and infants. *Cognition*, *127*(3), 427–438.

Fenson, L., Bates, E., Dale, P. S., Hartung, J. P., Pethick, S. J., Reilly, J. S., ... Thal, D. J. (1993). *The MacArthur communicative development inventories.* Singular Publishing Group.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The annals of statistics*, 209–230.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in childrens language processing* (pp. 97–135). Amsterdam: John Benjamins Publishing Company.

Figueroa, M., & Gerken, L. (2019). Experience with morphosyntactic paradigms allows toddlers to tacitly anticipate overregularized verb forms months before they produce them. *Cognition*, *191*, 103977.

Fillmore, C. J. (1968). The case for case. In E. Bach & R. Harms (Eds.), *Universals in Linguistic Theory.* New York, NY: Holt, Rinehart, & Winston.

Fillmore, C. J. (1970). The grammar of 'hitting' and 'breaking'. In R. A. Jacobs & P. S. Rosenbaum (Eds.), *Readings in English Transformational Grammar* (pp. 120–133). Waltham, MA: Ginn and Company.

Fillmore, C. J. (1986). Pragmatically controlled zero anaphora. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 12, pp. 95–107).

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 501–538.

Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive psychology*, *31*(1), 41–81.

Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(2), 143–149.

Fisher, C., Jin, K.-S., & Scott, R. M. (in press). The developmental origins of syntactic bootstrapping.

Fodor, J. D. (1998). Parsing to learn. *Journal of Psycholinguistic Research*, *27*(3), 339–374.

Fodor, J. D., & Frazier, L. (1980). Is the human sentence parsing mechanism an ATN? *Cognition*, *8*(4), 417–459.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*, *44*(03), 677–694.

Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and cognitive processes*, *4*(2), 93–126.

Frazier, L., & d'Arcais, G. B. F. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of memory and language*, *28*(3), 331–344.

Gagliardi, A. (2012). *Input and intake in acquisition* (Dissertation). University of Maryland.

Gagliardi, A., & Lidz, J. (2014). Statistical insensitivity in the acquisition of Tsez noun classes. *Language*, *90*(1), 58–89.

Gagliardi, A., Mease, T. M., & Lidz, J. (2016). Discontinuous development in the acquisition of filler-gap dependencies: Evidence from 15-and 20-month-olds. *Language Acquisition*, *23*(3), 1–27.

Geffen, S., & Mintz, T. H. (2015). Can You Believe It? 12-Month-Olds Use Word Order to Distinguish Between Declaratives and Polar Interrogatives. *Language Learning and Development*, *11*(3), 270–284.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*(6), 721–741.

Gerken, L., Jusczyk, P. W., & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition*, *51*(3), 237–265.

Gertner, Y., & Fisher, C. (2012). Predicted errors in childrens early sentence comprehension. *Cognition*, *124*(1), 85–94.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: abstract knowledge of word order in early sentence comprehension. *Psychological Science*, *17*(8), 684–691.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176.

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language acquisition*, *1*(1), 3–55.

Gleitman, L. R., Gleitman, H., Landau, B., & Wanner, E. (1988). Where learning begins: Initial representations for language learning. *Linguistics: the Cambridge survey*, *3*, 150–193.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in cognitive sciences*, *4*(5), 178–186.

Gordon, P. (2003). The origin of argument structure in infant event representations. In *Proceedings of the 28th Annual Boston University Conference on Language Development* (Vol. 1, pp. 189–198).

Gotowski, M. (2017). Wh-In Situ Production in Child French. *Linguistica Atlantica*, *36*(2).

Gotowski, M., & Becker, M. (2016). An information-structural account of childrens wh-in situ questions in French. In *Proceeding of the 6th Conference on Generative Approaches to Language Acquisition North America (GALANA 2015)* (pp. 24–35).

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language*,

51(4), 548–567.

Grimshaw, J. (1981). Form, function and the language acquisition device. In C. L. Baker & J. J. Mccarthy (Eds.), *The Logical Problem of Language Acquisition* (pp. 165–182). Cambridge, MA: MIT Press.

Gutman, A., Dautriche, I., Crabb, B., & Christophe, A. (2015). Bootstrapping the syntactic bootstrapper: probabilistic labeling of prosodic phrases. *Language Acquisition*, *22*(3), 285–309.

Hacquard, V., & Lidz, J. (2018). Children's attitude problems: Bootstrapping verb meaning from syntax and pragmatics. *Mind & Language*, *34*(1), 73–96.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109.

He, A. X. (2015). *Verb Learning Under Guidance* (Dissertation). University of Maryland.

He, A. X., & Lidz, J. (2017). Verb learning in 14-and 18-month-old English-learning infants. *Language Learning and Development*, 1–22.

Hicks, J., Maye, J., & Lidz, J. (2007). The role of function words in infants' syntactic categorization of novel words. In *Proceedings of the Linguistic Society of America Annual Meeting.* Anaheim, CA.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 105–124). Cambridge, MA: The MIT Press.

Höhle, B., Schmitz, M., Santelmann, L. M., & Weissenborn, J. (2006). The recognition of discontinuous verbal dependencies by German 19-month-olds: Evidence for lexical and structural influences on children's early processing capacities. *Language Learning and Development*, *2*(4), 277–300.

Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy*, *5*(3), 341–353.

Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, 251–299.

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development: An International Journal of Research and Practice*, *13*(4), 341–348.

Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Hudson Kam, C. L. H., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, *59*(1), 30–66.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in infancy research.*

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar.* MIT Press.

Jin, K.-S., & Fisher, C. (2014). Early evidence for syntactic bootstrapping: 15-month-olds use sentence structure in verb learning. In *Proceedings of the 38th Boston University Conference on Language Development.* Boston, MA: Cascadilla Press.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992, April). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*(2), 252–293.

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and philosophy*, *1*(1), 3–44.

Keenan, E. L. (1976). Towards a Universal Definition of 'Subject'. In C. Li (Ed.), *Syntax and Semantics: Subject and Topic.* New York: Academic Press.

Knowlton, T., Perkins, L., Williams, A., & Lidz, J. (2018). *Getting a grip on infants' event representations: Participant number in TAKE and PICK-UP* [Poster]. The XXI Biennial International Congress of Infant Studies (ICIS), Philadelphia.

Koenig, J.-P., & Mauner, G. (1999). A-definites and the discourse status of implicit arguments. *Journal of Semantics*, *16*(3), 207–236.

Kwiatkowski, T., Goldwater, S., Zettlemoyer, L., & Steedman, M. (2012). A probabilistic model of syntactic and semantic acquisition from child-directed utterances and their meanings. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 234–244).

Landau, B., & Gleitman, L. R. (1985). *Language and Experience: Evidence from the Blind Child.* Cambridge, MA: Harvard University Press.

Langacker, R. W. (1999). *Grammar and Conceptualization.* New York, NY: Walter de Gruyter.

Lasnik, H. (1989). On certain substitutes for negative data. In R. J. Matthews & W. Demopoulos (Eds.), *Learnability and linguistic theory* (pp. 89–105). Dordrecht: Kluwer.

Lee, J. N., & Naigles, L. R. (2005). The input to verb learning in Mandarin Chinese: a role for syntactic bootstrapping. *Developmental Psychology*, *41*(3), 529.

Lee, J. N., & Naigles, L. R. (2008). Mandarin learners use syntactic bootstrapping in verb acquisition. *Cognition*, *106*(2), 1028–1037.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation.* Chicago: University of Chicago press.

Levin, B., & Hovav, M. R. (2005). *Argument realization.* Cambridge: Cambridge University Press.

Lidz, J., & Gagliardi, A. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annual Review of Linguistics*, *1*(1), 333–353.

Lidz, J., & Gleitman, L. R. (2004a). Argument structure and the child's contribution to language learning. *Trends in cognitive sciences*, *8*(4), 157–161.

Lidz, J., & Gleitman, L. R. (2004b). Yes, we still need universal grammar. *Cognition*, *94*(1), 85–93.

Lidz, J., White, A. S., & Baier, R. (2017). The role of incremental parsing in syntactically conditioned word learning. *Cognitive Psychology*, *97*, 62–78.

MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.

Manning, C., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval.* Cambridge University Press.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, *164*(1), 177–190.

Maurits, L., Perfors, A., & Navarro, D. (2009). Joint acquisition of word order and word reference. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society.*

Maye, J., Werker, J. F., & Gerken, L. (2002, January). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), 101–111.

McCloskey, J. (2017). Resumption. In M. Everaert & H. C. Van Riemsdijk (Eds.), *The Wiley Blackwell Companion to Syntax, Second Edition* (pp. 1–30). John Wiley & Sons, Inc.

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental science*, *12*(3), 369–378.

Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental psychology*, *31*(5), 838.

Messenger, K., Yuan, S., & Fisher, C. (2015). Learning verb syntax via listening: New evidence from 22-month-olds. *Language Learning and Development*, *11*(4), 356–368.

Mintz, T. H. (2013). The segmentation of sub-lexical morphemes in English-learning 15-month-olds. *Frontiers in Psychology*, *4*(24).

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, *26*(4), 393–424.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, *96*(2), 143–182.

Morgan, J. L. (1986). *From simple input to complex grammar.* The MIT Press.

Morgan, J. L., & Demuth, K. (1996). Signal to syntax: An overview. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 1–22.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive psychology*, *19*(4), 498–550.

Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of verbal learning and verbal behavior*, *20*(1), 67–85.

Naigles, L. R. (1990). Children use syntax to learn verb meanings. *Journal of child language*, *17*(2), 357–374.

Naigles, L. R. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, *58*(2), 221–251.

Nazzi, T., Barrire, I., Goyet, L., Kresh, S., & Legendre, G. (2011). Tracking irregular morphophonological dependencies in natural language: Evidence from the acquisition of subject-verb agreement in French. *Cognition*, *120*(1), 119–135.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998, June). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(3), 756–766.

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: language input and acquisition.* Cambridge: Cambridge University Press.

Noble, C. H., Rowland, C. F., & Pine, J. M. (2011). Comprehension of Argument Structure and Semantic Roles: Evidence from English-Learning Children and the Forced-Choice Pointing Paradigm. *Cognitive science*, *35*(5), 963–982.

Omaki, A., Orita, N., & Lidz, J. (in prep). Beyond artificial language: Statistical learning of an English non-adjacent syntactic dependency.

Parisien, C., & Stevenson, S. (2010). Learning verb alternations in a usage-based Bayesian model. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 32).

Pearl, L., & Lidz, J. (2009). When Domain-General Learning Fails and When It Succeeds: Identifying the Contribution of Domain Specificity. *Language Learning and Development*, *5*(4), 235–265.

Pearl, L., & Sprouse, J. (2013). Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition*, *20*(1), 23–68.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2006). Poverty of the stimulus? A rational approach. In *Proceedings of the Cognitive Science Society* (Vol. 28).

Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of child language*, *37*(3), 607–642.

Perkins, L., Feldman, N. H., & Lidz, J. (2017). Learning an input filter for argument structure acquisition. In *Proceedings of the 7th Workshop on Cognitive Modeling and Computational Linguistics.*

Perkins, L., Knowlton, T., Williams, A., & Lidz, J. (2018). *Matching number vs. linking roles: Using 3-participant scene percepts to understand infants' bootstrapping* [Poster]. 43rd Boston University Conference on Language Development (BUCLD 43), Boston, MA.

Perkins, L., & Lidz, J. (under review). Filler-gap dependency comprehension at 15 months: The role of vocabulary.

Perry, J. W., Allen, K., & Berry, M. M. (1955). Machine literature searching x. Machine language; factors underlying its design and development. *American Documentation (pre-1986)*, *6*(4), 242.

Pinker, S. (1984). *Language Learnability and Language Development.* Cambridge, MA: Harvard University Press.

Pinker, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure.* Cambridge, MA: MIT Press.

Polinsky, M. (2017). Antipassive. In J. Coon, D. Massam, & L. d. Travis (Eds.), *The Oxford Handbook of Egrativity.* Oxford University Press.

Pozzan, L., Gleitman, L. R., & Trueswell, J. C. (2015). Semantic ambiguity and syntactic bootstrapping: The case of conjoined-subject intransitive sentences. *Language Learning and Development*.

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animateinanimate distinction. *Psychological bulletin*, *127*(2), 209.

Reinhart, T. (1998). Wh-in-situ in the framework of the Minimalist Program. *Natural language semantics*, *6*(1), 29–56.

Rizzi, L. (1986). Null objects in Italian and the theory of 'pro'. *Linguistic inquiry*, *17*(3), 501–558.

Ross, J. R. (1967). *Constraints on variables in syntax.* (Doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Rowland, C. F., Pine, J. M., Lieven, E. V., & Theakston, A. L. (2003). Determinants of acquisition order in wh-questions: Re-evaluating the role of caregiver speech. *Journal of Child Language*, *30*(3), 609–635.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*(4), 1144–1167.

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition*, *69*(2), 105–134.

Scott, R. M., & Fisher, C. (2009). Two-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and cognitive processes*, *24*(6), 777–803.

Seidl, A., Hollich, G., & Jusczyk, P. W. (2003). Early Understanding of Subject and Object Wh-Questions. *Infancy*, *4*(3), 423–436.

Shi, R., & Melanon, A. (2010). Syntactic Categorization in French-Learning Infants. *Infancy*, *15*(5), 517–533.

Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of child language*, *25*(01), 169–201.

Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy*, *10*(2), 187–198.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, *72*(2), B11–B21.

Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology*, *49*(4), 370–407.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1-2), 39–91.

Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, *35*(4), 869–902.

Soderstrom, M., Wexler, K., & Jusczyk, P. W. (2002). English-learning toddlers sensitivity to agreement morphology in receptive grammar. In *Proceedings*

of the 26th Annual Boston University conference on language development (Vol. 2, pp. 643–652). Somerville, MA: Cascadilla Press.

Soderstrom, M., White, K. S., Conwell, E., & Morgan, J. L. (2007). Receptive grammatical knowledge of familiar content words and inflection in 16-month-olds. *Infancy*, *12*(1), 1–29.

Stabler, E. P. (1998). Acquiring Languages with Movement. *Syntax*, *1*(1), 72–97.

Stromswold, K. (1995). The acquisition of subject and object wh-questions. *Language Acquisition*, *4*(1-2), 5–48.

Suppes, P. (1974). The semantics of children's language. *American Psychologist*, *29*(2), 103.

Sussman, R. S., & Sedivy, J. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, *18*(2), 143–163.

Suzuki, T., & Kobayashi, T. (2017). Syntactic Cues for Inferences about Causality in Language Acquisition: Evidence from an Argument-Drop Language. *Language Learning and Development*, *13*(1), 24–37.

Tincoff, R., Santelmann, L. M., & Jusczyk, P. W. (2000). Auxiliary verb learning and 18-month-olds' acquisition of morphological relationships. In *Proceedings of the 24th annual Boston University conference on language development* (Vol. 2, pp. 726–737). Cascadilla Somerville, MA.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454–475.

Valian, V. (1990). Logical and psychological constraints on the acquisition of syntax. In L. Frazier & J. G. De Villiers (Eds.), *Language Processing and Language Acquisition.* Dordrecht: Kluwer.

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition*, *40*(1-2), 21–81.

Van Heugten, M., & Shi, R. (2010). Infants sensitivity to non-adjacent dependencies across phonological phrase boundaries. *The Journal of the Acoustical Society of America*, *128*(5), EL223–EL228.

Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*, *113*(4), 569–586.

Wagner, S. H., & Sakovits, L. J. (1986). A process analysis of infant visual and cross-modal recognition memory: Implications for an amodal code. *Advances in infancy research.*

Wanner, E., & Gleitman, L. R. (1982). *Language acquisition: The state of the art.* CUP Archive.

Waxman, S. R., Lidz, J. L., Braun, I. E., & Lavin, T. (2009). Twenty four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, *59*(1), 67–95.

Wellwood, A., He, A. X., Lidz, J., & Williams, A. (2015). Participant structure in event perception: Towards the acquisition of implicitly 3-place predicates. *University of Pennsylvania Working Papers in Linguistics*, *21*(1), 32.

Wexler, K., & Culicover, P. (1980). *Formal principles of language acquisition.* Cambridge, MA.

Williams, A. (2015). *Arguments in syntax and semantics.* Cambridge: Cambridge University Press.

Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current directions in psychological science*, *18*(1), 53–57.

Yang, C. (2002). *Knowledge and learning in natural language.* Oxford: Oxford University Press.

Yuan, S., & Fisher, C. (2009, May). Really? She Blicked the Baby?: Two-Year-Olds Learn Combinatorial Facts About Verbs by Listening. *Psychological Science*, *20*(5), 619–626.

Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child development*, *83*(4), 1382–1399.