LEXICAL CATEGORY ACQUISITION VIA NONADJACENT DEPENDENCIES IN CONTEXT: EVIDENCE OF DEVELOPMENTAL CHANGE AND INDIVIDUAL DIFFERENCES

by

Michelle Sandoval

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PSYCHOLOGY

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2014

THE UNIVERSITY OF ARIZONA

GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Michelle Sandoval, titled Lexical Category Acquisition Via Nonadjacent Dependencies in Context: Evidence of Developmental Change and Individual Differences and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

____ Date: April 18, 2014

Rebecca Gómez

____ Date: April 18, 2014

LouAnn Gerken

____ Date: April 18, 2014

Cecile McKee

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_______________________________________ Date: April 18, 2014

Dissertation Director: Rebecca Gómez

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Michelle Sandoval

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Lexical categories like noun and verb are foundational to language acquisition, but these categories do not come neatly packaged for the infant language learner. Some have proposed that infants can begin to solve this problem by tracking the frequent nonadjacent word (or morpheme) contexts of these categories. However, nonadjacent relationships that frame categories contain reliable adjacent relationships making the type of context (adjacent or nonadjacent) used for category acquisition unclear. In addition, previous research suggests that infants show learning of adjacent dependencies earlier than learning of nonadjacent dependencies and that the learning of nonadjacent word relationships is affected by the intervening information (how informative it is and how familiar it is). Together these issues raise the question of whether the type of context used for category acquisition changes as a function of development. To address this question, infants ages 13, 15, and 18 months were exposed to an artificial language containing adjacent and nonadjacent information that predicted a category. Infants were then tested to determine whether they 1) detected the category using adjacent information 2) only detected the nonadjacent dependency, with no categorization, or 3) detected both the nonadjacent relationship and the category. The results showed high individual variability in the youngest age group with a gradual convergence towards detecting the category and the associated environments by 18 months. These findings suggest that both adjacent and nonadjacent information may be used at early stages in category acquisition. The results reveal a dynamic picture of how infants use distributional information for category acquisition and support a developmental shift consistent with previous infant studies examining dependencies between words.

CHAPTER 1

INTRODUCTION TO THE DISSERTATION

1.1 The Topic and Aim of this Dissertation

Lexical categories, such as noun and verb, are categories containing the units at the word level of language. They are a core part of language acquisition as they are essential to the formation and use of grammar. Specifically, they are instrumental in making the acquisition of grammar tractable as they reduce the learning problem from understanding the relationships between many words to understanding the relationships between a few categories. In turn, plugging words into these categories allows us to produce and understand an endless number of meanings. But how do infants acquire these categories when the words do not appear to be overtly or consistently marked with category features? In answer to this question language scientists have proposed that learners can use the phonological, morphological, and/or semantic properties of these categories and/or the categories' word/morpheme environments (Braine, 1987; Frigo & MacDonald, 1998; Gerken, Wilson, & Lewis, 2005; Gómez & Lakusta, 2004; Maratsos & Chalkley, 1980; Pinker, 1984; Mintz, 2003; Monaghan & Christiansen, 2008; Redington, Chater, & Finch, 1998; St Clair, Monaghan, & Christiansen, 2010). Corpus analyses and modeling studies have shown that several of these cues to category membership are available in the input to support learning, while experimental studies have shown that infants at varying developmental stages are sensitive to these cues (see Section 1.3). However, the empirical literature does not paint a complete picture of whether, when, or how the use of these cues changes over the course of acquisition for the infant language learner, leaving a gap between the theories and the phenomena these theories intend to explain. Theories of lexical category acquisition must consider whether, when, and how *infants* use these cues for the task at hand.

The aim of this dissertation is to contribute to a developmental picture of lexical category acquisition that takes into account the presence of multiple cues in the input and the infant's ability (or biases) for tracking these cues at different points in development. Although there are other bases for lexical categorization, in particular this dissertation examines the role of distributional information in early lexical category acquisition, namely the role of adjacent dependencies and nonadjacent word/morpheme contexts that frame categories when both are present and informative in the input.

In natural language the association between the category and its word/morpheme environment can typically be seen between categories and frequent function words/morphemes such as *the* occurring adjacent to nouns and *is_-ing* framing verbs. Artificial language studies mirror these patterns to investigate how these contexts are used for category acquisition by familiarizing infants to languages with aX phrases containing reliable adjacent dependencies (e.g., a-element reliably co-occurring with phonological cue) *or* to languages with aXb phrases containing reliable nonadjacent dependencies (a_b reliable).

Although several studies have investigated different facets of the use of distributional information for lexical category acquisition, we still do not have a clear picture of when infants use these different sources of information (adjacent vs. nonadjacent) or if nonadjacent relationships are a viable solution for infant learners. Part of the reason for this is that as illustrated in Table 1.1, nonadjacent dependencies that frame categories also contain adjacent dependencies, making the interpretations of studies examining these questions unclear (e.g., see Mintz, 2006).

Table 1.1

Predictive Relationships in Nonadjacent Dependencies that Frame Categories

	Nonadjacent information predicts category	Adjacent information predicts category
Artificial Language (aXb)	a_b↔X	a↔X, X↔b
Natural Language (isVERBing)	is_ing↔VERB	is↔VERB, VERB↔ing

Another reason is that the studies in which the interpretations are more clear are studies conducted with adults. If processing undergoes developmental change, adult data does not inform the question at hand. Indeed, with increases in processing, infants may move from detecting adjacent category contexts (aX) to considering nonadjacent contexts (aXb). The research described herein investigates *if* and *when* infants employ adjacent and nonadjacent information for category acquisition across development.

1.2 Dissertation Outline

In the following sections of this chapter I discuss two types of information available in the input argued to support category acquisition: 1) sound cues (acoustic, phonetic, and phonological) and 2) word or morpheme environments. In addition, I review two approaches to the lexical category acquisition problem, one suggesting word or morpheme environments (i.e., nonadjacent dependencies) as the entry point to lexical category acquisition and an alternative proposal that takes into account the role of multiple cues for categorization. In Chapter 2 I review evidence that speaks to 1) if and when, during development, infants are able to use these cues for category acquisition and 2) how learners process linguistic input that contains adjacent and nonadjacent relationships. In Chapter 3 I present my findings and show that developmental change occurs in infants' use of word environments for category acquisition. I also explore predictors of category acquisition, comparing the predictive power of language knowledge, age, and information processing variables. Finally, in Chapter 4 I conclude by integrating the results of this dissertation into the literature and discuss implications and future directions.

1.3 Cues that Distinguish Lexical Categories

Although categories do not come prelabeled for the infant learner, corpus analyses have found cues available in the input that may facilitate the acquisition of grammatical categories. Several of these studies have found speech/sound level characteristics and word/morpheme environments that while probabilistic, can predict category classification at levels greater than chance (see Sandoval, Gonzales, & Gómez [2012] for a review). The focus of the following sections is on phonological and word/morpheme environment cues that predict lexical categories like noun and verb.[1]

1.3.1 Word/Morpheme Environments

Lexical categories and function words/morphemes (e.g., inflections) do not appear to be ordered randomly in language and because of this, a learner could infer that words (or word stems) that are surrounded by the same context belong to the same category. For example, the morphosyntactic agreement between auxiliary and verb tense is a functional morpheme context that frames verbs as illustrated in 1a and 1b. When provided with this word/morpheme context

[1] Although this dissertation focuses on phonological and word/morpheme level cues, the role of semantics is not denied. For instance semantic information, like phonological information, could act as a correlated cue to category membership (as in Brain, 1987; for evidence of a semantic distinction between nouns and verbs see Sandhofer, Smith, & Luo [2000]). On the other hand, categorization via form based cues (e.g., phonological cues and word/morpheme environments) may play a more critical role at the beginning stages of category acquisition (e.g., see argument in Gómez & Lakusta [2004]).

information one could conclude that a novel word in the same context is a verb as illustrated in 1c.

1a. She **is** walk**ing**
1b. She **is** sing**ing**
1c. She **is** blick**ing**

But how reliable is this information and how useful is it for acquiring lexical categories? To test the utility of distributional word environments for category acquisition, Mintz (2003) analyzed several different English corpora to test how well frequent frames could classify nouns and verbs into distinct categories. Mintz (2003) defined *frequent frames* as “an ordered pair of words with any word intervening” (p. 95). Classification based on the 45 most frequent word frames in each corpus were exceedingly accurate, with word environments (e.g., *you*___*it*) predicting distinct category membership for nouns and verbs at over 90% accuracy (*accuracy* is the number of word pairs that were correctly grouped together divided by the number of word pairs that were correctly and incorrectly grouped together). Other analyses have revealed similar results and have shown that frequent frames are reliable predictors of lexical categories in several languages such as French, Spanish, Turkish, and German (Chemla, Mintz, Bernal, & Christophe, 2009; Wang, Höhle, Ketrez, Küntay, & Mintz, 2011; Weisleder & Waxman, 2010; but see Erkelens, 2009 and Stumper, Bannard, Lieven, & Tomasello, 2011). One of the factors that contributes to the high accuracy of the frame environment is the restrictive nature of the environment. However, this restrictive nature also results in several small categories (e.g., several verb categories) rather than one category that aligns with the target lexical category (i.e., low completeness). *Completeness* is the number of word pairs that were correctly grouped together divided by the total number of word pairs that were and that should have been grouped together.

Another type of word/morpheme environment is the adjacent word/morpheme context. Note that the *is_ing* example provided earlier contains adjacent environments that may be tracked in relation to the category. For instance, the inflection *-ing* on the word *blicking* may still lead one to infer *blick* is a verb. Adjacent word environments have higher completeness but this completeness comes at the expense of accuracy. In a corpus analysis using the 45 most frequent frames, in comparing adjacent and nonadjacent word environments as cues to category membership, St. Clair et al. (2010) found that because an adjacent distributional environment is not as restricted as a frame environment, more words can be classified as belonging to the same category: completeness for the frame based environments was approximately 19% of the noun/verb types in each corpus analyzed, whereas completeness for the adjacent environment was approximately 86%. Because this boost in input coverage comes at the expense of accuracy, with the adjacent environment categorizing words at 50% accuracy, St Clair et al. proposed that infants could gain the best of both worlds if they were able to use information from adjoining adjacent dependencies called "flexible frames." Specifically, the adjacent dependencies would be between a lexical element and the *preceding* frequent morpheme (e.g., *is*-*X*), and that same lexical element and the *following* frequent morpheme (e.g., *X*-*ing*). Indeed when St Clair et al. modeled categorization using flexible frames vs. fixed frames they found that flexible frames gained the accuracy of Mintz's frames (accuracy was at levels similar to frame categorization) and gained the coverage of adjacent environments (the number of words categorized exceeded the number of words categorized by frames).

While a flexible frame proposal may provide the best solution when the only cue at the learner's disposal is word/morpheme contexts, a number of studies have examined the acoustic distinctions between grammatical categories (e.g., Cassidy & Kelly, 1991; Kelly, 1992; Kelly &

Bock, 1998; Monaghan & Christiansen, 2008; Morgan, Shi, & Allopenna, 1996; Sereno & Jongman, 1990).

1.3.2 Sound Cues: Acoustic, Phonetic, and Phonological

There are several sound level cues that distinguish nouns and verbs in English and in other languages (e.g. Dutch, French, Japanese, Mandarin, and Turkish) (see Monaghan & Christiansen [2008] for a review). For instance in English, in contrast to verbs, nouns tend to have more syllables, be trochaically stressed, have word final voicing if the last phoneme is a consonant, have nasals, and have a greater proportion of back vowels and low vowels (Cassidy & Kelly, 1991; Kelly, 1992; Kelly & Bock, 1998; Sereno & Jongman, 1990; Monaghan, Chater, & Christiansen, 2005; Morgan et al., 1996) (see Table 1.2). Although these cues are listed as characteristics of English words, a subset of them apply to other languages as well, such as Dutch (Durieux & Gillis, 2001), French, Japanese (Monaghan et al., 2007), Mandarin, and Turkish (Shi, Morgan, & Allopena, 1998). Moreover, prosodic cues distinguishing nouns from verbs have been found in infant directed speech in English (Conwell & Morgan, 2012), French, (Shi & Moisan, 2008), and Mandarin (Li, Shi, & Hua, 2010) (e.g. duration of words, syllables, and vowels, and frequency changes within words). Finally, Farmer, Christiansen, and Monaghan (2006) have shown that the phonological properties of nouns and verbs lead to two separate clusters in phonological space.

Table 1.2

Table adapted from Monaghan and Christiansen (2008): Sound Cues that Distinguish Nouns and Verbs in English

Cue	Distinction	References
How long is the word in terms of:		
Number of phonemes?	Nouns > Verbs	(Kelly, 1992; Morgan, Shi, & Allopena, 1996)
Number of syllables?	Nouns > Verbs	(Kelly, 1992; Morgan et al., 1996; Cassidy & Kelly, 1991)
Does the first syllable receive stress?	Nouns > Verbs	(Kelly & Bock, 1988)
Is the stressed vowel a front vowel?	Nouns < Verbs	(Sereno & Jongman, 1990)
Are the vowels of the word:		
Front vowels?	Nouns < Verbs	(Monaghan, Chater, & Christiansen, 2005)
High vowels?	Nouns < Verbs	
Does the word end in a voiced consonant?	Nouns > Verbs	(Kelly, 1992)
What proportion of the consonants are nasals?	Nouns > Verbs	
Which consonants are more likely to occur in the word onset:		(Monaghan, Christiansen, & Chater, 2007)
Bilabials?	Nouns > Verbs	
Approximants?	Nouns < Verbs	
Are the consonants in the word velars?	Nouns < Verbs	

In addition to the presence of these sound level cues in language, very young infants show sensitivity to these differences. Newborns are sensitive to the acoustic cues that distinguish functional from lexical categories (Shi, Werker, & Morgan, 1999). And relevant to the problem addressed in this dissertation, 13-month-olds can discriminate between noun/verb homophones like *kiss* and *drink* using sound level differences alone as a basis for discrimination (Conwell & Morgan, 2012). Indeed, one of the challenges the infant begins to solve starting in the womb and continuing throughout out the first year of life, is learning the sound structure of their native language. If infants are already focused on these sound level cues and their importance to other linguistic tasks (e.g., segmentation), using these cues for category acquisition may be a logical next step.

In summary, there are several cues that could be used for lexical category acquisition. Although they are probabilistic, word/morpheme environments and sound cues appear to be reliable predictors of category membership. Below I review distributional proposals that posit different roles for these cues during category acquisition.

1.4 Distributional Accounts of Lexical Category Acquisition

Distributional-based proposals are proposals that view distributional properties as primary in the category acquisition process. These properties include but are not limited to the sound level and word/morpheme environments reviewed in Section 1.3. There are a few distributional-based proposals that have been put forth to explain how infants might acquire lexical categories (e.g., see Maratsos & Chalkley, 1980; Redington et al., 1998; Mintz, 2003; Monaghan & Christiansen, 2008). Many of these proposals suggest that infants can detect frequent elements such as functional morphemes. Using these frequent elements as a toe hold into the input, the infant can begin the process of detecting local dependency patterns (co-occurrences between words/morphemes and/or words/morphemes and other cues). By virtue of the fact that words/morphemes are not ordered randomly, this analysis should lead infants to form categories from the regular co-occurrence of particular categories around particular functional morphemes. This dissertation distinguishes between two types of distributional proposals: the first suggests that word/morpheme environments alone play a primary role during category acquisition (e.g., Mintz, 2003) and the second suggests that correlated cues, for instance the integration of word/morpheme environments with sound level cues, play a primary role during category acquisition (e.g., Monaghan & Christiansen, 2008).

One proposal, of the first type, relies on the learning of nonadjacent dependencies. Mintz (2003; 2006) proposes that infants are sensitive to the most frequent frames in their languages and argues that these relationships are used to categorize the intervening material at the outset of category acquisition. In languages with less reliable word order patterns, frequent frames might involve dependencies over inflections rather than words.

However, as mentioned previously, there exist other cues in the input and consistent with this observation, Monaghan and Christiansen's multiple-cue approach suggests that infants integrate all the reliable information they have at their disposal to form categories (Monaghan & Christiansen, 2004; 2008). Further, based on categorization studies with adults and infants, it has been argued that correlated cues not only facilitate categorization but are necessary for categorization and generalization to take place (Frigo & MacDonald,1998; Braine, 1987; Gerken et al., 2005; Gómez & Lakusta, 2004). In addition, Monaghan, Christiansen, and Chater (2007) have found that there is a balance in the reliability of sound-level information and word/morpheme environments such that cross-linguistically, when one type of cue is less reliable, the other is more reliable (also see Monaghan et al., 2005 for evidence that this balance is also present within languages for low and high frequency words). Finally, this approach suggests that categorization via sound level cues and adjacent word/morpheme environments is likely to precede categorization via fixed frame environments (St. Clair et al., 2010).

1.5 Summary

In the previous sections, I reviewed the corpus and modeling results showing that both the multiple cue integration (i.e., adjacently co-occurring word/morpheme contexts and acoustic/semantic-based cues) and nonadjacent-based approach are viable with respect to their reliability in the input. But which is more psychologically plausible as the entry point into category acquisition?

The nonadjacent dependencies that frame categories in Mintz's approach (Mintz, 2002; 2006), also contain adjacent dependencies; therefore the plausibility of this proposal relies heavily upon how infants process linguistic input that contains multiple cues such as reliable adjacent and nonadjacent relationships. Furthermore, the plausibility of both of these proposals relies upon if and when, during development, they are able to use these cues. In the following chapter, I review empirical evidence that suggests 1) infants integrate information across cues for category formation before they show evidence of learning nonadjacent dependencies and 2) learning of nonadjacent relationships is impacted by the intervening information.

CHAPTER 2

A REVIEW OF THE LITERATURE: CATEGORY ACQUISITION VIA ADJACENT & NONADJACENT WORD/MORPHEME CONTEXTS

By the time children are producing multi-word utterances, they seem to produce many of these utterances in a consistent fashion; for instance showing appropriate use of determiners in relation to nouns and noun subclasses such as *a* before singular nouns and *the* before count and mass nouns (Valian, 1986; Pine & Lieven, 1997). Some have argued that this reflects knowledge of syntactic categories and how these categories pattern in their language (Valian, 1986). Others assert that these utterances do not reflect any underlying syntactic competence, but rather emerge from the child's ability to track and imitate the frequencies of these patterns with respect to certain words (Pine & Lieven, 1997; Tomasello, 2000). Regardless of the theoretical interpretations, it seems clear that by 24 months of age children are at least aware of some of the most frequent distributional characteristics of the categories we call noun and verb. How do they come to acquire this knowledge? Below I review studies that speak to the two distribution-based proposals of lexical category acquisition discussed above, specifically studies relevant to adjacent and nonadjacent dependencies as entry points into category acquisition.

2.1 Adjacent Dependencies as a Cue for Category Acquisition

Gómez and Lakusta (2004) investigated learning of categories via adjacent dependencies when in addition to the adjacent context, infants are provided with a secondary cue to category membership. In this study 12-month-olds were familiarized to an artificial language with the structure aX bY, where a-words co-occurred with X-words and b-words co-occurred with Y-words. The secondary source of information or correlated cue was phonological in nature: in addition to being preceded by a specific word, X- and Y-words were distinguishable by syllable number (X words were monosyllabic and Y-words were disyllabic). Gómez and Lakusta (2004) found that 12-month-olds could use adjacent information to learn categories when 83% of the two-word phrases were consistently marked by the phonological cue. In addition, infants were able to associate the a- and b-words to the probabilistic phonological category cue on X and Y words and use this knowledge to generalize to aX bY pairings with novel one- and two-syllable X and Y category members.[2]

This evidence supports the multiple-cue integration account of lexical category acquisition as categorization required infants to integrate a syllable-number cue with the adjacent context. In addition, although corpus and modeling studies have shown that acoustic-phonetic and phonological information is probabilistic and thus does not perfectly predict lexical category membership (Farmer, Christiansen, & Monaghan, 2006; see also Monaghan & Christiansen, 2008 for a review), Gómez and Lakusta's study suggests that infants can use these cues for category acquisition even when they are not perfectly predictive. Moreover, there is evidence that 13-month-olds can distinguish noun and verb uses of the same word (e.g., *kiss* used a noun vs. *kiss* used a verb) based solely on the acoustic-phonetic differences of these two forms (Conwell & Morgan, 2012). Thus, the prerequisites for categorization via adjacent environments appear to be in place early on.

In addition, studies conducted using natural language suggest that infants use adjacent contexts to classify novel nouns but not verbs. By 14 to 16 months of age, German-exposed infants are able to detect when a novel word is placed in the incorrect syntactic position in continuous speech (i.e. a noun in a verb position) if during familiarization, the novel word

---

[2] Also see Gerken, Wilson, & Lewis (2005) for evidence that infants can use redundant morphophonological information to create word classes.

follows a determiner (Höhle, Weissenborn, Kiefer, Schulz, & Schmitz, 2004). Another study has shown that after Canadian-French learning 14-month-olds are familiarized to phrases containing a determiner followed by a novel word, they can discriminate between phrases where this novel word is preceded by a pronoun (ungrammatical) and phrases where the novel word is preceded by a different determiner not from familiarization (grammatical) (Shi & Melançon, 2010). Shi & Melançon (2010) also show that 14-month-olds do not use adjacent pronouns to classify novel verbs, a finding that they predicted given the low reliability of pronouns as indicators of verbs.

In summary, in artificial language studies learners are able to form categories using adjacent co-occurrence information if categories are signaled by correlated cues. Furthermore, the evidence suggests that infants are using adjacent word information for categorization in natural and artificial language studies sometime around 12 to 16 months of age (Gómez & Lakusta, 2004; Höhle et al., 2004; Shi & Melançon, 2010).

2.2 Nonadjacent Dependencies as Cues for Category Acquisition

Even if frequent nonadjacent word environments do not predict the category membership of most words in child directed speech, they are highly accurate predictors of the words that are classified (Mintz, 2003). Therefore, Mintz argues that they may be the starting point in lexical category acquisition. But before we can say infants use "frames" to learn lexical categories, we need to see that infants are able to learn frames (i.e. nonadjacent word/morpheme dependencies). In contrast to the findings on adjacent dependencies, sensitivity to nonadjacent dependencies at the word/morpheme level emerges sometime between 15 to 24 months, making the evidence regarding whether infants use frames before adjacent information for categorization seem unclear.

In one of the first nonadjacent dependency learning studies with infants, Santelmann and Jusczyk (1998) investigated the effects of infants' processing limitations on their detection of verb-tense agreement, specifically the relationship *is-X–ing*. In this study, 15- and 18-month-olds were presented with English sentences containing the grammatical relationship *is-X-ing* (e.g., *Grandma* ***is*** *sing****ing***) and sentences made ungrammatical by replacing *is* with the modal *can* (e.g., **Grandma* ***can*** *sing****ing***). Only 18-month-olds could track this relationship and only if the intervening material spanned three syllables or less (e.g., *Grandma* ***is*** *always sing****ing***, but not *Grandma* ***is*** *almost always sing****ing***). Santelmann and Jusczyk concluded that the processing of any more than three intervening elements interferes with the detection of nonadjacent relationships and thus the learning of these relationships is likely constrained by the infant's processing limitations (as suggested by Newport [1988; 1990]).

Several studies have subsequently shown the same general developmental timeline. Van Heugten and Johnson (2010) found that 24-, but not 17-month-old Dutch learning infants have learned the relationship between the definite article *het* and the diminutive *–je* as in ***het*** *hond****je*** (*the doggy*), but these same age infants still cannot detect violations between the definite article *de* and the plural marker *–en* as in ***de*** *hond****en*** (*the dogs*). There is also evidence that by 19 months, English learning infants track the dependencies in subject-verb agreement (e.g., ***A*** *team bake****s*** vs. ****A*** *team bake****Ø***) (Soderstrom, Wexler, & Jusczyk, 2002). Converging evidence is found with 17-month-olds tracking the singular/plural distinction in subject-verb agreement in French (Van Heugten & Shi, 2010) and with 18-month-olds tracking the singular/plural distinction in subject-irregular verb agreement, again in French (Nazzi, Barrière, Goyet, Kresh, & Legendre, 2011). However, 14-month-olds do not appear to track such relationships (Nazzi et al., 2011). Although there is evidence that may suggest nonadjacent dependency learning of subject-verb agreement and/or noun-phrase agreement in English learning 16-month-olds (Soderstrom et al., 2007), the two agreement types were presented in the same sentence and the experiment was not designed to test morphosyntactic dependency agreement uniquely.

Several of these studies have suggested that the type of information contained in the intervening elements may critically impact the detection of nonadjacent dependencies (Van Heugten & Johnson, 2010; Höhle et al., 2006; Soderstrom et al., 2007). For example, Höhle et al. have demonstrated that German 19-month-olds' detect the co-occurrence of an auxiliary and its corresponding past participle verb morphemes (e.g., discriminate sentences containing the grammatical co-occurrence of ***hat*** and ***ge***_heul_***t*** in _hat-X-geVERBt_ vs. ungrammatical co-occurrence _*_***kann*** _and_ ***ge***_heul_***t*** _in kann-X-geVERBt_). However, their ability to detect this nonadjacent relationship varies as a function of the type (not syllable length) of the intervening information. Specifically, they detect the _hat ge_-X-_t_ relationship when there are no intervening syllables between _hat_ and _ge_-, and when _hat_ and _ge_- occur across two intervening syllables that make a determiner phrase (e.g., across the DP _den Ball_ in ***hat*** _den Ball_ ***ge***_hol_***t***), but not across a two-syllable adverb (e.g., not across the adverb _leise_ in ***hat*** _leise_ ***ge***_quiek_***t***) (see Table 2.1 for a summary of their results).

Table 2.1
Summary of Höhle et al. (2006)

Intervening Element	Grammatical	Ungrammatical	Discrimination
None	*Das kleine unzufriedene Kind* <u>*hat ge-heul-t*</u> The little unhappy child has-AUX PST PTCP-cry-PST PTCP 'The little unhappy child has cried'	*Das kleine unzufriedene Kind kann geheult* The little unhappy child can-AUX PST PTCP-cry-PST PTCP 'The little unhappy child can cried'	Yes
Determiner Phrase (2 syllables)	*Das kleine phantasievolle Kind hat den Ball geholt* the little imaginative child has-AUX the ball PST PTCP-fetch-PST PTCP 'The little imaginative child has fetched the ball'	*Das kleine phantasievolle Kind kann den Ball geholt* the little imaginative child can-AUX the ball PST PTCP-fetch-PST PTCP ‘The little imaginative child can fetched the ball’	Yes
Adverb(2 syllables)	*Der Hamster hat leise gequiekt, weil er schlafen wollte* the hamster has-AUX softly-ADV PST PTCP-squeak-PST PTCP because it sleep wanted ‘The hamster has squeaked softly because it wanted to sleep’	*Der Hamster kann leise gequiekt, weil er schlafen wollte* the hamster can-AUX softly-ADV PST PTCP-squeak-PST PTCP because it sleep wanted ‘The hamster can squeaked softly because it wanted to sleep’	No

The difference in discrimination in their determiner phrase experiment and lack of discrimination in the adverb experiment cannot be explained by the distance between the dependent elements as this distance was constant (two syllables in both experiments). Instead, citing their study with 14-16-month-olds where infants show evidence of using determiners to categorize novel nouns (Höhle et al., 2004; reviewed earlier in Section 2.1), Höhle et al. argue that it is infants' familiarity with determiner phrases that allows them to track the nonadjacent dependency when it occurs over a determiner phrase. Infants may analyze this structure and treat it as one unit (DP) or their greater familiarity with specific determiners (*den*, *the* in English) may be reducing the processing load of the intervening information. Another interesting observation that Höhle et al. make is that the determiner phrases in the sentences provided were verb complements and thus, if 19-month-olds are able to connect the complement to the verb this nonadjacent dependency between the auxiliary and verb morphemes becomes an adjacent dependency between auxiliary and complement-verb (a↔[X↔b]).

The key message to take away from these studies is that the ability to track nonadjacent dependencies emerges after the age at which infants use adjacent information to form categories. In addition, this ability appears to be affected by processing constraints and the intervening information. Why might this be? One possibility is that infants face processing difficulties because the intervening elements are competing with the nonadjacent elements during processing. Researchers have addressed this issue of competition between adjacent and nonadjacent elements in studies manipulating the statistical predictability of adjacent vs. nonadjacent relationships (Gómez, 2002; Onnis, Christiansen, Chater, & Gómez, 2003; Onnis, Monaghan, Christiansen, & Chater, 2004; Gómez & Maye, 2005).

Gómez (2002) conducted a study with adults and 18-month-olds to examine the role of middle element variability in nonadjacent dependency learning. Participants listened to a made-up language with an aXb structure, where the nonadjacent elements (a_b) had a co-occurrence probability of 1 and the co-occurrence probabilities of the adjacent relations aX and Xb varied across conditions. Gómez manipulated the adjacent statistics by manipulating the variability of the middle elements; the X elements could either come from a set of 2, 6, 12, or 24 words for adults or from a set of 3, 12, or 24 words for infants. The condition with the largest set of middle elements was meant to mirror natural language where the X-element in constructions like *is-X-ing* comes from a large set of lexical items and the nonadjacent words come from a relatively small set (mimicking function morphemes). Gómez hypothesized that due to the low level of predictability of adjacent elements in the conditions where middle words came from a large pool, participants would ignore the adjacent dependencies and detect the invariance of the nonadjacent relationships. Adults learned best when the middle elements came from a large set (i.e., 24), obtaining a mean accuracy score of 90%. In contrast, they showed a comparatively low level of learning across the set sizes 2, 6, and 12 with accuracy scores of 60.5%, 66.5%, and 65.5%. Infants mirrored these results in showing discrimination for strings obeying the nonadjacent dependency vs. violating it after exposure to the large set size condition (set size 24), but not after exposure to set sizes 3 or 12. Gómez has since then replicated this finding and has shown a developmental shift similar to the one seen in sensitivity to nonadjacent morpheme relationships reported in studies using natural languages (Gómez & Maye, 2005). Gómez and Maye have shown that infants learn nonadjacent dependencies at 15 months, but fail to show learning at 12 months. Gómez has argued that reducing the statistical informativeness of adjacent relationships allows learners to find the reliable relationship among the nonadjacent units.

Given the general timeline of when infants become sensitive to nonadjacent dependencies, we should expect to see infants use this source of information sometime around 15 months or later. However, infants have shown categorization of novel words when these words are embedded in frame contexts as early as 12 months of age (Mintz, 2006).

In a study with 12-month-olds, Mintz (2006) demonstrated categorization of nonce words surrounded by frequent frames. Infants were first familiarized with noun and verb sentences containing various frequent frames surrounding two nonce nouns and two nonce verbs (e.g., the noun frame sentence *I see* ***the*** *gorp* ***in*** *the room* and the verb frame sentence *She wants* ***to*** *deeg* ***it***). After familiarization to these frame sentences infants discriminated between new grammatical sentences containing the learned nonce verbs and sentences made ungrammatical by placing the nonce verbs in noun frames (e.g., grammatical: ***I*** *deeg* ***you*** *now!* vs. ungrammatical: *I put* ***his*** *deeg* ***on*** *the box*). Because infants did not discriminate between sentences containing the novel nouns, it cannot be said that the infants learned the distributional regularities during the study, rather infants needed to have come to the lab knowing something about the verb frame environments and their connection to a category. It is unclear, however, whether the infants in this study discriminated based on frames or adjacent dependencies. As Mintz points out, to succeed in this task, infants could have been drawing on their knowledge of the two adjacent relationships in each frame, using *to-VERB* and *VERB-it* dependencies to categorize the novel words. Because Mintz aimed to show evidence of categorization via frames and not knowledge of the frames themselves, the ungrammatical test items violated both adjacent- and nonadjacent-category relationships. To understand this point take for example the adjacent relationships between auxiliaries and verbs and between verbs and the suffix *–ing*; if you know that *apple* is a noun and not a verb, the phrase *is-apple-ing* not only violates the connection between the nonadjacent relationship and the lexical category (i.e., the grammatical link between *is_ing* to *VERB*), it also violates two adjacent relationships (i.e., the grammatical links between *is-VERB* and *VERB-ing*). Therefore, it is possible that prior to participating in the study infants learned the frequent adjacent environments predictive of verbs in English and used this knowledge to succeed in this task. This interpretation of the findings is consistent with Christiansen and colleagues' (Monaghan & Christiansen, 2004; St Clair et al., 2010) proposal stating that categorization via adjacent contexts is likely to precede categorization via frames. In sum, we do not know whether the 12-month-olds in this study were bringing in their natural language knowledge of adjacent or nonadjacent dependencies to discriminate at test.

Studies with adults using artificial languages have also shown categorization of novel words when these words are embedded in frame contexts (Mintz, 2002; Vuong, Meyer, & Christiansen, 2011), but the methodology in these studies makes it unclear whether learning is the result of adjacent or nonadjacent dependency learning or both. To address this issue, Mintz (2011a) compared adults' categorization of intervening words when these words were preceded and followed by reliably co-occurring elements (i.e., two adjacent contexts that critically formed a frame) versus categorization of words that were preceded and followed by particular elements that did not form a frame. Mintz showed that adults were at chance in their discrimination of grammatical and ungrammatical utterances when the adjacent contexts did not reliably form a frame and above chance when the adjacent elements co-occurred to form a frame and argued that some aspect of the frame is necessary for categorization to take place. However, in the presentation of individual differences, some adults discriminated at levels above chance when adjacent dependencies did not form a reliable frame suggesting that there may be individual differences in the type of information used. This is consistent with adult data from Romberg and Saffran (2013) showing that when provided with three-word phrases that contain adjacent and nonadjacent dependencies, there are individual differences in the learning of these structures (although it is important to note that this experiment did not examine categorization).

Mintz's study (2011a) was conducted with adults and it did not employ the use of any phonological information that might allow more robust learning of adjacent dependencies, therefore a critical piece of the puzzle is still missing. We do not yet know what dependencies *infants* track when they are presented with informative adjacent and nonadjacent information or *whether* and *when* infants employ these different sources of information to form categories. If we are to provide a psychologically plausible theory of lexical category acquisition it is imperative we answer these questions. The present study attempts to address these questions. Before moving on to the study, Section 2.3 summarizes the general developmental trend present in the infant literature.

2.3 Summary: Placing the Findings on a Developmental Timeline

During the period of 12 to 14 months of age, infants begin to use adjacent word environments for categorization (Gómez & Lakusta, 2004; Höhle et al., 2004; Shi & Melançon, 2010), but they do not show evidence of learning nonadjacent dependencies or detecting nonadjacent dependency violations (Gómez & Maye, 2005; Nazzi et al, 2011; Santelmann & Jusczyk, 1998; see Figure 2.1). It is not until 15 months of age that we see evidence of nonadjacent dependency learning (Gómez & Maye, 2005). In natural language studies, infants' nonadjacent dependency violation detection is tenuous: during the period of 17 to 24 months of age we see detection of certain nonadjacent dependencies but not others (van Heugten & Johnson, 2010) and detection across certain intervening elements but not others (Santelmann & Jusczyk, 1998; Höhle et al., 2006). The prediction based on these findings is that infants should show developmental change in their use of adjacent and nonadjacent dependencies for category acquisition, a prediction that is most consistent with the multiple-cue integration story. The present study was designed to address this prediction and test Mintz's hypothesis that infants use nonadjacent dependencies at the beginning stages of category acquisition (Mintz, 2003; 2006; 2011b).

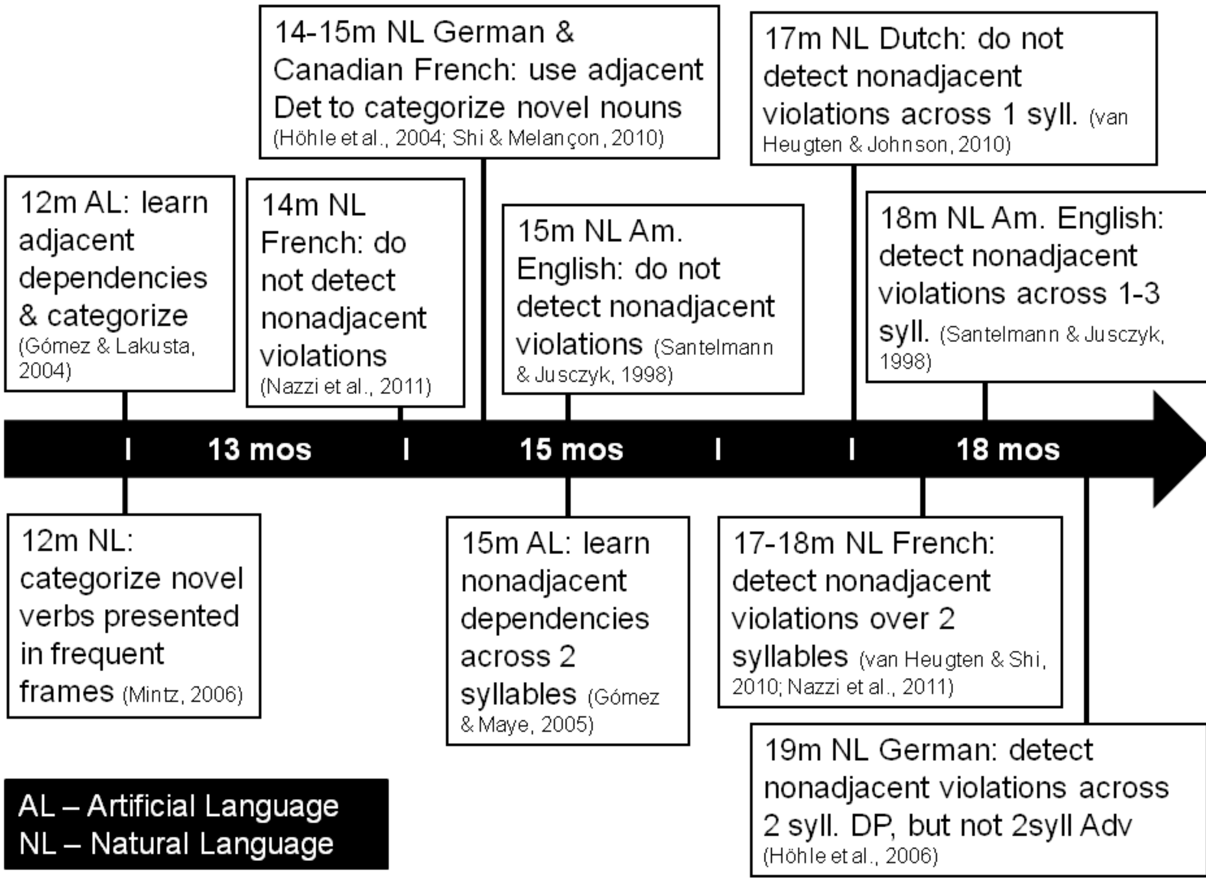

Figure 2.1. Developmental Timeline of the Findings on Adjacent and Nonadjacent Dependency Learning.

CHAPTER 3

THE USE OF DISTRIBUTIONAL CONTEXTS FOR CATEGORY ACQUISITION IN INFANCY

3.1 Methodological Choices

3.1.1 The Language

Language contains multiple cues that predict lexical categories (see Section 1.3 for a review). The present study was designed to investigate whether, when, or how the use of these cues changes over the course of category acquisition. Therefore, an artificial language was made of three-word phrases that contained both adjacent and nonadjacent dependencies where the relationship between the first and the last word was predictable (a_b), and the first and the last word each predicted the syllable-number of the intervening word (aXb, a predicts X and b predicts X). For instance, in aXb and eYf phrases *a* and *b* predict monosyllabic *X* words and *e* and *f* predict disyllabic *Y* words. Syllable-number, although probabilistic, is a cue that is present in natural language (Cassidy & Kelly, 1991). As mentioned previously, they are several phonological cues that distinguish nouns and verbs in natural language (Farmer, Christiansen, & Monaghan, 2006; see also Monaghan & Christiansen, 2008 for a review), but these other cues are probabilistic, typically predicting category membership at accuracy rates ranging from 60-90% (with this higher value being less common). In addition, 12-month-olds can learn adjacent dependencies under probabilistic conditions (Gómez & Lakusta, 2004). However, keeping the language as evenly informative across the dependency types was critical given that this study is the first to investigate infants' learning of adjacent and/or nonadjacent dependencies embedded in the same language. The language for the present study was thus designed with deterministic (one element predicting another with 100% consistency) rather than probabilistic dependencies.

Although this language does not reflect natural language as well as a probabilistic language would, if infants cannot learn one or more of these dependencies under these more optimal conditions, then they may not be able to use these dependencies in natural language given that more reliable co-occurrences are generally easier for infants to learn (Gerken et al., 2005; Gómez & Lakusta, 2004; Voloumanos, & Werker, 2009).

3.1.2 The Test

To assess whether infants were forming categories and to determine the type of information infants were learning across development, infants were provided with three test trials types: 1) trials that violated the adjacent dependencies, 2) trials that violated the nonadjacent dependencies, and 3) grammatical trials. A system was created based on the patterning of these three test types and this system was used to classify individual infants in the study as a) using adjacent information to form categories, b) learning nonadjacent dependencies but not categorizing, c) learning nonadjacent dependencies and categorizing, or d) not categorizing. More specifically, indiscriminate listening time to the three test types was interpreted as not categorizing and was labeled No Detection. Infants were labeled as Detect Adjacent Violation and Category learners if they treated Nonadjacent Violation and Grammatical trials similarly and Adjacent Violation trials as being different from the other two trial types. Similarly, infants were labeled Detect Nonadjacent Violation if they treated Adjacent Category Violation and Grammatical trials similarly and Nonadjacent Violation trials differently from the other two types. Finally, if infants treated Adjacent Category Violation and Nonadjacent Violation trials similarly, and Grammatical trials differently from the other two types, we concluded that infants were able to learn the nonadjacent dependency and the categories associated with the distributional environments (I elaborate on this description in Section 3.2 Method).

3.1.3 The Variables

Age. In the present study, infants ages 13, 15, and 18 months were habituated to an artificial language containing adjacent and nonadjacent information that predicted two categories. This age range was selected because given prior findings, older infants show fairly advanced category knowledge in their productions (e.g., Valian, 1986) and in their use of this knowledge for restricting word meanings (Bernal, Lidz, Millote, & Christophe, 2007), thus the findings with these younger infants should be most indicative of the mechanisms used during category acquisition. Valian (1986) has shown that by the time infants regularly combine words, they do so in a fashion that suggests they have acquired syntactic categories (Valian's sample ranged from 24 to 29 months of age). In addition, word learning studies have shown that beginning around 23 months of age, infants show evidence of using the sentential contexts of words to guide word learning (e.g., Bernal et al., 2007; Naigles, 1990; Robertson, Shi, & Melançon, 2012). Thus, given the aims of the current study it was important to assess infants before we see regular word combinations and before 23 months of age. Eighteen-month-olds were selected as our oldest age group was because they meet these requirements (i.e., less than 13% of 18-month-olds regularly combine words [Fenson et al., 1994]).

The specific age groups were selected because they correspond to previous studies examining adjacent and nonadjacent dependency learning in natural and artificial languages (e.g., Höhle et al., 2004; Gómez & Lakusta, 2004; Gómez & Maye, 2005; Santelmann & Jusczyk, 1998; Nazzi et al., 2011). Specifically, an age group between 12 to 14 months of age was selected because previous studies suggest that during this period infants can learn adjacent dependencies but not nonadjacent dependencies, and 18 months was selected because most of the studies show that infants can detect/learn nonadjacent dependencies by this age. Finally, the 15 month age group was selected because although 15-month-olds learn nonadjacent dependencies when categories are absent (Gómez & Maye, 2005), the natural language literature suggests that they cannot detect nonadjacent dependency violations until around 17 months of age (e.g., van Heugen & Shi, 2010). Although there are several possible reasons for why we see this difference between the natural and artificial language studies, one intriguing possibility is that the presence of reliable intervening categories in natural language dependencies disrupts learning or detection of nonadjacent dependencies at this age. This age group could thus speak to this possibility. In addition, if the predictions regarding 13- and 18-month-olds are supported, the 15-month-old age group could provide a look at the transition point between adjacent and nonadjacent dependency use for category acquisition. Examining these specific age groups, thus not only provided a skeleton onto which I could base specific predictions but also facilitates comparisons with prior studies in the literature.

Although there are individual differences in language development that may affect the uniformity of the results for each group, the studies on adjacent and nonadjacent dependency learning do not provide a reliable precedent for alternative predictors that would allow for fine-grained predictions regarding learning besides age. To the best of my knowledge, only Santelmann & Jusczyk (1998) have examined the relationship between language knowledge and nonadjacent dependency detection. They found greater discrimination between grammatical *is_ing* and ungrammatical *can_ing* sentences in 18-month-olds that produced word combinations than in 18-month-olds that did not produce word combinations. Because infants younger than 18 months rarely produce word combinations, this measure would not be informative for the present study. Therefore, age was selected as the organizing variable for the present study because the literature provided a background for more fine-grained predictions regarding this variable as opposed to other variables.

Vocabulary. Prior studies examining vocabulary size and grammatical development (e.g., Snedeker, Geren, & Shafto, 2007; 2012) do not suggest or lead to specific predictions regarding how infants should be distributed in the four learner types. However, these factors may be related to a subset or superset of the outcomes (e.g., whether or not infants categorize).

Specifically, according to several accounts of syntactic development, vocabulary is a precursor to the acquisition of grammar (Bates & Goodman, 1997; Tomasello, 2000; Snedeker et al., 2007). Indeed, studies with adopted children have shown that vocabulary size (as opposed to age, cognitive development, or other biological factors) correlates with the onset of grammatical development (Snedeker et al., 2007; 2012). In addition, vocabulary size is also correlated with speech processing (Fernald, Perfors, & Marchman, 2006). Given these findings there may be a difference in vocabulary size among categorizers and non-categorizers.

Habituation. The habituation paradigm is based on the observation that repetition in the stimulus leads to progressively less processing/attention, therefore it is possible that habituation time might be associated learning or learner type in the present study. Although Arterberry and Bornstein (2002) found that increases in habituation time decreased the odds of an infant being classified as a categorizer and habituation trials increased the odds, Gerken, Balcomb, and Minton (2011) found a different pattern of results. Gerken et al. (2011) found greater habituation time and habituation trials in infants that were exposed to a learnable version of a language as compared to infants exposed to an unlearnable version of the same language. Gerken et al. (2011) suggest that learner's have an internal metric of whether or not they are making progress towards solving a learning problem. Although all the infants in the present study were exposed to the same language, infants may differ in what they are able to extract from the input and this could potentially be reflected in their habituation times (e.g., some infants may not yet be able to track discontinuous relationships without first processing adjacent relationships). If this is the case, we might expect that increases in habituation would be predictive of the level or number of relationships the infant learned during habituation.

3.1.4 Analyses and Predictions

Developmental Change Prediction. To assess the question of whether there is developmental change in the type of information infants detect when frames predict categories, that reflects a change in the use of particular cues for category acquisition, the learning outcomes will be organized according to age and vocabulary quartile.

Consistent with the multiple-cue integration account and the literature reviewed in Chapter 2, three predictions were made:

1) A majority of the infants in the youngest age group (or with the smallest vocabulary size) should show evidence of using the adjacent information for category formation. This result would suggest that infants rely on adjacent dependency information to form categories before they rely on nonadjacent dependencies.

2) Because previous studies have shown that 18-month-olds are capable of learning nonadjacent dependencies (Gomez, 2002), a majority of the infants in the oldest group (or with the highest vocabulary size) should show sensitivity to the nonadjacent dependencies and the category.

3) Infants in the 15-month-old group (or in the middle vocabulary size quartiles) should show more individual differences in learner type than either of the other two age groups

as would be expected by a period of developmental transition (and for possibly for the reasons discussed earlier in Section 3.1.3).

Given this last prediction and the number of learnable patterns in the input, it was important to examine individual infants. Research with adults has shown that there individual differences in whether learners track adjacent or nonadjacent dependencies when both are present and informative in the input (Romberg & Saffran, 2013). If infants also show variability in learner type, the specific group predictions would not to be supported, but the data from individual infants would still be informative if they followed a general developmental trend.

It should be noted that my specific predictions regarding developmental change are in contrast to the frequent frame prediction. If infants first learn nonadjacent dependencies (or frames) and use these dependencies to learn categories (e.g., as suggested by Mintz, 2006), then given this experimental design, infants should show uniformity across age and/or levels of vocabulary in their ability to connect nonadjacent dependencies to their respective categories. A change (across age groups or vocabulary) going from nonadjacent dependency violation detection to nonadjacent and category violation detection would also be consistent with this hypothesis. However, evidence of categorization via adjacent dependencies would be inconsistent with this story.

Predictors of Learning Outcome. Using logistic regression, the final section in this chapter explores age, vocabulary size and habituation as potential predictors of learning outcome. Because the present study was not organized or designed around vocabulary or habituation (trials and time), findings from these analyses will need to be expanded upon by future research. Although these analyses are largely exploratory, given the correlations between vocabulary and grammatical development and vocabulary and processing efficiency, I predict that vocabulary size will increase the odds of an infant being classified as a categorizer. In addition, given Gerken et al.'s (2011) results, habituation trials and time may be predictors of level of categorization (or level of learning).

3.2 Method

3.2.1 Participants

In order to examine developmental changes in the type of distributional information used for category abstraction 13-month-olds, 15-month-olds, and 18-month-olds were tested. Each infant participated in the study only once. Infants were recruited from the Tucson metropolitan area. Only infants who met the following inclusion criteria were included in the final data set: weighed 5.5 pounds or more at birth with a term of 36 weeks or more, exposed to a weekly language environment that was at least 70% American-English, completed at least two trials for each test stimulus type, had no family history of speech or language impairment, and successfully habituated to the training materials (see Procedure for details regarding the habituation criterion).

13-month-olds. Twenty-one 13-month-old infants (10 female) with a mean age of 13 months, 7 days (SD = 14 days; range: 12 months, 8 days to 14 months, 6 days) were tested. Eleven other 13-month-olds were tested but were excluded from the final data set for the following reasons: weighing less than 5.5 pounds at birth or born before 36 weeks gestation (n = 1), exposed to less than 70% English in their daily environment (n = 3), excessive fussiness (n = 4), family history of speech or language impairment (n = 1), failure to habituate (n = 1), and technical difficulties (n = 1).

15-month-olds. Twenty 15-month-old infants (10 female) with a mean age of 15 months, 2 days (SD = 17 days; range: 14 months, 3 days to 16 months, 14 days) were tested. Twelve other infants were tested but were excluded from the final data set for the following reasons: exposed to less than 70% English in their daily environment (n = 1), excessive fussiness (n = 4), family history of speech or language impairment (n = 3), failure to habituate (n = 1), falling asleep during the study (n = 1), and caregiver affecting test completion (n = 2).

18-month-olds. Twenty 18-month-old infants (10 female) with a mean age of 18 months, 5 days (SD = 10 days; range: 17 months, 17 days to 18 months, 28 days) were tested. Eighteen other infants were tested but were excluded from the final data set for the following reasons: weighing less than 5.5 pounds at birth or born before 36 weeks gestation (n = 1), exposed to less than 70% English in their daily environment (n = 5), less than two data points per trial type (n = 1), excessive fussiness (n = 9), family history of speech or language impairment (n = 1), and excessive squirming and attempts to leave caregiver's lap (n = 1).

3.2.2 Stimulus Materials

To determine whether infants use adjacent and/or nonadjacent information to form categories, we created a language with three-word phrases (aXb) that contained both adjacent and nonadjacent dependencies: the relationship between the first and the last word was predictable (a↔b), and the first and the last word each predicted the middle word in the phrase (a↔X and X↔b). In addition we created three test types to assess whether infants in our study were a) using adjacent information to form categories, b) learning nonadjacent dependencies but not categorizing, c) learning nonadjacent dependencies and categorizing, or d) not learning. In order to keep the adjacent and nonadjacent information on a relatively even playing field, various factors that have been shown to facilitate learning of these dependencies were incorporated into the training language. To facilitate nonadjacent dependency learning the language contained high variability of the middle element (Gómez, 2002); assuming infants detect the nonadjacent relationships first, the variability of the elements from the different categories should be high (i.e., set-size=24 across nonadjacent word environments). To test generalization and facilitate adjacent dependency learning intervening words were distinguishable by syllable number (mono- and disyllabic words were used) and reliably co-occurred in specific word environments (e.g., pel_rud and vot_jic with monosyllablic words and fim_sog and dak_tup with disyllablic words). As mentioned previously, phonological differences (such as differences in syllable number) and acoustic phonetic differences between nouns and verbs have been documented (see Monaghan & Christiansen, 2008 for a review). Not only do these differences lead nouns and verbs to pattern differently in phonological space (Farmer, Christiansen, & Monaghan, 2006), but 13-month-olds can discriminate between nouns and verbs using speech-based differences alone as a basis for discrimination (Conwell & Morgan, 2012). Moreover, Gómez & Lakusta (2004) found that infants as young as 12 months learn categories using adjacent word contexts such as in aX bY if the Xs and Ys are distinguishable by syllable number. Thus, even though infants younger than 15 months have failed to show sensitivity to nonadjacent dependencies in previous studies, even the youngest infants in this study should be able to succeed at test when provided with this language.

To create the materials, a female speaker of American-English recorded the phrases in infant-directed speech. Using Sound Studio software (version 3.5.5), the phrases were deconstructed and the best exemplars of each word were used to create new phrases. The silence between words within a phrase varied to maintain a similar duration and rhythm across phrases. Note that inserting this silence between the word/morpheme elements in aXb phrases is also thought to facilitate nonadjacent dependency learning (Peña, Bonatti, Nespor, & Mehler, 2002). Because this silence is present, I refer to these elements as words. Phrases were separated by .75 seconds of silence and were approximately 2.75 seconds long.

Training. Two versions of the training language were created, Language 1 and Language 2 (hereafter L1 and L2). L1 took the form aXb, cXd, eYf, gYh, while L2 took the form aYb, cYd, eXf, gXh. The languages contained the same 32 unique words strung together into three-word phrases. The first and the third word of each phrase formed a nonadjacent dependency and this dependency framed a word belonging to a larger category. Four words occurred in first position, four occurred in last position and 24 occurred in the middle position. Twelve of the 24 words were monosyllabic (X words) and the other 12 were disyllabic (Y words). The four nonadjacent dependencies were: 1) fim_sog, 2) dak_tup, 3) pel_rud, and 4) vot_jic.

The sole difference between L1 and L2 was in the assignment of the intervening categorical elements. In L1, the dependencies fim_sog and dak_tup framed monosyllabic words and pel_rud and vot_jic framed disyllabic words. This syllable-frame relationship was reversed in L2 such that fim_sog and dak_tup framed disyllabic words and pel_rud and vot_jic framed monosyllabic words. This resulted in 48 unique phrases for each language (see Table 3.1 for example training phrases).

Table 3.1

Example Training Phrases from Language 1 and Language 2

L1						
a $X_{1\text{-}12}$ b	*fim* ***hos*** *sog*	*fim* ***tam*** *sog*	*fim* ***bap*** *sog*	*fim* ***kif*** *sog*	...	a X_{12} b
c $X_{1\text{-}12}$ d	*dak* ***hos*** *tup*	*dak* ***tam*** *tup*	*dak* ***bap*** *tup*	*dak* ***kif*** *tup*	...	c X_{12} d
e $Y_{1\text{-}12}$ f	*pel* ***loga*** *rud*	*pel* ***feenam*** *rud*	*pel* ***malsig*** *rud*	*pel* ***roosa*** *rud*	...	eY_{12} f
g $Y_{1\text{-}12}$ h	*vot* ***loga*** *jic*	*vot* ***feenam*** *jic*	*vot* ***malsig*** *jic*	*vot* ***roosa*** *jic*	...	g Y_{12} h
L2						
a $Y_{1\text{-}12}$ b	*fim* ***loga*** *sog*	*fim* ***feenam*** *sog*	*fim* ***malsig*** *sog*	*fim* ***roosa*** *sog*	...	a Y_{12} b
c $Y_{1\text{-}12}$ d	*dak* ***loga*** *tup*	*dak* ***feenam*** *tup*	*dak* ***malsig*** *tup*	*dak* ***roosa*** *tup*	...	c Y_{12} d
e $X_{1\text{-}12}$ f	*pel* ***hos*** *rud*	*pel* ***tam*** *rud*	*pel* ***bap*** *rud*	*pel* ***kif*** *rud*	...	e X_{12} f
g $X_{1\text{-}12}$ h	*vot* ***hos*** *jic*	*vot* ***tam*** *jic*	*vot* ***bap*** *jic*	*vot* ***kif*** *jic*	...	g X_{12} h

Test. Three different test trial types were created: 1) Adjacent Violation, 2) Nonadjacent Violation, and 3) Grammatical. For all test trial types the X and Y words from training were replaced with 12 novel X and 12 novel Y words to test category formation and generalization. This resulted in 48 unique phrases for each test trial type. To describe the test trial types, tests used for L1 will be used as an example but versions consistent with testing L2 were also created. See Table 3.2 for a summary of structural and phrase differences between training and test materials for L1.

***Adjacent Violation*.** Adjacent Violation trials were created to test adjacent dependency learning. Infants trained on L1: aXb, cXd, eYf, gYf received Adjacent Violation phrases with the structure aYb, cYd, eXf, gXh. This test type preserves the nonadjacent dependency learned during training (e.g., a_b, c_d, e_f, g_h), but violates the association between the category and the distributional environment (e.g., a- and -b no longer correspond to X). If infants were able to learn the category and the distributional environment it occurred in, hearing this test item would be analogous to an adult hearing *is APPLE-ing.* Critically the Xs and Ys were novel at test, thus a better analogy would be is-$\text{NOUN}_{\text{NOVEL WORD}}$-ing, with the novel word possessing phonological properties that made it sound noun-like.

***Nonadjacent Violation*.** Nonadjacent Violation trials were created to determine if infants were learning only the nonadjacent dependency during training. Infants trained on L1: aXb, cXd, eYf, gYf received Nonadjacent Violation phrases with the structure aXd, cXb, eYh, gYf. These phrases thus preserved the association between the category and its adjacent environments (e.g., aX, Xb, cX, Xd, eY, Yf, gX, Xh), but violated the nonadjacent dependencies from training (e.g., a_d did not occur during L1 training). If infants learned the nonadjacent dependency during training, hearing this test item would be similar to hearing *can SING-ing*.

Grammatical. Grammatical trials followed the same structure as those heard during training and thus did not violate any relationships. If infants learned the categories and the distributional environments of those categories, then hearing this item would be similar to hearing *is_ing* framing a novel verb with similar sound properties as other English verbs even though the items contained novel middle words.

Table 3.2

Phrase Examples and Structure of Language 1 Training and Test Stimuli

Stimulus Type	Structure	Example Phrase
Train	$aX_{1\text{-}12}$ b, $cX_{1\text{-}12}$ d, $eY_{1\text{-}12}$ f, $gY_{1\text{-}12}$ h	*pel* ***loga*** *rud*
Test		
Adjacent Violation	$aY_{13\text{-}24}$ b, $cY_{13\text{-}24}$ d, $eX_{13\text{-}24}$f, $gX_{13\text{-}24}$ h	*pel* ***nim*** *rud*
Nonadjacent Violation	$aX_{13\text{-}24}$ d, $cX_{13\text{-}24}$ b, $eY_{13\text{-}24}$h, $gY_{13\text{-}24}$ f	*pel* ***kaspit*** *jic*
Generalization	$aX_{13\text{-}24}$ b, $cX_{13\text{-}24}$ d, $eY_{13\text{-}24}$f, $gY_{13\text{-}24}$ h	*pel* ***kaspit*** *rud*

3.2.3 Procedure

The experiment was conducted at the Child Cognition Lab at the University of Arizona. Infants sat in a sound-proof booth on their caregiver's lap for the duration of the experiment. Caregivers listened to masking music through headphones and were instructed to refrain from influencing their infant's reactions. The experimenter sat outside the booth and watched the infant via a TV monitor while tracking the infant's looking behavior using Habit X software (Cohen, Atkinson, & Chaput, 2004). Each infant was trained on a single language (L1 or L2) and then tested with the corresponding test types. Training and test phases consisted of 60-second trials. Each trial began with a bullseye image on the screen. Once the infant fixated on the bullseye, the experimenter initiated the start of a trial. During trials, a cartoon jumping baby was projected on the screen and the audio stimuli were played until the infant looked away for 2 consecutive seconds or until the trial timed out at 60 seconds. At this point the trial ended and the bullseye image reappeared. The program immediately replayed trials with a fixation duration of less than 1 second: this ensured that every infant heard at least one full phrase before the experiment moved onto the next trial.

Training phase. There were four unique training trials for each language; each a pseudo-randomized subset of the 48 training phrases. The program was instructed to present the trials randomly with the constraint that every set of four trials contain one of four unique training trials. Training continued until the infant habituated as determined by the habituation criterion. The habituation criterion was set at 50% of the sum of the infant's listening time to the first three trials. Training ended when the sum of three consecutive trials fell below this criterion.

Test phase. After habituation, infants were exposed to up to nine test trials (three of each test type). Each trial was a pseudo-randomized subset of the 48 phrases of each test type. Trials were split into blocks of three trials with each block containing one of each of the test trial types. Only infants who completed two blocks of test or more were included in the final data set.

3.3 Results and Discussion

3.3.1 Classification System

We designed the test stimuli to reveal four different learning outcomes: 1) No Detection as shown by indiscriminate listening to the three test types, 2) Detect Adjacent Category Violations as shown by discrimination of Adjacent Violation trials from other test trials, 3) Detect Nonadjacent Violations as shown by discrimination of Nonadjacent Violation trials from other test trials, and 4) Detect Nonadjacent and Category Violations as shown by discrimination of Grammatical trials from other test trials. See Figure 3.1 for ideal examples of the different learning outcomes.

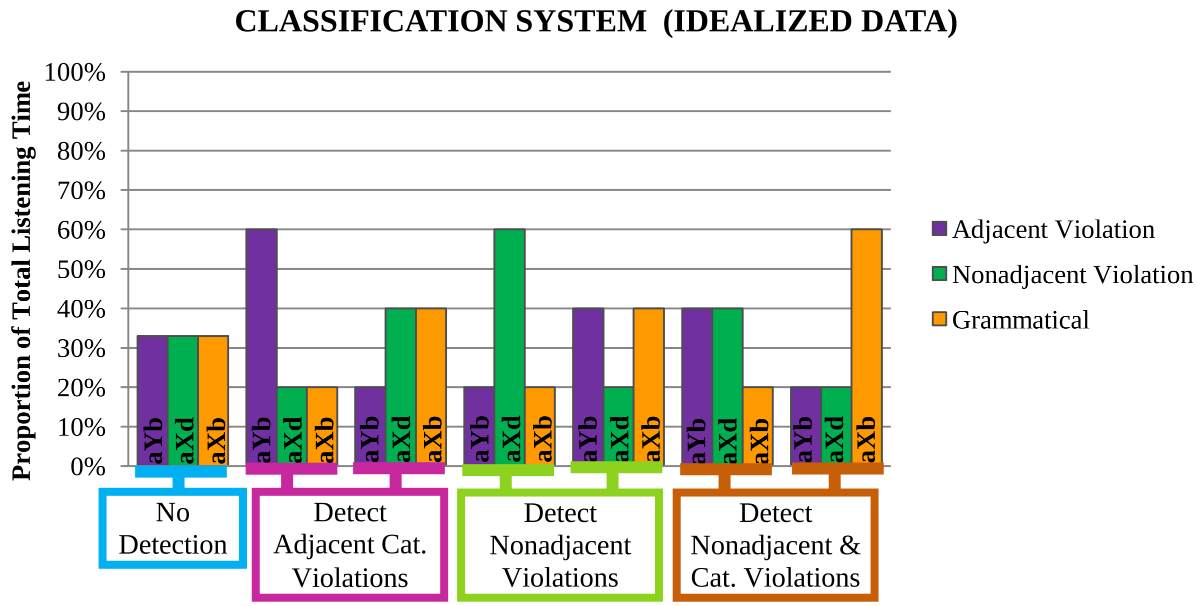

Figure 3.1. Idealized Data Showing the Four Different Learning Outcomes: 1) No Detection, 2) Detect Adjacent Category Violations, 3) Detect Nonadjacent Violations, and 4) Detect Nonadjacent and Category Violations. The colored bars indicate trial type. Example stimulus structures are provided for L1 for comparison of the different test trial types.

No Detection (Outcome 1). If infants failed to learn the nonadjacent dependency and/or failed to associate the category with the distributional environment (adjacent and/or nonadjacent), infants should show indiscriminate listening time to the three test types operationalized as statistically equal listening to the three types. We should also see a No Detection pattern if infants learned adjacent relationships between specific words during training, because there were novel X- and Y-words at test.

Detect Adjacent Category Violations (Outcome 2). If infants treated Nonadjacent Violation and Grammatical trials similarly and Adjacent Violation trials as being statistically different, we concluded that infants were only detecting the violation of the adjacent relationship. In other words, for an L1 learner percentage of total listening time to aYb (Adjacent Violation) should not be the same as the percentage to aXd (Nonadjacent Violation) or aXb (Grammatical) and the percentage of total listening time to aXd (Nonadjacent Violation) and aXb (Grammatical) should be similar. Nonadjacent Violation trials and Grammatical trials display legal adjacent relationships between the first (or last) word in the phrase and the category feature of the middle word. Therefore, infants displaying this listening pattern are determined to have learned the category using the adjacent information in the input.

Detect Nonadjacent Violations (Outcome 3). If infants treated Adjacent Category Violation and Grammatical trials similarly and Nonadjacent Violation trials as being statistically different, we concluded that infants were only detecting the violation of the nonadjacent relationship. For an L1 learner percentage of total listening time to aXd (Nonadjacent Violation) should not be the same as the percentage to aYb (Adjacent Violation) or aXb (Grammatical), and the percentage of total listening time to aYb (Adjacent Violation) and aXb (Grammatical) should be similar. Note that unlike the Nonadjacent Violation, Adjacent Violation trials and Grammatical trials display legal nonadjacent relationships between the first and the last word in the phrase (also note the category difference that is being ignored in the Adjacent Violation test type). In sum, infants showing this pattern of listening are learning the nonadjacent dependencies but have not learned that they predict the intervening categories.

Detect Nonadjacent and Category Violations (Outcome 4). If infants treated Adjacent Category Violation and Nonadjacent Violation similarly and Grammatical trials as statistically different, we concluded that infants were able to learn the distributional environments and the categories associated with these environments. For an L1 learner, the percentage of total listening time to aXb (Grammatical) should not be the same as the percentage to aYb (Adjacent Violation) or aXd (Nonadjacent Violation), and the percentage of total listening time to aYb

(Adjacent Violation) and aXd (Nonadjacent Violation) should be similar. Note that our current test does not reveal whether infants in this classification are forming the link between 1) a category and adjacent environment or 2) a category and nonadjacent environment, but it is clear that they learned the nonadjacent dependencies (a_b, c_d, e_f, g_h) *and* acquired the category and its associated distributional environment because aXd contains a nonadjacent violation and aYb contains a categorical violation. In other words, although it is clear that infants showing this pattern of listening are categorizing above and beyond the Detect Adjacent Category Violations infants, it is not clear whether infants in this group used the adjacent information to form the category and arrived at the frame knowledge (a_b) by forming a higher order dependency between adjacent dependencies (aX, Xb) or whether their performance is driven by nonadjacent frames. We return to this issue in the General Discussion.

3.3.2 Determining Chance vs. True Discrimination in Individual Infants

One of the primary predictions for this study was that we would see variability across development and across individual infants in the type of distributional information used for category acquisition. In order to assess the infants on the same scale, instead of using raw scores as the dependent variable we used the proportion of total listening time to a given test type. The proportion of total listening time was calculated by taking the sum of an infant's listening time to a given test type and dividing it by the sum of the infant's total listening time across test trials. We also needed a method to display and detect learning for individual infants. We assessed infants using a criterion similar to one developed by Arterberry and Bornstein (2002). Their criterion assesses whether a given infant is displaying chance versus true discrimination. Below is an excerpt from their article on infant categorization describing the formula and its application.

> In order to classify individual infants as categorizers, we identified a categorization criterion based on novelty preference [NP] scores. To determine a novelty preference (NP) that was significantly above chance (.50), we used a medium effect size of .50 (Cohen, 1988) and the standard deviation (SD) of the NP for the group. (Arterberry & Bornstein, 2002, pp. 522)

Thus, the measure they used is as follows:

$Effect\ size = (NP_{above\ chance} - NP_{at\ chance})/SD$ (Arterberry & Bornstein, 2002).

In their study, Novelty Preference (NP) was calculated by dividing total looking time to the novel stimulus trials by total looking time to all test trials:

$$NP = Looking\ Time_{Novel\ Stimulus} / Looking\ Time_{Total}$$

However, our test contains three test types and "novelty" is not a measure of learning in our study. We therefore modified the Arterberry and Bornstein formula to fit our experimental design.

1. We changed chance from .5 to .33 given that we had three test trial types and not two.
2. Because we did not predict novelty scores for a given test type we had to go beyond analyzing scores to one test trial type and instead we had to analyze scores for each trial type. Thus, instead of NP we calculated Proportion of Total Listening Time to Adjacent Violation trials, Nonadjacent Violation trials, and Grammatical trials by dividing the infant's total listening time to a given stimulus type by their total listening time to all test trials. For example, $\text{Proportion of Total Listening Time}_{\text{Adj Violation}} = \text{Looking Time}_{\text{Adj Violation}} / \text{Looking Time}_{\text{Total}}$.
3. Related to novelty, we predicted patterns that were irrespective of novelty (i.e., same, same, different) thus using the standard deviation of the group's novelty scores was

> not as appropriate as using the standard deviation of an individual infant's scores to a given trial type. Using the SD of the group's novelty scores (or in our case, scores for each trial type) introduces variability associated with novelty/familiarity effects. For example, two infants who show the same learning pattern can show this pattern with a novelty effect or a familiarity effect (see Figure 3.1). And even if the absolute differences were equal, the group SD would be affected more so than if the infants both showed a novelty effect. Given that our learning predictions go beyond novelty/familiarity effects we decided to use the SD for each infant for each trial type.

Therefore, to determine the Above Chance Criterion for each infant for each trial type we used the following formula with Effect Size set at .5, Preference At Chance set at .33, and SD as the standard deviation for each infant for each trial type:

$$\textit{Effect Size} = (\textit{Proportion of Total Listening Time}_{\textit{Above Chance}} - \textit{Proportion of Total Listening Time}_{\textit{At Chance}}) / \textit{SD}$$

or solved for *Proportion of Total Listening Time* $_{\textit{Above Chance}}$...

$$\textit{Proportion of Total Listening Time}_{\textit{Above Chance}} = \textit{SD}(\textit{Effect Size}) + \textit{Proportion of Total Listening Time}_{\textit{At Chance}}$$

To illustrate this formula, let us consider an infant's score of .52 total listening time (SD = .07) to the Grammatical test type. Using the SD for this infant (.07), chance set at .33, and the medium effect size of .50, our formula provides us with the Above Chance criterion .37 (see the calculation below).

$$\textit{Proportion of Total Listening Time}_{\textit{Above Chance}} = .07_{\textit{SD}}(.50_{\textit{Effect Size}}) + .33_{\textit{Proportion of Total Listening Time At Chance}}$$

$$\textit{Proportion of Total Listening Time}_{\textit{Above Chance}} = .37$$

Therefore, for this trial type, the infant's score (.52) exceeds chance because it exceeds the above chance criterion (.37).

Applying the Formula to the Classification System. To label individual infants we solved for the Above Chance Criterion (i.e., the Proportion of Total Look Time $_{\text{Above Chance}}$) for each infant for each trial type. We then applied these values to our classification system to label individual infants as following one of the four outcome types. Infants were classified as

1. *No Detection* if the proportion of total listening time failed to exceed the criterion for each trial type.

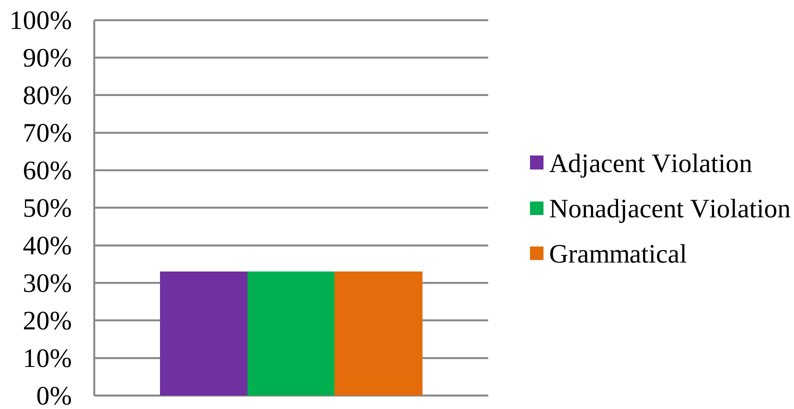

Figure 3.2. Idealized Data Showing the No Detection Outcome.

2. *Detect Adjacent Category Violations* if (A) both Nonadjacent Violation and Grammatical test types did not exceed their trial type criteria *and* the Adjacent Violation test type did exceed its trial type criterion or if (B) the Nonadjacent Violation and Grammatical test types each exceeded their criterion *and* the Adjacent Violation test type did not exceed its trial type criterion.

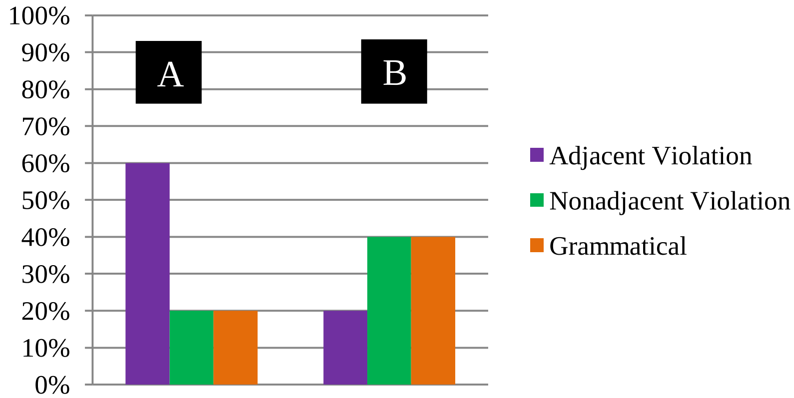

Figure 3.3. Idealized Data Showing the Detect Adjacent Category Violation Outcome.

3. *Detect Nonadjacent Violations* if (A) both Adjacent Violation and Grammatical test types did not exceed their trial type criteria *and* the Nonadjacent Violation test type did exceed its trial type criterion or if (B) the Adjacent Violation and Grammatical test types each exceeded their trial type criteria *and* the Nonadjacent Violation test type did not exceed its trial type criterion.

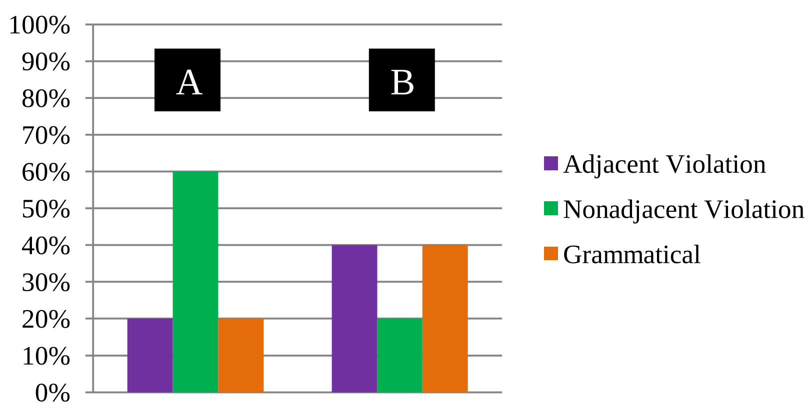

Figure 3.4. Idealized Data Showing the Detect Nonadjacent Violation Outcome.

4. *Detect Nonadjacent and Category Violations* if (A) both Adjacent and Nonadjacent Violation test types did not exceed their trial type criteria *and* the Grammatical test type did exceed its trial type criterion or if (B) the Adjacent and Nonadjacent Violation test types each exceeded their criterion *and* the Grammatical test type did not exceed its trial type criterion.

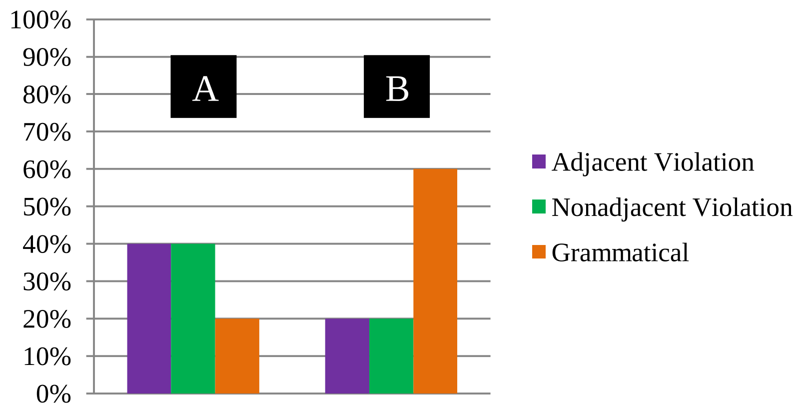

Figure 3.5. Idealized Data Showing the Detect Nonadjacent And Category Violations Outcome.

3.3.3 Individual Differences Organized By Age[3]

The looking patterns for individual 13-month-olds are shown in Figure 3.6 below. Of the twenty-one 13-month-olds, 14.3% showed a No Detection pattern, 28.6% displayed the ability to generalize the category using adjacent information (i.e., the Detect Adjacent Category Violations pattern), 28.6% displayed sensitivity to the nonadjacent dependencies only (i.e., the Detect Nonadjacent Violations pattern), and the remaining 28.6% showed sensitivity to the nonadjacent dependency and the association between the category cue and the word environment (i.e., the Detect Nonadjacent & Category Violations pattern).

[3] For individual differences organized by learner type and the specific characteristics of individual infants see Appendix A.

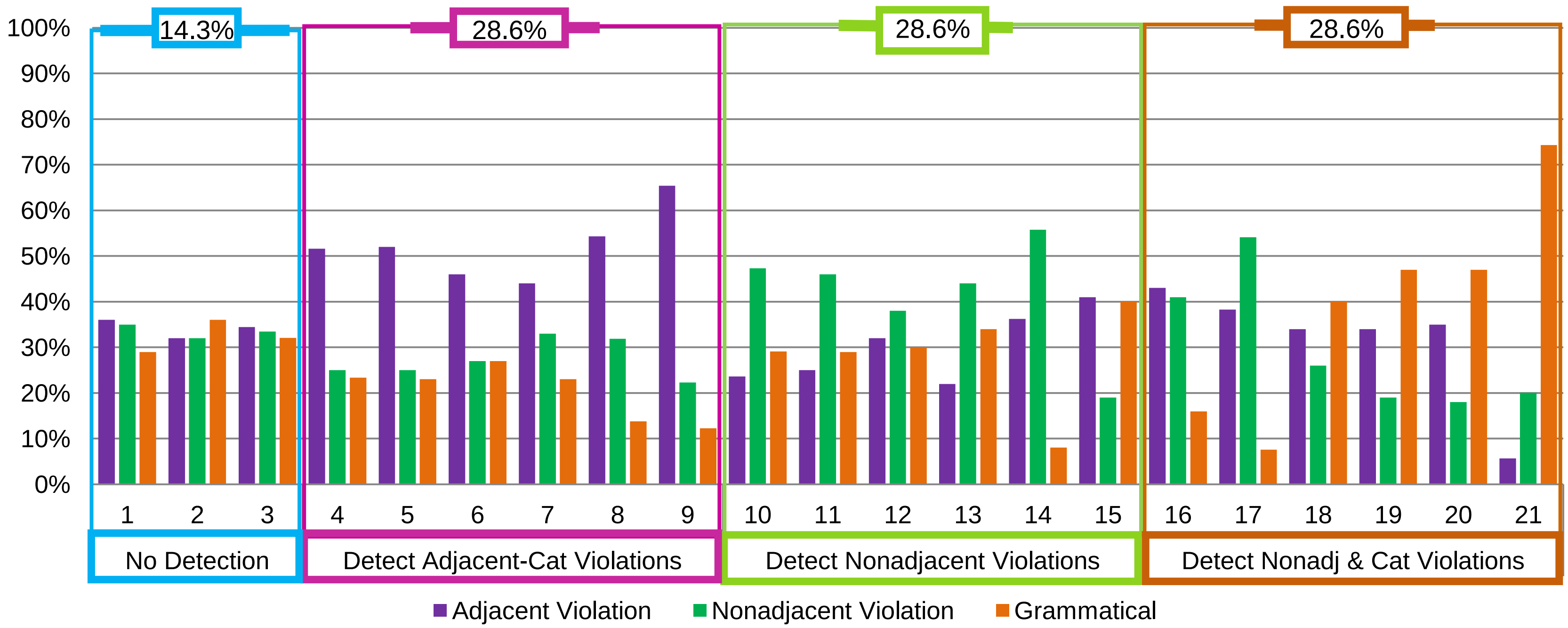

Figure 3.6. Listening Patterns for Individual 13-month-olds. Individual infants are plotted along the X-axis to show their proportion of total listening time to each trial type (Y-axis). The three trial types are displayed by different colored bars. The percentage of infants displaying each of the four learning outcomes can be seen above each group.

The looking patterns for individual 15-month-olds are shown in Figure 3.7 below. Of the twenty 15-month-olds, 0% showed a No Detection pattern, 50% displayed the ability to generalize the category using adjacent information (i.e., the Detect Adjacent Category Violations pattern), 20% displayed sensitivity to the nonadjacent dependencies only (i.e., the Detect Nonadjacent Violations pattern), and the remaining 30% showed sensitivity to the nonadjacent dependency and the association between the category cue and the word environment (i.e., the Detect Nonadjacent & Category Violations pattern).

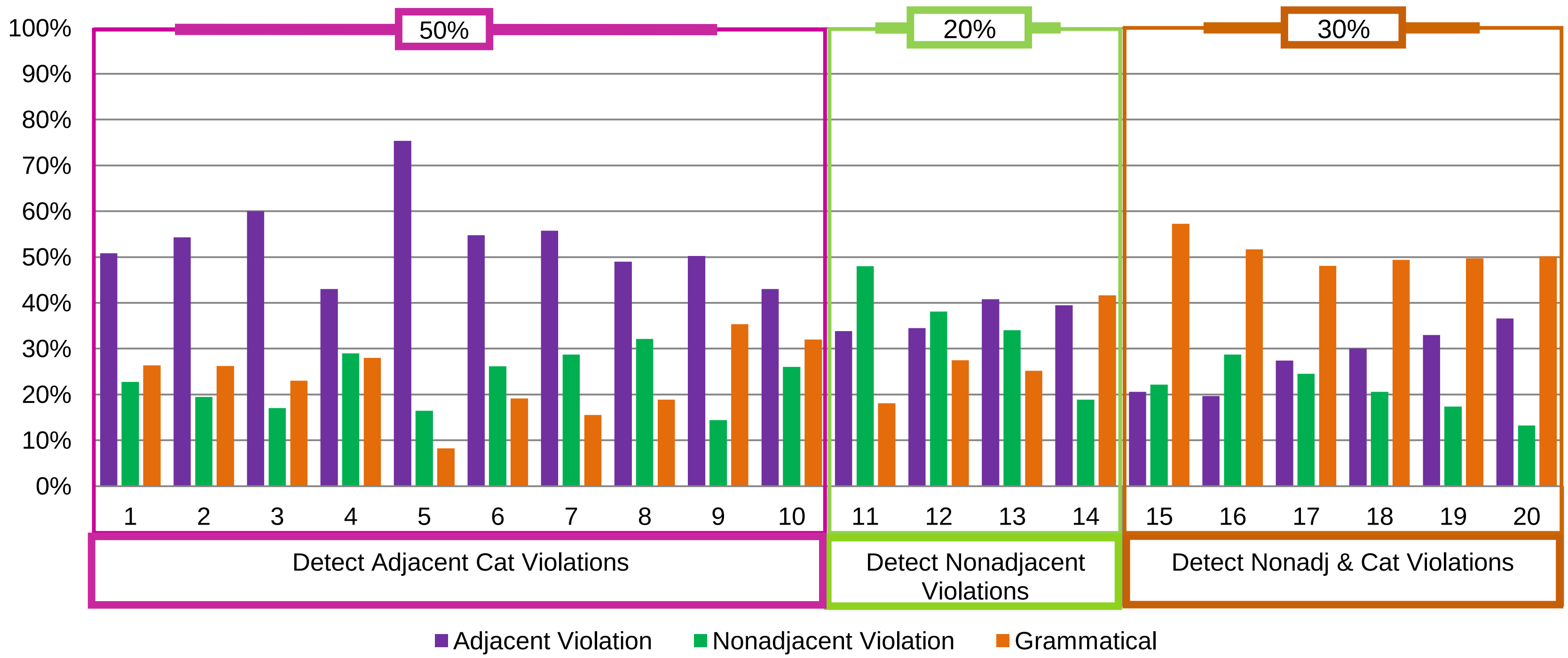

Figure 3.7. Listening Patterns for Individual 15-month-olds. Individual infants are plotted along the X-axis to show their proportion of total listening time to each trial type. The percentage of infants displaying the different learning outcomes can be seen above each group.

The looking patterns for individual 18-month-olds are shown in Figure 3.8 below. Of the twenty 18-month-olds, 5% showed a No Detection pattern, 25% displayed the ability to generalize the category using adjacent information (i.e., the Detect Adjacent Category Violations pattern), 10% displayed sensitivity to the nonadjacent dependencies only (i.e., the Detect Nonadjacent Violations pattern), and 60% showed sensitivity to the nonadjacent dependency and the association between the category cue and the word environment (i.e., the Detect Nonadjacent & Category Violations pattern).

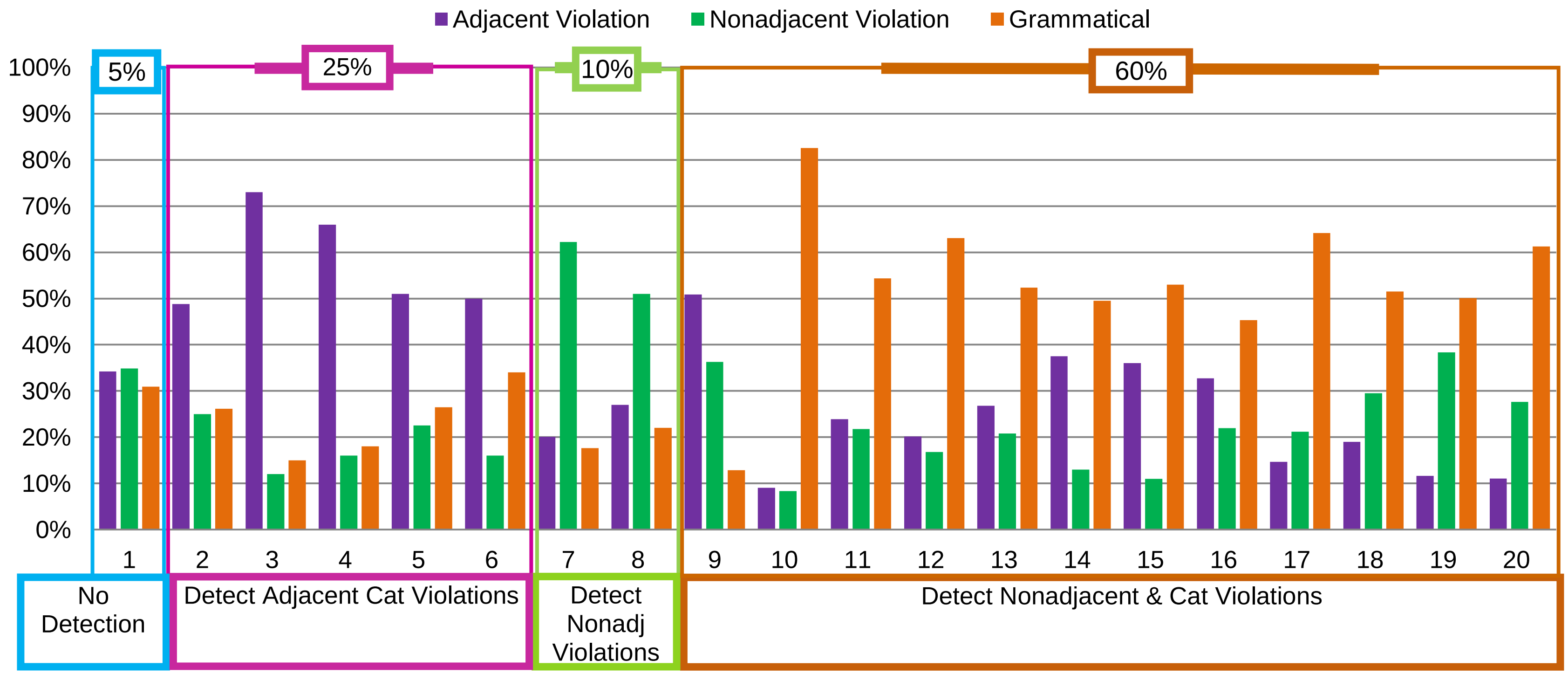

Figure 3.8. Listening Patterns for Individual 18-month-olds. Individual infants are plotted along the X-axis to show their proportion of total listening time to each trial type. The percentage of infants displaying the different learning outcomes can be seen above each group.

3.3.4 Assessing the Developmental Change Prediction

Mintz (2006; 2011b) has suggested that frames may be a primary cue to category membership and that infants use frames for categorization as early as 12 months of age. If, at the earliest stages of category acquisition, frames are used as primary cue for category formation (as opposed to other cues), then the presence of reliable adjacent dependency information should not impact their use of nonadjacent dependencies for category formation. Support for this proposal would be displayed in the present study as a minority of infants classified as Detect Adjacent Category Violations. Infants in the Detect Adjacent Category Violations group have learned the category but have failed to detect the violation in the nonadjacent dependency, therefore evidence that infants fall under this classification as opposed to the others would suggest that when provided frames that predict categories (and necessarily adjacent and nonadjacent dependencies that predict a category), infants can learn the categories without using frames. **In contrast to this prediction, I hypothesized that infants use of either adjacent or nonadjacent dependencies for category formation would change across development.**

Age. My specific predictions for the age measure in the present study were as follows:

1) Most 13-month-olds should show categorization via adjacent information. This would be displayed by a large proportion of infants being classified as Detect Adjacent Category Violations for their similar listening times to Nonadjacent Dependency Violation and Grammatical trials but different listening times to Adjacent Dependency Violation trials.
2) By 18 months, most infants should show knowledge of the nonadjacent dependency and the category environment. This would be displayed by most of the infants being classified as Detect Nonadjacent and Category Violations because of their similar listening times to

Adjacent Dependency Violation and Nonadjacent Dependency Violation trials but different listening times to Grammatical trials.

3) Finally, because 15-month-olds would be in a state of transition from using adjacent information for categorization at 13 months to detecting nonadjacent information at 18 months, I predicted variability in this age group with some infants learning categories via adjacent information, some learning only nonadjacent dependencies, and some learning the nonadjacent dependencies and the category environments.

The number of infants displaying each learner type in each age group can be seen in the 4 X 3 contingency table below (Table 3.3).

Table 3.3

The Number of Infants in Each Age Group Displaying the Different Learner Types

	No Detection (Outcome 1)	Detect Adjacent Category Violations (Outcome 2)	Detect Nonadjacent Dependency Violations (Outcome 3)	Detect Nonadjacent & Category Violations (Outcome 4)	Total
13 months	3	6	6	6	21
15 months	0	10	4	6	20
18 months	1	5	2	12	20
Total	4	21	12	24	61

A correlational analysis for ordinal data (Kendall's tau-b) was used to look at the association between Age and Learner Type (classifications rank coded as 1, 2, 3, or 4 in accordance with Table 3.3 and the hypotheses regarding ease of learning). This analysis was used because it tests the prediction that increases in age correlate with increases in learner classification. The analysis revealed a positive association between the infant's age and their learner type, $r_\tau = .192$, $p < .05$, one-tailed, which suggests that an increase in age is associated with an increase in learner type with the oldest infants being classified as the highest learner type (Outcome 4). Mintz's hypothesis would predict developmental continuity in the number of infants in Outcome 4. If this hypothesis were correct we should see no change in Learner Type as a function of Age but the data seem to reflect an increasing number of infants in the Detect Nonadjacent and Category Violations group with most infants classified in this group at 18 months as predicted by our Hypothesis 3.

However, although the results are significant, age only accounts for 8.6% of the variation in learner type (see Table 1 from Gilpin, 1993 for converting $r_\tau = .192$ to r^2). This result is likely due to the individual differences apparent across the age groups (see Figure 3.9). Although, individual differences were predicted for the 15-month-olds, the variability in learner type at 13 months was unexpected. Therefore although it seems clear that infants in these younger age groups are not as a group using frames for category acquisition, we need to formulate new hypotheses that take into account this variability. The presence of individual differences across the age groups suggests that both dependency types play a role in categorization.

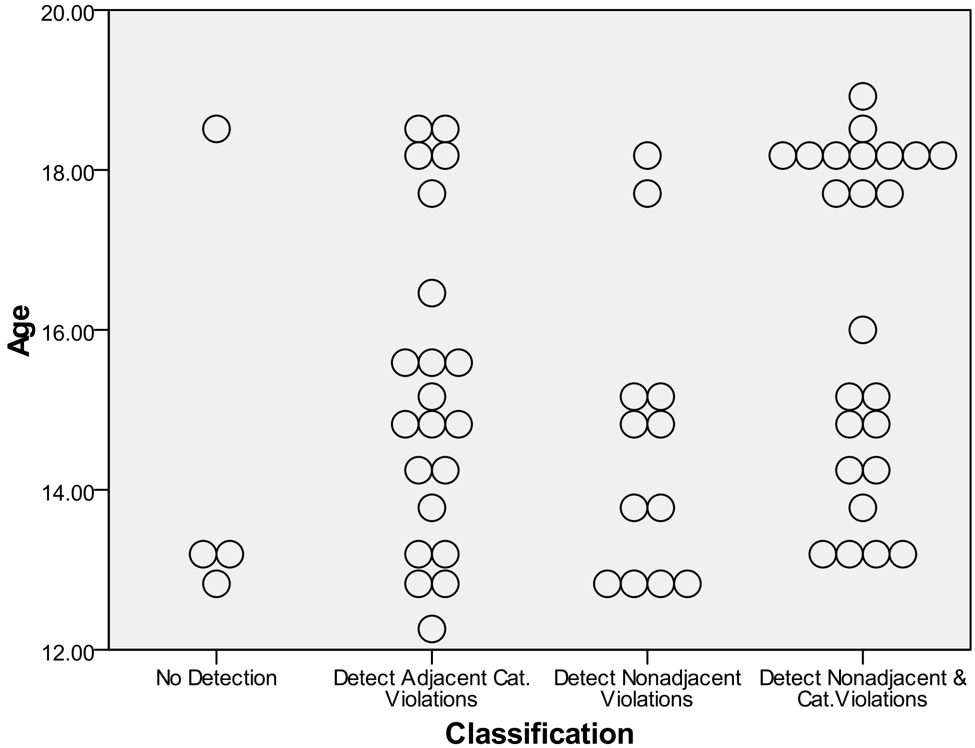

Figure 3.9. Individual Differences in Learner Type Across Age In Months. Individual infants plotted as dots.

Vocabulary Quartile. The high variability in learner type within age groups may reflect the high variability in language development within age groups. To assess whether the prediction of development change might be best assessed in terms of linguistic development, the data were organized according to parental estimates of productive vocabulary size. This measure of productive vocabulary was obtained at the time of the study by asking parents to report the words their child produced. Parents provided this report after the consenting process and before infants were taken to the habituation booth to complete the experimental portion of the study. Although this report consisted of parents' free recall and may have been incomplete, the reliability of these reports was assessed for a subset of the infants (n =20) from the larger sample. This smaller sample was comprised of participants who had also completed the MacArthur Communicative Development Inventory: Words and Gestures at 14 months of age. Thus to validate use of parental free report, a correlation analysis was performed on the number of words produced according to the parents' free recall at the time of the study and the infants' 'Words Says' percentile according to the MCDI norms. The Pearson correlation coefficients for each age group were either significant or approaching significance and the explained variance (r^2) values indicated a strong relationship between the free recall measure and the standardized MCDI measure (see Table 3.4 for a summary of the correlation results). Thus, the free recall measure of production was used for 59 of the 61 infants in subsequent analyses; two infants were missing this production measure (no estimate provided by parent) and were thus not included in the analyses.[4]

Table 3.4

Relationship Between Words Produced at the Time of Study and MCDI: Words & Gestures 'Words Says' Percentile

	N	*r*	r^2
13- month-olds	8	.706†	.498
15-month-olds	6	.907*	.823
18-months-olds	6	.861*	.741

† $p < .1$
* $p < .05$

The remaining 59 data points were then divided into quartiles (1 = infants in the lowest vocabulary quartile/infants with the smallest vocabularies). The number of infants displaying

[4] For one of the two infants, the parent answered questions about grammatical development and indicated the child was combining words. This child was 18-months-old and was classified as Detect Nonadjacent and Category Violations.

each learner type in each vocabulary quartile can be seen in the 4 X 4 contingency table below (Table 3.5).

Table 3.5

The Number of Infants in Each Vocabulary Quartile Displaying the Different Learner Types

	No Detection (Outcome 1)	Detect Adjacent Category Violations (Outcome 2)	Detect Nonadjacent Dependency Violations (Outcome 3)	Detect Nonadjacent & Category Violations (Outcome 4)	Total
1	2	5	7	1	15
2	1	4	4	6	15
3	0	6	1	8	15
4	1	6	0	7	14
Total	4	21	12	24	59

Another correlational analysis for ordinal data (Kendall's tau-b) was conducted but this time to examine the association between Vocabulary Quartile (1-4) and Learner Type (classifications rank coded as 1, 2, 3, or 4 in accordance with Table 3.5 and the hypotheses regarding ease of learning). The correlation between vocabulary quartile and learner type was not significant, $r_\tau = .155$, $p = .08$, one-tailed. In the lowest vocabulary quartile, there appear to be individual differences in whether infants are classified as Detect Adjacent Category Violations or Detect Nonadjacent Violations. In the highest quartile, most infants appear to be categorizing, but there are individual differences in whether infants are Detect Adjacent Category Violations or Detect Nonadjacent and Category Violations. As for the age variable, there is high variability when the data were organized by vocabulary (see Figure 3.10 below).

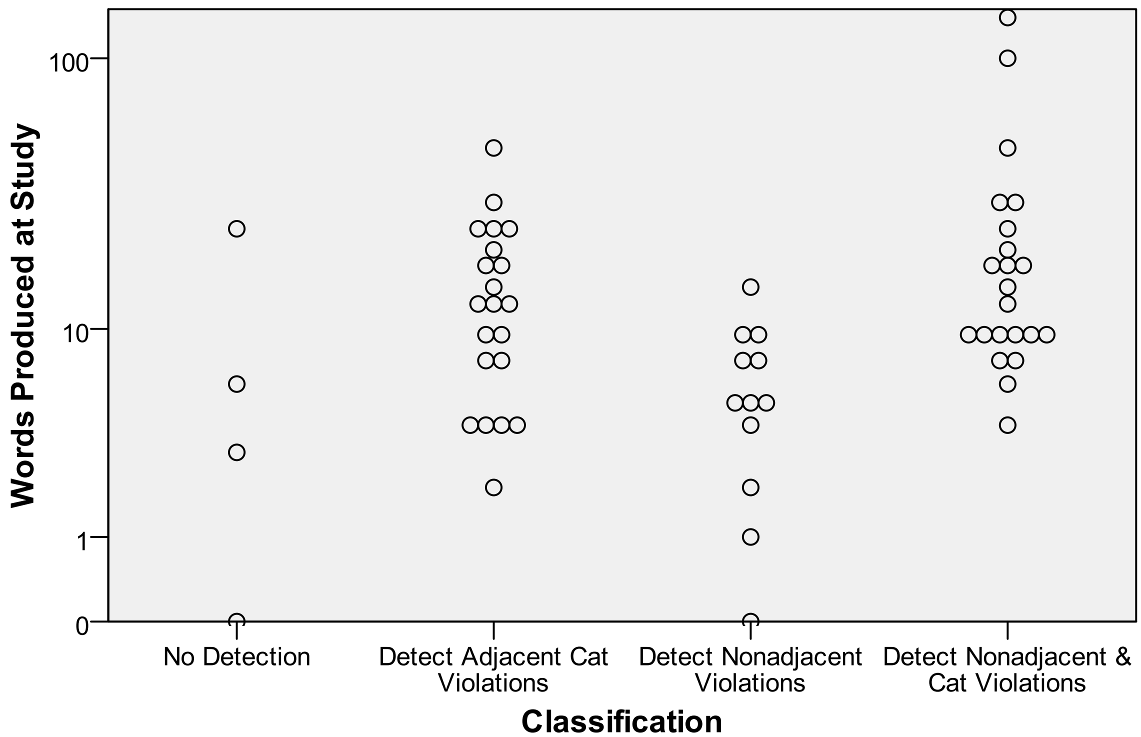

Figure 3.10. Individual Differences in Learner Type Across Productive Vocabulary. Individual infants plotted as dots.

Taken together, the results suggest that not all infants use fixed frequent frames at the beginning stages of category acquisition. However, there was also not uniformity in the use of adjacent dependencies for categorization in the 13-month group. Therefore, it is not clear whether all infants begin the process as adjacent dependency trackers. Instead the results suggest that when multiple sources of information cue category membership such as phonological cues and adjacent and nonadjacent word information, as was provided in this study and as is provided in natural language, the information infants track at younger ages (and at earlier stages of lexical development) differs. In addition individual difference persist throughout development, however infants become more uniform in their ability to categorize.

3.3.5 Exploring Predictors of Categorization

Although this study was not designed to address predictors other than age, other factors such as the child's auditory processing and memory, language processing ability, and/or language knowledge (e.g., productive and/or receptive vocabulary size) may predict the learning mechanisms underlying category acquisition. This section provides an alternative analysis of the data to explore how well age, language knowledge in the form of productive vocabulary, and information processing in the form of habituation measures, predict learner type.

Lexical and grammatical development are tightly correlated (Bates & Goodman, 1997). Indeed there are various 'contingent acquisition' theories of grammar that suggest the acquisition of grammatical knowledge is contingent upon and emerges from lexical knowledge (Bates & Goodman, 1997; Pinker, 1984; Snedeker, Geren, & Shafto, 2007; 2012; Tomasello, 2000). In addition, previous studies have linked infants' lexical development to their use of natural language distributional word contexts for word learning (e.g., Fisher, Klinger, & Song, 2006; Lew-Williams & Fernald, 2007). However, these studies do not examine specific context types, (adjacent or nonadjacent). Thus, the most specific prediction that can be made from the literature is that greater lexical knowledge should predict grammar acquisition. In the present study, the two groups that can be said to have acquired 'grammar knowledge' are the Detect Adjacent Category Violations infants (Outcome 2) and the Detect Nonadjacent and Category Violations infants (Outcome 4). These infants learned the categories and their associated distributional environments (e.g., a↔X, X↔b and/or a_b↔X) and generalized this knowledge to phrases containing novel category members (i.e., novel X and Y words). In contrast, infants classified as Detect Nonadjacent Violations (Outcome 3) only detected violations of the co-occurrence relationships between words (e.g., a_b) and did not show evidence of using this information to form a categories. In addition, infants classified as No Detection (Outcome 1) may have learned the co-occurrence relationships between adjacent words such *pel feenam*, but critically, did not show evidence of using this information to form categories. Although it can be argued that learning co-occurrence relationships between words is grammatical in nature, connecting these co-occurrences to categories is stronger evidence of grammatical knowledge and/or a higher level of grammar learning. Therefore, a general prediction was made that Words Produced would predict category acquisition. To test this prediction, Outcome 2 and 4 infants were combined and labeled as Categorizing and Outcome 1 and 3 infants were combined and labeled as Not Categorizing. In addition, given Gerken et al.'s (2011) results, habituation trials and time may also be predictors of categorization.

Analysis 1. A logistic regression analysis was performed with Age, Total Habituation Time, Trials to Habituate, and Words Produced at the Time of the Study entered as continuous predictors, with Learner type (Categorizing = 1 or Not Categorizing = 0) entered as the dependent variable. To remain conservative predictors were entered into the model at the same time (Enter method in SPSS) rather than in steps or in blocks. Categorizer classification (Categorizing or Not Categorizing) was significantly improved by the inclusion of these variables, χ^2 (4, N = 58) = 20.869, $p < .001$ (note that the Enter method does not allow us to provide the significance of the model given a specific predictor).[5] However, only the word production measure produced a significant coefficient (see Wald test in Table 3.6) suggesting that vocabulary size was more indicative of an infant being classified as a categorizer.

[5] No issues of collinearity were present among the predictor variables (VIF < 10 for all variables). One influential data point was found (Cook's Distance > 1; Standardized Residual > 4); this variable was removed from the analysis. The reported results reflect the model without this data point, but note that the model and the Wald statistic for vocabulary are both significant with this data point included.

Specifically, the odds ratio informs us of the direction and indicates that for every one word increase in production, the odds of Categorizing are 1.33 times as likely.

Table 3.6

Logistic Regression Results for Predictors of Categorization

Variable	*B*	SE	Wald test	Odds ratio
Age	.031	.240	.017	1.03
Words Produced at the Time of Study	.286	.108	7.055**	1.33
Total Habituation Time	-.004	.007	.280	.996
Trials to Habituate	.174	.153	1.296	1.190

** $p < .01$

Note: The model's $R^2_{Nagelkerke}$ = .444 and $R^2_{Cox\ \&\ Snell}$ = .302.

Analysis 2. A logistic regression analysis was performed with the same continuous predictors as in Analysis 1, however this time the dependent variable comparison was between Outcome 3 (Detect Nonadjacent Dependency Violations Only = 0) and Outcome 2 (Detect Adjacent Category Violations = 1). This analysis was important to conduct given that previous studies have shown adjacent dependency learning (Outcome 2) at earlier ages than nonadjacent dependency learning (of the type measured by Outcome 3). Again, predictors were entered into the model at the same time rather than in steps or blocks. The model was significantly improved by the inclusion of the predictor variables, χ^2 (4, N = 33) = 14.65, p < .01. However, as in the previous analysis, only the production measure produced a significant coefficient (see Wald test in Table 3.7) suggesting that an increase in the production measure was more indicative of learner type. Specifically, every one-word increase in words produced made the odds of being classified as Detect Adjacent Category Violations (instead of Detect Nonadjacent Violations) 1.362 times as likely.

Table 3.7

Logistic Regression Results: Predictors of Classification into Detect Adjacent Category Violations vs. Detect Nonadjacent Violations

Variable	*B*	SE	Wald test	Odds ratio
Age	-.286	.349	.673	.751
Words Produced at the Time of Study	.309	.132	5.476*	1.362
Total Habituation Time	-.016	.011	1.943	.985
Trials to Habituate	.399	.278	2.068	1.491

* $p < .05$
Note: The model's $R^2_{Nagelkerke}$ = .491 and $R^2_{Cox \,\&\, Snell}$ = .359.

Analysis 3. When infants were organized according to their productive vocabulary quartiles (see Section 3.3.4), the lowest quartile contained a higher proportion of Detect Adjacent Category Violations infants (n = 5) than Detect Nonadjacent and Category Violations infants (n = 1), but the other three quartiles contained roughly equal numbers of these two learner types. Is there a variable that can account for these individual differences? To assess whether age, the habituation measures, or vocabulary size could predict classification into the two categorizer types one final regression analysis was needed. The other analyses do not speak to this question as Detect Adjacent Category Violations and Detect Nonadjacent & Category Violations were grouped together in Analysis 1, and the latter group of infants were excluded from Analysis 2. Therefore one final logistic regression analysis was performed with Age, Total Habituation

Time, Trials to Habituate, and Words Produced at the Time of the Study entered as continuous predictors, and this time the dependent variable comparison was between Outcome 2 (Detect Adjacent Category Violations = 0) and Outcome 4 (Detect Nonadjacent & Category Violations = 1). Again, the predictors were entered into the model at the same time rather than in steps or blocks. The model was significantly improved by the inclusion of the predictors, χ^2 (4, N = 43) = 9.763, *p* = .045. However in this analysis, Total Habituation Time produced a marginally significant coefficient, all other coefficients were not significant (see Wald test in Table 3.8) suggesting a trend of the habituation measure as indicative of learner type. This pattern is evident in the Total Habituation Time averages for the two groups (Detect Adjacent Category Violations *M* = 78.75 s, SE = 8.88 s; Detect Nonadjacent & Category Violations M = 131.08 s, SE = 17.04 s). However, the confidence interval for this measure crosses 1 suggests that that the direction of this relationship is unstable and cannot be generalized to the general population.

Table 3.8

Logistic Regression Results: Predictors of Classification into Detect Adjacent Category Violations vs. Detect Nonadjacent and Category Violations

Variable	*B*	SE	Wald test	Odds ratio
Age	.210	.20	1.101	1.233
Words Produced at the Time of Study	.006	.021	.086	1.006
Total Habituation Time	.015	.009	2.724†	1.015
Trials to Habituate	.026	.149	.029	1.026

† $p < .1$
Note: The model's $R^2_{Nagelkerke}$ = .271 and $R^2_{Cox\ \&\ Snell}$ = .203.

In summary, the results from the first analysis suggest that an increase in productive vocabulary significantly increases the odds that an infant will be classified as a categorizer. In addition, the second analysis comparing Detect Adjacent Category Violations and Detect Nonadjacent Dependency Violations revealed that an increase in productive vocabulary significantly increases the odds that an infant will be classified as Detect Adjacent Category Violations. Finally, neither habituation, age, nor vocabulary provided significant predictions to distinguish the two types of categorizers (Detect Adjacent Category Violations vs. Detect Nonadjacent & Category Violations). The stable of effect of vocabulary as a predictor of categorization is consistent with previous reports that show a stronger link between vocabulary size and grammar acquisition than age and grammar acquisition (Snedeker et al., 2007; 2012). In addition, these results expand upon previous findings by connecting vocabulary size to infants' use of distributional word contexts for *category acquisition*. However, the results of the third regression analysis indicated that the individual differences in the *type* of word contexts used for category acquisition could not be adequately explained by habituation measures, age, or productive vocabulary. Factors that might account for these individual differences will be explored in Chapter 4.

CHAPTER 4

GENERAL DISCUSSION & FUTURE DIRECTIONS

4.1 Summary

There is an ever growing literature on the precocious abilities of infant language learners with several studies indicating early knowledge of the association between categories and syntactic contexts. How infants come to acquire this knowledge is unknown. This dissertation investigated how the use of word/morpheme contexts facilitates acquisition of categories across development. Specifically, the present research investigated how adjacent and nonadjacent dependencies are used to solve the problem of lexical category acquisition when both patterns are present and informative in the input. In addition, the experiments assessed the proposal that infants use nonadjacent dependencies to form categories (Mintz, 2003; 2006; 2011b). To assess this proposal and embed it in a developmental framework, 13-, 15-, and 18-month-old infants were provided input that contained reliable adjacent and nonadjacent information, information that is found in natural language frames that predict noun and verb categories. As illustrated in Chapter 2, previous studies have investigated the use of adjacent and nonadjacent contexts for category formation with infants, but they have not investigated the contribution of each context when both sources of information are present in the input as would be expected in natural language. Further, although there are such studies in the adult literature, they do not directly inform the question of development especially when we consider the developmental changes that take place in the ability to detect nonadjacent dependencies.

The present study did not reveal evidence that infants uniformly use nonadjacent dependencies across development. Nor did the data reveal that infants younger than 15 months of age, uniformly learn categories using adjacent information. Instead, the data revealed considerable individual differences in the 13- and 15-month-old age groups, with a majority of the infants in the 18-month-old age group learning nonadjacent dependencies and category environments. In the sections below, I integrate these findings into the literature on adjacent and nonadjacent dependency learning and discuss future experiments that can further test how frame-based and multiple-cue based theories scale up to the problem of natural language acquisition. Finally, I discuss the implications of the learner type predictor analyses for theories of category acquisition.

4.2 Integrating the Findings of this Study Into the Literature

The current findings revealed individual differences across the 13- , 15- , and 18-month-old age groups with an increasing number of infants showing nonadjacent dependency detection and categorization across development. What do these results mean in the context of previous findings?

Previous nonadjacent dependency studies have found null results with younger infants (e.g., null results found with 12-month-olds in Gómez & Maye, 2005 and 15-month-olds in Santelmann & Jusczyk, 1998), while adjacent dependency studies have shown robust findings in younger and older infants (e.g., see adjacent dependency learning in the face of probabilistic cues by 12-month-olds in Gómez & Lakusta, 2004 and by 22-month-olds in Lany & Saffran, 2011). Although, it is possible that the younger infants in previous nonadjacent dependency studies (e.g., 15-month-olds in Santelman & Jusczyk, 1998; 14-month-olds in Nazzi et al., 2011) may have displayed mixed preference patterns but uniformly discriminated between grammatical and ungrammatical material (the interpretation of Mintz, 2011b), given the present results it is more likely that null results were found in previous studies in the younger age groups because as a group, younger infants do not show learning of nonadjacent dependencies. But why if nonadjacent dependency learning appears more challenging, did the present findings not reveal uniformity in adjacent dependency learning in the 13-month-old age group? And why did the 15-month-olds not show uniformity in nonadjacent dependency detection in this study when they have shown nonadjacent dependency learning in previous artificial language studies (e.g., Gómez & Maye, 2005)? Below I discuss how frames that predict categories might lead to earlier nonadjacent dependency learning by virtue of their containing adjacent dependencies and how this might be the reason for a lack of uniformity in the 13-month-old group. I also explain how this characteristic might lead infants who can learn nonadjacent dependencies to show adjacent dependency learning (i.e. 50% of the 15-month-olds and 25% of the 18-month-olds learned adjacent dependencies).

Before discussing these questions it is important to review how the language in the present study differed from the stimuli in previous artificial language studies. In the present study infants were provided with aXb strings containing predictable aX and a_b dependencies, while in other artificial language studies infants were only provided predictable aX or a_b dependencies. Specifically, in Gómez's studies (Gómez, 2002; Gómez & Maye, 2005), infants were provided with nonadjacent dependencies that did not reliably predict middle words, and the middle words did not come from two distinct sets as they did in the present study (1- vs. 2-syllable words; words that occur with pel_rud and vot_jic vs. words that occur with fim_sog and dak_tup).

The syllable number cue, provided in the current study, was added to facilitate categorization by acting as a feature that unifies and distinguishes the two sets of intervening words. The presence of this cue may reduce the intervening information infants are processing across phrases from 24 different words to one distinguishing feature (syllable number). In addition to information reduction across phrases, the syllable number cue allows information from adjoining adjacent dependencies. Thus, this environment changes the learning problem and may change the learning processes used to solve this problem into something that is not as straightforward as adjacent versus nonadjacent dependency learning. That is, given that the frequent frames predicted categories (a_b↔X) they also contained predictable adjacent relationships (a↔X). This makes it is possible that adjacent dependency learning was instrumental in the infants nonadjacent dependency learning.

As St. Clair et al. (2010) suggest, learners may be capable of building frames from their composite parts (e.g., aX + Xb = aXb). Another possibility is that by learning the category via adjacent dependencies, the intervening material becomes familiar and the adjacent dependencies predictable. This could facilitate nonadjacent dependency learning if learners are driven by information gains as suggested by Onnis et al. (2003) as learners would be driven to finding new patterns in the input after discovering the invariant adjacent dependency patterns.

Another way adjacent dependencies may have facilitated nonadjacent dependency learning is via reduction in the attentional demands of the middle elements. Pacton and Perruchet (2008) have proposed an attention-based account to explain and predict how learners process input that contains reliable adjacent and nonadjacent dependencies. They argue that the learning of these dependencies is a product of the joint attentional processing of the dependent elements. Adjacent dependency learning is possible and often more robust than nonadjacent dependency learning because units that are temporally or spatially contiguous are more likely to be attended to and processed together. In contrast, if nonadjacent units are in attentional focus the relationship should be just as easily learned as the relationship among adjacent units. In a study with adults, Pacton and Perruchet (2008) have demonstrated that when both adjacent and nonadjacent relationships occur in number sequences such that both are equally informative, experimental manipulations designed to bring nonadjacent co-occurrences into focus result in robust learning of nonadjacent numbers and little to no learning of adjacent numbers and vice versa. Thus, applying this attention-based explanation to the current findings would involve infants first learning adjacent dependencies because of their bias to attend to elements that are temporally contiguous and/or because novelty drives attention to middle elements. Learning of adjacent dependencies would have in turn, reduced the task demands and freed up attentional resources allowing for attention to shift to the competing nonadjacent dependency.

In addition, Newport's (1988; 1990) "less is more" hypothesis would predict that the infant's developing memory and processing capabilities should guide and constrain infant attention during language acquisition. This proposal aligns with the explanations described above but specifies that infants might initially attend to adjacent morpheme environments because of their limited processing windows.

If infants learn a nonadjacent dependency by first learning adjacent dependencies, it is possible that the ability to learn nonadjacent dependencies without first forming a category is on a different developmental timeline. Interestingly, the results from the present study more closely match the developmental timeline in morphosyntactic dependency studies with natural languages (e.g., Santelmann & Jusczyk, 1998: detection at 18, but not 15 months of age; Nazzi et al., 2011: detection at 18, but not 14 months of age) than they do the artificial language studies looking solely at nonadjacent dependency learning (Gómez & Maye, 2005; Gómez, 2002: learning at 15 and 18 months of age). In short, the individual differences in the present study could reflect different steps along the path to discovering multiple dependencies (adjacent and nonadjacent) and thus, the adjacent dependency first hypothesis may be one that applies during a given learning situation rather than across development.

On the other hand, the individual differences may reflect that some infants indeed learn the nonadjacent dependency first while others learn the adjacent dependency when both are present and informative in the input. For example, individual differences in memory and processing capabilities may have allowed some of the younger infants to learn nonadjacent dependencies earlier than their same age peers but by 18 months these individual differences largely subside when dependencies occur across 1-2 syllables. This explanation is consistent with the results from the morphosyntactic dependency studies showing that infants' processing of intervening lexical elements affects the trajectory of morphosyntactic dependency learning (Santelmann & Jusczyk, 1998; Van Heugten & Johnson, 2010; Höhle et al., 2006; Soderstrom et al., 2007). Testing younger infants or providing infants with a greater distance between nonadjacent words may reveal fewer individual differences and a greater uniform reliance on adjacent dependencies. A related explanation is that we see individual differences because there are differences in what is salient to different infants. Because there is evidence that 15- and 18-month-olds can learn nonadjacent dependencies without needing to learn adjacent dependencies first (Gómez, 2002; Gómez & Maye, 2005), it may be that for the Detect Nonadjacent Dependency infants (Outcomes 3 and 4) attention was first pulled to the highly frequent framing elements. And before attention was diverted to the intervening words, infants were able to learn the nonadjacent dependency. While for the Detect Adjacent-Category Dependency infants, attention was first focused and maintained on the adjacent co-occurrences because of the presence of novel words across phrases and because of the temporal contiguity of these elements.

Although the data and these interpretations may align more with the multiple cue integration approach (i.e., the proposal that infants exploit all the cues they can detect and integrate the cues to solve the learning problem at hand), a frame-based approach cannot entirely be rejected. What may be needed to account for infant learning is a more dynamic version of these two proposals; for instance, a multiple cue integration proposal that would predict frames as a primary cue for categorization for some infants (or some categories), while for other infants (or for other categories) frames would be secondary cue, serving to facilitate further learning and refinement of categories.

4.3 Future Research: Scaling Up Distribution-Based Proposals of Category Acquisition

The idea that different constellations of cues may be used for different categories and differently by different infants is especially important when we consider that in natural language, cues are probabilistic and not all are created equal within a given language or across languages. Corpus and modeling studies have revealed that acoustic and distributional cues are probabilistic: individually cues do not predict categories 100% of the time but they do predict categories at levels greater than chance (e.g., see Shi, Morgan, & Allopena, 1998; Monaghan, Chater, & Christiansen, 2005; Monaghan, Christiansen, & Chater, 2007). When aligned, cues more accurately predict category membership (Shi, Morgan, & Allopena, 1998; Monaghan et al., 2005) but which cues are most informative may change as a function of lexical frequency (Monaghan et al., 2005), lexical category (Mintz, 2006; Monaghan et al. 2005), and language (Erkelens, 2009; Stumper et al., 2011; Monaghan et al., 2007). For instance, Maratsos (1990) has argued that across languages, the distributional environments of verbs are much more predictive of category membership, than noun environments because verbs have more reliable local environments that apply to most of the words in the category and that do not apply to other categories (e.g. tense-markers). Indeed the frequent frames that frame verbs predict the verb category at a greater accuracy than the frequent frames that frame nouns (Mintz, 2003; 2006) and infants' categorization when provided with these frames reflects this difference (Mintz, 2006). Mintz (2006) showed that infants categorized novel words when the words were embedded in verb frames during familiarization but they did not show categorization for words embedded in noun frames.

The present study was a first step in investigating infants' learning biases at different points in development. The aim was to assess their base level biases in an "all else being equal" fashion. Thus, both dependency types were made deterministic to keep the learnability of adjacent and nonadjacent dependencies equal (or as equal as possible given the current knowledge of the learning requirements for both dependency types). However, as just discussed, this does not reflect the highly probabilistic nature of cues in natural language. Although, previous studies have shown adjacent dependency learning in the face of noise (Gómez & Lakusta, 2004), we do not know if infants would use adjacent over nonadjacent dependency information for categorization in the face of probabilistic cues. Therefore to better test and scale-up the frame-based and multiple-cue integration proposals of lexical category acquisition, future studies should examine learning by using probabilistic aXb languages.

For instance, it would be informative to know whether probabilistic phonological cues (e.g., a probabilistic X/Y syllable-number cue) would lead more infants to nonadjacent dependency first categorization when provided with the type of aXb language used in the present study. This would be an important test for understanding distributional mechanisms as phonological cues are not as predicative of category membership as are distributional cues for high frequency words (Monaghan et al., 2005). Future studies will also need to show that infants can learn nonadjacent dependencies even under probabilistic conditions. Manipulating probabilities of nonadjacent dependencies would be informative as some categories (which varies according to language) are marked more reliably by acoustic/phonological cues than they are predicted by distributional word/morpheme contexts (Monaghan et al, 2007). On the other hand, the differences we see in cue reliability in corpus and modeling studies may be negligible to the infant language learner during category acquisition. For instance, Curtin, Campbell, and Hufnagle (2012) have shown that 16-month-olds are able to map actions with iambic labels and objects with trochaic labels but not the reverse (also see Graf Estes & Bowen, 2013), even though in natural language nouns follow the trochaic pattern far more reliably than verbs follow the iambic pattern (94% of English disyllabic nouns are trochaic, whereas only 69% of disyllabic verbs are iambic [Kelly & Bock, 1988]).

In addition, it could be expected that by manipulating cue reliability, we can manipulate the infants' attention to alternative sources of information. However, this assumes that infants are able to shift their attention to alternative dependencies (including nonadjacent dependencies). As we have seen, there were individual differences in infants' abilities to track the nonadjacent dependencies that were present in the deterministic language used in this study. Therefore, it will be important for future studies to continue to explore predictors of categorization in natural and artificial languages to determine how the input and the learner's non-linguistic and linguistic abilities impact the process of category acquisition.

4.4 Predictors of Learner Type

The alternative analyses provided in Section 3.3.5 revealed that vocabulary, and not age, significantly predicted whether or not infants would learn the categories and their distributional environments in the present study. The analyses further indicated that neither age nor vocabulary could significantly predict whether infants would be classified as Detect Adjacent Category Violations or Detect Nonadjacent & Category Violations. Although these analyses should be interpreted with caution because the study was not designed to include vocabulary as an independent variable, the results are consistent with previous findings indicating a strong link between lexical and grammatical development (Bates & Goodman, 1997). What do these results mean for stories of lexical category acquisition?

4.4.1 Predictors of Categorizing and Not Categorizing

The result showing that productive vocabulary predicts categorization is consistent with contingent-acquisition theories that specify lexical development as a critical precursor to syntactic development (e.g., Bates & Goodman, 1997; Pinker, 1984; Tomasello, 2000). Some of these theories involve some critical mass of words (and or phrases) that are then analyzed and compared by the child (e.g., Tomasello, 2000). However, we are seeing this relationship between productive vocabularies that contain less than 100 words and 'grammatical' knowledge acquired from an artificial language that does not contain meaning. Therefore, this finding may be more indicative of the underlying processes that are shared between word learning and grammatical development, than it is of some directional relationship as implied by some contingent-acquisition theories. For instance, both the task in the present study and word learning could be said to involve cue integration: learners have to integrate labels with meanings to learn words. Thus, if infants are better cue integrators we would expect them to be better at learning the categories in the present study and better at learning words in the real world. In addition, this link could be due to differences in speech/acoustic processing as infants' speech processing abilities have been linked to their language learning abilities (e.g., see Fernald, Perfors, & Marchman, 2006; Fernald & Marchman, 2012). Infants who are better speech processors may have learned more quickly and/or more relationships in the present study. These and possibly other processes that may be more specific to language development could account for the observed relationship between vocabulary production and categorization in the present study. The role of these factors can be investigated in future research. If they account for greater variability than vocabulary (receptive or productive) then this would be strong evidence for the interpretation that shared underlying processes, rather than lexical development itself, account for the connection between lexical and grammatical development.

On the other hand, if vocabulary itself is a stronger predictor of grammatical development than shared learning mechanisms as has been suggested by Snedeker et al. (2012), then the results in the present study would indicate that words have allowed infants to learn something very general about word co-occurrences and how they can connect to word categories before 18 months of age. This knowledge would have to go beyond detecting how distributional patterns connect to categories in their native language and would have to be general enough that it could be carried over to the artificial language presented to them in the lab. Thus, this hypothesis would predict that native language grammar knowledge should better account for artificial language grammar learning than vocabulary size and shared learning mechanism factors.

4.4.2 Predictors of Detect Adjacent Category Violations and Detect Nonadjacent & Category Violations

The results of the final regression analysis revealed that the Detect Adjacent Category Violations infants and the Detect Nonadjacent & Category Violations were indistinguishable given our predictor factors age and productive vocabulary size. What can we conclude about frame-based and multiple-cue integration proposals from these results?

The data made it clear that different infants tap into different structures over the course of development. Thus, rather than rejecting or supporting either of these theories, the data call for a more dynamic theory and further investigation that can help explain under what circumstances infants use these different distributional environments. Cognitive constraints such as attention and processing windows have been discussed as key mechanisms in nonadjacent dependency learning (e.g., see Santelmann & Jusczyk, 1998; Pacton & Perruchet, 2008). Individual differences in these factors could account for the differences evident in the present study. Thus, future studies could test whether this predictor can distinguish these two learner types. But more importantly, this predictor should be investigated in studies with languages that contain probabilistic cues matching the variability present in natural language. Not only would such studies help us understand individual differences in learning but they would also lead to a more psychologically valid theory that could shed light on how lexical category acquisition varies as a function of the input and as a function of the learner's cognitive and linguistic constraints.

4.5 Conclusion

In conclusion, by unveiling individual differences we have found a complex picture of category acquisition that points to mixed beginnings but uniform endings. Although the data do not support a nonadjacent dependency first hypothesis for all infants, the data do not deny that some infants may be able to use information from nonadjacent dependencies to form categories. Rather they reveal that some infants may use both dependency types, while others may only use one at different stages in development. In addition, the results support a developmental shift consistent with previous infant studies examining dependencies between words (e.g., Gómez & Maye, 2005; Santelmann & Jusczyk, 1998; Nazzi et al., 2011; Van Heugten & Johnson, 2010).

We are also reminded that infants' abilities are not static and thus the findings we see in adult studies need to be replicated with infants if we are to truly assess theories of language acquisition. Further, not only is it likely that infants process the input differently than adults (Newport 1988; Newport, 1990), our study shows that some infants process the input differently than their same age peers when the input contains multiple sources of reliable information as in natural language. More work can be done to examine how other factors such as the infant's processing and language capabilities predict learning (as opposed to age); assessing these other factors would allow us to build a more informative story of how infants use distributional patterns to form categories.

APPENDIX A: CHARACTERISTICS OF INDIVIDUAL INFANTS

The following four sections display the characteristics of individual infants organized according to classification type: 1) No Detection, 2) Detect Adjacent Category Violations, 3) Detect Nonadjacent Violations and 4) Detect Nonadjacent and Category Violations. Each of these sections includes: 1) a graph displaying the listening patterns of individual infants (numbered for reference) and 2) a table that includes a reference number for each infant that matches the number on the graph, the infant's age, reported vocabulary size, habituation information (total time and trials), and other information that might be of interest to the reader.

CHARACTERISTICS OF INFANTS CLASSIFED AS NO DETECTION

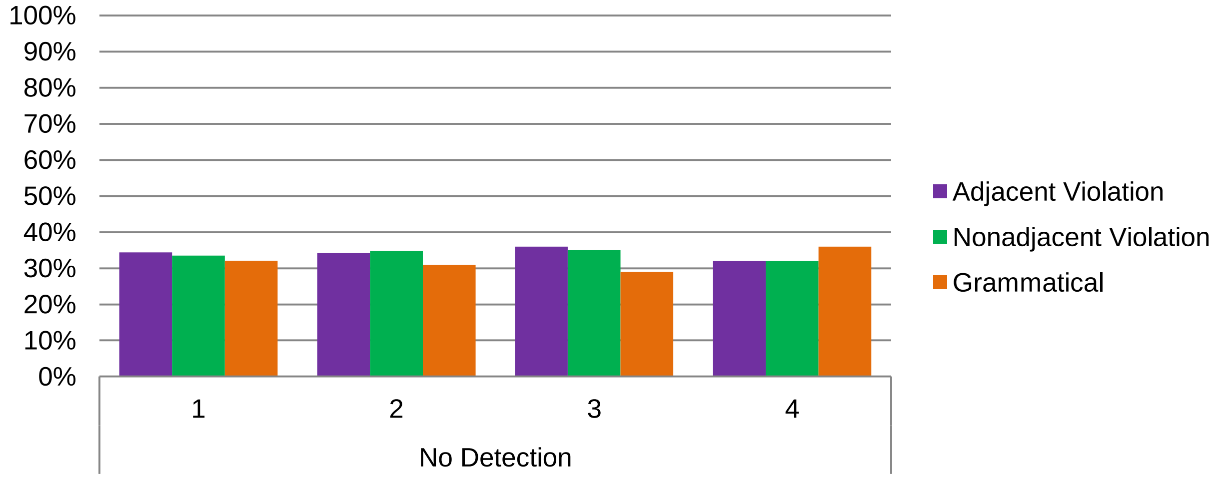

Figure A.1. No Detection Infants. Individual infants are plotted along the X-axis to show their proportion of total listening time to each trial type.

Table A.1

Characteristics of Infants Classified as No Detection

	Age	Words Produced at Time of Study	Total Time to Habituate	Trials to Habituate	Other
1	13.36	6	60	7	male
2	18.43	25	179	8	male; influential data point (see Section 3.3.5)
3	13.36	0	102.7	16	female
4	12.99	3	75.4	4	female
***N* = 4**	***M* = 14.54**	***M* = 8.5**	***M* = 104.28**	***M* = 8.8**	**female (N = 2)**

CHARACTERISTICS OF INFANTS CLASSIFED AS DETECT ADJACENT CATEGORY VIOLATIONS

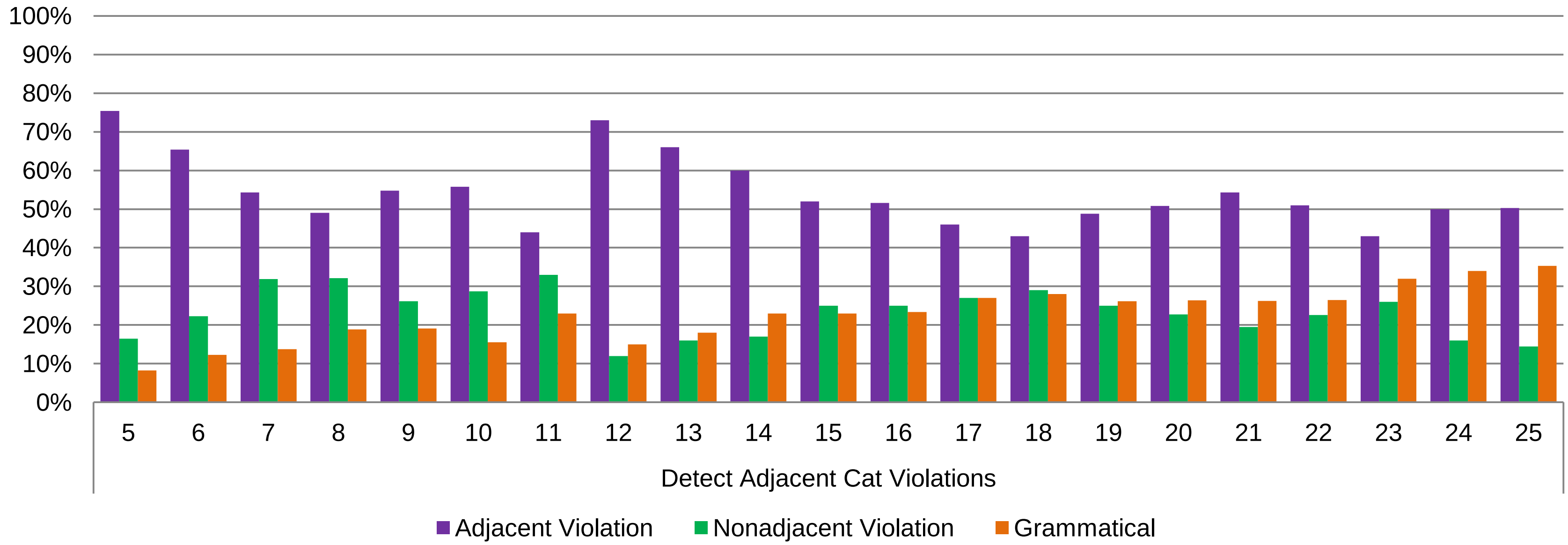

Figure A.2. Detect Adjacent Category Violations Infants. Individual infants are plotted along the X-axis to show their proportion of total listening time to each trial type.

Table A.2.

Characteristics of Infants Classified as Detect Adjacent Category Violations

	Age	Words Produced at Time of Study	Total Time to Habituate	Trials to Habituate	Other
5	16.46	25	93.7	4	female
6	13.89	9	89.9	4	female
7	13.03	4	92.3	7	male
8	14.39	15	47.5	6	female
9	14.72	4	144.8	8	female
10	15.72	7	55.7	7	female
11	13.36	8	36	7	male
12	18.26	13	143	10	male
13	18.33	30	105.4	4	female
14	15.46	23	56.2	6	female
15	12.79	4	37.1	5	female
16	12.26	2	46.4	10	male
17	12.89	10	157.5	15	female
18	14.89	12	138.8	5	male
19	18.46	20	26.9	4	male
20	14.1	12	64.7	5	male
21	15.13	18	72.2	6	female
22	17.56	23	51.6	4	female
23	15.49	18	74.6	7	female
24	18.39	50	95.4	4	female
25	14.89	4	24	5	male
N = 21	***M* = 15.26**	***M* = 14.81**	***M* = 78.75**	***M* = 6.3**	**female (N = 13)** **novelty (N = 21)**

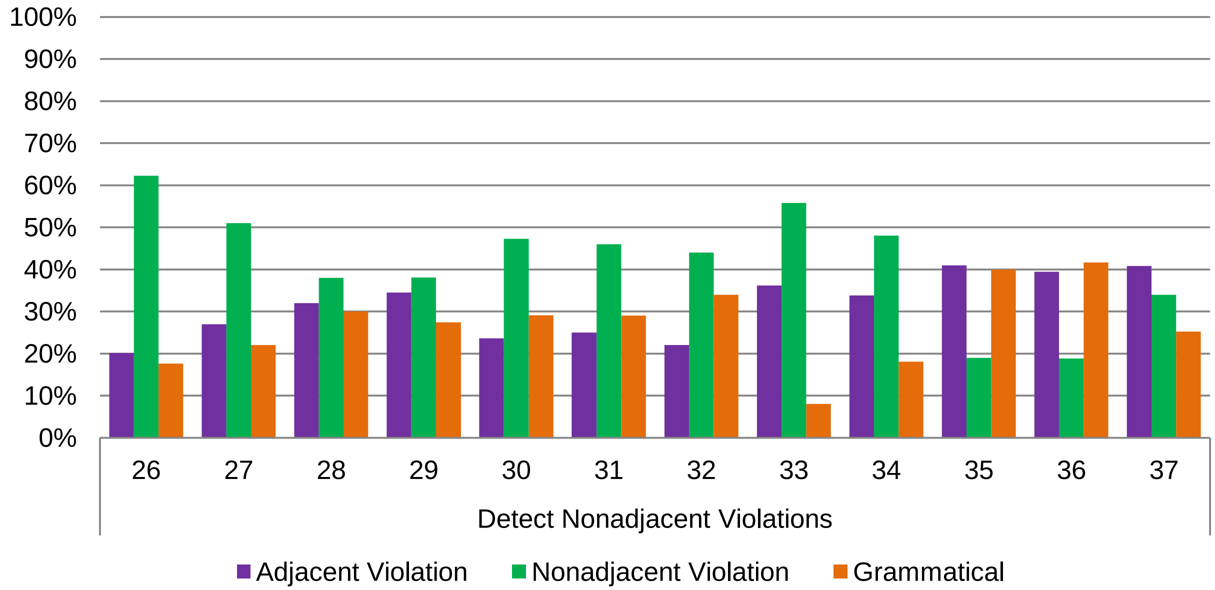

Figure A.3. Detect Nonadjacent Violations Infants. Individual infants are plotted along the X-axis to show their proportion of total listening time to each trial type.

Table A.3

Characteristics of Infants Classified as Detect Nonadjacent Violations

	Age	Words Produced at Time of Study	Total Time to Habituate	Trials to Habituate	Other
26	18.03	10	78.6	5	male
27	17.59	8	125.2	6	male
28	13.69	7	42	7	female
29	14.76	10	176.1	7	male
30	12.66	0	41.1	6	female
31	12.92	4	35.3	7	female
32	12.92	5	27	4	male
33	12.92	1	202.9	7	female
34	15.1	5	57.5	4	female
35	13.66	2	79.7	9	male; familiarity
36	15.03	5	131.6	5	female; familiarity
37	14.95	14	144.7	6	female
***N* = 12**	***M* = 14.52**	***M* = 5.92**	***M* = 95.14**	***M* = 6.1**	**female (N = 7)** **novelty (N = 10)**

CHARACTERISTICS OF INFANTS CLASSIFED AS DETECT NONADJACENT AND CATEGORY VIOLATIONS

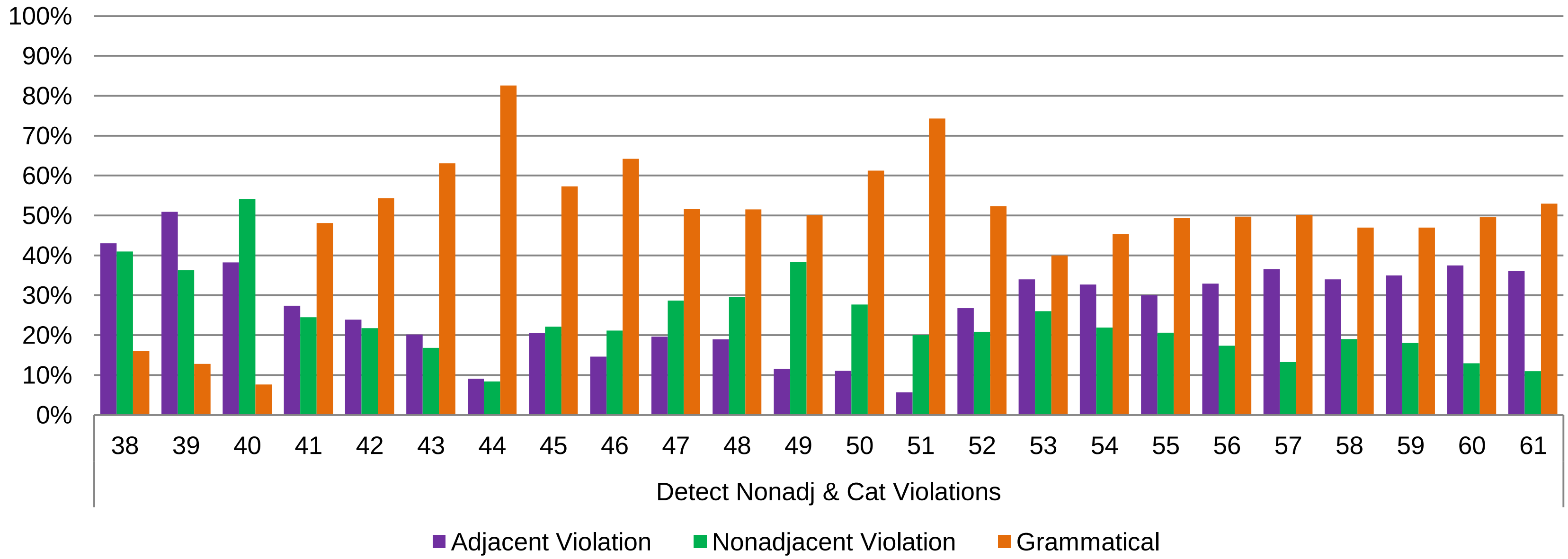

Figure A.4. Detect Nonadjacent and Category Violations Infants. Individual infants are plotted along the X-axis to show their proportion of total listening time to each trial type.

Table A.4.

Characteristics of Infants Classified as Detect Nonadjacent and Category Violations

	Age	Words Produced at Time of Study	Total Time to Habituate	Trials to Habituate	Other
38	13.33	9	128.8	9	male; novelty
39	18.1	45	89	9	male; novelty
40	14.2	10	152.8	17	female; novelty
41	14.92	17	96.1	5	male
42	18.1	100	291.4	8	male
43	17.85	30	49.8	4	female
44	18.33	15	180.9	8	male
45	15.3	10	98.4	8	male
46	18.33	--	86.6	4	female
47	15.03	18	58.1	4	male
48	18.92	7	37	4	male
49	18.16	13	188.1	6	male
50	18.3	9	87	4	female
51	13.7	10	87.4	6	male
52	17.69	140	152.6	5	female
53	13.1	25	74.9	5	male
54	18.13	20	85	5	female
55	16.00	6	161.1	9	male
56	14.33	8	313.2	9	male
57	14.69	4	71.5	5	male
58	13.26	--	105.5	11	male
59	13.3	9	144	11	male
60	17.79	17	68.4	8	female
61	18.56	30	338.3	11	female
***N* = 24**	***M* = 16.23**	***M* = 25.09**	***M* = 131.08**	***M* = 7.3**	**female (N = 8) novelty (N = 3)**

REFERENCES

Arterberry, M. E., & Bornstein, M. H. (2002). Variability and its sources in infant categorization. *Infant Behavior & Development*, *25*, 515-528.

Bates, E., & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language and Cognitive Processes*, *12*, 507–584.

Bernal, S., Lidz, J., Millote, S., & Christophe, A. (2007). Syntax constrains the acquisition of verb meaning. *Language Learning and Development*, *3*, 325–341.

Braine, M. D. S. (1987). What is learned in acquiring word classes—A step toward an acquisition theory. In Brian MacWhinney (Ed.), *Mechanisms of Language Acquisition* (pp. 65-87). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, *14*, 333-352.

Chemla, E., Mintz, T.H., Bernal, S., Christophe, A. (2009). Categorizing words using 'frequent frames': what cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental Science*, *12*, 396-406. doi: 10.1111/j.1467-7687.2009.00825.x

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2000). Habit 2000: A new program for testing infant perception and cognition. (Version 2.2.5c) [Computer software]. Austin, TX: The University of Texas.

Conwell, E., & Morgan, J. L. (2012). Is it a noun or is it a verb? Resolving the ambicategoricality problem. *Language Learning and Development*, *8*, 87-112.

Curtin, S., Campbell, J., & Hufnagle, D. (2012). Mapping novel labels to actions: How the rhythm of words guides infants' learning. *Journal of Experimental Child Psychology*, *112*, 127-140.

Erkelens, M. A. (2009). *Learning to Categorize Verbs and Nouns: Studies on Dutch*. Utrecht, The Netherlands: LOT.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences of the USA*, *103*, 12203-12208.

Fenson, L., Dale, P., Reznick, J.S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, Serial No. 242. Vol 59, No. 5.*

Fernald, A., & Marchman, V.A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Children Development*, *83*, 203-222.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*, 98-116.

Fisher, C., Klinger, S.L., & Song, H. (2006). What does syntax say about space? 2-year-olds use sentence structure to learn new prepositions. *Cognition, 101*, B19-B29.

Frigo, L., & McDonald, J. L. (1998) Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language*, *39*, 218-245.

Gerken, L.A., Balcomb, F. K., Minton, J. L. (2011). Infants avoid 'labouring in vain' by attending more to learnable than unlearnable linguistic patterns. *Developmental Science*, *14*, 972-979.

Gerken, L.A., Wilson, R., & Lewis, W. (2005). Seventeen-month-olds can use distributional cues to form syntactic categories. *Journal of Child Language, 32*, 249–268.

Gilpin, A. R. (1993). Table for conversion of Kendall's Tau to Spearman's Rho within the context of measures of magnitude of effect for meta-analysis. *Educational and Psychological Measurement*, 53, 87-92.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431–436.

Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction in 12-month-old infants. *Developmental Science, 7*, 567–580.

Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7*, 183–206.

Graf-Estes, K., & Bowen, S. (2013). Learning about sounds contributes to learning about words: Effects of prosody and phonotactics on infant word learning. *Journal of Experimental Child Psychology*, *114*, 405-417.

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 & Development*, *2*, 277-300. doi:10.1207/s15473341lld0204_3

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*, 341-353.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, *99*, 349-364.

Kelly, M. H. & Bock, J. K (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 389-403.

Lany, J., & Saffran, J.R. (2011). Interactions between statistical and semantic information in infant language acquisition. *Developmental Science*, *14*, 1207-1219.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*, 193–198.

Maratsos, M. (1990). Are actions to verbs as objects are to nouns? On the differential semantic bases of form, class, category. *Linguistics*, *28*, 1351-1380.

Maratsos M., & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In: Nelson K, ed. *Children's Language Vol. 2.* New York: Gardner Press, 127-213.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition*, *30*, 678-686.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*, 91-117.

Mintz, T. H. (2006). Finding the verbs: Distributional cues to categories available to young learners. In: Hirsh-Pasek K, Golinkoff R. M., Eds. *Action Meets Word: How Children Learn Verbs*. NewYork: Oxford University Press, 31–63.

Mintz, T. H. (2011a). Comparing the efficacy of bigrams and frames in cuing lexical categories for human learners. Poster presented at: Expanding the Space of Cognitive Science. 33rd Annual Cognitive Science Conference, 2011 Jul 20-23, Boston, MA.

Mintz T. H. (2011b). Fifteen-month-old infants can categorize words using distributional information alone and retain the categories after 1 week. *Paper presented at the 36th Annual Boston University Conference on Language Development,* Boston, MA.

Monaghan, P., Chater, N., & Christiansen, M.H. (2005). The differential role of phonological and distributional cues in grammatical categorization. *Cognition, 96,* 143-182.

Monaghan, P., & Christiansen, M. H. (2004). What distributional information is useful and usable for language acquisition. In *Proceedings of the 26th annual conference of the Cognitive Science Society* (pp. 963-968).

Monaghan, P., & Christiansen, M. H.(2008). Integration of multiple probabilistic cues in syntax acquisition. In Shanley, E., Allen, M., Rowland, C. F., De Houwer, A. & Gillis, S. (Series Eds.) and H. Behrens (Vol. Ed.), *Trends in language acquisition research: Vol. 6. Corpora in language acquisition research: History, methods, perspectives* (pp. 139-163). Amsterdam: John Benjamins.

Monaghan, P., Christiansen, M.H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, *55*, 259-305.

Morgan, J.L, Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Towards a broader conceptualization of bootstrapping. In: Morgan JL, Demuth K, Eds. *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates, 263-283.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language, 17*, 357–374.

Nazzi, T., Barrière, 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*, 119-135. doi:10.1016/j.cognition.2011.03.004

Newport, E. L. (1988). Constraints on learning and their role in language acquisition. *Language Science*, *10*, 147–172.

Newport EL. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*, 11–28.

Onnis, L., Christiansen, M. H, Chater, N., & Gómez, R. L. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial language learning. In: Alterman, R., & Kirsh, D. Eds. *Proceedings of the 25th Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum, 886-891.

Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In: Forbus, K., Gentner, D., Regier, T., Eds. *Proceedings of the 26th Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum, 1047-1052.

Pacton, P., & Perruchet P. An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *34*, 80-96. doi:10.1037/0278-7393.34.1.80

Peña, M., Bonatti, L. L, Nespor, M., & Mehler, J. (2002). Signal driven computations in speech processing. *Science*, *298*, 604-607.

Pine, J. M., & Lieven, E. V. M. (1997). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics*, *18*, 123-138.

Pinker, S. (1984). Language learnability and language development. Cambridge, MA: Harvard University Press.

Redington, M., Chater, N., & Finch, S. P. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*, 425-469.

Robertson, E. K., Shi. R, & Melançon, A. (2012). Toddlers use the number feature in determiners during online noun comprehension. *Child Development, 83*, 2007-2018.

Romberg, A. R., & Saffran, J. R. (2013). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science. Advance online publication.* doi:10.1111/cogs.12050.

Sandhofer, C., Smith, L. B., & Luo, J. (2000). Counting nouns and verbs in the input: Differential frequencies, different kinds of learning? Journal of Child Language, 27, 561-585.

Sandoval, M., Gonzales, K., & Gómez, R. L. (2012). The road to word class acquisition is paved with distributional and sound cues. In P. Rebuschat & J. Williams (Eds.), *Statistical Learning & Language Acquisition*, pp. 145-170. De Gruyter Mouton Press.

Santelmann, L. M, & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition*, *69*, 105–134.

Sereno, J. A., & Jongman, A. (1990). Phonological and form class relations in the lexicon. *Journal of Psycholinguistic Research*, *19*, 387-404.

Shi, R., & Melançon, A. (2010). Syntactic categorization in French-learning infants. *Infancy, 15*, 517-533.

Shi, R., Morgan, J. L., & Allopena, P. (1998). Phonological and acoustic bases for early grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language*, *25*,169-201.

Shi, R., Werker, J., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. Cognition, 72, B11-21.

Snedeker, J., Geren, J., & Shafto, C. (2007). Starting over: International adoption as a natural experiment in language development. *Psychological Science, 18*, 79–87.

Snedeker, J., Geren, J., & Shafto, C. (2012). Disentangling the effects of cognitive development and linguistic expertise: A longitudinal study of the acquisition of English in internationally-adopted children. *Cognitive Psychology*, *65*, 39-76.

Soderstrom, M., Wexler, K., & Jusczyk, P.W. (2002). English-learning toddlers' sensitivity to agreement morphology in receptive grammar. In Skarabela, B., Fish, S. A, Do, A. H. J, Eds. *Proceedings of the 26th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 643–652.

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,* 1-29.

St. Clair, M. C., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition*, *116*, 341-360.

Stumper, B., Bannard, C., Lieven, E., & Tomasello, M. (2011). ‘‘Frequent frames’’ in German child-directed speech: A limited cue to grammatical categories. *Cognitive Science, 35*, 1190-1205. doi: 10.1111/j.1551-6709.2011.01187.x

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, *74,* 209 - 253.

Valian, V. (1986). Syntactic categories in the speech of young children. *Developmental Psychology, 22*, 562-579.

Van Heugten, M., & Johnson, E. K. (2010). Linking infants' distributional learning abilities to natural language acquisition. *Journal of Memory & Language*, *63*,197-209. doi:10.1016/j.jml.2010.04.001

Van Heugten, M., & Shi, R. (2010). Infants' sensitivity to non-adjacent dependencies across phonological phrase boundaries. *Journal of the Acoustical Society of America, 128,* EL223-EL228. doi:10.1121/1.3486197

Voloumanos, A. & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology, 45*, 1611-1617.

Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2011). Simultaneous online tracking of adjacent and non-adjacent dependencies in statistical learning. In: Carlson, L., Hoelscher, C., & Shipley, T. F., Eds. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum, 964-969.

Wang, H., Höhle, B., Ketrez, N. F., Küntay, A. C., & Mintz, T. H. (2011). Cross-linguistic distributional analyses with frequent frames: The cases of German and Turkish. In: Danis, N., Mesh, K., Sung, H., Eds. *Proceedings of the 35th annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 628-640.

Weisleder, A., & Waxman, S. R. (2010). What's in the input? Frequent frames in child-directed speech offer distributional cues to grammatical categories in Spanish and English. *Journal of Child Language, 37*, 1089-1108.