



ELSEVIER

Contents lists available at ScienceDirect

## Cognitive Psychology

journal homepage: [www.elsevier.com/locate/cogpsych](http://www.elsevier.com/locate/cogpsych)



# Getting it right by getting it wrong: When learners change languages

Carla L. Hudson Kam<sup>a,\*</sup>, Elissa L. Newport<sup>b</sup>

<sup>a</sup>Department of Psychology, University of California, Berkeley, 3210 Tolman Hall, #1650, Berkeley, CA 94705, United States

<sup>b</sup>Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, United States

### ARTICLE INFO

#### Article history:

Accepted 12 January 2009

Available online 25 March 2009

#### Keywords:

Language acquisition

Linguistic input

Probability learning

Miniature artificial languages

Second language learning

Regularization

### ABSTRACT

When natural language input contains grammatical forms that are used probabilistically and inconsistently, learners will sometimes reproduce the inconsistencies; but sometimes they will instead regularize the use of these forms, introducing consistency in the language that was not present in the input. In this paper we ask what produces such regularization. We conducted three artificial language experiments, varying the use of determiners in the types of inconsistency with which they are used, and also comparing adult and child learners. In Experiment 1 we presented adult learners with *scattered inconsistency* – the use of multiple determiners varying in frequency in the same context – and found that adults will reproduce these inconsistencies at low levels of scatter, but at very high levels of scatter will regularize the determiner system, producing the most frequent determiner form almost all the time. In Experiment 2 we showed that this is not merely the result of frequency: when determiners are used with low frequencies but in consistent contexts, adults will learn all of the determiners veridically. In Experiment 3 we compared adult and child learners, finding that children will almost always regularize inconsistent forms, whereas adult learners will only regularize the most complex inconsistencies. Taken together, these results suggest that regularization processes in natural language learning, such as those seen in the acquisition of language from non-native speakers or in the formation of young languages, may depend crucially on the nature of language learning by young children.

© 2009 Elsevier Inc. All rights reserved.

\* Corresponding author.

E-mail address: [clhudson@berkeley.edu](mailto:clhudson@berkeley.edu) (C.L. Hudson Kam).

## 1. Introduction

In the past half-century, great strides have been made in documenting the linguistic accomplishments of children as they learn their native languages. Despite this increased base of knowledge regarding *what* learners know as their language abilities develop, we are only beginning to understand *how* this learning takes place. This paper addresses questions about the learning mechanisms themselves, focusing on the limits of the system by examining cases where learners acquire something different than the patterns in their input.

We examine what human language learners can (and cannot) acquire when the input is abnormal in a particular way. Occasionally, language learners are exposed to input that contains grammatical inconsistency, what we call *probabilistic grammatical tendencies*: a form is used some percentage of the time in a particular context, with its occurrence not predictable on the basis of any features of the context. This kind of input is unusual in that probabilistic grammatical tendencies of this sort are not typically found in human languages. However, they do occasionally occur, for instance when learners are acquiring their language from non-native speakers. Evidence suggests that this unpredictable variation disappears as the language is learned; it is *regularized* (Newport, 1999; Ross & Newport, in prep; Singleton & Newport, 2004). The result of this change is a language that is no longer abnormal; the language as spoken by the learner is like other natural human languages. Learners, it seems, are able to 'fix' or repair this kind of abnormal input.

Here we examine how this change is accomplished, by examining the kinds of information that human learners extract from inconsistent linguistic input. Our interests are broader, however. We are not simply interested in characterizing the learning mechanisms involved in abnormal situations; we are interested in gaining a greater understanding of the mechanisms involved in language acquisition more generally. We submit that understanding the performance of learning mechanisms when the input is atypical can contribute to our understanding of the learning mechanisms involved in typical language acquisition as well; the input may be unusual, but the learners are not.

We present results from three experiments investigating how humans learn from languages containing inconsistent or probabilistic grammatical tendencies, asking about the circumstances under which they succeed at learning the variation veridically, and when and why they make the language more regular and more like other natural languages as it is learned. In previous research we have shown that, at least in one circumstance, adults reproduce the inconsistencies they are exposed to, while children do not. In the present work we investigate this question more comprehensively, examining both the age of learners and also the nature and complexity of the inconsistencies to which they are exposed, to see how these variables affect what is learned. In Experiment 1 we present adult learners with inconsistency that is either relatively simple or more complex, to investigate whether they might be less likely to learn veridically, and more likely to regularize, when the inconsistencies are complex. In Experiment 2 we present adult learners with a complex but consistent language, to see whether complexity in and of itself is enough to induce regularization in adult learners. In Experiment 3 we vary the age of the learner, testing both adults and children to investigate whether children are more likely to regularize a range of types of inconsistencies than are adults. To foreshadow, our results suggest that humans can learn from inconsistent linguistic input, but also that they do indeed make it more consistent under certain circumstances. Importantly, the degree to which they regularize depends on both the age of the learner and the presence and nature of the inconsistency. In the discussion we will return to the question of how our findings fit into broader issues, and particularly what we think our findings say about the mechanisms involved in language learning and the circumstances under which natural languages become less inconsistent and more regular.

### 1.1. Linguistic variation and accompanying change by learners

Typically, language learners are exposed to input that contains very consistent grammatical patterns. Sometimes these patterns are deterministic: a particular grammatical form is used every time a particular meaning is expressed. The regular plural form in English is an example of a deterministic

pattern (though even such patterns can have lexical exceptions).<sup>1</sup> Nouns take *-s*, *-z*, or *-iz* in the plural, depending on the phonological form of the noun (and those that are exceptions are always exceptional, in every context in which they occur). Other patterns are variable but are nonetheless grammatically predictable: the same form does not always occur with a particular meaning, but the variation among forms is contextually dependent and consistent across speakers (Labov, 1969, and papers in Chambers, Trudgill, & Schilling-Estes, 2003). As Smith, Durham, and Fortune (2007) put it, the alternation between forms is 'variable but highly structured'. For example, the pronunciation of *-ing* in English varies between *'-ing'* and *'-in,'* with *'-in,'* being the casual form, used in faster speech, less formal contexts, and more often by younger rather than older speakers. However, sometimes learners are exposed to linguistic input that contains grammatical patterns that are truly inconsistent tendencies. These patterns are unpredictable and probabilistic in nature; a form is used some percentage of the time in a particular context, with its occurrence or non-occurrence not predictable on the basis of any features of the context (see e.g. Newport, 1999; Singleton & Newport, 2004). We call these *inconsistent* or *probabilistic grammatical tendencies*.

Probabilistic grammatical tendencies are very common in the speech of late learners of a second language (Adamson, 1988; Goldowsky & Newport, 1993; Johnson, Shenkman, Newport, & Medin, 1996; Newport, 1990, 1999). Late learners are generally not as proficient as early learners. The most noticeable difference is often their accent, but late learners also have problems with grammatical devices like tense and aspect, agreement marking, and case marking (Birdsong, 1999; Johnson & Newport, 1989). Speakers who have learned a language as adults may simply omit the grammatical marking altogether, but often they will use a grammatical device inconsistently and probabilistically. This probabilistic usage can take different forms: the probabilistic usage of a single form; an unpredictable alternation between several forms (only one of which would be considered correct in the native form of the language); or a combination of the two, with the speaker sometimes using the correct form, sometimes an incorrect form, and sometimes using no form at all. Probabilistic usage is seen in second language interlanguage (during learning, Kanno, 1998), as well as in fossilized asymptotic second language grammars (Sorace, 2000). Thus, although probabilistic usage is atypical of mature native speakers, it is not simply a characteristic of language learning in progress. Furthermore, the specific probabilities and patterns of usage differ between individual second language speakers, even when they share the same native language, with the result that there may be no consistency across speakers within the same community (see, for example, Meisel, Clahsen, & Pienemann, 1981; Wolfram, 1985).

A question of interest is therefore how children might behave if they had to acquire their own native language from parents (or other adults) whose usage contained such inconsistencies. Children are known to have difficulty acquiring lexical exceptions to grammatical rules (Marcus et al., 1992). Linguistic input of the kind described above, with true inconsistencies, seems as if it should be particularly difficult to learn from. There are a few recorded instances of learners facing just this kind of input (see e.g. Aitchison, 1996; Kotsinas, 1988; Newport, 1999; Sankoff, 1994; Sankoff & Laberge, 1973; Singleton & Newport, 2004). The outcomes suggest that, while learners can acquire language from this kind of input, they do not acquire such inconsistent variation veridically. In contrast with the more typical situation of structured variation in the input, which is learned correctly (Kovac, 1981; Labov, 1989; Roberts, 1997; Smith et al., 2007), learners exposed to inconsistent input appear to change the language as they learn it, making it more regular.

#### 1.1.1.1. Children learning from non-native input

Singleton and Newport have conducted very detailed studies of the acquisition of American Sign Language (ASL) by a deaf child they called Simon, whose only input source was his deaf parents who were late learners of the language (Newport, 1999; Ross & Newport, 1996; Singleton, 1989; Singleton & Newport, 2004). Like other late learners of ASL, the parents' signing contained many errors and was governed by probabilistic, rather than deterministic, rules. That is, they would use complex morphemes each some percentage of the time in the obligatory context, with its occurrence or non-

<sup>1</sup> Although there is variation in the pronunciation of the plural form (*cats* versus *dogs* versus *canvases*), it is completely consistent and predictable, and depends on phonological form of the noun.

occurrence inconsistent and not predictable on the basis of features of the context. Simon did not reproduce the unpredictable inconsistency present in his input, however; he changed the system, using the forms most frequently present in his input almost categorically. This is most compellingly demonstrated by one particular regularization that Simon made. His parents most frequently used a handshape for vehicles that is not typical in ASL; Simon used their *incorrect* handshape consistently. As Singleton and Newport suggested, this result indicates that he is regularizing his parents' inconsistent system, and not somehow secretly learning ASL from another source (Singleton & Newport, 2004). He received variable input and made it far more consistent and predictable. Ross and Newport (1996 and in preparation; Ross, 2001) verified this claim in longitudinal analyses of Simon's ASL usage, and also demonstrated a similar outcome in three other deaf children learning ASL from their late-learning hearing parents.

Kotsinas (1988) reports on a similar case involving children who were immigrants to Sweden. These children lived in immigrant communities in Stockholm and learned their Swedish primarily from their parents and other immigrants to the community, all of whom were late learners of Swedish. According to Kotsinas, the parents' speech contained "considerable variation among the speakers' varieties" (p. 133). The children's productions, however, displayed more consistency, indicating the emergent grammaticalization of some of the forms present probabilistically in the parent's speech. Importantly, the non-standard Swedish the children spoke was not the same as the vernacular Swedish that ethnic Swedes speak – it was a modified version of the pidgin-like variety spoken by their parents. Notably, although some of the children also spoke Standard Swedish learned through exposure at school, their vernacular was well stabilized prior to exposure to the standard variety.

#### 1.1.2. *The language of the first generation of native speakers of a new language*

Similar phenomena also occur during the emergence of a new language. A number of researchers have documented the changes that occurred in two related pidgin languages as they were acquired by the first generation of native speakers. The first, Tok Pisin, a contact language spoken in Papua New Guinea, has been extensively studied by Sankoff and her colleagues, among others. They were particularly interested in the presence and form of grammatical devices (such as tense and aspect marking) and the clause and sentence structure in the speech of those who learned the language as adults, as compared with the speech of those who learned the language as a native language.

Though the language was in the early formational stages, they did find grammatical structures in the speech of the adults; but they also found some variability in the use of those structures. This variability took several forms, with two of these most relevant to the present work. First, as would be expected in a late-learned language, there was variation in occurrence: any particular form occurred in its appropriate context probabilistically (Aitchison, 1996; Sankoff, 1994).<sup>2</sup> Second, there were meanings that could be expressed in any of several ways (Sankoff, 1979).<sup>3</sup>

Importantly, the speech of the children learning Tok Pisin as their native language from these late-learning models contained less unpredictability (Romaine & Wright, 1987; Sankoff, 1979, 1994; Sankoff & Laberge, 1973). For instance, although the native speakers produced the preverbal form *i* in the same locations as did their non-native-speaker elders, the frequencies of use in the various syntactic environments differed between the children and adults (Aitchison, 1996; Sankoff, 1994).

Similar results have been found in Solomon Islands Pijin (Jourdan & Keesing, 1997), another descendant of Melanesian Pidgin (Keesing, 1988), and these kinds of changes have also been proposed, although not directly witnessed, in French-based creoles (Becker & Veenstra, 2003).

<sup>2</sup> It should be noted that there was also variation between speakers that was semi-predictable (Sankoff, 1994). Some of the interspeaker variation was conditional on age, such that speakers who learned the language at different times spoke a little differently from each other. Some of it was correlated with location of origin, such that speakers from different regions of the country had different typical speech patterns. There were also sex differences. However, even beyond this, there was a great deal of unpredictability in the speech of any one individual.

<sup>3</sup> Non-creole languages of course also often possess multiple ways of expressing the same meaning. For instance, the English future tense can be expressed using 'to be going to' or 'will', or in short colloquial forms 'be gonna' or 'll' (e.g., I am going to go to South Carolina, I will go to South Carolina, I'm gonna go to South Carolina, I'll go to South Carolina). These are not truly in free variation with each other, however; they differ in certainty and formality, and possibly in focus. The Tok Pisin forms are not in free variation with each other either, but they differ more according to context. See Sankoff (1979) for a more complete discussion.

### 1.1.3. Regularization by adult learners

Thus far, all the examples we have given involve regularization by children. However, there are reasons to believe that adults may also regularize variability in languages. Aitchison (1996; Aitchison & Agnihotri, 1985) points out the tendency of adult language learners to overregularize morphological as well as syntactic patterns. For example, one adult learner of German described by Klein and Perdue (1993) always used *eine* as the indefinite article (when he used an article with indefinites), despite the fact that German has multiple indefinite article forms that vary by gender of the noun (as does Italian, his native language). This can be seen as a reduction or regularization of variation (even though the variation in German is predictable). Unlike overregularization in children's first language acquisition from native input (e.g. the overregularization of *-ed* to irregular verbs), much of the adult learner's overregularization remains in their system as it fossilizes; they do not outgrow it (Adamson, 1988; Sorace, 2000).

### 1.1.4. What kind of mechanism is likely to produce regularization?

As presented, the evidence suggests that both adults and children can, at least in principle, introduce greater regularity into languages. Are the learning mechanisms that produce regularization specific to language learning, or might they be more general in scope? Some results from studies of probability learning suggest that it might be the latter. The general aim of probability learning studies was to describe what participants learn when they are provided with information that is probabilistic in nature. For instance, participants are asked to watch two lights that flash, one at a time. The participant's task is to make a prediction about which of the two lights will flash just before each flash event. Which light actually flashes is probabilistically determined so that the overall probability is within a pre-determined range. For instance, in a 70/30 experiment, light A flashes 70% of the time, and light B flashes 30% of the time, entirely probabilistically. (The particular ratio can, of course, differ by experiment.)

Most experiments in this literature show that after very little exposure, adults' predictions begin to match the exposure probabilities. For example, in the 70/30 example, participants predict that light A will flash next on 70% of the trials and that light B will flash next on 30% of the trials (Estes, 1964, 1976). This kind of response pattern is called *probability matching*. (Note that probability matching is not the optimal response for success in prediction or in securing reinforcement, as the paradigm is run in animals, since predicting A on 70% of the trials, when it does in fact flash unpredictably with a .70 probability, will lead the participant to be correct only 58% of the time. Nonetheless, probability matching is the most common response pattern seen in these experiments.) However, under certain circumstances another type of response appears. For instance, when participants are asked to attend to their level of correctness on blocks of trials rather than for each individual trial, they tend to *overmatch*, selecting the more frequent alternative more often than it actually occurs. Of particular interest to us, some experiments suggest that one can induce overmatching by changing features of the presentation. Gardner (1957) and Weir (1972), for example, found that adults overmatched when presented with more than two alternatives. For example, if light A flashes 70% of the time, and lights B and C each flash 15% of the time, participants guess light A *more* than 70% of the time. This literature thus suggests that adults will regularize non-deterministic information under some conditions.

Similar experiments conducted with young children show that they are less likely than adults to probability match and more likely to *regularize*. (This is our term, which we use to suggest similarity to regularization in language acquisition. In the terminology of this literature, children are more likely to overmatch or even to pick the more frequent item all the time, called *maximizing*.) However, the degree to which they regularize, and the age at which they stop regularizing and begin to probability match, varies across studies. The general trend in the child literature is that younger children are more likely than older children to overmatch (see discussion in Hudson, 2002). The upper limit of regularizing behavior is dependent on the task, with more complex tasks producing regularizing at older ages than easier tasks. Bever (1982) found high degrees of overmatching in 2- and 3-year old children in a two-choice task, and very little maximizing in 4-year olds in the same experiment. Kessen and Kessen (1961) found probability matching by age 3;7. However, overmatching has been found in children as old as 5 and 7 using a slightly different task that included three choices (Stevenson & Weir, 1959; Stevenson & Zigler, 1958; Weir, 1964). The findings, then, are similar to those in the adult studies, where

increased complexity produces less probability matching and increased regularizing behavior. However, the literature also shows quite a bit of variation in the tasks (especially with respect to the amount of information available to the learner) and in the details of their outcomes.

#### 1.1.5. Predictions for language learning

Taken together, the findings from studies of language acquisition and probability learning reviewed above provide some suggestions about variables that might lead to the regularization that occurs in the acquisition of non-native language input or in the emergence of a new language. These findings suggest that the nature of the inconsistencies themselves may play a role in regularization, and also that children may be more likely than adults to regularize inconsistencies; but neither of these variables has previously been investigated systematically. Indeed, as discussed above, typical human languages do not contain much unpredictable inconsistency, so it is difficult to answer this question by examining normal language acquisition. There are, as discussed above, natural situations where linguistic input is provided by non-native speakers. However, in such cases many variables are confounded, making it difficult to determine which of them lead learners to change the language. We have therefore developed a miniature artificial language paradigm for use in investigating this question. In previous work using this paradigm, we presented learners with simple linguistic input that contained one inconsistent part of the grammar. We found that adult learners did tend to reproduce (or probability match) the inconsistencies in their input; but, in contrast, children turned inconsistent forms into rules (Hudson Kam & Newport, 2005).

In that study we exposed learners to an artificial language with probabilistically occurring determiners (articles, like *'the'* and *'a'* in English). Nouns were accompanied by determiners some percentage of the time; the rest of the time determiners were absent, and the nouns were alone in the noun phrase. When nouns had determiners and when they did not was determined randomly: there were no differences in meaning or other aspects of context when the nouns appeared with versus without determiners. In this way the variation was unlike that typically present in natural human languages, though much like that in the speech of late learners, adult speakers of emerging contact languages, and the parents of children like Simon. At testing, adult participants produced about as many determiners in their own productions as they had heard in the exposure (that is, they probability-matched the use of determiners). In contrast, children's productions were more systematic than their input.

However, the type of inconsistency investigated in that study – variation in the presence versus absence of a form – is not the only type of inconsistency that occurs in the natural language phenomena we are interested in understanding. In the present series of studies, we examine a different type of inconsistency, more like that found in the natural input received by Simon and by those exposed to an emerging language, to see how this type of inconsistency affects adult learning; and we also observe child learners exposed to the same type of inconsistency, to see how age differences interact with this. In Experiment 1 we ask whether adult language learners will regularize more when the language they are exposed to contains more complex variation than when it contains simple two-alternative variation. In Experiment 2 we consider in greater detail the character of the complexity and inconsistency required to produce regularization. In Experiment 3 we compare the performance of adult learners with that of children exposed to these types of inconsistencies, to see if they regularize in the same or different ways.

## 2. Experiment 1

In this experiment we examine whether adult learners regularize complex variable input more than the simple presence/absence input. To investigate this we used an artificial language paradigm to expose participants to miniature languages containing unpredictable inconsistency that was more complex than the inconsistency we had previously found to produce probability matching in adult learners (Hudson Kam & Newport, 2005). We did this by increasing the number of options in a particular grammatical category. (Exactly what this means is described below.) This is much like increasing the number of lights over which participants make predictions, a manipulation that produces overmatching in probability learning studies, and thus should produce overmatching in our language-

learning task if the two result from the same mechanisms. It is also very much like the kind of inconsistency to which Simon was exposed and which he regularized. We exposed participants in different conditions to increasing numbers of options, reasoning that increased options equals increased complexity in the nature of the probabilistic variation and therefore perhaps increased regularization.

As in our earlier studies, we exposed participants to a language in which all the elements displayed regular properties except the determiners. Determiners were selected because we wanted to study the learning of an inconsistent functional category, and in a short period of time. This makes most other functional categories unsuitable, given the complexities of the meanings they encode. As in Hudson Kam and Newport (2005), we examined what participants had learned about the language using several different measures, including production and grammaticality judgment tasks.

## 2.1. Method

### 2.1.1. Participants

Participants were students at the University of Rochester at the time of the study. Average age was 19.8 years. Forty-one women and 19 men participated. They were paid daily for their participation and received a bonus after the final session for completing the entire experiment. They were recruited primarily from the department's subject pool list via an email describing the study and inviting them to participate.

### 2.1.2. The Language

The basic language to which we exposed participants was small, consisting of 51 words: 36 nouns, 7 intransitive verbs, 5 transitive verbs, 1 negative, and 2 main determiners, 1 for each of 2 noun classes. These words were in the input of participants in all conditions. In addition, there were another 16 noise determiners used in the experimental manipulation. These did not vary by noun class, but the exact number of noise determiners varied across the experimental conditions. (This is described in greater detail below.)

The language was presented in conjunction with a small world of videotaped events showing objects and actions, whose permissible combinations restricted the number of possible sentences. Even with these semantic restrictions there were over 13,200 possible sentences in the language. The grammatical structure of the language is shown in Fig. 1. The basic word order is (NEG) V-S-O. As is typical for a real VSO language (Greenberg, 1963), the determiner follows the noun within the NP. This word order was selected to be quite different from that of English, and also to permit the use of a sentence completion task (see below) that would readily elicit NPs (the crucial portion of the artificial language sentences) from participants during testing. This basic grammatical structure permits four possible sentence types: intransitive, transitive, negative intransitive, and negative transitive.

Although the vocabulary is relatively small in comparison to full natural languages, we took great care to make the language as realistic as possible. (See MacWhinney, 1983, for a discussion of this point with respect to research using miniature artificial languages). Complete sentences can be produced in the language, and there are different kinds of sentences (e.g. negative and positive, transitive and intransitive). The word order and functional category properties were modeled after those of natural languages (though unlike English, in order to avoid simple transfer), and the sentences expressed meanings. In addition, while some of the natural language cases we were modeling did not contain their inconsistencies within the determiner system, the details of their probabilistic variations are very similar to the inconsistencies of our artificial languages.

$$\begin{aligned}
 S &\longrightarrow (\text{NEG}) V + \text{NP}_{\text{Subj}} + (\text{NP}_{\text{Obj}}) \\
 \text{NP}_{\text{NC1}} &\longrightarrow N + \text{DET}_1 \\
 \text{NP}_{\text{NC2}} &\longrightarrow N + \text{DET}_2
 \end{aligned}$$

Fig. 1. Grammatical structure of the language.

The nouns were divided into two classes. Nouns were assigned to classes on a completely arbitrary basis, with 20 nouns in class 1 and the remaining 16 nouns in class 2. This was done to keep the methods as similar as possible to those in earlier studies to permit comparisons of previous and current results. The only grammatical consequence of noun class membership in the language is determiner selection: each class of nouns takes a different main determiner. (A word list and gloss for each word can be found in Hudson Kam & Newport, 2005.) This division into two noun classes, with a different main determiner for each, is similar to the division of nouns into two gender or declension classes in many natural languages. The exact nature of the linguistic input received by a participant varied according to consistency condition assignment and is described below.

### 2.1.3. Presentation

Participants were exposed to the language by videotape for eight sessions, each lasting 25–29 min. Participants were seated in front of a video monitor, on which they watched a scene or event. They then heard a sentence in the miniature language that described the scene. Sentences were spoken at a normal rate with English prosody and phonology and sounded very natural and fluent. There was no explicit instruction in grammar or vocabulary, and they never saw anything written. Participants learned the language solely from the auditory exposure to the sentences. For example, a participant saw a toy boat hitting a girl-figurine and heard:

(1)	/flɪm mæ:wzɪnɛr kɑ hit boat DET <sub>1</sub> 'The boat hits the girl.'	fɛrluk <sub>A</sub> pɔ/ girl DET <sub>2</sub>
-----	--	--

The exposure set contained 230 sentences and their corresponding videotaped scenes. Half the exposure set sentences were intransitive and the other half were transitive. Negative sentences were included to help the participants learn the meaning of the verbs, especially the intransitives, as well as to expand the number of possible sentences in the language. Overall, however, there were relatively few negative sentences in the presentation set (seven transitive sentences and 43 intransitive sentences).

Verb frequencies varied due to the importance of keeping noun occurrences balanced, along with the constraints arising from the meanings and associated selectional restrictions of the verbs. Each intransitive verb occurred 15–18 times in the 115 intransitive sentences. Each transitive verb occurred 14–27 times in the 115 transitive sentences. Individual verbs were presented either in both negative and positive sentences or in only positive sentences; no verb was presented in only negative sentences. Each noun in the language occurred 3–4 times in the intransitive sentences, and 3–4 times in each syntactic position (subject and object) in the transitive sentences. Like the verbs, each noun could appear in both positive and negative sentences, or only in positive sentences; no noun appeared in only negative sentences. (The exact number of times any particular word occurred in the exposure set is listed in Appendix B of Hudson Kam & Newport, 2005).

Each exposure session contained a different set of approximately 115 sentences drawn from the 230 sentence exposure set. Each sentence (and scene) was presented four times over the course of the eight exposure sessions. Participants were asked to repeat each sentence after hearing it. They were told that this was pronunciation practice which would be helpful since they would have to produce their own sentences at the end of the experiment. The entire experiment took nine sessions to complete (the eight exposure sessions and one test session). Participants completed the experiment in 9–12 days. All exposure and testing was done individually.

### 2.1.4. Experimental manipulation

In this experiment as well as those that follow, equal numbers of subjects were assigned to each of the five conditions (input groups) described below. Participants in all five conditions were exposed to the same basic sentences, and therefore their exposure was equivalent in almost all aspects of sentence structure. Input sentences differed between conditions only in the occurrence of determiners



in the noun phrases. All participants were exposed to sentences containing the main determiner forms (those agreeing with the class of the noun) 60% of the time. Although the sentences containing the main determiner forms were selected randomly from the exposure set, they were the same for all participants, regardless of complexity condition.

The experimental manipulation occurred in the remaining 40% of noun phrases. Participants in the control condition heard no determiner form in those noun phrases. These participants were exposed to the same input as in our previous study (Hudson Kam & Newport, 2005), a kind of inconsistency we call *presence/absence inconsistency*. This inconsistency exhibited the least complex variation, and on the basis of previous results participants were expected to probability match the occurrence of the determiners in their own usage. All other participants received more complex determiner variation, of a type we call *scattered inconsistency*. In scattered inconsistency there is one main form that occurs a majority of the time, but a number of other forms may occur instead of this main form, inconsistently and at a lower frequency. The degree of scatter differed across conditions. Like the control participants, those in the 2<sup>ND</sup> group (ND = noise determiner) heard the main determiner forms 60% of the time, but in the remaining 40% of noun occurrences, one of two other determiner forms (hereafter called 'noise' forms) occurred, each in 20% of the noun occurrences of each noun class. For noun class 1, 60% of the noun phrases occurred with the main determiner form, /kɑ/, 20% occurred with the noise determiner form /te/, and the remaining 20% with the noise form /mɛg/. The same was true for noun class 2: 60% of the noun phrases had the main form /pɒ/, 20% had /te/, and 20% had /mɛg/. Note that the noise forms occurred with both noun classes, unlike the main forms, which were restricted in their distribution to the main noun class. Noise forms were thus both lower in frequency and more unpredictable in the context in which they occurred.

The input for the three other complexity groups was similar, but contained more noise determiners that each occurred with lower frequency. The 4<sup>ND</sup> group heard the main determiner form 60% of the time and 4 noise determiner forms, /te/, /mɛg/, /li/, and /kum/, that each occurred with 10% of the noun phrases within a noun class. Again, the noise forms occurred with nouns in both classes. The 8<sup>ND</sup> condition heard the main determiner 60% of the time and 8 noise forms that each occurred 5% of the time. In addition to the 4 noise forms listed above, they heard /su/, /gɪ/, /lɛt/, and /bɑn/. The final condition was the 16<sup>ND</sup> condition. They heard the main determiner form 60% of the time and 16 noise forms that each occurred 2.5% of the time. The additional noise forms they heard (over and above those present in 8<sup>ND</sup> condition) were /bɪp/, /fu/, /zæɪ/, /zɒ/, /sɛp/, /mɪb/, /lɒm/, and /dæf/. This input is represented graphically in Fig. 2, which shows the percentage of occurrence of each determiner form within each noun class across the different conditions. Each fill pattern represents one determiner form.

As noted, the noise forms are less predictable than the main forms in two ways. First, they are less frequent. Second, they occur with nouns in both classes. This is not true of the main determiner forms. Note that the percentages given above are true within each noun class, but change when considering

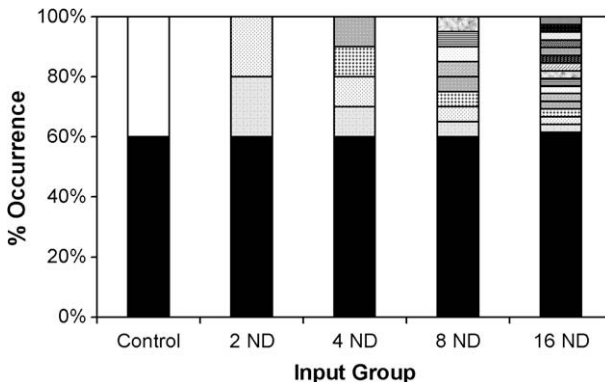


Fig. 2. Occurrence of determiner forms within each noun class in the input languages, across the five different conditions.

nouns overall. In the 2<sup>ND</sup> condition, for example, the percentage of all nouns occurring with each main determiner form is 30%, whereas the percentage of all nouns occurring with each noise form is 20%. While these overall percentages are much closer together than 60% and 20% (the percent occurrences for the main and noun determiners within each noun class), the main forms are still more frequent than the noise forms even when considering nouns overall.

In all other ways the input languages across the groups were as similar as possible. As mentioned, the noun phrases containing the main determiner forms were the same for all participants. Similarly, all noun phrases occurring with /te/ and /mɛg/ in the input of the 16<sup>ND</sup> participants also occurred with /te/ and /mɛg/ in the input of the 8,4, and 2<sup>ND</sup> groups. Similarly, a noun phrase containing /kum/ in the 16<sup>ND</sup> input also contained /kum/ in the 8 and 4<sup>ND</sup> input sets. All other parts of the grammar were the same, and completely consistent, in all four input groups.

In order to be sure that the various determiner forms were actually inconsistent and probabilistic in their appearance (and not accidentally associated with a syntactic function or with a particular lexical item), the occurrence percentages for each determiner for noun phrases in general were also maintained for each syntactic position and for each noun. For example, in the presentation set of the 2<sup>ND</sup> group, 60% of intransitive subjects, transitive subjects, and transitive objects occurred with main determiner forms, and the 2 noise forms were similarly evenly distributed across each syntactic position. While it was not possible to maintain precisely the same occurrence percentages for each noun, individual nouns occurred with particular determiner forms within a range centered around the condition percentage. For example, in the 2<sup>ND</sup> condition, the main forms occurred 41–74% of the time with individual nouns, with an average of 60% across nouns, and the noise determiners occurred with particular nouns 9–33% of the time, with an average of 20%. In addition, each presentation of a particular sentence within and across sessions could be different from the other three. This last point is particularly important, because it ensures that there were no consistent conditioning contexts. Thus there was no pattern of determiner use available to be learned from the input data, other than the percentages of use of the various forms.

### 2.1.5. Tests

Participants were given four different types of tests to evaluate their performance. Tests were given in the order in which they are described below.

**2.1.5.1. Vocabulary.** A vocabulary test was given twice. The first was administered after participants watched the videotape in the fourth session. In this task, participants were tested on their knowledge of twelve vocabulary items. They were told that this test was designed to give them some idea of how they were doing up to this point – that it was for their own benefit and would not be analyzed. Participants were asked to provide a name for each object as it appeared on a video monitor and were given as much time as they needed to respond. All responses were videotaped, but (in accord with the instructions) the results were not analyzed.

A second vocabulary test was used to evaluate whether participants had learned enough vocabulary to be tested on more complex aspects of the language and was administered with the other tests in the final session. Participants were tested on the same 12 items as in the first vocabulary test, but the order in which the items appeared was different. We tested the same nouns twice for one principal reason: these are the nouns required to complete the sentences in the production task, and we therefore wanted to direct attention to them in an implicit way. Post-test debriefing indicated that very few participants had noticed this. Presentation and recording were the same as in the first vocabulary test.

**2.1.5.2. Sentence completion task.** The test of primary interest was a sentence completion task. This task was designed to evaluate the participants' production of determiners – the inconsistent part of the language. Participants saw a novel scene on the video monitor and heard the first word of the corresponding sentence. They were then asked to produce the complete sentence and were given as much time as they needed to provide an answer. For example, a participant sees a toy bird jump around and hears the word /mɔrt/ 'move'. She should then say /mɔrt fʊmpɔgɔ po/ 'move bird det'. Because the language is V-S-O, participants were always given the verb and had to produce the whole sentence, thus generating the NP(s) themselves.

There were 24 test sentences (12 transitive and 12 intransitive), resulting in 36 possible NPs and therefore 36 possible determiners. Participants were first tested on the transitive sentences and then on the intransitives. The test set was designed so that 12 nouns each appeared once in each possible syntactic position (intransitive subject, transitive subject, and transitive object). The first use of the individual noun varied between subject and object position in the transitive sentences; some nouns were first used as subjects and others as objects. Participants were asked to indicate where a noun they could not recall should go in the sentence (for instance, by saying X instead of the noun). This allowed us to include the data from incomplete responses. Responses were videotaped and later transcribed for analysis. All sentences used in this and other tests were novel to the participants and were not part of the exposure set.

*2.1.5.3. Determiner judgment task.* The third test was a grammaticality judgment task that also examined participants' knowledge of determiner usage, but through judgment rather than production. Participants were asked to listen to 48 sentences one at a time and judge each of them on a four-point scale according to how much they 'liked' or 'disliked' the sentence. Participants were instructed to respond that they really liked a sentence when it sounded like a sentence from the language that they had been learning, and to respond that they really disliked a sentence when it sounded completely unlike a sentence from the language. They were also told that if they thought a sentence was mostly, but not completely, like or unlike sentences from the language, they should use the middle of the scale. Participants responded by pointing to one of four different 'happy' or 'sad' faces. The experiment was designed in this fashion so that it also could be done with children without changing the tasks.

The 48 test sentences consisted of four variations of 12 base sentences: one form contained the main determiner form appropriate to the noun, one contained a noise determiner (one of the forms to which all subjects other than the control subjects had been exposed), one had the determiner in the wrong location (preceding the noun), and one had no determiner at all. The sentences were randomly ordered, with the constraint that two versions of the same base sentence could not follow each other. The four variations of one base sentence are shown in (2):

(2)	a.	/gɛrn fɛrluka po/ 'fall girl det'	(correct: det follows noun)
	b.	/gɛrn fɛrluka mɛg/ 'fall girl det'	(infrequent/incorrect: noise det) <sup>a</sup>
	c.	/gɛrn pɔ fɛrluka/ 'fall det girl'	(incorrect: main det precedes noun)
	d.	/gɛrn fɛrluka/ 'fall girl'	(incorrect/infrequent: no det) <sup>b</sup>

<sup>a</sup> This is a lower frequency form for all participants exposed to noise determiners, although the frequency varies by condition. For the control participants, however, it is an incorrect form which they have never heard.

<sup>b</sup> As with example (b), the type of this example varies in its correctness by input group. For the control participants it is the lower frequency form in their exposure. For all other participants this type of sentence is an incorrect form to which they were not exposed.

Four of the 12 base sentences varied the determiner occurring with the transitive subject, four varied the determiner occurring with the transitive object, and four were intransitive (and therefore varied the determiner occurring with the subject). Sentences were presented on a Sony™ minidisk deck MDS-S38 through headphones, preventing the experimenter from hearing the sentence to which the participant was responding. This prevented the experimenter from being able to inadvertently cue the participant to any particular response. Responses were recorded on an answer sheet by the experimenter. Participants had 3 s to respond to each test item. Again, all sentences were novel.

*2.1.5.4. General Grammar Test.* The fourth test, also a grammaticality judgment task, examined what participants had learned about the rest of the language. Participants listened to 16 pairs of sentences and were asked to select the sentence from each pair that sounded most like a sentence from the

language that they had been learning. The two sentences in each pair were versions of the same sentence, one grammatical, the other ungrammatical. Test sentences were presented using the Sony™ minidisk deck MDS-S38 over headphones. Participants listened to both versions of the sentence and circled 1 or 2 on an answer sheet, indicating whether they preferred the first or second sentence in the pair. Half of the sentence pairs tested participants' knowledge of verb subcategorization, that is, whether they knew that transitive verbs required two nouns and intransitives only one. The remaining sentence pairs tested whether participants knew that a verb was required in every sentence. These rules of the grammar were tested for both transitive and intransitive sentences. For the transitive sentences with missing arguments, either the subject or the object could be the missing argument. Which sentence (first or second) in the pair was grammatical was randomized, as was the ordering of sentence pairs in the test, with the constraint that no more than two sentences could occur in a row that tested the same rule and were of the same valence. There was a 1-s pause between the two sentences that formed a pair and a 5-s pause between pairs. Pairs were not identified as such, except by the occurrence of the longer pause. Again, all test sentences were novel; none appeared in the exposure set.

## 2.2. Results

We begin by reporting results for tests that demonstrate more general knowledge of the language, and then move on to describe the results on determiners that are of primary interest.

### 2.2.1. Vocabulary tests

In accord with the instructions given to participants, the data from the first vocabulary test were not tabulated. On the second vocabulary test, all participants scored at least 5 out of a possible 12 (the criterion we have previously used for deciding whether participants would be given certain of the remaining tests), with a mean of 8.88 items ( $SD = 2.39$ ) for participants overall. Means varied slightly across the input groups, ranging from 7.83 (16 ND) to 10 (2 ND). A one-way ANOVA with input condition as a between-subjects factor indicated that the differences in vocabulary scores were not significant.

### 2.2.2. General Grammar Test

This test examined participants' knowledge of parts of the grammar other than determiners. We conducted this test to ensure that learners in all conditions successfully acquired those parts of the grammar represented consistently in the input. The test examined participants' knowledge of sentence construction (did they know that a verb is required in every sentence) and verb subcategorization (did they know that one set of verbs is transitive, requiring two nouns, and another intransitive, requiring only one noun).

Fig. 3 shows the mean score by noise condition. The overall mean was 14.5 out of a possible 16,  $SD = 1.66$ . This was significantly and substantially above chance ( $t_{(59)} = 30.29, p < .001$ ). We conducted

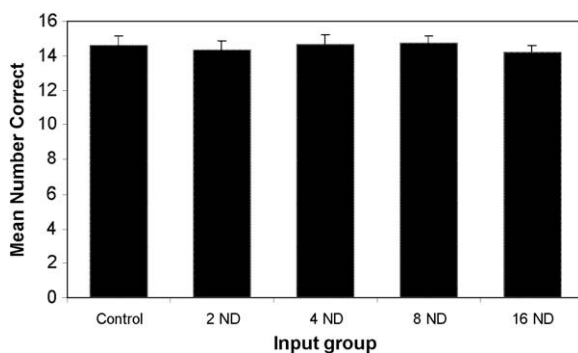


Fig. 3. Mean score on general grammar test by input group. (In this and all other figures, error bars represent standard error.)

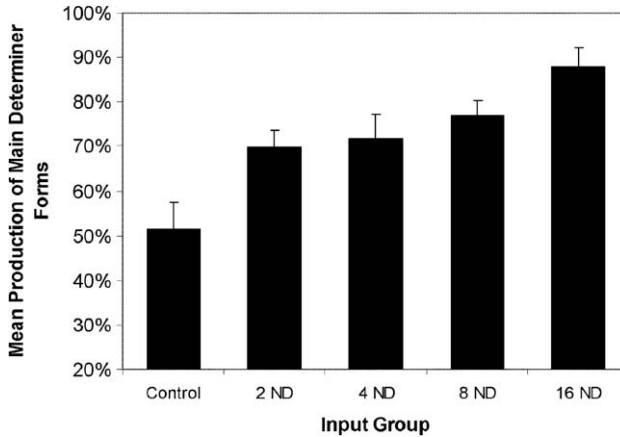


Fig. 4. Mean production of main determiner forms by input group.

a repeated measures ANOVA with rule type (two levels) and transitivity (two levels) as within-subject factors and noise condition (five levels) as a between-subjects factor. The main effect for noise condition was not significant, indicating that the manipulation of determiners had no effect on participants' learning of the other, consistent parts of the grammar. There was a significant effect of rule type ( $F_{(1,55)} = 9.2, p = .004$ ), with participants scoring slightly higher on knowledge of basic sentence structure (every sentence must have a verb) than on verb subcategorization (7.55 vs. 6.95 out of a possible 8) – not surprising since the former requires very general knowledge of the language, while the latter depends on knowledge of particular verbs. There was also a significant main effect of transitivity ( $F_{(1,55)} = 16.55, p < .001$ ), with participants performing slightly better on transitive test items than intransitive ones (7.63 vs. 6.87 out of 8). There were no significant interactions.

Overall, participants performed very well on this test, indicating that they had learned these consistent facets of the grammar. Moreover, their performance was not affected by the amount of inconsistency of the determiners: participants performed equally well in all input conditions.

### 2.2.3. Sentence completion task

The results of this test were of primary interest. It permitted us to observe the effect of the presence and amount of scatter in the linguistic input on the production of determiners. In particular, we wanted to know whether, when exposed to scattered inconsistency in determiner usage, participants would reproduce the inconsistency present in their input or would regularize the more frequent forms to which they were exposed. For each participant we computed the percentage of main determiner production (the number of correct main determiners used by the participant, divided by the number of possible determiner usages, multiplied by 100). The number of possible determiner usages was simply the number of correct nouns produced by the participant in this task. Fig. 4 shows the mean percentage of main determiner production for the five input groups. Recall that the percentage of main forms in the input was the same – 60% – for all input groups.

As can be seen in Fig. 4, participants exposed to scattered inconsistency (input containing noise determiners – all of the ND conditions) produced more main determiner forms than those exposed to presence/absence inconsistency (main determiner forms alternating with determiner omission – the Control condition). A one-way ANOVA confirms that there is a significant effect of input condition on the production of determiners ( $F_{(4,55)} = 8.27, p < .001$ ), and a  $t$ -test comparing the control group with the four scatter groups is significant ( $t_{(55)} = -4.889, p < .001$ , with pooled variance estimate).<sup>4</sup>

<sup>4</sup> Tests for homogeneity of variance were not significant, indicating that it is licit to pool the variances. The contrast is also significant using a separate variance estimate ( $p < .001$ ).

We were particularly interested in whether increasing the noise, or scatter, would induce increased regularization behavior in adult learners. The data indicate that this is indeed the case: As participants are exposed to increasing numbers of noise determiners, they produce increasing percentages of main determiner forms ( $F_{\text{linear}(1,57)} = 25.92, p < .001$ ).<sup>5</sup> This effect shows that extensive scattered inconsistency does produce regularization, with participants in the 16 ND producing almost 90% main determiner forms, even though their input contained these forms only 60% of the time.

The design of the test allowed us to examine whether participants were using determiner forms differently according to the syntactic position of the noun (subject of an intransitive verb, subject of a transitive verb, or object of a transitive verb). A repeated-measures ANOVA with input condition as a between-subjects factor and syntactic position as a within-subject factor showed no effect of syntactic position and no significant interaction between input and syntactic position. This indicates that participants were not imposing on the input a more deterministic, linguistically-based rule, such as using the main determiners in association with subjects versus objects or with transitives versus intransitives.

What were participants doing when not using the main determiner forms? Control participants used no determiner form at all. Thus they were replicating their input, as we found in our earlier study. Participants in the noise determiner input groups primarily used noise forms, with many participants using 1 or 2 noise forms to the exclusion of the others. (This is the only possibility in the 2 ND condition. However, this was common among participants in the 8 and 16 ND groups as well.) This use of a few noise determiners did not reduce the amount of inconsistency present in the language, however, as different speakers preferred different noise forms. Most participants preferred the phonologically simpler CV forms, but this trend did not hold for all participants; some used a CVC form as their preferred noise form. There were also some who used the incorrect main form (the main determiner form for the other noun class), and some who created a novel determiner form (usually blends of one or more existing forms). Occasionally, a participant in a scatter condition used a bare noun, something not present in their input. Thus, participants were not regularizing the noise or scatter in their input; their non-main form productions remained inconsistent and noisy. Regularizations, when they occurred, involved more frequent use of the main determiner forms.

In sum, speakers in the control group basically reproduced their input. In contrast, participants in noise groups showed a tendency to regularize their input, using the most frequent forms more often than they had heard them. Moreover, increased complexity of variation in the input resulted in increased regularization, with participants producing increasingly more main forms in their speech as the number of noise forms in their input increased. Participants exposed to a few noise forms produced the main determiner forms only slightly more often than they had heard them, though much more often than participants who heard no noise forms at all (presence/absence inconsistency); participants exposed to 16 noise determiners produced the main determiner forms almost 90% of the time, a full 30% more often than they heard them.

#### 2.2.4. Determiner judgment task

This task was designed to assess participants' knowledge of determiners in a different way – through grammaticality judgments. As described above, participants were asked to rate 48 novel sentences, one at a time. Twelve of the test sentences were correct, 12 contained a noise determiner (to which all subjects other than the control subjects had been exposed), 12 had the determiner in the wrong location (preceding the noun), and 12 had no determiner at all.

Fig. 5 shows the mean ratings given by participants to each kind of sentence for the five input groups. A MANOVA with sentence type and syntactic position as within-subject repeated-measures variables and input groups as a between-subjects variable was conducted on the data.

<sup>5</sup> This  $F$  was computed adjusting for unequal spacing between the categories of the factor (input condition). Because it is not clear whether the control condition falls along a continuum with the scatter conditions, we also conducted the linear trend analysis with only the 4 noise conditions included, and it too is significant ( $F_{(1,45)} = 11.21, p = .002$ ).

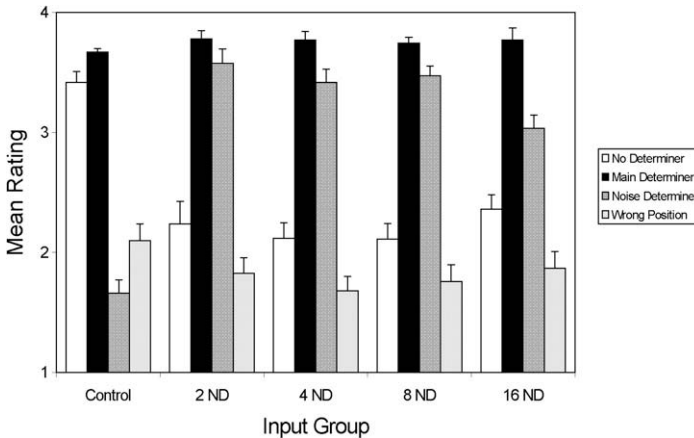


Fig. 5. Mean ratings by input group.

The primary variables of interest are the effects of input group and determiner manipulation. The main effect of input group is not significant. The main effect of determiner manipulation is significant ( $F_{(2.4, 130.9)} = 368.31, p < .001$ ).<sup>6</sup> This is modulated by a significant interaction between the two variables ( $F_{(9.5, 130.9)} = 35.29, p < .001$ ), reflecting the fact that all participants liked sentences with main determiners and disliked sentences with the determiner in the wrong location, but the groups differed in their ratings of sentences without determiners and with noise determiners. We were particularly interested in the ratings given by participants to sentences which they had encountered in their input. For the control group these were sentences with main determiners and those without determiners. For all other groups these were sentences with main determiners and those with noise determiners. For each input group, we compared the ratings given to these two sentence types. The results show that all participants reliably rated the more frequent sentences higher than the less frequent sentences. Repeated-measures ANOVAs with sentence type as a within-subjects repeated-measure were significant for all five input groups: control group ( $F_{(1,11)} = 12.19, p = .005$ ); 2-noise ( $F_{(1,11)} = 7.79, p = .018$ ); 4-noise ( $F_{(1,11)} = 20.51, p = .001$ ); 8-noise ( $F_{(1,11)} = 12.20, p = .005$ ); 16-noise ( $F_{(1,11)} = 60.71, p < .001$ ).

We also examined the data to see whether the degree of difference between the ratings given to the two types of sentences increased as the number of noise determiners increased. That is, did the participants exposed to 16 noise determiners distinguish between sentences with main determiner forms and those with noise determiner forms to a greater degree than participants exposed to fewer (or no) noise forms? To examine this, we computed a difference score for each participant between the mean rating given to sentences with main determiners and the mean rating given to whichever kind of sentence was the other one in her input (sentences with no determiners for control participants and sentences with noise determiners for all others). We then performed a trend analysis on the difference scores. This analysis is significant ( $F_{(1,56)} = 22.89, p < .001$ ), indicating that the difference in ratings does indeed increase as the number of noise determiners increases, mirroring the trend we found in the production data.

Also as in the production task, we asked whether participants would judge determiner forms differently according to the syntactic position of the noun, indicating that they might be imposing on the input a more deterministic linguistic rule, such as using the determiner forms in association with subjects versus objects or with transitive but not intransitive sentences. Such a tendency would be reflected in a significant interaction between syntactic position and sentence type. However, this was not the case. Although the main effect of syntactic position is significant ( $F_{(2,110)} = 4.59, p = .012$ ), there

<sup>6</sup> The degrees of freedom used in the sentence/determiner manipulation analyses are adjusted using the Huynh-Feldt Epsilon due to a significant test for heterogeneity of variance.

is not a significant interaction between sentence type and syntactic position, nor is there a significant three-way interaction between sentence type, syntactic position, and input group.

In sum, to some degree participants' judgments reflected their input. Sentences with the determiner in the wrong location were universally disliked by all groups. None of these sentences occurred in the input of any group. Likewise, all participants preferred the sentences with the main determiner forms, which were the most common sentence form in the input for all participants. Where participants' ratings differed was in the ratings given to sentences they had heard, but less frequently (that is, those with noise determiner forms or those without determiners, depending on condition). Participants rated those sentences they had heard less frequently substantially higher than those they had never heard. Importantly, however, participants also showed the same regularizing tendency in their judgments as was reflected in their productions: with increasing numbers of noise determiners, they showed an increasing tendency to prefer the main determiner forms over less frequent forms.

### 2.3. Discussion

The data from Experiment 1 clearly demonstrate that regularization behavior can be induced in adult language learners when they are given input that contains what we have called 'scattered inconsistency.' This contrasts with the results from our earlier work (Hudson Kam & Newport, 2005), where adult participants given input containing presence/absence inconsistency did not regularize or over-use the main determiner forms, but instead used these forms with almost exactly the probabilities with which they appeared in the input. Interestingly, this parallels the findings seen in natural language acquisition in children like Simon, who regularized the use of his parents' most frequently used morphemes (Singleton & Newport, 2004). Of note is that Simon's parents did not show a simple alternation between the presence and absence of required ASL morphemes, but instead showed variation of forms that was more like scattered inconsistency: they used the correct form most of the time, but with low and variable frequency might replace this form with any of several different incorrect forms. It also parallels the hints of similar effects seen in probability learning, where adults usually probability-matched, but when asked to make predictions over more than two lights, often displayed over-matching (Gardner, 1957; Weir, 1964, 1972). This similarity suggests that perhaps the same mechanisms are at work in response to inconsistencies in natural language acquisition, in the language learning modeled in this paper, and in the kind of learning investigated in basic probability-learning experiments.

However, a number of questions remain regarding this phenomenon, and especially how and why we have been able to induce regularization in adult learners. It is important to note that extensive regularization occurred in this experiment only in the extreme case where there were 16 noise determiners, each appearing only 2.5% of the time, varying with main determiners that appeared 60% of the time. In subsequent experiments we will address two important questions about this finding. First, perhaps this apparent regularization occurred only because of the low frequency of the noise determiners, and not because of the inconsistency of main and noise determiner use that characterizes the natural language phenomena with which we began our studies. To address this question, we will compare performance in the 16<sub>ND</sub> condition of Experiment 1 with a frequency-matched but differently structured condition in Experiment 2, in which the same number of determiners are used with the same overall frequencies, but where their appearance is perfectly regular and consistent. Subsequently, in Experiment 3 we will investigate learning with scattered inconsistency in children as compared with adults, to see whether scatter has the same effects in children, or rather whether child learners show tendencies to regularize that are more independent of the nature and extent of the inconsistency than is the case for adults.

## 3. Experiment 2

The purpose of Experiment 2 was to assess whether the regularization found in adult learners in the 16<sub>ND</sub> condition of Experiment 1 was due to the scattered inconsistency of the noise determiners as compared with the main determiners, as we hypothesized, or rather whether it resulted more sim-



ply from the difficulty of learning any forms that occur with the low frequency of these noise determiners. In the present experiment, two main determiner forms occurred 60% of the time and 16 lower frequency determiner forms occurred 2.5% each, but their appearance was strictly conditioned by the occurrence of particular nouns with which they were associated. The question of interest was whether adult learners also regularized the main determiners under these circumstances, or rather whether they were able to learn the low frequency determiners as well as the main determiners when each occurred in structured contexts.

### 3.1. Method

#### 3.1.1. Participants

Eleven adults, mean age 20.9 years, participated in this study. All were students at the University of California, Berkeley at the time of the study. Participants were recruited through flyers posted on campus. They were paid daily for their participation and received a bonus upon completion of the experiment.

#### 3.1.2. The Language

We used the same basic language as in the 16<sub>ND</sub> condition of Experiment 1, with one important difference. In this study, though the two main determiners and the 16 low frequency determiners occurred with the same frequencies as in Experiment 1 (that is, in 60% of the noun phrases for each of the two main determiners and in 2.5% of the noun phrases for each of the 16 low frequency determiners), their appearance was perfectly regular and consistent. To achieve this consistency, each of the determiners was assigned to particular nouns and occurred every time these nouns occurred (that is, the determiners were lexically consistent). However, by varying the number of nouns assigned to each determiner, we could create the same high and low frequencies for the determiners as was the case for the 16<sub>ND</sub> condition. Nouns were divided into 18 arbitrary classes, 2 large and 16 small. One large class contained 11 nouns, the other contained 9. The small classes each contained a single noun. As before, there were no differences in meaning or phonology between the nouns in different classes; the only grammatical consequence of noun class was determiner selection: each class of nouns took a different determiner.

#### 3.1.3. Presentation

Presentation was the same as in Experiment 1. As before, exposure and testing were conducted individually for each participant.

#### 3.1.4. Experimental manipulation

As mentioned above, nouns were divided into 18 classes, 2 large classes and 16 small ones containing a single noun, with each class taking a different determiner. This particular division allowed us to present input sentences that were exactly the same as in the 16<sub>ND</sub> condition of Experiment 1, except with respect to which determiner occurred with which noun phrases. Moreover, the input set contained the same overall distribution of determiners as in the 16<sub>ND</sub> condition. That is, the probability of any individual determiner given a noun (any noun) was almost exactly the same in the two experiments ( $p_{\text{DET}_i|\text{noun}}$  in Experiment 2 =  $p_{\text{DET}_i|\text{noun}}$  in 16<sub>ND</sub> condition).<sup>7</sup>

In terms of the number and overall distribution of the determiners, the languages are equally complex. Importantly, however, the occurrence of the determiner with any particular noun is more consistent and predictable in the present language ( $p = 1$ ) than it was in the previous experiment ( $p = 0\text{--}0.722$ ). All other aspects of the grammars were the same, and completely consistent, in the two experiments.

<sup>7</sup> To be precise, the percentages for each determiner differed by .1 or .2% for 8 of the low frequency determiners (and otherwise were exactly the same) and differed by 1% for one of the high frequency determiners. These tiny differences between Experiment 1 and 2 arose from using the same input set and tests for the two experiments and are extremely unlikely to be responsible for any differences in results.

### 3.1.5. Tests

To evaluate what they had learned about the language, participants were given three of the five tests used in Experiment 1: a vocabulary test, the sentence completion task, and the forced-choice general grammar test. The sentence completion task contained five large class nouns and seven small class nouns. Tests were constructed and presented as in Experiment 1. All tests were administered in the final ninth session.

## 3.2. Results

Recall that the language to which we exposed participants in this experiment contained the same overall distribution of determiners as the condition in Experiment 1 in which participants regularized most (the 16<sub>ND</sub> condition). In many ways, then, the two languages are equally complex, though one is completely regular and the other contains scattered inconsistency. To assess the degree to which complexity alone leads to particular learning outcomes, and to assess whether regularization of the dominant determiners might be due purely to the low frequency of the noise determiners, we compared the data from Experiment 2 with those from the 16<sub>ND</sub> condition in Experiment 1. We begin with the general tests and proceed to the ones of principal interest. Note that in the following analyses, whenever there is a reference to Experiment 1 it is to the 16<sub>ND</sub> condition only.

### 3.2.1. Vocabulary test

Participants in this experiment knew an average of 9.82 ( $SD = 2.04$ ) vocabulary items (out of 12). The minimum score was 7, the maximum was 12. The vocabulary score for participants in Experiment 1 was slightly lower at 7.83 ( $SD = 1.9$ ), a significant difference ( $F_{(1,21)} = 6.782, p = .017$ ). However, the higher vocabulary performance for participants in Experiment 2 is within the range of condition means from Experiment 1.

### 3.2.2. General Grammar Test

Participants performed very well on this test, indicating that they did indeed learn the language. The mean scores for the two groups of participants were 13.82 (Exp. 2,  $SD = 2.79$ ) and 14.12 (Exp. 1,  $SD = 1.47$ ) out of a possible 16. We subjected the data to repeated-measures ANOVA with rule type (two levels) and transitivity (two levels) as within-subject factors and experiment (two levels) as a between-subjects factor. The main effect of experiment was not significant, indicating that the two groups of participants learned the language equally well. The overall mean score of 14.00 out of a possible 16 ( $SD = 2.15$ ) was significantly above chance (two-tailed  $t_{(22)} = 13.36, p < .001$ ). As in Experiment 1, transitivity was significant ( $F_{(1,21)} = 5.32, p = .031$ ): participants performed slightly better on test items involving transitive sentences than those involving intransitives (7.43/8 vs. 6.57/8). Also as in Experiment 1, none of the interactions were significant. Unlike Experiment 1, rule type was not significant; participants performed equally well on items testing both rules.

### 3.2.3. Sentence completion task

This test examined participants' use of determiners. In particular, we were interested in whether participants exposed to a complex but consistent pattern of determiners would regularize the language like participants in Experiment 1. Fig. 6 shows the mean percentage of correct determiner uses for nouns from the large and small classes. Participants overwhelmingly used the determiner form that was correct for the noun. This is true for nouns from both the large and small classes, although participants made fewer determiner errors with nouns in the large classes ( $t_{(10)} = 2.39, p = .038$ ).<sup>8</sup>

While these results suggest that participants in Experiment 2 are not regularizing or over-using the main determiners, it is necessary to score the results in a different way to see this more clearly. In order to compare the productions of participants in the two experiments, we need a metric that is comparable and equally meaningful for each of the conditions. Overall production of large and small class

<sup>8</sup> The figure and related analyses are based on 11 nouns, not 12, for participants in this experiment only, because all participants consistently got one of the small class nouns wrong (and therefore could not be scored on usage of its determiner).

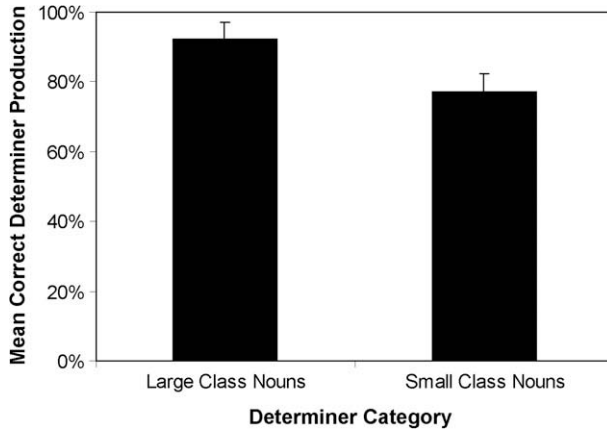


Fig. 6. Mean production of correct determiner forms for nouns in large and small classes.

determiners versus main and noise determiners is not such a measure. While the nouns do not matter for participants in Experiment 1, since the proportion of each kind of determiner produced or observed in the input does not depend on them, they do for participants in Experiment 2, since determiner choice depends crucially on the identity of the noun. To obtain a score that is comparable for these two different circumstances, we computed two proportion scores for each participant, one for main or large class determiners and the other for small class or noise determiners. Each score reflects the proportion, given what would be expected from the input, of correct determiner production for that type of determiner. A proportion of 1 represents exactly matching the input, while anything above or below 1 is a deviation from the input.<sup>9</sup> What constitutes ‘correct’ is different for participants in the two experiments. For participants in Experiment 1, correct responding is probability matching. For participants in Experiment 2, correct responding is getting the determiner correct for the specific noun. For example, if a participant in Experiment 1 produced 10 nouns, 8 with the correct main determiner and 2 with a noise determiner, the proportion of main determiners produced would be 1.33, since they are producing a third more main determiners than expected, and the proportion of noise determiners produced would be .5, one half of the four that would be expected were the participant matching the probabilities.

Fig. 7 shows the mean proportions on this measure for participants in the two experiments, for large class or main determiners and small class or noise determiners, with correct responding (1.0) indicated by a dotted line. It is clear from the figure that participants in Experiment 1 are producing far more main determiners than they heard and far fewer noise determiners. Participants in Experiment 2, by contrast, are slightly under-producing both types of determiners (large and small class) but otherwise approximately producing the language that they were exposed to. The difference between the two groups is significant for both types of determiners (Main/Large class:  $F_{(1,21)} = 41.67$ ,  $p < .001$ ; Noise/Small class:  $F_{(1,21)} = 53.75$ ,  $p < .001$ ).<sup>10</sup> Thus, despite hearing the same proportions of 18 different determiners, participants in Experiment 2 are much better able to reproduce what they’ve heard.

A complementary view of this difference is provided by examining what participants are doing when they do not match their input. The relevant productions for participants in Experiment 2 are er-

<sup>9</sup> This measure corrects for the fact that different participants produced different nouns, and thus each had different underlying probabilities of production. Let’s imagine two participants who both produced six nouns. One produced four large class and two small class nouns, and failed to produce any small class determiners, the other produced two large class and four small class nouns, and produced half of the small class nouns with the correct determiners. Both would have undershot the target percentage by 33.33%. However, one is getting rid of the small classes, the other is not. Using a proportional measure captures this.

<sup>10</sup> These two measures are not necessarily reciprocal because these are not the only possible types of productions.

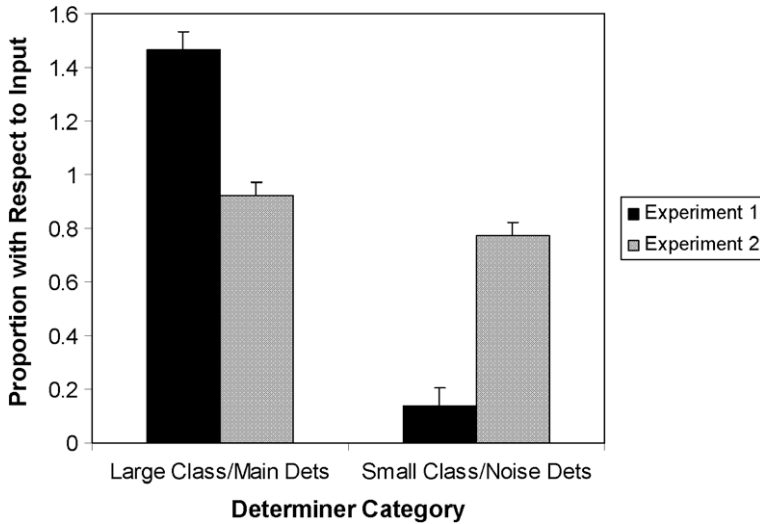


Fig. 7. Mean over- and under-production of main/large class and noise/small class determiners – Experiments 1 and 2.

rors. For participants in Experiment 1, ‘errors’ include overproduction of the noise and main forms, as well as true errors. As already noted, participants in Experiment 2 had far fewer errors; on average, only 16.3% of their noun phrases fell into this category ( $SD = 4.4$ ), compared to 44.12% ( $SD = 3.02$ ) in Experiment 1. Here we examine the proportions of the different types of errors participants made; despite their low rate of errors, it is possible that participants in Experiment 2 might have shown some regularization, for example, preferentially extending large class determiners when they did make errors.

Fig. 8 shows these errors for the two experiments, divided into four types: (1) overuse of the correct main determiner (only possible for Exp.1, not possible for Exp. 2), (2) use of the incorrect main determiner (possible in both experiments), (3) incorrect noise forms (only possible in Exp 2), and (4) zero forms (determiner omission). The figure clearly shows that the participants in Experiment 2 did not regularize like participants in Experiment 1. While they did make errors, these were few, and impor-

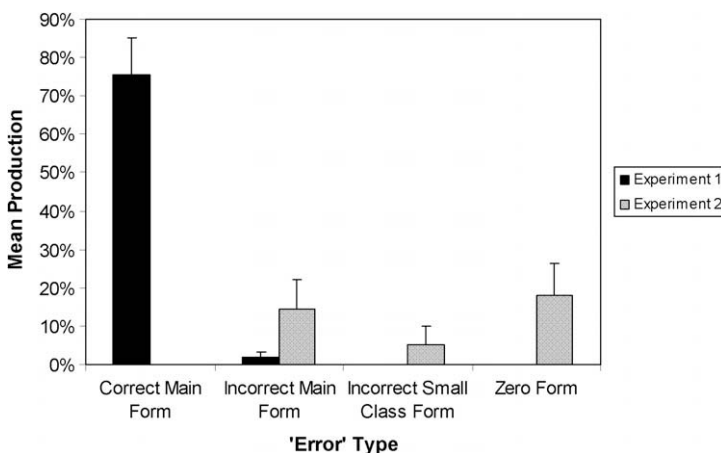


Fig. 8. Mean proportion of ‘errors’ of different types – Experiments 1 and 2.

tantly did not involve preferentially overextending the more frequent large class determiners. For participants in Experiment 2, the percentage of errors that were incorrect uses of the large class determiners was not significantly different from incorrect uses of the less frequent small class determiners or from using no determiner at all.

### 3.3. Discussion

Overall, these results indicate that distributional complexity per se is not enough to cause adult learners to regularize a grammatical form. The results also suggest that it is not merely the low frequency of the noise determiners that produced regularization in Experiment 1. Rather, it appears that inconsistency in how a form is used, in combination with inconsistency as well as low frequency in how competing forms are used (what we have called *scattered inconsistency*), are necessary for adult learners to regularize and extend the form's usage.

The question remains, however, whether the regularization seen in adult learners of Experiment 1 demonstrates the same mechanism involved in cases of actual language change, particularly creolization (see Hudson Kam & Newport, 2005, for a discussion of the hypothesis that adult learners might be responsible for regularization in creole languages). Although the participants exposed to scattered inconsistency in Experiment 1 did produce the main determiner forms more often than they had heard them, they did not, in most of the noise conditions, fully regularize the inconsistent forms (that is, use them virtually all the time in a given context). Only the participants exposed to 16 noise determiners approached truly rule-like usage of the main determiner forms, using them close to 90% of the time. Although pidgins and incipient creoles may exhibit scattered inconsistency like that modeled here, they do not, as a general rule, contain nearly so many forms in competition with one another, even across speakers. In the Tok Pisin example mentioned earlier, for instance, there were five forms in competition for marking continuous aspect (Sankoff, 1979); but the adult learners in Experiment 1 who were exposed to this amount of scattered inconsistency produced the main determiners only slightly more than they heard them in the input. Given these results, it does not seem likely that adult learners of a pidgin or incipient creole would regularize one of these forms to the degree that would be necessary to explain the rapid linguistic change hypothesized to occur in creolization. This in turn suggests that, although adults can regularize under certain circumstances, it is unlikely that they are the primary agents of regularization in the circumstances of real language change. In the next experiment, we investigate this question further, by directly comparing child learners with adult learners, to see whether children regularize more readily than adults and do so under circumstances closer to those of natural language change.

## 4. Experiment 3

In previous work we found that children regularized inconsistencies that adults reproduced, suggesting that children might be more likely than adults to regularize (Hudson Kam & Newport, 2005). Work by Newport and colleagues on children learning American Sign Language from non-native input also found that children regularize more than adults; the children they studied regularized to a much higher degree than most of the adults in Experiment 1 (Newport, 1999; Singleton & Newport, 2004, Ross & Newport, in prep). The latter differences in outcome could be due to the difference in language modality, but a more likely possibility is that it is due to a difference in the ages of the learners: children might regularize more readily than adults across a variety of circumstances. Experiment 3 explicitly compares children and adults when learning several different types of inconsistent languages. In particular, we ask whether children and adults differ or look the same in learning versus regularizing inconsistencies, examining both presence/absence inconsistencies and scattered inconsistencies. In addition, we ask, when adults and children do regularize inconsistency, whether they do so in the same way (though perhaps to a different degree), or rather whether they seem to be performing different processes.

Importantly, in this study we compare children's and adults' learning of the type of inconsistency that we know (from Experiment 1) adults regularize, what we have called *scattered inconsistency*.

However, we introduce only the lower levels of scatter – the 2<sup>ND</sup> and 4<sup>ND</sup> conditions – which the adult participants of Experiment 1 did not regularize very much. This allows us to see if children are more likely than adults to be systematic, perhaps doing so at lower degrees of scatter than adults, or whether they regularize under different circumstances than adults. Our own previous work suggests that this is a possibility. In Hudson Kam and Newport (2005), we found that children will regularize presence/absence inconsistency, a type of inconsistency that adults do not regularize but instead reproduce quite accurately in their speech.

We exposed children ages 5–7 to languages containing presence/absence inconsistency and some limited scattered inconsistency and then tested them to see what they had learned about the consistent and inconsistent facets of the language. In order to allow the languages to be mastered by children, the methods used in Experiment 3 were simplified from those in Experiment 1 in several ways and so are described in some detail. Because of these differences, we also tested a small number of adults with the same procedures, to ensure that any differences in results we found between adult and child learners are actually due to the age differences and not to differences in methods.

#### 4.1. Method

##### 4.1.1. Participants

Forty children participated in the study. Of those, 30 completed the study. Three children stopped attending child care in the middle of the study (two children left to go on holidays, one got sick) and seven children either did not know enough nouns to complete the production task or could not produce any sentences and so did not complete the study. Mean age of the 30 children was 5 years, 10.6 months. Thirteen of the children were male, 17 were female. Sixteen adults participated. They had a mean age of 20 years, 2.25 months. Two of the adult participants were male, 14 were female.

Child participants were recruited through local daycares and preschools that had agreed to participate. Parental consent was first obtained, and then each child was asked whether they would personally assent to participate. Most received a small toy at the end of each session. (This was against the policy of one of the preschools.)

All adult participants were students at the University of Rochester or University of California, Berkeley, at the time of the study. Adult participants were recruited through posters (Berkeley) or emails sent to people in the department subject pool (Rochester) describing the study and inviting them to participate. All were paid daily for their participation and received a bonus upon completion of the entire experiment.

##### 4.1.2. The Language

The basic language contained 17 words: 4 verbs, 12 nouns, and 1 determiner. Unlike the language in Experiment 1, there was only one noun class and therefore only one main determiner. (The vocabulary with glosses is given in Hudson Kam & Newport, 2005). Although this is a larger vocabulary than is often used with children in artificial language experiments (cf. Moeser & Olson, 1974, but see also Braine et al., 1990 and MacWhinney, 1983), it was learnable to some extent by almost all of the children. The lexicon, in conjunction with the objects used and the constraints they impose, result in 99 semantically possible sentences. Participants were exposed to only a sample of these sentences; the rest were reserved for testing.

##### 4.1.3. Presentation

Participants were typically run in groups of two or three as they were available to us. Given changes in the daily availability of the children, the grouping of the children could change from day to day. This allowed us to run numerous children at the same site within as short a time as possible. Adult participants were also run in groups, but always of two and always with the same partner. However, as described below, all testing was done with participants individually.

The exposure set consisted of 12 intransitive sentences and 12 transitive sentences. Each of the twelve nouns in the language appeared once in each syntactic position (intransitive subject, transitive subject, transitive object) in the exposure set. The intransitive sentences were split equally between the two intransitive verbs: six sentences were ‘fall’ events, six were ‘move’ events. The transitive sen-

tences consisted of three ‘inside-of’ events and nine ‘hit’ events. This reflects the fact that there are more possible ‘hit’ events than there are ‘inside-of’ events.

Pilot work suggested that the videotaped exposure method used in Experiments 1 and 2 was ineffective for use with children, so in this experiment we used live exposure. Children also found it difficult to learn the vocabulary from their presentation in sentences, so we directly taught participants the vocabulary items as well as their meanings. However, there was no explicit teaching of the grammatical aspects of the language; as in the previous experiments, participants acquired the grammar entirely through exposure to sentences and their accompanying events.<sup>11</sup> Importantly, the same methods were used with the adult participants in this experiment as were used with the child participants.

There were six exposure sessions, each of which lasted approximately 10–20 min. The seventh session was a test session. The seven sessions were completed over nine days by all participants.

Exposure proceeded as follows: The experimenter began by explaining to the participants that she was going to teach them a new language called Sillyspeak; first, they would learn some new words for things, and then some new ways to say things. For the adults, exposure began at this point. For the children, the experimenter would often chat for a few moments with the children, explaining what it means to learn another language. Most children did not know, or at least did not really understand, the words ‘word’, ‘sentence’, or ‘language’, so the concept of other languages often revolved around other people talking in ways that they did not understand. (Many of them had grandparents or parents who spoke other languages, and one day-care was teaching the children signs.). After this chat, exposure began.

On the first day participants were taught the vocabulary, excluding the determiners. The entire list was run through four times. Each run through the vocabulary began with the four verbs, always in the same order. The experimenter would say “if you want to say ‘hit’ in Sillyspeak you say /flim/,” then the same thing for /præg/ ‘inside of’, /mært/ ‘move’ and /gærn/ ‘fall’. Participants were asked to repeat the Sillyspeak word after they heard it. Each of the verbs was accompanied by a gesture, and participants (especially the children) often also repeated the gestures, although they were not asked to do so. After running through the verbs, they were taught the nouns. On the first run through the noun vocabulary, participants were shown a toy and asked to name it, and were corrected if required. This was done to ensure that they were encoding the intended meaning. The experimenter then told participants how to say the word in Sillyspeak. This was repeated three more times without having the participant name the object. The nouns were presented in a randomized order each time that was the same for all participants.

Sentences were first introduced in the second session. This session began by going through the vocabulary list once. The experimenter then demonstrated how to ‘put words together’ to ‘say bigger things.’ The experimenter began with the 12 intransitive sentences. She showed the participants an event involving the toys and then said the corresponding sentence out loud (read from a piece of paper on her lap). Participants were asked to repeat the sentence after hearing it. After the intransitive sentences, the experimenter went through the vocabulary a second time, and then went on to the 12 transitive sentences in the exposure set, done the same way as the intransitives. The exposure sentences were always performed in the same order. The third and fourth days proceeded in exactly the same way: vocabulary, intransitive sentences, vocabulary, transitive sentences. Day five consisted of one pass through the vocabulary, then the intransitive sentences, then the transitives, and then the intransitives again. Day six consisted of one pass through the vocabulary, then the transitive sentences, the intransitive sentences, and finally a second pass through the transitives. This design allowed 12 passes through the vocabulary and six through each kind of sentence. Participants were allowed to help the experimenter act out the sentences to maintain their interest and attention.

Occasionally participants had difficulty repeating a sentence. When this happened, the experimenter said the sentence a second time. This happened almost exclusively with the children and was most common on days 2 and 3, the first and second times they heard the sentences. When it oc-

<sup>11</sup> The children did frequently ask questions about the language, such as what the function of the determiner was. The experimenter told the children that she did not speak the language, however, and so could never answer their questions. They seemed to believe this, frequently making comments to each other about how the experimenter had to read the words off of a piece of paper, as if to confirm that indeed, she did not speak the language and so really could not answer their questions.

curred later, it was usually due to inattention: the children were often run in rooms where other children were playing and sometimes were distracted by their activity.

#### 4.1.4. *Experimental Manipulation*

All participants were exposed to the same basic sentences; input sentences differed across conditions only in the use of determiners. There were four determiner conditions in this experiment: completely consistent use of the determiner (100%), 60% presence/40% absence of the determiner (0 ND), 60% occurrence of the main determiner form, alternating with 2 noise determiners that each occurred 20% of the time (2 ND), and 60% occurrence of the main form and 4 noise determiners, each occurring 10% of the time (4 ND). As in Experiment 1, these percentages were true for nouns overall; individual nouns occurred with the various determiners within a range that averaged to 60%, 20%, and 10% as appropriate. Note that these percentages were true only of the nouns occurring within sentences. During vocabulary training, nouns were presented without determiners. This was the same for participants in all four input conditions.

#### 4.1.5. *Tests*

In the final session participants were given three different tests to evaluate their performance. Two tests were given to evaluate participants' knowledge of determiners; one test examined their knowledge of the consistent aspects of the grammar. Testing was always done individually. Tests were given in the order in which they are described below.

**4.1.5.1. *Sentence completion task.*** This task was designed to elicit the production of noun phrases, the part of the sentence containing the inconsistency, in order to evaluate whether participants' determiner productions varied with the type of inconsistency present in the input.

As in Experiment 1, we used a sentence completion task to accomplish this. First, the participant was shown a series of toys and asked to name them. This continued until she had named five to seven objects or it became clear that she did not know any more nouns, whichever came first. Objects which had been named became part of the participant's test set. Objects were selected (for showing) in two ways. First, the participant was always shown at least two of the three container objects (cup, barrel, and truck), since only these objects can be used with the verb /*prag*/ 'inside of.' Second, toys that had been remembered by previous participants were shown early. Often the participant would begin to spontaneously produce words she knew (both children and adults did this), and when this happened the experimenter asked what the word meant. If the participant produced the correct English word or retrieved the correct object, the object was included in the test set.

Once a set of objects that the participant could name had been selected, the sentence completion task began. Using the objects that the participant had named, the participant was shown an event or scene and was told what the sentence should mean in English, and was told what the first word of the corresponding SillySpeak sentence was. For example, the experimenter would wind the bear up (which made it move) and put it down in front of the participant and say something like, "OK, I want you to tell me how to say 'the bear moves' in SillySpeak. The first word would be /*mərt*/, right?" If the participant had difficulty, they were reminded that they knew how to say things like 'the bear falls' and 'the rhinoceros moves' in SillySpeak, but they were not reminded how to say these familiar sentences in SillySpeak. The only SillySpeak they were given by the experimenter was the relevant verb.

Many children expressed a lack of confidence with the transitive (long) sentences, so we always began with the intransitive (short) sentences. This allowed the children to gain confidence with the task before attempting the longer transitive sentences. Transitive and intransitive sentences were interspersed. The experimenter wrote down each response before moving on to the next sentence. A subset of participants was videotaped and their productions later coded for reliability by a second coder who was blind to experimental condition. As in all tests, the test sentences were novel; they had not occurred in the exposure set.

**4.1.5.2. *Determiner judgment task.*** The second test administered was a grammaticality judgment task that also examined participants' knowledge of determiner usage. Participants listened to 18 sentences and judged each of them on a four-point scale, according to how much they 'liked' or 'disliked' the sen-



tence, by pointing to one of four faces ranging from happy to sad. Participants were instructed to respond that they really liked a sentence (picking a happy face) when it sounded like a SillySpeak sentence, and to respond that they really disliked a sentence (picking a sad face) when it sounded completely different from SillySpeak. They were also told that if they thought a sentence was mostly, but not completely, like or unlike sentences from the language, they should use the middle of the scale (slightly happy and slightly sad faces).

The 18 test sentences consisted of three variations on six base sentences. The form of the variations differed for participants in different input conditions. For all participants, one form of the sentence was correct, and one had no determiner at all. For participants who had heard either completely consistent input or input containing presence/absence inconsistency, the third variant had the determiner in the wrong location (preceding the noun). For participants whose input had included noise determiners, the third variant contained the noise determiner /te/.<sup>12</sup>

Two of the six base sentences varied the determiner occurring with a transitive subject, two varied the determiner occurring with a transitive object, and two were intransitive (therefore varying the determiner occurring with the subject). Sentences were presented by audiotape recorder, and the experimenter recorded responses on a response sheet. Participants had 4 s in which to respond to each test item (although participants were allowed a little extra time by pausing the tape player). All test sentences were novel and had not occurred in the exposure set.

*4.1.5.3. General Grammar Test.* The third test examined what participants had learned about aspects of the language that were always represented consistently in the input. Specifically, this test examined whether participants thought that sentences required verbs, and if they knew that some verbs (the transitives) required two nouns and others (the intransitives) allowed only one noun. In this task, participants listened to 16 sentences and were asked to judge each using the same set of faces used in the previous task. The 16 sentences were actually two versions of each of eight sentences, one grammatical and the other ungrammatical. For the transitive sentences with missing arguments it was always the object that was missing. All nouns occurred with determiners. Test sentences were randomized with the constraint that the two versions of the same sentence could not follow each other. Randomization was the same for all participants. Sentences were played on an audio tape recorder, and the experimenter recorded the responses on a response sheet. As in the determiner manipulation judgment task, participants were given 4 s in which to provide a rating, although the tape was paused to allow them to respond if needed. Again, all test sentences were novel; none appeared in the exposure set.

## 4.2. Results

As above, we present the results from the general grammar test first.

### 4.2.1. General Grammar Test

As in Experiment 1, this test examined participants' knowledge of parts of the grammar other than determiners and served to ensure that learners successfully acquired those parts of the grammar that were represented consistently in the input. The test examined participants' knowledge of sentence construction (did they know that a verb is required in every sentence?) and verb subcategorization (did they know that one set of verbs is transitive, requiring two nouns, and another intransitive, requiring only one noun?).

Fig. 9 shows the mean ratings given to grammatical and ungrammatical strings for child and adult participants. Two children, one in the 100% condition, one in the 60% + 4<sup>ND</sup> condition, did not contribute data for this task or the determiner judgment task. Both responded prior to hearing each sentence and so their data were excluded. We subjected the data to a repeated-measures ANOVA with rule type

<sup>12</sup> While it would have been desirable to have all four types of test items in all conditions, children were unable to attend to such a lengthy test. We retained test items with no determiner for all participants, since this would permit comparison with Experiment 1 and also would allow us to assess how children responded to having vocabulary training of nouns with no determiners. Across conditions, we can assess how children respond to all four types of test items.

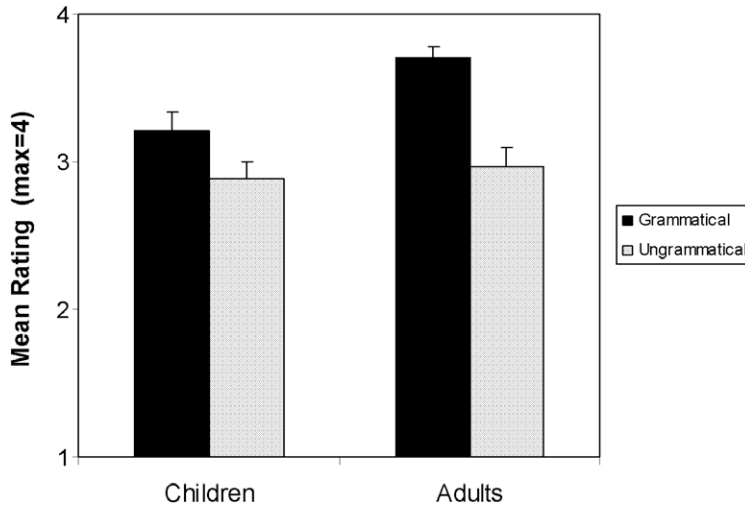


Fig. 9. Mean ratings given to grammatical and ungrammatical strings by child and adult participants.

(two levels) and grammaticality (two levels) as within-subject factors and input type (four levels) and age group (two levels) as between-subjects factors. Grammaticality was significant ( $F_{(1,36)} = 64.69$ ,  $p < .0001$ ); participants rated grammatical strings higher than ungrammatical ones. Rule type was also significant ( $F_{(1,36)} = 18.23$ ,  $p < .0001$ ), with mean ratings for items testing verb subcategorization higher than those testing sentence structure. The main effect of age group was significant ( $F_{(1,36)} = 4.48$ ,  $p = .041$ ); reflecting the fact that children's overall mean ratings were lower than adults'. The interaction between age and grammaticality was significant ( $F_{(1,36)} = 7.70$ ,  $p = .009$ ) because children differentiated less between grammatical and ungrammatical strings than adults, particularly for the strings testing basic sentences structure (rule  $\times$  grammaticality  $\times$  age  $F_{(1,36)} = 5.65$ ,  $p = .023$ ). However, these were differences of degree of differentiation – grammatical strings were always rated higher than ungrammatical strings by both ages and for both rules.

In contrast to Experiment 1, input type was significant ( $F_{(1,36)} = 3.63$ ,  $p < .022$ ). In general, however, this does not result from participants in different input groups showing different abilities to differentiate grammatical and ungrammatical strings. Rather, participants in the 100% condition gave higher mean ratings overall. Simple contrasts with rule type and grammaticality as within-subjects variables reveal that grammaticality is a significant variable for each input condition (100%:  $F_{(1,11)} = 31.73$ ,  $p < .0001$ ; 60%:  $F_{(1,10)} = 9.38$ ,  $p < .012$ ; 60% + 2ND:  $F_{(1,10)} = 6.12$ ,  $p < .033$ ; 60% + 4ND:  $F_{(1,9)} = 9.46$ ,  $p < .013$ ). (The age by input type interaction was not significant, so these contrasts were conducted with data from adults and children pooled together.) Overall, then, participants performed very well on this test, and all participants, regardless of their age or the quality of the inconsistency in their input, learned the consistent parts of the grammar.

#### 4.2.2. Sentence completion task

This test examined participants' own use of determiners. In particular, we examined whether participants' use of determiners differed with the quality of inconsistency in their input.

4.2.2.1. *Reliability.* Agreement was 100% between the live transcriptions and those produced from videotapes by a second coder who was blind to the experimental condition.

4.2.2.2. *Production.* Fig. 10 shows the mean percentage production of the main determiner forms for adults and children in each of the four input groups (100%, 60% + 0ND, 60% + 2ND, 60% + 4ND). In an ANOVA with age group and input type as between-subjects factors, only input type was significant

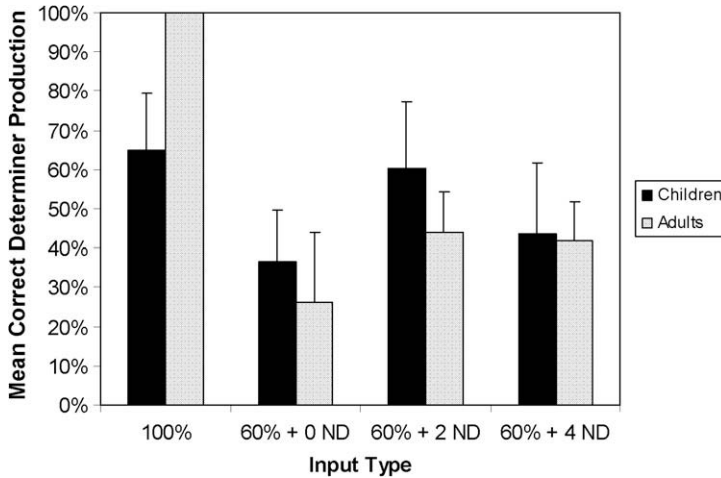


Fig. 10. Mean main determiner production for adults and children in each input group.

( $F_{(3,38)} = 3.5, p = .025$ ), suggesting that adults and children performed very similarly and were not regularizing the inconsistency present in the language.

However, this analysis potentially hides a difference between adult and child learners; it is possible that individual participants were using determiners in consistent ways not evident in the overall means. We therefore examined the data for each individual, to see if there was any evidence for individual systematicity. We examined each participant's productions for patterns in her speech and then classified each participant, according to the presence or absence of a pattern, as a systematic or unsystematic speaker. Classification as displaying a pattern in determiner production required the speaker to meet a very strict criterion, with all or all but one of the productions by that speaker observing the pattern. In this analysis we found a small number of production categories: Systematic speakers used the main determiner form all the time (*systematic users*), never used any determiners (*systematic non-users*), used one of the noise determiner forms all the time (*systematic noise users*: this had to be the same noise form all the time), or were systematic in some other way (*systematic other*: see below for examples). Unsystematic speakers fell into two categories. *Variable users* were participants who used the main determiner forms inconsistently, probability matching the inconsistency of their input. *Scatter users* were participants who used main and noise determiner forms inconsistently and in variation with each other.

Table 1 shows the number of speakers falling into each category for adults and children in each input group. Clearly, the children are performing very differently than the adults. Among adults, the only systematic use of determiners occurred in the 100% consistent input condition. Once input became inconsistent, adults became either non-users (two in the 60% presence/absence condition) or, more frequently, used determiners in a variable or scattered way, as in their input. The children, in contrast, are virtually always systematic, regardless of the input condition. Children's patterns are distributed throughout the various systematicity categories, with children displaying not only systematic use of the main determiner (12 of the 30 children), but also other types of systematic production. Most interesting are the three children classified as 'Systematic Other,' who imposed their own systematicity on the determiner system of the language. Two of these children produced determiners with nouns in transitive sentences but not intransitive sentences. The child exposed to consistent input used the main form /pə/; the child exposed to 2 noise determiners used an idiosyncratic form (/me/) that appears to be a blend of the 2 noise forms, /mæg/ and /te/. The third child produced the main determiner form with object nouns, but never with subjects, transitive or intransitive. There was no evidence for any such patterns in the input data; these patterns were introduced by the children. No adults produced any similar patterns.

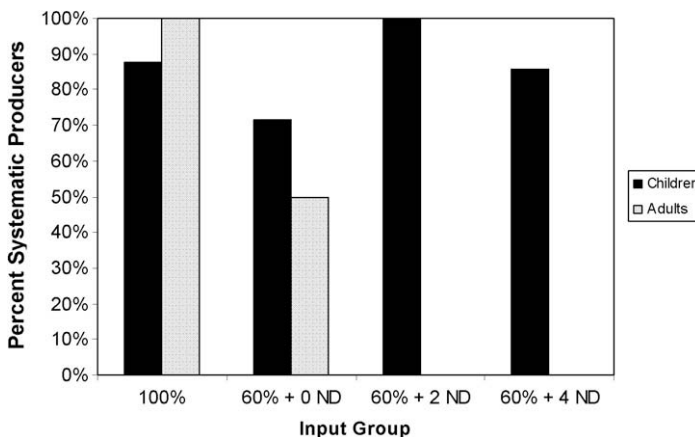
**Table 1**

Number of participants in each production sub-category by input group.

Input group	Production Type							
	Systematic user	Systematic non-user	Systematic noise user	Systematic other	Variable user	Scatter user	Systematic total	Unsystematic total
<i>Children</i>								
100%	4	2	0	1	1	0	7	1
60%	1	4	0	0	2	0	5	2
60% + 2 <sup>ND</sup>	5	1	1	1	0	0	8	0
60% + 4 <sup>ND</sup>	2	3	0	1	1	0	6	1
							Total 26	4
<i>Adults</i>								
100%	4	0	0	0	0	0	4	0
60%	0	2	0	0	2	0	2	2
60% + 2 <sup>ND</sup>	0	0	0	0	0	4	0	4
60% + 4 <sup>ND</sup>	0	0	0	0	0	4	0	4
							Total 6	10

Fig. 11 shows the overall percentage of child and adult participants in each of the input groups that were systematic speakers. Chi-square analyses show that input is a significant determiner of systematicity for adults (Pearson Chi-square (1,3) = 11.773,  $p = .008$ ), but not for children (Pearson Chi-square (1,3) = 2.648,  $p = .449$ ). As already noted, adults were all systematic users of determiners in the 100% condition; but in this condition they were merely reflecting the systematic appearance of determiners in their input. In the 60% presence/absence condition, two adults were systematic non-users, omitting all determiners; but all other adults in all conditions were inconsistent users of determiners, either probability matching with the main determiner or variably alternating between the main and noise determiners. Overall, then, adults were reflecting the inconsistency of determiners in their input. Because the levels of scattered inconsistency in this experiment were never as complex as the most extreme conditions of Experiment 1, adults did not display here the tendency to regularize that was seen there with 16 noise determiners. In sharp contrast, children were virtually always using determiners systematically, regardless of their input condition, and did so just as often when the input displayed scattered inconsistency as they did when the input was perfectly consistent.

Fig. 12 shows the children from Experiment 3 compared with the results of the same type of analysis on the adult data from Experiment 1. (Note that there are no data for children in the 8<sup>ND</sup> and 16<sup>ND</sup> conditions, of course, because children were not run in these conditions.) Again we find that the de-

**Fig. 11.** Percentage of child and adult participants in each input group classified as systematic.

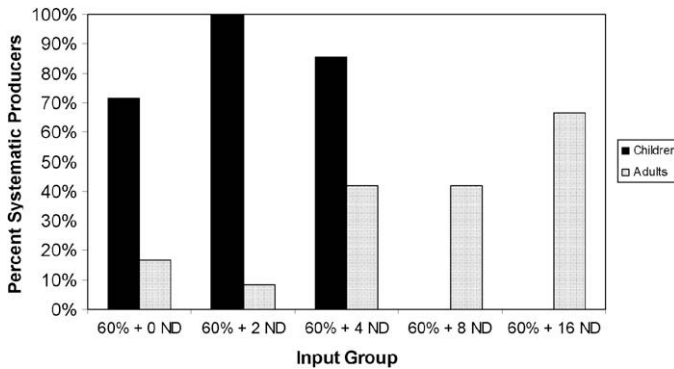


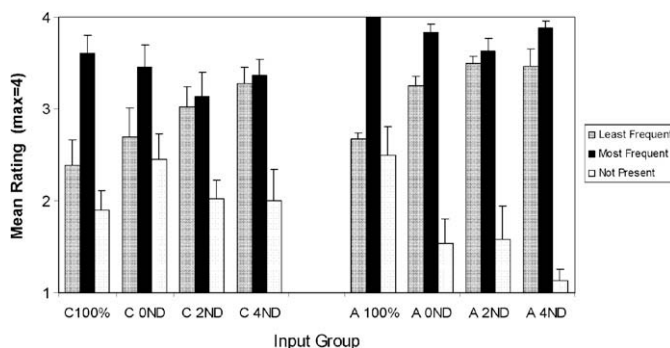
Fig. 12. Percentage of child (Exp 2) and adult (Exp 1) participants given inconsistent input classified as systematic.

gree of regularization differs between the two age groups. As already noted, the children were very systematic, whatever their input. The adults, in contrast, were much less likely to be systematic, only approaching child-like levels with 16 noise determiners in the input. Some adults were more systematic in their productions when they heard small numbers of noise forms, but very few. Although there were no adults classified as ‘Systematic Other’ in Experiment 3, there were five adults in Experiment 1 who imposed their own systematic rules on the language. (One was in the 0<sub>ND</sub> group, one was in the 4<sub>ND</sub> group, two were in the 8<sub>ND</sub> group, and one was in the 16<sub>ND</sub> group.) However, the rules imposed by these adults were very different in nature from those of the children: all five used individual determiners systematically with individual nouns. That is, they were systematic with respect to individual lexical items, not in terms of higher level or more general categories.

#### 4.2.3. Determiner judgment task

This task was designed to assess participants’ knowledge of determiners in a different way, through judgments. As described above, participants were asked to rate 18 novel sentences one at a time. The sentence types were defined by their frequency of occurrence in the input. Six of the test sentences were the type most frequently encountered in the input (main determiner form), six were the type less frequently encountered in the input (no determiner or noise determiner), and the remaining six were a type not encountered in the input (determiner in the wrong location or no determiner). (For participants in the 100% input groups there was only one type of sentence in the input and thus only two categories of item types, present and not present. For ease of presentation, however, the two kinds of non-present sentences are shown separately for these participants as well.) This allowed us to assess how participants’ judgments would be affected by the frequency of the sentence type in the input.

Fig. 13 shows the mean ratings given by child and adult participants in each input group to sentences from the three item types. To ask whether the ratings given to the three types of sentences differed and whether input type and/or age affected those ratings, we entered the rating data into a repeated-measures ANOVA with sentence type as a within-subjects factor and age group and input type as between-subjects factors. Neither the age of the learner nor the input type had a significant main effect (Age group:  $F_{(1,36)} = 1.26$ , *ns*; Input Type:  $F_{(3,36)} = .038$ , *ns*). Sentence type had a significant effect on the ratings given by participants ( $F_{(2,72)} = 118.04$ ,  $p < .001$ ), but this effect was modulated by significant interactions between sentence type and input type ( $F_{(6,72)} = 4.72$ ,  $p < .001$ ), and between sentence type and age ( $F_{(2,72)} = 2.29$ ,  $p < .001$ ). The first interaction reflects the fact that all participants liked best sentences that had occurred frequently in their input, but differed in their ratings of the other sentence types. The latter interaction reflects the fact that the children give lower high ratings and higher low ratings than the adults. The three-way interaction between sentence type, age group, and input type was not significant ( $F_{(6,72)} = 1.86$ , *ns*). The ratings, then, do not reflect the same differences between adults and children that we found in the production data. Children apparently do recognize the sentence forms with noise determiners, even if they do not produce them.



**Fig. 13.** Mean ratings (Min = 1, Max = 4) for Least Frequent, Most Frequent, and Non-occurring sentence types, adult and child participants by input group.

## 5. Discussion

Previous research has suggested that, in a number of important circumstances, language learners exposed to inconsistent use of grammatical forms may *regularize* these usages – that is, they may turn these inconsistencies into new rules of the language (Bickerton, 1981; Newport, 1999; Singleton & Newport, 2004; Traugott, 1977). In the present studies we have investigated the factors that may produce such regularization. In Hudson Kam and Newport (2005) we found that adult learners, exposed to varying proportions of present versus absent determiners (ranging from 45% to 100% determiners present in the language), never regularized the appearance of the determiners but rather acquired and reproduced in their own speech the same probabilities that were present in their input. In Experiment 1 of the present paper we presented adult learners with a different type of inconsistency, more like that found in natural language circumstances producing regularization in children. Here we exposed learners to scattered inconsistency, in which a main determiner occurred probabilistically and was always present in 60% of the noun phrases, but other determiner forms (like errors produced by late learning parents) occurred at lower frequencies, also inconsistently. Across conditions in which these noise determiners varied from two forms each at 20% to 16 forms each at 2.5%, adult learners produced the main determiners more and more regularly; in the 16<sup>ND</sup> condition, they produced the main determiner almost 90% of the time. We thus demonstrated that, at least under conditions of extremely complex variation, adult learners will begin to regularize inconsistent grammatical forms. In Experiment 2 we showed that this pattern of regularization is not due merely to the low frequency of the noise determiners: when determiners were used with exactly the same low frequencies as in Experiment 1, but quite consistently (in perfect association with particular nouns in the language), adults did not regularize but instead reproduced all of the determiner probabilities fairly accurately. These two sets of results thus show that it is the combination of inconsistency and a particular pattern of high and low frequency forms that leads to regularization in adult learners.

In Experiment 3 we investigated similar conditions in child learners, as compared with adults, and found that, unlike adults, children almost always regularized the use of inconsistent forms: when there was a fairly simple variation in the presence or absence of a form, variation among one main form and 2 noise determiners, and also one main form and 4 noise determiners. Indeed, for children, there was no change in the tendency to regularize across these conditions: they did so equally strongly across all conditions of inconsistent input. They did not always regularize use of the main determiner. While many children did do this (12 out of 30), others regularized the inconsistencies by omitting all determiners (10 out of 30), one regularized the use of noise determiners, and a few (3 out of 30) formed other regular patterns, such as using determiners with nouns in transitive but not intransitive sentences (2) or using determiners with object nouns but not subjects (1). The important generalization about child learners thus seems to be that they make inconsistent input more regular. Further research is needed to clarify the direction of these regularizations and the degree to which they can be

pushed by details of the input and learning circumstances (Austin, Newport & Wonnacott, in progress).

In the present research, these regularizations in production are not always matched by regularization in children's ratings of the familiarity of sentence forms containing the same determiners; in some conditions, children are able to reflect in their ratings the more graded statistics of their input. However, in previous research we have obtained regularization effects in ratings as well as in production (Hudson Kam & Newport, 2005). Further research will therefore be required to determine whether children's regularizations are especially characteristic of production or are characteristic of all measures of their knowledge of the language.

If these results reflect the tendency of children to regularize inconsistencies in natural language acquisition, why do children regularize non-native input or input in emerging contact languages, but do not (permanently) regularize irregular morphemes or variable rules elsewhere? As we have noted earlier, variable rules (such as *-ing/-in* variation) describe variation among forms that is predictable, contextually dependent, and consistent across speakers (Labov, 1969, 1994). Similarly, irregular morphemes (such as the past tense morpheme in English), while irregular across verbs, are entirely consistent for individual lexical items (*went* is always the past tense of *go*). Under such circumstances, learners will apparently master the variation (Labov, 1994), though they may produce some regularization errors along the way (Marcus et al., 1992). In contrast, it is particularly the inconsistent variation characteristic of non-native input, which we have modeled in our experiments, that children apparently regularize.

To consider these results further, we turn next to two related questions: Are adults and children displaying similar tendencies to regularize and differing only in the degree to which the complexity of inconsistencies affects their behavior, or do they differ qualitatively in their tendencies to regularize? More generally, what are the types of mechanisms that could underlie regularization processes and the differences between children and adults?

### 5.1. Mechanisms of change

The important question raised by these results concerns the nature of the learning mechanism that, under certain circumstances when exposed to inconsistently used grammatical forms, results in the formation of regular, rule-like processes. Two types of mechanisms are discussed prominently in the language acquisition literature.

One possibility often suggested by those interested in language change (e.g. Bickerton, 1981, 1984; Traugott, 1973, 1977) is that children impose these kinds of changes on languages because they have access to innate domain-specific knowledge about the structure of languages. Bickerton suggests that, when children receive unnatural input, they change it in ways that accord with what they know about natural language structure. Most relevant to the present case, natural languages contain consistent, regular rules that apply obligatorily in specific contexts, but they do not typically contain processes or forms that appear unpredictably and entirely probabilistically. On this view, children would change inconsistent usages into consistent grammatical rules due to domain-specific tendencies to form languages in this way; but adults would not do so, having passed a domain-specific critical period for language acquisition. This type of hypothesis has been invoked to account for the creolization of young languages (Bickerton, 1981; Lumsden, 1999) and for certain phenomena in the process of historical language change (Kiparsky, 1971; Slobin, 1977; Traugott, 1973, 1977).

A different possibility, which we have suggested in earlier work, is that learners change inconsistent input when they find it too complex to learn veridically (Hudson Kam & Newport, 2005; Newport, 1999). On this view, children should regularize more than adults because they can be overwhelmed by much simpler input than are adults, but it should be possible to induce adults to regularize if they are presented with complex enough input. This hypothesis accords well with hints of similar effects in the literature on non-linguistic probability learning (Bever, 1982; Gardner, 1957; Stevenson & Weir, 1959; Weir, 1964, 1972) and with previous results on non-linguistic pattern learning (Goldowsky, 1995), as well as with two related hypotheses in the literature on age effects in language acquisition.

Newport's (1990) Less-is-More hypothesis suggests that the well known differences in adults' and children's language learning abilities are due to children's more limited memory capacities: children

may have an advantage in learning componentially organized forms (such as morphology) due to limitations on their ability to store complex forms holistically. In a similar vein, Elman (1993) has suggested that limitations on short-term memory capacity might help children to learn long-distance dependencies, by focusing their learning first on the local instances of these dependencies. Here we suggest another version of such a 'less-is-more' notion, that limitations in children's cognitive abilities might lead to increased regularization.

How exactly might differences in cognitive capabilities between adults and children lead to differences in regularization? One possibility is that children are worse at directed memory search than adults. Another possibility is that children are less efficient at laying down memory traces, with the consequence that they have more difficulty retrieving specific forms (therefore especially those that are lower in frequency or less broadly or consistently used). (See Gathercole, 1998, and papers in Cowan, 1997, and Weinart & Schneider, 1995, for perspectives on children's memory.) In either case, the result will be the same: children will over-produce some forms and lose or fail to retrieve others, whereas adults will be more capable of storing and retrieving most or all the forms they were exposed to (though according to the results of Experiment 1, only up to some limit, beyond which they will begin to show the same losses of low frequency inconsistent items, and a resulting overproduction of higher frequency items, that children display).

A similar explanation has been proposed for children's over-regularizations in typical language acquisition, namely, that they result from failures to retrieve exceptional forms (Marcus et al., 1992). However, for lexical exceptions, (e.g., *went*, not *goed*, as the past tense of the verb *go*), the ambient language contains positive evidence of the correct forms used in consistent contexts, and so the occasional retrieval errors are not incorporated into the child's grammar. In contrast, when the ambient language contains the kind of inconsistent and scattered variability we are modeling, the child's own productions may become canalized over time and the child's grammar may come to reflect the child's regularized productions.

This hypothesis also makes predictions about conditions under which we would expect more or less regularization from both types of learners. For example, if we could tax adults' capacities so that they experienced retrieval difficulty, we would expect to see increased regularization; and the form of this regularization should primarily follow broad patterns and be less item-based than is otherwise typical for adults. Likewise, if we could reduce the cognitive load for children, we should see reduced regularization in their productions. There is some experimental evidence supporting the first prediction. Bybee and Slobin (1982) found that adults show overregularization of morphological forms even for words they already know when speaking under less than ideal conditions (such as severe time constraints on production). Also suggestive is a study by Pitts Cochran, McDonald, and Parault (1999). They found that adults learning ASL while performing a secondary task, although showing poorer overall learning as compared to a control group with no secondary task, showed more evidence of having learned the regularities and patterns underlying the sentences in their input. These two studies, while consistent with our hypothesis, are only suggestive, and more research is clearly required.

On the other hand, some aspects of the results from Experiments 1 and 3 suggest that the story might be a bit more complicated. In particular, while the cause of regularization might be differences in complexity caused by the interaction of age and input as we have suggested, the result of the regularization appears to differ in adult and child learners. Recall that there were three children in Experiment 3 who imposed their own systematic patterns on the language. Studies of probability learning in children have found that children are prone to display non-random patterns in their responses – for example, using a left, middle, right prediction strategy in a 3-light random probability task (Bogartz, 1965; Craig & Myers, 1963; Stevenson & Weir, 1959; Weir, 1964). Goldowsky (1995) found the same type of result in a probabilistic visual feature-prediction task modeled after the inconsistent structure of Simon's ASL input. However, in the present study these systematic 'other' patterns in children's productions appear to be based on linguistic categories, such as *subject* and *object* or *transitive* and *intransitive*. We say 'appear' because they could alternatively be viewed as rules like 'if there are two nouns in the sentence, use a determiner on the last one,' or 'if there are two nouns in the sentence, use determiners; if there is one noun, do not.' We cannot distinguish these possibilities in the present data. However, they are very different from the kinds of patterns we find in the adult learners. None of the adults in Experiment 3 showed evidence of imposing their own systematic rules on the language



(that is, no participants were classified as ‘Systematic Other’); but there were five adults in this category in the reanalysis of the data from Experiment 1. Interestingly, all of them showed quite different types of patterns than the children: all five used individual determiners systematically with individual lexical nouns; none formed general and potentially specifically linguistic patterns, such as distinguishing subject versus object or transitive versus intransitive sentences.<sup>13</sup>

Moreover, adult learners boost the frequency of usage of the main determiner forms, but not necessarily to the point of complete systematicity. As the number of noise forms increases in the input (and with them the complexity of variation), the amount by which adults boost the frequency of the main forms gradually increases, but few of them meet our strict criteria for systematic usage of determiners. It is possible that, with even more complex input, adult learners might be systematic at the same levels as children, but from the data at hand it appears that the manner in which adults and children change inconsistent languages may not be entirely the same: adults regularize, while children systematize.

## 5.2. Broader implications

Recent research in language acquisition has demonstrated that human learners are incredibly sensitive to the statistics present in linguistic input (Chambers, Onishi, & Fisher, 2003; Gómez & Gerken, 1999; Maye, Werker, & Gerken, 2002; Mintz, 2002; Newport & Aslin, 2004; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Thompson & Newport, 2007) and that they can use such distributional information to acquire aspects of both natural and artificial languages (Gerken, Wilson, & Lewis, 2005; Graf Estes, Evans, Alibali, & Saffran, 2007; Jusczyk, Hohne, & Bauman, 1999; Mattys & Jusczyk, 2001), as well as non-linguistic patterns (Creel, Newport, & Aslin, 2004; Fiser & Aslin, 2001, 2002; Hunt & Aslin, 2001; Saffran, Johnson, Aslin, & Newport, 1999; Turk-Browne, Junge, & Scholl, 2005). The central claim of statistical learning approaches to language acquisition is that the statistics of linguistic input can be used by learners to acquire the regularities of languages. However, an open question in most of this research concerns the outcome of such learning. One might imagine that statistical learning would always produce veridical outcomes, reproducing in output the statistics provided in the input. However, learning (including statistical learning) is not always veridical (Newport & Aslin, 2000, 2004; Seidenberg, MacDonald, & Saffran, 2002). We believe the present examples of regularization may be important instances of probabilistic or statistical learning in which learners will sometimes change their languages as they learn. Both production and judgment measures show that in our studies, learners do track the statistics or probabilities of the input they receive. At the same time, under specifiable circumstances the outcome of this learning is a regular, rule-like product. These findings suggest that statistical learning can entail shifts and sharpening of the input statistics, particularly when the input is inconsistent and the learners are children. A variety of phenomena related to producing and learning from inconsistent input – creolization, historical language change, and age differences in language acquisition – have often been cited as evidence for a domain-specific mechanism responsible for learning languages differently from other types of patterns. However, here we have tried to suggest that at least some aspects of these phenomena might arise from the nature of statistical learning itself.

At the same time, while we are proposing that regularization may reflect the influence of domain-general cognitive processes (such as retrieval and statistical learning), we are not claiming that language and language learning have no domain-specific components. Human languages exhibit a wide array of structured properties that as yet have no explanation or analogue in accounts of cognitive processes; indeed, there was some evidence in our child learners’ regularizations for the possible influence of such linguistic constructs. The issue of where these representations come from is well beyond the scope of this paper. It is worth noting, however, that certain aspects of these representations may be more innate and possibly domain-specific in their source. Work with home signers, for instance – deaf children or adults who have received no conventional linguistic input and are forming

<sup>13</sup> A few adult participants in Wonnacott and Newport (2005) showed their own systematic patterns like those of the children, but these were in an experimental condition where they had to produce utterances using completely novel lexical items. Whether such patterns in adults are limited to such circumstances requires further research.

a gestural language for communication with their families – has shown that their productions are highly structured according to abstract, language-like categories (Goldin-Meadow, 2003; Goldin-Meadow & Mylander, 1984). Coppola and Newport (2005) have found evidence for the grammatical categories Subject and Object, which are not present in the gestures of their parents. An important question for future studies is how biases toward such categories combine with more domain-general tendencies to regularize and systematize language input, to produce the types of patterns that recur across languages of the world.

### 5.3. Summary

Our experiments have shown that, under certain circumstances, language learners exposed to inconsistent input will regularize the inconsistencies, producing the same forms in rule-like and systematic ways. While adults most often reproduce inconsistencies, the results of our experiments demonstrate that, when the inconsistencies are great enough (when alternate forms are numerous and low frequency, as well as inconsistent), adults will begin to regularize. However, the strongest effects of systematizing and regularizing inconsistent input appear in children: children regularize under a much wider range of circumstances than adults, and also sometimes produce systematic uses of determiners that are characteristic of languages but are not patterns of their input. We have suggested that, while these results are compatible with a domain-specific account of language acquisition, they may be more readily accounted for by limitations on learners' abilities to store or retrieve forms undergoing complex variation – limitations that are typically more severe in children than adults, and may therefore lead to more regularization in child learners. We believe these phenomena may fit within a new approach to language acquisition, known as 'statistical learning,' which can capture not only the veridical learning of distributional details of languages but also a number of complexities and constraints on learning. While further research is certainly needed, we hope that these studies contribute to our growing understanding of important (and previously puzzling) phenomena of language change.

### Acknowledgments

This research was supported by NIH grant DC 00167 to E. Newport and T. Supalla, and by NIH grant HD 048572 and two NSERC post-graduate scholarships to C. Hudson Kam.

We wish to thank all the children and adults who participated in the study, and the parents and staff at Care-a-lot Childcare (North Center), Child's Play Day Care, Kids First Childcare, Mendon Child Care Center, Our Savior Child Development Center, and the Harold E. Jones Child Study Center for their cooperation. Thanks also to Joanne Esse, Joanne Wang, and Xi Sheng for assistance in conducting the experiments, and to Mike Tanenhaus, Jeff Runner, and the members of the Newport/Aslin lab for helpful comments at all stages of this project.

### References

- Adamson, H. D. (1988). *Variation theory and second language acquisition*. Washington DC: Georgetown University Press.
- Aitchison, J. (1996). Small steps or large leaps? Undergeneralization and overgeneralization in creole acquisition. In H. Wexler (Ed.), *Creole languages and language acquisition* (pp. 9–31). Berlin: Mouton de Gruyter.
- Aitchison, J., & Agnihotri, R. K. (1985). 'I deny that I'm incapable of not working all night': Divergence of negative structures in British and Indian English. In *International conference on English historical linguistics (4th) Papers from the 4th International conference on English historical linguistics* (pp. 3–14). Amsterdam: John Benjamins.
- Becker, A., & Veenstra, T. (2003). The survival of inflectional morphology in French-related creoles. *Studies in Second Language Acquisition*, 25, 283–306.
- Bever, T. G. (1982). Regression in the service of development. In T. G. Bever (Ed.), *Regression in mental development: Basic properties and mechanisms* (pp. 153–188). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bickerton, D. (1981). *Roots of language*. Ann Arbor, MI: Karoma.
- Bickerton, D. (1984). The language bioprogram hypothesis. *The Behavioral and Brain Sciences*, 7, 173–221.
- Birdsong, D. (1999). Introduction: Whys and why nots of the critical period hypothesis for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 1–22). Mahway, NJ: Lawrence Erlbaum Associates Publishers.
- Bogartz, R. S. (1965). Sequential dependencies in children's probability learning. *Journal of Experimental Psychology*, 70, 365–370.

- Braine, M. D. S., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., et al (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language*, 29, 591–610.
- Chambers, J. K., Trudgill, P., & Schilling-Estes, N. (Eds.). (2003). *The handbook of language variation and change*. Malden, MA: Blackwell Publishers.
- Bybee, J., & Slobin, D. (1982). Why small children cannot change language on their own: Evidence from the English past tense. In A. Alqvist (Ed.), *Papers from the 5th international conference on historical linguistics* (pp. 29–37). Amsterdam: John Benjamins.
- Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, 87, B69–B77.
- Cowan, N. (Ed.). (1997). *The development of memory in childhood*. Hove, East Sussex: Psychology Press.
- Coppola, M., & Newport, E. L. (2005). Grammatical subjects in home sign: Abstract linguistic structure in adult primary gesture systems without linguistic input. *Proceedings of the National Academy of Sciences*, 102, 19249–19253.
- Craig, G. J., & Myers, J. L. (1963). A developmental study of sequential two-choice decision making. *Child Development*, 34, 483–493.
- Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1119–1130.
- Elman, J. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48, 71–99.
- Estes, W. K. (1964). Probability learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 89–128). New York: Academic Press.
- Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review*, 83, 37–64.
- Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12, 499–504.
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 458–467.
- Gardner, R. A. (1957). Probability-learning with two and three choices. *American Journal of Psychology*, 70, 174–185.
- Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry*, 39, 3–27.
- Gerken, L. A., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, 32, 249–268.
- Goldin-Meadow, S. (2003). *The resilience of language*. New York: Psychology Press.
- Goldin-Meadow, S., & Mylander, C. (1984). Gestural communication in deaf children: The effects and noneffects of parental input on early language development. *Monographs of the Society for Research in Child Development*, Vol. 49 (3–4, Serial No. 207).
- Goldowsky, B. (1995). Learning structured systems from imperfect information. (Doctoral dissertation, University of Rochester, 1995). *Dissertation Abstracts International*, 56-05B, 2910.
- Goldowsky, B. N., & Newport, E. L. (1993). Modelling the effects of processing limitations on the acquisition of morphology: The less is more hypothesis. In E. Clark (Ed.), *The proceedings of the twenty-fourth annual child language research forum* (pp. 124–138). Stanford, CA: CSLI.
- Gómez, R. L., & Gerken, L. A. (1999). 11-month-olds are sensitive to structure in an artificial grammar. *Cognition*, 70, 109–135.
- Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, 18, 254–260.
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of language* (pp. 73–113). Cambridge, MA: MIT Press.
- Hudson, C. L. (2002). *Pidgin, creoles, and learners: How adults and children contribute to the formation of creole languages*. Doctoral Dissertation, University of Rochester, 2003.
- Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language variation and change. *Language Learning and Development*, 1, 151–195.
- Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, 130, 658–680.
- Johnson, J., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.
- Johnson, J., Shenkman, K., Newport, E. L., & Medin, D. (1996). Indeterminacy in the grammar of adult language learners. *Journal of Memory and Language*, 35, 335–352.
- Jourdan, C., & Keesing, R. (1997). From Fisin to Pijin: Creolization in process in the Solomon Islands. *Language in Society*, 26, 401–420.
- Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, 61, 1465–1476.
- Kanno, K. (1998). Consistency and variation in second language acquisition. *Second Language Research*, 14, 376–388.
- Keesing, R. M. (1988). *Melanesian Pidgin and the oceanic substrate*. Stanford, CA: Stanford University Press.
- Kessen, W., & Kessen, M. L. (1961). Behavior of young children in a two-choice guessing problem. *Child Development*, 32, 779–788.
- Kiparsky, P. (1971). Historical linguistics. In W. O. Dingwall (Ed.), *A survey of linguistic science* (pp. 576–642). College Park: University of Maryland Press.
- Klein, W., & Perdue, C. (1993). Utterance structure. In C. Perdue (Ed.), *Adult language acquisition: Cross-linguistic perspectives. The results* (Vol. 2, pp. 3–40). Cambridge: Cambridge University Press.
- Kotsinas, U.-B. (1988). Immigrant children's Swedish: A new variety? *Journal of Multilingual and Multicultural Development*, 9, 129–140.
- Kovac, C. (1981). Children's acquisition of variable features. *Dissertation Abstracts International*, 42(2), 687A.
- Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language*, 45, 715–762.
- Labov, W. (1989). The child as linguistic historian. *Language Variation and Change*, 1, 85–97.

- Labov, W. (1994). *Principles of linguistic change. Internal factors* (Vol. 1). Malden MA: Blackwell.
- Lumsden, J. S. (1999). Language acquisition and creolization. In M. Degraff (Ed.), *Creolization, diachrony, and language acquisition* (pp. 161–178). Cambridge, MA: MIT Press.
- MacWhinney, B. (1983). Miniature linguistic systems as tests of the use of universal operating principles in second-language learning by children and adults. *Journal of Psycholinguistic Research*, 12, 467–478.
- Marcus, G., Pinker, S., Ullman, M., Hollander, M., Rosen, T., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development* (Serial no. 228).
- Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, 78, 91–121.
- Maye, J., Werker, J., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, 101–111.
- Meisel, J., Clahsen, H., & Pienemann, M. (1981). On determining stages in natural second language acquisition. *Studies in Second Language Acquisition*, 3, 109–135.
- Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition*, 30, 678–686.
- Mooser, S. D., & Olson, A. J. (1974). The role of reference in children's acquisition of a miniature artificial language. *Journal of Experimental Psychology*, 17, 204–218.
- Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, 14, 11–28.
- Newport, E. L. (1999). Reduced input in the acquisition of signed languages: Contributions to the study of creolization. In M. Degraff (Ed.), *Creolization, diachrony, and language acquisition* (pp. 161–178). Cambridge, MA: MIT Press.
- Newport, E. L., & Aslin, R. N. (2000). Innately constrained learning: Blending old and new approaches to language acquisition. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *BUCLD 24: Proceedings of the 24th annual Boston University conference on language development* (pp. 1–21). Somerville, MA: Cascadia Press.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance 1. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127–162.
- Pitts Cochran, B., McDonald, J. L., & Parault, S. J. (1999). Too smart for their own good: The disadvantage of a superior processing capacity for adult language learners. *Journal of Memory and Language*, 41, 30–58.
- Roberts, J. (1997). Acquisition of variable rules: A study of (-t, d) deletion in preschool children. *Journal of Child Language*, 24, 351–372.
- Romaine, S., & Wright, F. (1987). Short forms in Tok Pisin. *Journal of Pidgin and Creole Languages*, 2, 62–66.
- Ross, D. S. (2001). Disentangling the nature–nurture interaction in the language acquisition process: Evidence from deaf children of hearing parents exposed to non-native input. (Doctoral dissertation, University of Rochester, 2001). *Dissertation Abstracts International*, 62-07B, 3402.
- Ross, D. S., & Newport, E. L. (1996). The development of language from non-native linguistic input. In A. Stringfellow, D. Cahana-Amity, E. Hughs, & A. Zukowski (Eds.), *Proceedings of the 20th annual Boston University Conference on language development* (pp. 623–645). Boston: Cascadia Press.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science*, 274, 1926–1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27–52.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606–621.
- Sankoff, G. (1979). The genesis of a language. In K. C. Hill (Ed.), *The genesis of language* (pp. 23–47). Ann Arbor, MI: Karoma Publishers.
- Sankoff, G. (1994). An historical and evolutionary approach to variation in the Tok Pisin verb phrase. In K. Beals, J. Denton, R. Knippen, L. Melnar, H. Suzuki, & E. Zeinfeld, (Eds.), *CLS 30 - Papers from the 30th regional meeting of the Chicago Linguistic Society, 1994, Volume 2: The parasession on variation in linguistic theory* (pp. 293–320). Chicago: Chicago Linguistic Society.
- Sankoff, G., & Laberge, S. (1973). On the acquisition of native speakers by a language. *Kivung*, 6, 32–47.
- Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002). Does grammar start where statistics stop? *Science*, 298, 553–554.
- Singleton, J. L. (1989). Restructuring of language from impoverished input: Evidence for linguistic compensation (Doctoral dissertation, University of Illinois at Urbana-Champaign, 1989). *Dissertation Abstracts International*, 50-11B, 5349.
- Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American Sign Language from impoverished input. *Cognitive Psychology*, 49, 370–407.
- Slobin, D. (1977). Language change in childhood and history. In J. Macnamara (Ed.), *Language learning and thought* (pp. 185–214). New York: Academic Press.
- Smith, J., Durham, M., & Fortune, L. (2007). 'Mam, my trousers is fa'in doon!' Community, caregiver and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change*, 19, 63–99.
- Sorace, A. (2000). Syntactic optionality in non-native grammars. *Second Language Research*, 16, 93–102.
- Stevenson, H. W., & Weir, M. W. (1959). Variables affecting childrens' performance in a probability learning task. *Journal of Experimental Psychology*, 57, 403–412.
- Stevenson, H. W., & Zigler, E. F. (1958). Probability learning in children. *Journal of Experimental Psychology*, 56, 185–192.
- Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, 3, 1–42.
- Traugott, E. C. (1973). Le changement linguistique et sa relation à l'acquisition de la langue maternelle. *Langages*, 32, 39–53.
- Traugott, E. C. (1977). Pidginization, creolization, and language change. In A. Valdman (Ed.), *Pidgin and creole linguistics* (pp. 70–98). Bloomington, IN: Indiana University Press.
- Turk-Browne, N. B., Junge, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, 134, 552–564.
- Weinart, F. E., & Schneider, W. (Eds.). (1995). *Memory performance and competencies: Issues in growth and development*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Weir, M. W. (1964). Developmental changes in problem-solving strategies. *Psychological Review*, 71, 473–490.
- Weir, M. W. (1972). Probability performance: Reinforcement procedure and number of alternatives. *American Journal of Psychology*, 85, 261–270.

- Wolfram, W. (1985). Variability in tense marking: A case for the obvious. *Language Learning*, 35, 229–253.
- Wonnacott, E., & Newport, E. L. (2005). Novelty and regularization: The effect of novel instances on rule formation. In A. Brugos, M. R. Clark-Cotton, & S. Ha (Eds.), *BUCLD 29: Proceedings of the 29th annual Boston University conference on language development* (pp. 663–673). Somerville, MA: Cascadilla Press.