Starting Over:
International Adoption as a Natural Experiment in Language Acquisition

Jesse Snedeker (snedeker@wjh.harvard.edu)
Joy Geren (geren@wjh.harvard.edu)
Department of Psychology, 33 Kirkland St. 
Cambridge, MA 02138 USA

Abstract

Language development is characterized by predictable shifts in the words that children learn and the complexity of their utterances. Because language acquisition typically occurs simultaneously with cognitive development and maturation, it is difficult to determine the causes of these shifts. We explored how acquisition precedes in the absence of possible cognitive or maturational roadblocks, by examining the acquisition of English in internationally-adopted preschoolers. Like infants, and unlike other L2 learners, these children acquire a language from child-directed speech in the home, without access to bilingual informants. Parental reports (CDI-2) and speech samples were collected from 14 preschoolers, 3 to 18 months after they were adopted from China. These children made rapid progress in acquiring English and showed the same developmental patterns as monolingual infants (matched for vocabulary size). Early on, their lexicons were dominated by nouns, their utterances were short, and function morphemes were almost entirely absent. Children at later stages of development had more diverse lexicons and produced longer utterances with more closed-class morphemes.

Introduction

Language development is marked by a series of qualitative shifts. Infants speak in single-word utterances for several months before beginning to combine words. Young children learn a disproportionate number of nouns before acquiring a balanced complement of verbs, adjectives and prepositions. Young English speakers typically omit function morphemes from their early word combinations, and then gradually begin to add them in. A central question in language acquisition is what causes children to move through these phases (Lenneberg, 1967, Bloom, 1973, Gleitman & Newport, 1995, Ganger, Pinker, Chawla & Baker, 2004). Are the early stages reflections of cognitive immaturity or do they represent necessary steps in decoding the target language? Is the emergence of new linguistic abilities driven by maturation or by the child’s growing knowledge of the language? These questions are difficult to answer because language acquisition is confounded with cognitive development and maturation in typically developing children.

However much of what we know about language development comes from studies of atypical populations. Research on aphasics suggested there might be a critical period for language acquisition (Lenneberg, 1967), a hypothesis confirmed by studying adults deprived of linguistic experience in childhood (Newport, 1990). Research on blind children (Landau & Gleitman, 1985), linguistic isolates (Goldin-Meadow & Feldman, 1977), and children with limited language models (Singleton & Newport, 2004) has been central in demonstrating the resilience of children’s linguistic abilities. These natural experiments allow us to explore the effects of variables that are impossible or unethical to manipulate.

International adoption provides the opportunity for just such a natural experiment. Over 20,000 internationally-adopted children enter the U.S. each year. While most are infants or toddlers, thousands of older children are also adopted. Typically these children rapidly lose their birth language (Glennen & Masters, 2002). By adulthood, they fail to distinguish speech in their birth language from an unknown language and show no cortical activations that distinguish them either (Pallier et al., 2003). Most of these children are adopted well within the sensitive period for language acquisition (Newport, 1990) and become fluent speakers of their new language (Pallier, et al., 2003). But we know almost nothing about how they get there. Recent studies show that internationally-adopted infants make rapid progress in acquiring English (Glennen & Masters, 2002). But there is no existing data on older adoptees.

The learning problem faced by these children is broadly similar to that of infants learning their first language: they are exposed to child-directed speech in the context of daily routines; they must learn the new language to communicate with their families; they have little access to text or bilingual informants; and they lack many of the metalinguistic abilities available to older children and adults (Gombert, 1992). However, these children are more cognitively and physically mature than their infant counterparts and have already started to learn one language.

The current study compares language acquisition in internationally-adopted preschoolers and monolingual infants. Our goal is to explore the role that cognitive development and maturation play in shaping the course of first-language acquisition by examining how acquisition proceeds when these road blocks have been removed. By doing this we hope to tease apart two very broad kinds of explanation for systematic changes during language acquisition:

1) Maturational-Cognitive Hypotheses: Theories of this kind attribute the order of acquisition or the emergence of new abilities to changes in the learner which are independent of the child’s experience with
a given language. Immaturity constrains language acquisition, limiting the kinds of words that a child can learn, the kinds of representations she can create or the kinds of utterances she can produce. When these roadblocks are removed, either by maturation or cognitive development, children can acquire new linguistic abilities.

2) **Contingent Acquisition Hypotheses:** These theories attribute the order of acquisition to the interdependence of different linguistic elements, representations, or processes. The emergence of new abilities is driven by the child’s growing knowledge of the language. If knowledge of form A is necessary for acquiring form B, then the acquisition of B will have to await the acquisition of A.¹

Critically, this distinction is completely orthogonal to the nativist/empiricist and domain-specific/domain-general dichotomies that organize theoretical discourse in this field. Contrast, for example, Wexler’s maturational theory (1998) with the cognitive theory of Shore (1986) or the nativist contingent acquisition hypothesis of Snedeker & Gleitman (2004) with the proposal of Bates & Goodman, (1997).

This paper explores just two patterns in early language development that could be explained by either maturational-cognitive hypotheses or contingent acquisition. We begin by describing these two patterns, then we briefly examine why prior research on second-language acquisition has not resolved these issues.

**Changes in Vocabulary Composition**

Children’s early vocabularies are dominated by nouns that refer to people, animals, and moveable objects. Although adults speak to children in full sentences, complete with verbs and function words, these elements are massively underrepresented in children’s early vocabularies (Gentner, 1982, Bates, Dale & Thal, 1995). This is true not only in English but also in languages like Mandarin and Korean where verbs frequently occur in perceptually salient positions (for review see Gentner & Boroditsky, 2001). This input-output disparity can be plausibly attributed to the conceptual limitations of young children (Macnamara, 1972; Huttenlocher, Smiley & Ratner, 1983). Perhaps the relative dearth of verbs and adjectives is attributable to the infant’s inability to conceive of relations, states or actions, while the overabundance of nouns is attributable to the conceptual primacy of object categories.

However, the changing composition of children’s lexicons could reflect linguistic rather than conceptual growth (Gillette, Gleitman, Gleitman, & Lederer, 1999; Snedeker & Gleitman, 2004). An infant who is just breaking into language has to learn the meanings of words by observing the situational contexts in which they are used. Older children, who have already acquired sizeable vocabularies, can also use the sentence in which the word appears.² To simulate the effects of linguistic development in the absence of cognitive limitations, Gleitman and her colleagues have asked adults to identify words from different representations of the contexts in which they occurred in child-directed speech (Gillette et al., 1999; Snedeker, Li & Yuan, 2003; Snedeker & Gleitman, 2004). When the adults were limited to situational cues, they could only identify the concrete nouns. But when given information about the linguistic context, they were able to learn the verbs as well.

These human simulations demonstrate that changes in vocabulary composition are not necessarily attributable to changes in the learner’s conceptual repertoire. But there are several differences between these studies and the experiences of young language learners which might limit the validity of the simulations. In contrast, the task and input of internationally-adopted children appear to closely parallel those of infant learners. Like infants, adopted children get prolonged exposure to their new language in the context of meaningful social interactions. Like infants they must simultaneously isolate the words and determine what they mean. However, like the adults in Gleitman’s simulations, adopted children are more cognitively mature than infants.

If shifts in vocabulary composition primarily reflect the changing cognitive capacities of the learner, then adopted children should acquire words from a variety of categories, much like their monolingual age mates. If vocabulary composition is largely a function of the input and children’s linguistic knowledge, adopted children should initially be restricted to the types of words learned by infants.

**Early Grammatical Development**

Similar questions have been raised about the role of maturation and cognitive development in children’s early combinatorial speech. For months after they begin speaking, infants are typically limited to one-word utterances. The appearance of word combinations has been attributed to motor and cognitive development and linguistic maturation, as well as the accumulation of linguistic knowledge (Bloom, 1973; Shore, 1986, Bates, et al., 1995). At around 24 to 30 months, children show a second burst of syntactic activity, adding determiners, auxiliaries and inflectional markers to their formerly sparse utterances (Brown, 1973).

Both of these shifts are strongly correlated with productive vocabulary size raising the intriguing possibility that lexical growth is causally related to syntactic development (Bates & Goodman, 1997). In support of this hypothesis, Bates and her colleagues demonstrated that

¹ The contingent acquisition hypotheses under consideration make the weaker claim that one type of knowledge is needed for efficient acquisition or utilization of another type. This is desirable since the phenomena under consideration under consideration are strong but violable. For example, children do learn some verbs early on (Bates, Dale & Thal, 1995).

² Note we are not drawing a contrast between conceptual and linguistic explanations for early noun-dominance. Both hypotheses claim that verbs are difficult to learn because of the kinds of concepts that they encode. The theories differ in their explanations for why this conceptual difference impedes acquisition and how this changes over development.
these correlations hold up in atypical populations such as early-talkers, late-talkers and children with Williams Syndrome (Bates & Goodman, 1997). However, these studies cannot rule out the possibility that both lexical and syntactic acquisition depend upon the development of some other cognitive ability, one which is accelerated for early-talkers, delayed for late-talkers, and selectively spared in Williams Syndrome (e.g., auditory memory). We can test this hypothesis by examining the relation between lexical and grammatical development in adopted preschoolers. If they are causally linked, then the relationship should persist in maturationally-advanced learners. In contrast if the correlation is created by rate-limiting development in another domain, then it should be possible to find disassociations in older learners.

Comparing first and second language acquisition

Many other researchers have explored the parallels between first and second language acquisition, finding both similarities and discrepancies (see Clahsen, 1990; Ellis, 1994; Freeman & Freeman, 2001 for discussions). But none of this existing work addresses the questions that motivate this study. L2 researchers have typically focused on the development of specific syntactic constructions that appear somewhat later in first-language development (e.g., negation). There is little work on the composition of the lexicon in L2 learners and no work that examines the relationship between early lexical and grammatical development. Furthermore, the most commonly studied populations, students receiving formal instruction or immigrants learning a language in the workplace or playground, are in language environments that are radically different from infants, making it unclear whether differences in acquisition are due to maturity or to differences in the learners’ input and motivations.

Finally, most L2 studies have examined adults or children over six. Our goal is to find out whether cognitive changes occurring between 16 and 30 months of age shape early language acquisition. When we compare infants with adults we cannot isolate these effects from age-related changes that occur during middle childhood and adolescence. Since these later changes are known to alter acquisition (see e.g., Johnson & Newport, 1989), we have chosen to limit our study to children who begin acquiring English before 6.

Methods

14 families with children adopted from China participated. The children were adopted between 2;7 and 5;1 (M = 4;2). Length of residence at the first session was 3 to 16 months (M = 8.1). Parents were invited to participate every 3 months until their child had been in the U.S. for 18 months. Thus each child participated in 1 to 5 sessions (M=2.5) and a total of 35 data points were collected. Children were excluded from participation if they had sensory, motor, or developmental disorders that might affect language.

Materials for the study were mailed to the parents who completed the Bates-MacArthur Communicative Development Inventory 2 (CDI-2) (Fenson, et al., 1993) and recorded a language sample in their home. The CDI-2 includes a 680-item vocabulary checklist and a 37-item forced-choice sentence-complexity measure, which asks about the child’s use of inflectional morphemes and closed-class words. The CDI-2 is normed for children 16-30 months (Fenson, et al., 1993), but has also been used with older children with limited English language skills (Berglund, Eriksson & Johansson, 2001). The language sample consisted of an hour-long recording of the parent and child playing with a standard toy set.

Parental reports for the adopted preschoolers were compared with those of monolingual infants who participated in earlier studies. Each session from an adopted child was matched to an infant with the same vocabulary size (±7%) on the parental report. The infant controls were 1;7 to 2;6 (M = 2;2).

Results

Children’s Spontaneous Speech

The speech sample from the first session was transcribed, and the first 100 utterances meeting standard criteria were analyzed using CLAN (MacWhinney, 2000). These analyses validate the use of the CDI-2 with this population. The number of word types in the speech sample is highly correlated with CDI-2 vocabulary size (R²=.56, p <.001). Parental reports also accurately reflect the kinds of words that the children use; the number of different nouns, verbs, and closed class items used by each participant correlates with the number that their parent endorsed on the CDI-2 (R²=.28, p=.053 for nouns; R²=.43, p <.05 for verbs; and R²=.66, p <.001 for closed-class words). Furthermore, the children’s MLU is correlated with their score on the CDI-2 sentence-complexity metric, demonstrating that parents were sensitive to differences in the children’s syntactic abilities (R²=.37, p <.05).
Rate of Acquisition

Since the number of sessions varied across participants, we conducted the CDI-2 analyses on the first data point contributed by each child, as well as on the entire data set. The results of the two analyses were quite similar and statistics for both are presented below.

Unsurprisingly, vocabulary size increased with the length of time that the child had spent in the U.S. As Figure 1 suggests, one participant was an outlier in this analysis, acquiring reliably fewer words than would be predicted at both observation sessions. With this participant removed, there is a robust logarithmic relation between time and vocabulary ($R^2=.64$, $p<.001$ first session, $R^2=.61$, $p<.0001$ all). Age of arrival was not a significant predictor of vocabulary size ($R^2<.01$, $p’s>.25$) indicating that older adoptees did not learn words any faster than young ones.

This logarithmic, or decelerating, growth curve contrasts sharply with the accelerating curve observed in infant learners (Fenson et al., 1994). We suspect that this reflects properties of the instrument rather than an actual deceleration in vocabulary growth. After 12 months many adoptees have developed beyond the point where the CDI provides an accurate estimate of their vocabulary size; in 54% of these sessions the children knew over 90% of the words and had presumably reached the ceiling of this instrument. The secondary y-axis of Figure 1 allows us to compare vocabulary growth in the adopted children to the CDI-2 norms for infant learners (Fenson et al., 1993). After 3 months in the U.S., adopted preschoolers have vocabulary scores that rival 24-month-olds, who have been speaking for about a year. Thus, internationally-adopted preschoolers initially acquire a productive vocabulary at roughly four times the rate of infants. This suggests that development or prior experience with a language can accelerate the initial pace of word learning. But we have no evidence that this advantage persists. Between month 3 and month 9, adopted children make as much progress as the average infant does between 24 and 30 months, though ceiling effects in both groups make this finding difficult to interpret.

Vocabulary Composition

When children are near the ceiling of the CDI-2, vocabulary composition necessarily reflects the composition of the checklist. To ensure this did not unduly influence our findings, we removed observations where children had acquired over 90% of words from the analyses of all sessions. The remaining sample included 23 observations.

The adopted preschoolers show the same shifts in vocabulary composition as the infant learners (Figures 2-4). The proportion of nouns decreases linearly as the children’s vocabularies grow ($R^2=.49$, $p<.005$ first session, $R^2=.54$, $p<.001$ all). In contrast the proportion of verbs increases.

---

3 When this participant is included, her observations have standardized residuals less than -2 and the logarithmic relation between time and vocabulary is considerably weaker ($R^2=.21$, $p=.06$ first session, $R^2=.24$, $p<.005$ all). This child did not appear to have any perceptual limitations, cognitive delays, or background characteristics that might explain her relatively slow acquisition of English. We suspect she may have a language impairment.
logarithmically with vocabulary size ($R^2=.39$, $p < .05$ first, $R^2=.50$, $p < .001$ all) and the proportion of closed-class items increases linearly ($R^2=.72$, $p < .001$ first, $R^2=.69$, $p < .001$ all). For each lexical class we compared the adoptees and controls with a paired t-test and a stepwise regression with vocabulary size, group, and group * vocabulary as predictors. We found no reliable differences between the adoptees and infant controls in vocabulary composition or its relation to vocabulary size ($t's<1.1$, $p's >.25$).

**Grammatical Development**

Infants typically begin combining words when they have a vocabulary of 50 to 200 words (Bates & Goodman, 1997). Since all of our participants had vocabularies over 150 words and were combining words, we were unable to examine whether these events were linked in the adopted preschoolers. Instead, our analyses focused on the relation between vocabulary size and sentence-complexity scores, which are robustly correlated in infants when CDI-2

![Graph showing the relationship between CDI-2 vocabulary size and sentence complexity scores.](image)

Figure 5: Total Score on Sentence Complexity Measure as a Function of Vocabulary Size

vocabulary is between 300 to 500 words (Bates & Goodman, 1997). As Figure 5 illustrates, this correlation was confirmed in both the adopted children ($R^2=.60$, $p < .001$ first, $R^2=.67$, $p < .001$ all) and the controls and there were no reliable differences between the two groups ($t's<1$, $p's >.25$).

**Discussion**

We found that internationally-adopted preschoolers go through many of the same shifts in early language development as typically-developing infants. Like infants, these preschoolers initially learned a disproportionate number of nouns, developing a more balanced lexicon over time. Like infants, their early utterances lacked inflectional morphemes and closed-class words, which were gradually added in as vocabulary grew. In infants we can credit these shifts to general cognitive or maturational changes. The preschoolers however, have passed the developmental milestones that typically coincide with early language development. Thus our results strongly suggest that at least some features of early language production are due to the nature of the learning problem rather than the limitations of infant learners.

For over twenty years, research on early vocabulary composition has largely centered on Gentner’s noun-dominance hypothesis (1982). Thus it is reasonable to ask how this work bears on that hypothesis. We chose not to frame the study in this way, primarily because we believe that the noun-dominance hypothesis lives a dual life. Many readers, including most text-book writers, have interpreted it as a maturational-cognitive hypothesis. For example, Hoff states (2001, p.157): “According to Gentner the relational meanings encoded in verbs are less available to young children through nonlinguistic experience. Thus, children acquire nouns before verbs because the concepts encoded by nouns are earlier cognitive developments than the concepts encoded by verbs.” Our results clearly speak to this interpretation of the noun-dominance hypothesis. Presumably preschool-aged children with prior linguistic experience have developed the concepts encoded by verbs that are typically learned by toddlers. Yet they go through the same shifts in vocabulary composition, suggesting that there is no need to invoke cognitive change to explain this phenomenon.

In subsequent writings, however, Gentner herself has clearly rejected this maturational/cognitive account and has argued instead that the shift in vocabulary composition is caused by the child’s growing knowledge of the language-specific conflation patterns that organize verb meanings (Gentner & Boroditsy, 2001). While this is clearly a contingent acquisition hypothesis, it is not clear what predictions it makes for second-language acquisition in childhood. If children simply attempt to map second-language labels onto the conceptual confluations provided in their first language, then we would expect to see precocious verb learning, to the degree that verb semantics in the two languages are aligned. We found no evidence of this in the adopted children, despite the fact that many of the common verbs in English and Chinese languages appear have similar meanings (Snedeker, Li & Yuan, 2003). On the other hand, if child learners attempt to map second-language labels directly to prelinguistic representations of event components, then the noun-dominance hypothesis would predict that second-language verb learning, like first language verb learning, should initially be slow and effortful, accelerating as the child learns more about the language and its conflation patterns. The current data are consistent with this reading of noun-dominance hypothesis as well as Gleitman’s informational-change hypothesis.

Our ongoing research extends this work in several ways. First, we are collecting data from Eastern-European

---

4 We confirmed this by having some families complete parental reports of developmental milestones that typically coincide with language acquisition. The internationally-adopted preschoolers (n=9) passed 88% of these milestones, while the vocabulary-matched infants (n=9) passed only 49%.
adoptees to explore whether the child’s birth language affects the acquisition of English. Second, we are conducting a longitudinal study of the first 6 months after arrival, so we can learn more about the earliest phases of acquisition and track the development of individual children. Finally, we are conducting fine-grained transcript analyses to get a more detailed picture of the children’s syntactic development.

Acknowledgments

We thank Isabel Martin and Laura Ax for assistance in recruitment and parent contact; Caroline Whiting, Nicole Gavel and Rachel Springer for help with transcription and data entry; and Lila Gleitman for conversations which led to this study. Data collection was supported by the Roger Brown Fund. Preparation of this paper (and the ongoing work) is funded by National Science Foundation (BCS-0418423). Finally, we thank the families who participated for being so generous with their time.

References


