The effects of age and memory on the ability to adjust to novel (artificial) dialect forms for L1 Spanish speakers*

Daniel Ross
University of Illinois at Urbana-Champaign
djross3@illinois.edu

There is general agreement in the Second Language Acquisition literature that there are age effects, and that children tend to outperform adults in ultimate attainment. Despite this, previous research has shown that adults tend to outperform children in short-term outcomes. To address the differences in those two findings, the study presented here was designed to show a young-age advantage, by measuring “flexibility,” which is to say their ability to adjust to novel forms present in the input of a novel (artificial) dialect. The results show that, although adults did well in all tasks, in some of the tasks young children also performed well: both extreme age groups outperformed the middle. This is interpreted in the Declarative/Procedural model (Ullman et al. 1997; Ullman 2012), which is based on the distinction of two memory systems: the procedural memory system, which explains the advantage in young children, and the declarative system, which explains the advantage in adults. Procedural memory is central in the processing of systematic grammar in language, while declarative memory is used for other tasks like remembering events and learning new vocabulary. This difference may help explain why children, are often found to learn languages better than adults. The paper concludes that age effects are at least partly explained by a gradual shift in memory systems, which correlates with age.

1. Introduction

This study addresses age effects on language acquisition within the Declarative/Procedural (DP) model introduced by Ullman (Ullman et al. 1997; Ullman 2001a, 2001b, 2001c, 2004, 2005, 2012). The DP model proposes that language acquisition relies on

---

* This research was approved by the UIUC Institutional Review Board, protocol number 12709.
I wish to thank everyone who has helped me in this research. First, to my advisor for the project, Tania Ionin for continued input and challenging my ideas as the developed. Second, to the native speakers who recorded the stimulus for the experiment, and to Elias Shakkour for his help setting up the experiment in Ecuador. I am especially grateful to my two undergraduate assistants, Kinga Wolska and Sara Londono who helped me to score the many hours of recordings used in the analysis of results, and to Micah Coston who provided the illustrations used for the stimulus stories. The advice from the UIUC ATLAS statistics lab was central in shaping the statistical analysis in this research. Feedback from others also helped me along the way, including the participants of the UIUC SLA group, and Professors Duane Watson and José Ignacio Hualde. Most importantly, this research was only possible due to all of the participants in Ecuador and to my friends there who helped me to schedule everything during my short trip and allowed me to use their space to perform the research. I also appreciate the help of those who participated in my pilot study to establish the methodology.
two long-term memory systems: the declarative memory system that stores facts and events
(2012: 160), while the procedural memory system stores systematic, non-conscious perceptual-
motor and cognitive skills (2012: 161). Thus, the declarative memory system is primarily
responsible for lexical knowledge that cannot be determined compositionally, and the procedural
memory system is primarily responsible for productive, instinctual grammar rules. It has been
proposed (Ullman 2001c, 2004, 2005) that differences in L1 and L2 acquisition can be explained
by differences in these underlying memory systems, with increased reliance on declarative
memory in adults for grammatical processes that are stored for children in procedural memory
(2001c: 108-109). Ullman's research has focused on relatively sharp distinction between child L1
and adult L2, without specifically addressing the gradual changes from early childhood through
adulthood, and “the precise relation between late SLA on the one hand, and both native language
acquisition and early SLA on the other, remains to be determined” (2005: 163).

This study investigates age effects in the DP model, comparing children from five
through fifteen years of age and adults. The research involved an experiment with native
speakers of Spanish repeating an artificial dialect of Spanish. Different experimental
manipulations were applied in order to test for the effects of procedural and declarative memory
systems. As predicted, the results show the effects of two underlying factors, one giving an
advantage to young children and the other giving an advantage to adults. These results support
and can be explained by the DP model with the relative use of its two memory systems gradually
shifting throughout childhood. Young children have an advantage due to procedural memory, and
that advantage decreases over time. Conversely, adults have an advantage due to declarative
memory, which increases with age throughout childhood. Because of this, young children and
adults show an advantage over intermediate ages. It is argued that the advantage for young
children from the procedural memory system corresponds to full nativelike acquisition, while the advantage for adults from the declarative memory system is effective for the short-term testing in the experiment but would not lead to the typical limited linguistic ability characteristic of adult L2 language, rather than full nativelike acquisition.

The rest of this section will discuss age effects in language acquisition and several proposed explanations, as well as provide an overview of the results and methodologies used in previous research as it relates to the current study. This section will conclude with the research goals and hypotheses for this study. Section 2 will outline the methodology used in the study including a detailed explanation of the Artificial Dialect Repetition task. Section 3 will report the results from the study, with a focus on the use of the declarative and procedural memory systems in the experimental tasks. Section 4 will look at the results in a broader context and present an explanation for age effects within the DP model, as well as suggest possibilities for future research.

1.1 Age effects in second language acquisition

Substantial previous research on the long-term results of both first and second language acquisition has shown an advantage in ultimate attainment for children over adults (Lenneberg 1967; Long 1990; Newport 1990; Slavoff & Johnson 1995; Muñoz & Singleton 2011; inter alia). Children appear to learn languages effortlessly, while adults learning a second language tend to struggle. Various explanations have been proposed, and there may be a complex set of factors involved including, in no particular order, motivation (cf. Dörnyei 1998), differences in implicit and explicit learning (cf. DeKeyser 2003; Ellis 1995; De Jong 2005; Lichtman 2012) or differences in procedural and declarative memory (cf. Ullman 2001c, 2005), cognitive
development (cf. Newport 1990), and biological changes that may be associated with puberty (cf. Lenneberg 1967), which may be directly connected with brain plasticity (Zhang & Wang 2007). Although this is one of the most discussed topics in Language Acquisition, there is still no definite answer, with many competing hypotheses and sometimes conflicting experimental results.

A popular explanation has been the Critical Period Hypothesis (Lenneberg 1967), which was developed to 1) suggest that for a first language to be learned fully, it must be learned early in life and 2) explain why we do not find any children who reach linguistic maturity in the first years of early childhood. There is a biological window for language acquisition to occur. After normal development through puberty, an adolescent or adult will not longer be able to acquire language for biological reasons. The original idea comes from similar patterns in other animals such as songbirds (Lenneberg 1967: 175; Doupe & Kuhl 1999), as well as from tenuous case studies of children who were deprived of language and never became fluent speakers as adults (Lenneberg 1967: 141-142). More recently, studies with late learners of sign language, in the case of deaf children with hearing parents, have also supported the hypothesis (Newport 1990: 12-18; Mayberry 1993).

The Critical Period Hypothesis has also been applied extensively to Second Language Acquisition with essentially the same predictions as for a first language (cf. Long 1990; Newport 1990; Slavoff & Johnson 1995; Muñoz & Singleton 2011; inter alia). Second language learners who begin early in life should be able to reach fluency and native or near-native levels, while learners who begin later will not become as proficient. As opposed to the strong critical period claims for first language acquisition (cf. Lenneberg 1967), second language acquisition data shows a “sensitive period” in second language acquisition, during which the acquisition of
language is natural and proceeds most easily, with a gradual transition to later in life when language learning proceeds with difficulty and requires effort (cf. Johnson & Newport 1989; Long 1990; Newport 1990; Slavoff & Johnson 1995; Birdsong & Molis 2001; Muñoz & Singleton 2011; *inter alia*).

One important related idea is that of *fossilization* (cf. Han 2003, 2004; Franceschina 2005; *inter alia*), which describes the state of a learner who does not continue to become more nativelike. The term itself is problematic as it has been applied inconsistently in the literature (see Han 2003 & 2004 for discussion). Broadly speaking, fossilization often occurs for older learners and rarely for younger learners, suggesting an association with ultimate attainment explained by the sensitive or critical period hypotheses. For the purposes relevant to this research, there are three ways to define fossilization. First, fossilization can be seen as a process in which a learner fossilizes, which is to say that they become less able to adjust their linguistic system based on input. Second, it may describe an imperfect final state, which, by definition, is an endpoint of ultimate attainment with no further improvement. Third, it may describe a property of a learner who is unable to learn anything new in the language due to fossilization; this third question will be most relevant to the current research. The question to be asked is what exactly predicts and defines such a property of a learner: what is it about these learners such that they are unable to continue improving their language toward nativelike-ness?¹

1.2 Increased use of declarative memory in adults

Ullman explains the differences in children and adults by a “shift of reliance from procedural memory in L1 to declarative memory in L2” (2001c: 108). This can be interpreted as

¹ This does not suggest that every language learner intends to become nativelike rather than just able to communicate effectively, but that for those who do wish to do so face a great challenge; likewise, children tend to naturally reach nativelike levels, regardless of motivation.
an increased use of declarative memory in adults, regardless of whether the language in question is an L1 or L2, for there is no reason to believe that there is a fundamental difference in memory systems in place for learning a second language. This will also allow the model to be compatible with the evidence for higher ultimate attainment in children learning an L2, and with the gradual shift from childhood to adulthood. Although procedural memory cannot plausibly store opaque non-compositional information such as lexical items, declarative memory can alternate with procedural memory in storing systematic grammatical knowledge (2001c: 109; 2005: 151).

Because adults do not reach the same levels of ultimate attainment as children, this use of declarative memory in place of procedural memory is not as effective. Although the adults can use declarative memory to imitate the performance of the children, their linguistic performance is based on a weaker foundation of memorized forms and limited analogical productivity. Despite the increased use of the declarative memory system, the procedural memory system is not entirely unavailable and with more time procedural learning can still occur (2005: 152-153).

Despite the clear evidence that children are, at least in some sense, better at learning languages than adults, it is also well established that adults are shown to learn faster than children in linguistic experiments and classroom instruction, as originally discussed in detail by Krashen, Long & Scarcella (1979) and supported by substantial empirical data (Saffran 2001; Boyd & Goldberg 2009; Wonnacott, Boyd, Thomson & Goldberg 2012). One explanation for this effect could be that adults are good at metalinguistic tasks. This interpretation is supported by Slavoff & Johnson (1995), who argue that the evidence is misleading because the effect is only observed when the adults are tested within a short time after exposure (1995: 3). They further argue that immersion contexts differ from formal instruction and linguistic experiments,

A similar explanation comes from the DP model. Declarative memory is dominant in adults, and declarative memory also may first store new information before it enters into procedural memory (Ullman 2005: 147). This means that adults learn faster because they have stronger declarative memories, while children reach higher ultimate attainment with stronger procedural memories. It is not completely understood why declarative memory becomes dominant in adults, although there is evidence for critical period effects for procedural memory (Ullman 2005: 151). Another possibility is that declarative memory improves and begins to take over some of the functionality of procedural memory. This may be useful for adults who already have a linguistic system in place but are still learning new vocabulary items, which would be facilitated by improved declarative memory. The result is that adults outperform children in learning new vocabulary (Ullman 2005: 148), while children reach higher ultimate attainment for systematic aspects of the grammar.

A possible explanation for this lower ultimate attainment in adults is Newport's (1990) “Less is More” hypothesis, which suggests that children may learn languages better than adults specifically because they do not have the cognitive resources available that adults do. Interpreted within the DP model, the claim would be that over-reliance on declarative memory in adults is at the expense of procedural learning. This hypothesis is also supported by empirical evidence such as in Conway, Ellefson & Christiansen (2003). Enhanced cognitive functions such as executive control, memory and metalinguistic awareness may be detrimental to adult learners; the fact that adults are faster than children in experiments suggests that they do rely on these enhanced abilities, while the poorer ultimate attainment suggests that this is detrimental to actually
becoming a fully nativelike speaker of the language. Although it is possible that the changes in
the child's cognitive system may only coincidentally coincide with age effects on language, this
is predicted in the DP model due to competitive interaction between the two memory systems
(Ullman 2005: 151).

Hudson Kam & Newport (2009) tested the “Less is More” hypothesis with an
exceptionally irregular artificial grammar, children were significantly less accurate than adults at
learning the irregular forms. However, they argued that poorer performances by children were
actually better in the long run because they showed regularization of the input. In this sense, the
children acted as a sort of linguistic filter, regularizing the input, while the adults were able to
perform better and maintained the irregular properties of the grammar. Although on the surface
this suggests that adults were better at learning the language, the results also show that children
may naturally respond to input differently than adults. By responding in this way and
regularizing the input, the children are showing a natural ability to learn patterns in languages.
By extending this idea to diachronic linguistics, what is found is that children may be responsible
for internal linguistic change and the regularization of pidgins to creoles, while adults tend to
simplify languages and have trouble learning certain patterns (cf. McWhorter 2007; Senghas &
Coppola 2001; Senghas, Kita & Özyürek 2004; inter alia).

The current study addresses age effects within the DP model by comparing how adults
and children from five to fifteen perform when presented with novel input and asked to repeat
the novel forms. The goal is to measure their ability to adapt their existing linguistic systems to
match the target forms in the input. This is intended to show an advantage for younger
participants who can more easily access their procedural memory systems and do not have
substantial interference from other cognitive systems used by adults.
1.3 Research with second dialects and artificial grammars

Because a dialect can be understood without extensive exposure and because the goal is to measure the ability of individuals to adjust their existing linguistic systems, participants in this study were exposed to a novel dialect rather than a completely unfamiliar language. Siegel (2010) provides a comprehensive overview of what research has been done in the small subfield of second dialect acquisition. Age effects are broadly equivalent in second dialect acquisition and second language acquisition, with the same tendency that a five year old would likely fully acquire a new dialect (as with a new language), while an adult would not and a twelve year old might fall somewhere in the middle (2010: 84). At the same time, this is based on limited studies to date and factors other than age may also be strong predictors of ultimate attainment (2010: 88-89). Strong conclusions are not possible at this point, but Siegel does note that “it appears that those who arrive in a new dialect area before adulthood will almost certainly use some features of the dialect, although not necessarily consistently,” while later arrivals may show no signs of the second dialect; phonology may be hardest for older learners, while lexical items can be learned throughout adulthood (2010: 91).

The most significant difference between learning a second language and second dialect is the amount of first language transfer involved in learning a second dialect. Unlike in the case of a second language, the learner of a second dialect can understand almost everything immediately. Transfer can be very strong, even to the point where no acquisition of the dialect occurs (2010: 139). A cognitive explanation may be perception, that adults do not perceive the new differences so that only children acquire them: “Once categories have become firmly established, learners tend to perceive many new sounds in terms of established categories instead of forming new
categories for them” (2010: 99). Speakers generally are aware of differences in other dialects, but that does not mean that speakers are able to identify the details of these features to the degree necessary to replicate them.

Other research on learning second dialects includes a detailed longitudinal case study involving three Canadian English speaking children acquiring British English (Tagliamonte & Molfenter 2007) that is useful in its level of description but difficult to use for generalizations across larger populations. The study did show that the acquisition a second dialect is a slow, gradually evolving process over years, which may be especially hard to predict in detail due to the lack of communicative need. Sharma (2005a, 2005b) considers the sociolinguistic dimension of second dialect acquisition and suggests that ultimate attainment, at least for adults, is primarily mediated by motivation and linguistic identity. Kerswill (1996) discusses the relationship between the difficulty of linguistic features and age at which they are most easily acquired, as well as the impact that children and adolescents have on language change.

For practical reasons, this study uses an artificial dialect, because the exact form of an artificial grammar can be precisely controlled for variation compared to the original language. Although artificial dialects have not been used extensively for research, previous research using artificial grammars has been popular for experimental approaches to language acquisition, and it has been used to test age effects (Saffran 2001; Gómez & Gerken 2000; Fedzechkina, Jaeger & Newport 2012; see Folia, Uddén, de Vries, Forkstam & Peterssen 2010 for an overview) and the “Less is More” hypothesis (Hudson Kam & Newport 2009; Conway, Ellefson & Christiansen 2003).

Warker et al. (Warker, Dell, Whalen & Gereg 2008; Warker 2009) used an artificial grammar to investigate the extent to which adult learners could learn and generalize
phonological rules from written input as an example. Although the results were mixed, there was substantial evidence for generalization of some rules, including some that were fairly complex over several sessions. Of course this approach can only test phonological rules, as the participants did not understand the content of the words based on read-aloud methods alone (similar to the oral repetition method in the current research). Other experiments with artificial languages designed to be understood, however, have succeeded with adults and children (Saffran 2001; Lichtman 2012), but they require practice or training beyond what is possible from limited exposure to novel input alone.

Overall, the results from research with artificial languages suggest that they can inform us about natural languages (Folia, Uddén, de Vries, Forkstam & Petersson 2010) and are therefore valid for experimental research on language acquisition, including the current study. I want to emphasize, however, that this study does not use an artificial dialect in the same sense as many other studies that use artificial languages, due to two specific differences. First, the material was presented to the participants as a real dialect of Spanish and intended to be found plausible by the participants; second, as an artificial dialect, it was relatively easily understood by the participants, either immediately or after a short adjustment period for most of the content. The result is that although it seems to be possible to use artificial languages to consider the effects of natural languages, this particular research method may be even more like natural languages in the first place.

Another line of research related to artificial dialects involves slight modifications to natural languages, primarily in the domains of syntax or lexical items, which is very similar to the current study but usually varies only in one way from the natural language in question. For example, Becker & Estigarribia (2011) investigate whether adults can generalize and
differentiate novel control and raising verbs in a modified version of English, with successful results intended to inform the field on how these verbal structures may be learned. Although the question of child language acquisition does interest the authors, they did not conduct the experiment with children yet for practical reasons, especially because it is unclear at which age these verbs are acquired in the first place (2011: 25). In the domain of phonology, Maye, Aslin & Tanenhaus (2008) used an artificial accent to demonstrate altered responses to a lexical decision task after participants were exposed to twenty minutes of speech containing altered vowels. Although their experiment did not involve production, it did show that the participants were processing and adjusting to the novel input.

A larger body of similar work that includes child and adult comparisons is that of Boyd, Goldberg et al. These studies introduce new verbal constructions (with novel object arguments). Boyd, Gottschalk & Goldberg (2009) perform several of these experiments with adults; Boyd & Goldberg (2009) reports results with the same experiments performed with children; and Wonnacott, Boyd, Thomson & Goldberg (2012) compares the performance of children and adults. Although it is shown that all participant groups can successfully acquire the novel constructions, the findings are consistent with previous research that adults acquire them faster than children, and in fact the adults are more accurate than children, at least within the limited time-span of the experiment. The goal of the research was not to prove “that the children's knowledge of the novel construction's linking rules is particularly good” but that they could perform the task at all (Wonnacott et al. 2012: 464), and although the children did not perform like the adults, “the fact that they showed significant improvement across days indicates incremental learning of the mapping from syntactic categories to argument structure” (2012: 467), while the adults behaved deterministically, consistently applying the new forms (2012:
474-476). In line with the results found by Hudson Kam & Newport (2009), Wonnacott et al. suggest that the conservativeness shown in the lack of generalization by the children may be part of what is responsible for children eventually outperforming adults in ultimate attainment by building up reliable grammatical procedures rather than generalizing based on minimal input.

In summary, using a naturalistic artificial dialect is likely to lead to generalizable results about natural language acquisition because previous research with artificial grammars and second dialect acquisition has demonstrated the same age effects known in the second language acquisition literature.

1.4 Defining flexibility, and the use of repetition tasks in language acquisition

In order to understand how languages are learned, it is necessary to understand what the process of learning a language involves. The perspective adopted in this paper is one in which learning is a process by which a linguistic system is changed over time based on noticing and reacting to input. Thus the definition of flexibility is based on how able a learner is to adjust their system based on input. A fossilized learner would be one who cannot adjust their linguistic system, while an optimal learner would respond to input by adjusting their linguistic system to produce output similar to that input. This interpretation of learning is compatible with models of language as a “complex adaptive system” (cf. Larsen-Freeman & Cameron 2008; Beckner et al. 2009, Boyd, Gottschalk & Goldberg 2009, Ellis & Larsen-Freeman 2009, and the full volume edited by Ellis & Larsen-Freeman 2009). These models predict that a linguistic system is constantly adjusting with input and that grammar in the traditional sense emerges from this input as an abstraction through analogy and statistical mechanisms. Pierrehumbert (2003), for example, gives evidence that the initial acquisition process for phonology may be active through
the first twenty years of life. Although the linguistic system is constantly adjusting, especially in
the early stages of development, in the complex adaptive system approach it is active throughout
a speaker's lifetime. In the DP model, the adjustments would take place in both memory systems,
with systematic grammar developing in the procedural memory system and adapting most
extensively in childhood and lexical information developing in declarative memory throughout
life.

By not conceptualizing of a linguistic system as a frozen entity, we can investigate age
effects by presenting participants with novel input and measuring their reactions. The method
selected to do this in the current experiment is an audio repetition task. This has been a popular
method in second language acquisition research, often used for assessing proficiency and
implicit knowledge (Erlam 2006; Gaillard 2012; see Jessop, Wataru & Yasuyo 2007 for an
overview). For example, Gaillard (2012) showed a strong correlation between French L2
proficiency and the amount of material learners could repeat. Oral repetition has even been used
successfully for pedagogical purposes (Jensen & Vinther 2003; Trofimovich & Gatbonton 2006).
For an overview of the role of input in language acquisition, see Zyzik (2009) with a thorough
discussion of the possible mechanisms in place for learners to generalize grammatical rules from
input including nativist Universal Grammar and usage-based (emergentist) explanations.

One significant concern in oral repetition experiments is that participants may be
repeating verbatim, having memorized the input sounds, rather than processing through their
linguistic systems (cf. Jessop et al. 2007). However, Acheson, Hamidi, Binder & Postle (2011)
show through a neuroimaging study that verbal working memory actually is connected to the
linguistic system. This means that therefore even material repeated verbatim is still filtered
through linguistic knowledge, at least to some degree. With the potential disadvantage of rote
repetition in mind, oral repetition will be utilized in this research as an effective way to study immediate input-output processing through the participants' adjusting linguistic systems.

1.5 Research goals and hypotheses

The goal of this study is to measure age effects in the repetition of novel linguistic forms. The results will be evaluated in the context of the DP model, which means that the results will be interpreted as due to the use of declarative and procedural memory. The hypotheses are as follows:

1) Verbal working memory increases with age due to increased use of the declarative memory system.

2) There is a positive correlation between age and linguistic performance based on the use of declarative memory, such as the ability to remember information about events and new lexical items, and this is mediated by verbal working memory.

3) There is a negative correlation between age and linguistic performance based on the use of procedural memory, such as the ability to learn systematic grammatical processes, and this is not mediated by verbal working memory.

2. Method

This study involved the repetition of stories in an artificial dialect designed to introduce novel linguistic forms to native speakers of Spanish. In total throughout the experiment, they were exposed to 7 minutes of input. Age effects were expected in how well the participants were able to notice, remember and repeat these novel forms.
2.1 Participants

In total, there were 72 participants in the study. Due to technical recording problems or participants who did not wish to finish the experiment, the results from 5 participants were excluded, with 67 subjects in total for the analysis, as shown in Table 1:

<table>
<thead>
<tr>
<th>Age</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>Adult</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>male</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>female</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>8</td>
</tr>
</tbody>
</table>

Adult ages: 16, 20, 31, 36, 41, 55 (male); 18, 21, 22, 30, 36, 38, 52, 60 (female)

All participants were native speakers of Ecuadorian Spanish. The majority of the children were tested at one public grade school in Machala, an industrial city on the southern Pacific coast of Ecuador. The school is located in a poor area, with working class parents, especially fishermen. Due to the demographics of the school, which was a boys only school until a few years ago, more boys were included than girls as shown above. The children were selected by their teachers (by distributing permission forms) and by the consent of their parents. From those with permission to participate, the researcher selected children while keeping age and gender distributions in mind. Beyond that, the selection was random. Adults and some additional children were tested in other private locations in the city, including private homes and a downtown office. The adults and children were recruited by word of mouth through friends in the area, and all who were available were included in the research.

The particular location in Ecuador is interesting for linguistic reasons. It appears to be on the border of two prominent South American dialects (cf. Esbobar 2011: 324), with the coastal dialect influenced by Andalusian Spanish as the primary local variety but also with speakers of

---

2 More precise ages (in months) were not collected from all participants, so the distribution is only accurate to the year; age 5, for example, corresponds to the range of 5;0 to 5;11.
the more conservative Andean variety from the mountains further inland. In fact, local children at the same public school from the same neighborhoods were found to have features of different dialects, presumably due to the dialects their parents spoke. For practical reasons, adults in the study were not necessarily from the local area, but all were from Ecuador, so there may have been more dialectal variation in that group.

All of the participants had very minimal experience with English, as instructed in Ecuador public schools. For most participants this meant they were far from conversational and perhaps knew a few lexical items with a strong accent. In the case of more exposure to English, this was recorded in the background information. Beyond this, almost all participants were completely monolingual, although several had experience with other languages such as Catalan and French, and this was included in the background information. Likewise, exposure to non-Ecuadorian dialects of Spanish was recorded. None of the participants had any formal linguistics training, although several of the adults had taken English classes at a university.

2.2 Procedure

All of the participants completed the same 30-45 minute experiment in a single session. The entire experiment was audio-recorded at 48KHz uncompressed on a midrange semi-professional microphone connected to a digital audio recorder for later analysis. There were four tasks total, in the following order: Language Sample Elicitation, Digit Span Memory Task, Artificial Dialect Repetition, Swahili Repetition Memory Task, in addition to Background and Debrief Interviews. The entire experiment was conducted in Spanish by the researcher. After

---

There were three exceptions for five year olds with especially short attention spans: one boy answered the Background Interview one day and returned several days later for the rest of the experiment; and a boy and a girl took a one hour break at the midpoint of the Artificial Dialect Repetition. This is unlikely to have affected the results because they would have been more distracted completing the experiment in a single sitting.
collecting informed consent (and assent from children), the researcher conducted a Background Interview to gather demographic information such as age, grade, dialect, socioeconomic status and exposure to other languages and dialects. For the rest of the experiment it was emphasized that there were no correct or incorrect responses and to do their best. Certain sections were especially difficult, so participants were constantly encouraged and told they were doing a good job, even though objectively they were performing, as expected, with great difficulty on very challenging tasks.

### 2.3 Language Sample Elicitation

The experiment began with a Language Sample Elicitation designed to elicit dialectal features and an example of the participant's normal speech in Spanish. The first component was a set of images that targeted specific everyday lexical items for pronunciation, and the second was a short story with pictures for the participants to narrate. The story showed a young boy who found a hedgehog, who he later had to set free again, which took about 90 seconds to tell. The writing in the story was in Hebrew, a language none of them knew, and the participants were asked to guess what it said. Any interpretation of the story was allowed; on occasion, if the participant stopped telling the story, the researcher asked them a content question (e.g., *What happened next? What did he do*?).

### 2.4 Digit Span Memory Task

Working memory tasks were chosen to be as similar to the experiment as possible, because it is unclear exactly what accounts for the verbal working memory system used in language production (Scott 1994; Gathercole, Pickering, Ambridge & Wearing 2004; Kane,
This memory task was adapted from the digit span task described by Gathercole et al. (2004). Participants repeat numbers digit-by-digit of increasing length. The digits used are presented in Appendix 1. Each column contains random numbers of increasing length, of which they must repeat four out of six correctly to move on to the next column. The final score reflects the number of digits the participant could reliably repeat. For example, a score of 4.75 digits would indicate that the participant completed the fourth column and was able to repeat three out of the required four five-digit numbers in the fifth. All of the numbers were played on a speaker from single-digit audio files to limit prosodic hints and to make the task equivalent for all participants.

2.5 Artificial Dialect Repetition

The main task testing the hypotheses in this research was the repetition of an artificial dialect. Participants were told that they would be listening to stories in a “different” dialect of Spanish, and the content of the stories implies that it may be a dialect from Africa (with the intention of choosing a location with which the participants would not be familiar).

2.5.1 Artificial Dialect Repetition Materials

The stories were read by trained native speakers of Peninsular Spanish. Ten systematic grammatical manipulations were also applied to create the artificial dialect. There were also ten novel lexical items in the stories, as discussed below. The grammatical manipulations are summarized in Table 2.

---

4 To save time, if the participant repeated the first three numbers correctly in a column, they moved on, assuming they would have repeated one of the last three correctly as well, as in Gathercole et al. (2004); additionally, if the first three columns appeared easy for the participants, they moved on after one number, with subsequent backtracking if they had difficulty on the longer numbers.
Table 2. Summary of manipulations by domain with examples and source language.

<table>
<thead>
<tr>
<th>#</th>
<th>Domain</th>
<th>Description</th>
<th>Example</th>
<th>Source Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phonetics</td>
<td>Palatalization: &lt;y,ll&gt; was pronounced as [ʃ], and &lt;j,g&gt; was pronounced as [ʒ].</td>
<td>allí [aji] → [aʃi] 'there' girafa [xirafa] → [ʒirafa] 'giraffe'</td>
<td>Argentinian Spanish for [ʃ]; Portuguese for [ʒ]</td>
</tr>
<tr>
<td>2</td>
<td>Phonetics</td>
<td>Diphthongs made into monophthongs: &lt;ue&gt; → [o] and &lt;ie&gt; → [e].</td>
<td>fuerte → [forte] 'strong' miedo → [medo] 'fear'</td>
<td>Originally in Latin, still found in other Romance languages</td>
</tr>
<tr>
<td>3</td>
<td>Phonology</td>
<td>Pragmatically salient information given a rising pitch (F0). Applied digitally.</td>
<td>Es solamente un esPEjo. 'It's only a MIRror.' Rising pitch in esPEjo</td>
<td>(For example, expressing surprise in English.)</td>
</tr>
<tr>
<td>4</td>
<td>Phonology</td>
<td>Word-final /o/ becomes [u].</td>
<td>miro → [miru] 'I see'</td>
<td>Brazilian Portuguese</td>
</tr>
<tr>
<td>5</td>
<td>Phonology</td>
<td>&lt;c,z&gt; pronounced as [θ].</td>
<td>cinco → [θinco] 'five'</td>
<td>Peninsular Spanish</td>
</tr>
<tr>
<td>6</td>
<td>Morphology</td>
<td>Adverbial suffix -mente truncated to -mén.</td>
<td>solamente → solamén 'only'</td>
<td>From Catalan -ment and Old French -men</td>
</tr>
<tr>
<td>7</td>
<td>Morphology</td>
<td>Irregular verbs regularized based on normal patterns.</td>
<td>dijó → deció 'said' soy → so 'I am'</td>
<td>(For example, in creolization.)</td>
</tr>
<tr>
<td>8</td>
<td>Morphology</td>
<td>Regularized masculine singular articles.</td>
<td>un → uno 'a(n)' el → lo 'the'</td>
<td>(Based on analogy to Portuguese.)</td>
</tr>
<tr>
<td>9</td>
<td>Syntax</td>
<td>All noun modifiers except articles (demonstratives, quantifiers, possessives, numbers) go after the noun.</td>
<td>Este idioma 'this language' → idioma este 'language this'</td>
<td>Based on head-initial languages like Swahili.</td>
</tr>
<tr>
<td>10</td>
<td>Syntax</td>
<td>A secondary verb goes at the end of a sentence.</td>
<td>podemos cruzar el río 'we can cross the river' → podemos el río cruzar 'we can the river cross'</td>
<td>Like underlyingly verb-final “V2” in German.</td>
</tr>
</tbody>
</table>

The manipulations ranged from Phonetics through Syntax, and they were applied consistently across all stories, without exception, to the artificial dialect. Sometimes this caused a layering effect with several changes for a single word or phrase, as would normally be encountered in a real-life second dialect situation. All of the manipulations were designed to be naturalistic and based on real languages. The result was intended to be a plausible dialect, to create a similar experience to what it would be like for the participants to learn a distant dialect or perhaps Portuguese.
All of the manipulations were pronounced by the trained native speakers, with the exception of Manipulation #3, which was done digitally in Praat for consistency by adjusting the pitch so that it was obviously different but not so much that it sounded artificial based on the judgment of the researcher. For the complete overlap between Manipulation #4 and Manipulation #8, the forms were produced with both changes, but the articles were not considered in the scoring of Manipulation #4, which already had many other instances in content words.

There were a total of four stories, each composed of a series of related sentences that were accompanied by an image to help the participants understand and remember the content and to keep the younger children focused on the experiment. The text of the stories is included in Appendix 1, and the details are summarized in Table 3:

Table 3. Stories used in the Artificial Dialect Repetition task.

<table>
<thead>
<tr>
<th>Story</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentences</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Syllables per Sentence</td>
<td>15, 17, 17</td>
<td>21, 17, 24, 21</td>
<td>18, 17, 17, 21, 17, 18</td>
<td>21, 20, 19, 19, 20, 19, 20, 24, 20</td>
</tr>
<tr>
<td>Recording Time</td>
<td>10 seconds</td>
<td>21 seconds</td>
<td>25 seconds</td>
<td>48 seconds</td>
</tr>
</tbody>
</table>

Due to the nature of the different domains in which the manipulations were applied, the frequencies were not equivalent, and the total number of instances in all four stories is shown in Table 4.

Table 4. Distribution of tokens for manipulations in artificial dialect stories.

<table>
<thead>
<tr>
<th>Manipulation</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Lexical Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instances</td>
<td>21</td>
<td>14</td>
<td>15</td>
<td>60</td>
<td>21</td>
<td>10</td>
<td>14</td>
<td>22</td>
<td>9</td>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>
There was at least one instance of each manipulation in each story. Those with the lowest frequencies were due to low type frequencies in texts; in fact, the stories were designed to maximize occurrences of these manipulations. Due to the high frequently occurrence of word-final -o in Spanish (the most common masculine suffix and a common verbal conjugation as well), it occurs nearly three times as often as any of the other manipulations.

Additionally, ten novel lexical items (one token each) were inserted in the stories, either proper names (people, places, and a name for a plant), or nonce words based on Spanish roots that were meant to be plausible and potentially comprehensible, especially given context. These were Adila (or Adilu), the name of a woman (or man); Kitale (the name of an African town); despaz ('slow'); nanasi (the name for a type of plant), tesoradores ('treasure hunters'); Isa (a girl's name); Samira (a girl's name); nacera ('of birth', as a type of party: 'birthday party'), Mamba (the name of a river); and Usena (the name of a god).

2.5.2 Artificial Dialect Repetition Procedure

All participants heard exactly the same input in the same order, with one exception. Due to the first-person perspective in the first (personal introduction) story, male participants heard a male voice and female participants heard a female voice, with the only change being in the name of the narrator to conform to common -o masculine and -a feminine endings on names in Spanish, while other no other potentially agreeing forms (such as adjectives) were present for the narrator. These two versions of the story were recorded to sound alike, with one native speaker modeling her speech on the other. After that they were digitally retimed and spliced from several recordings for maximum alignment and correlation between the production of the two stories.
Participants wore over-the-ear headphones, and the experiment was presented on a computer monitor. The sentences were accompanied by the images, and there was a brightly colored progress bar to indicate how much time was remaining in each audio clip. Additionally, icons of an ear and a face speaking indicated when the participant should listen and speak respectively. These were not explicitly introduced to the participants, but they were told to look at the monitor and oral instructions were given on how and when to repeat the material.

The task consistent of two rounds, and the participants heard each story twice in each round. The order is summarized in Table 5.

Table 5. Rounds and instructions for the Artificial Dialect Repetition task.

<table>
<thead>
<tr>
<th>Round</th>
<th>Instructions</th>
<th>For each story...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Iteration</td>
</tr>
<tr>
<td>1</td>
<td><em>Try to understand the stories, and then repeat what you hear in whatever way is natural.</em></td>
<td>A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td><em>Try to learn this dialect; repeat the style, sounds and way they speak, as if you speak like they do.</em></td>
<td>A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
</tr>
</tbody>
</table>

In Round 1, participants were expected to repeat the stories in the most natural way for them – in most cases this meant that they repeated the content primarily in standard Spanish rather than the form of the artificial dialect. When this round began, they first listened all of Story 1 for comprehension. After this, they listened to each sentence and repeated it. Then they moved on to the next story, listened to all of it for comprehension, then repeated each sentence individually. Because there is only one repetition in Round 1, this first repetition will be labelled as Round 1 in the rest of this paper. After completing all four stories this way, participants moved on to Round 2.
In Round 2, participants were told to now try to “learn” the dialect including the style and sounds, as if they spoke like the speakers on the recordings. Participants began by listening to and repeating Story 1 sentence by sentence. Then they listened to all of Story 1 and repeated the whole story, with the help of a printed copy of the images for each sentence. After this, they did the same for the other stories, first repeating sentence by sentence, then repeating the whole story. Because the stories were repeated twice this round, the first will be labelled Round 2A (sentence-by-sentence repetition), and the second will be labelled Round 2B (whole story repetition). A summary of the repetitions is provided in Table 6:

<table>
<thead>
<tr>
<th>Round</th>
<th>Focus on instructions for repeating the artificial dialect by Round.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round 1</td>
<td>Focus on content, repetition by sentence.</td>
</tr>
<tr>
<td>Round 2A</td>
<td>Focus on form, repetition by sentence.</td>
</tr>
<tr>
<td>Round 2B</td>
<td>Focus on form, repetition by full story.</td>
</tr>
</tbody>
</table>

It was expected that a practice effect would be seen from Round 1 to Round 2A, and that there would be a significant drop in performance from Round 2A to Round 2B because repeating the full stories is much more taxing on working memory. Round 1 was intended as practice as well as to record how the participants would naturally react to novel input. Round 2 was the central experimental focus because the participants were expected to try to repeat the artificial dialect accurately. Round 2A was designed to elicit relatively high performances after having already practiced once and having heard the stories three times. Round 2B was designed to eliminate any effects of verbatim repetition and demonstrate the participants using their linguistic systems to reproduce input after a significant delay. The focus in the analysis will be on Round 2A and Round 2B.
As shown above in Table 3, the length of all four stories was 1 minute, 45 seconds. After hearing each story four times, this means that they were exposed to a total of only 7 minutes of input from the artificial dialect, with repetition beginning within 15 seconds of initial exposure.

For the manipulations, scores for each participant were measured by calculating the number of instances in which they repeated the novel forms out of the total number of instances in which they repeated any form (either in standard Spanish or in the artificial dialect), in short the percentage of times that they repeated the novel form rather than the standard Spanish form. Collectively, the scores from the manipulations will represent the measure of flexibility for this study. Scores for the repetition of novel lexical items was based on the number of times the novel lexical items were repeated correctly out of the total number of novel lexical item instances in the input. Additionally, the repetition of semantic content was measured based on the percentage of information that was repeated by the participant for each sentence from the input, regardless of whether it was repeated in the novel dialect or standard Spanish, which was used to indicate comprehension and the ability to recall events using declarative memory. The details of the scoring procedure for each manipulation are included in the results section below.

2.6 Swahili Repetition Memory Task

The second verbal working memory task was designed to be as close to the Artificial Dialect Repetition as possible to reflect the same kind of processing used in that task. This task was the repetition of Swahili, a language unknown to the participants. They were told that this was another task to test their memory and that they would be listening to Swahili. Specifically, although Swahili and Spanish are similar in their phonological typologies, it was Swahili altered to conform to the phonological patterns of Spanish without violating any of the phonological
rules of the artificial dialect either, and the items were read by a native Spanish speaker. Therefore, there was no bias of learning new phonetic forms in the memory task, and what was measured was the participant's ability to repeat sound segments verbatim. Specifically, syllables were measured, and all words consisted of CV and V syllables only. There were fourteen stimuli, ranging from 2 through 12 syllables, with two stimuli each for syllable lengths of 6-8 syllables. In Swahili, these were meaningful phrases and proverbs, and the native speaker reading them was aware of these meanings for a natural prosodic effect. Participants heard each item once, in increasing length, and repeated it.

The score on this task was based on the number of syllables the participants could produce accurately. This task was expected to show similar results to the Digit Span Task. Additionally, it was expected that performance on this task would demonstrate that it was not possible for the participants to repeat the content of the artificial dialect stories verbatim, when the syllable counts (listed in Table 3) were substantially longer than even the longest Swahili items; this is especially true for the Round 2B full story repetition, with the stories ranging from 49 to 192 syllables in length.

2.7 Debrief Interview and Compensation

After completing the experiment, the participants were asked whether they enjoyed it (to gauge motivation) and approximately how much of the content they were able to understand, based on their own self-assessment. At this point, participants were given additional information about the methodology and motivation of the experiment if they were interested. Children were compensated with snacks, something to drink, and several small toys. Adults received five dollars in compensation for their time.
2.8 Predicted results

Based on the hypotheses given at the end of the last section, the following results are expected from this study:

1) Performance on the verbal working memory tasks will correlate with age, showing that adults have stronger declarative memory systems than children.

2) Older participants will outperform younger participants in repeating both new lexical items and semantic content in the artificial dialect, which will correlate with the verbal working memory.

3a) Younger participants will have an advantage in repeating the novel systematic grammatical properties of the artificial dialect, showing that children have stronger procedural memory systems than adults. This will not correlate with verbal working memory.

3b) Older participants may also show an advantage in repeating the novel systematic grammatical properties of the artificial dialect, by relying on the declarative memory system, which will correlate with verbal working memory.

3c) If there are any differences across manipulations relating to (3a) and (3b), they will relate to how declarative and procedural memory are used to process the specific manipulations.

Because Prediction 3a and Prediction 3b make opposite claims, it will be important to evaluate them based on the performance on other tasks, especially if adults outperform children on some tasks but not others. Additionally, if there is a nonlinear relationship between age and performance, this will suggest interaction between the two memory systems in the DP model.
3. Results

3.1 Scoring procedure

All of the results from this study were scored using a custom-built website and database system that performed the necessary calculations and stored the values. The Background Interview and Debrief Interview information was coded categorically based on the assumed honest and accurate responses from the participants. The Language Sample was scored descriptively to gather comparison data for their later performance in the Artificial Dialect Repetition and to eliminate any participants who might unexpectedly already speak with the novel forms. The Digit Span Task was scored based on the procedure explained in Section 2.2.3 above. The Artificial Dialect Repetition and Swahili Repetition Memory Task were scored from the audio recordings by the researcher and two undergraduate research assistants who were trained on the experimental methods and received course credit. For the sound-based manipulations in the Artificial Dialect Repetition (Manipulations #1, #2, #3, #4, & #8), the native Spanish speaker undergraduate research assistant was assigned. The rest of the scoring was divided up evenly across the three scorers. Due to time limitations, only the semantic content measure, novel lexical items measure, Manipulation #1 and Manipulation #2 from the Artificial Dialect Repetition and the Swahili Repetition Memory Task were scored by multiple scorers, in which case the mean across scorers was used as the final value. For each aspect of scoring, each scorer completed all of the participants for that measure, so that all participants are scored consistently, even if the style of each scorer might vary from that of the others.
3.1.1 Swahili Repetition Memory Task scoring

The Swahili Repetition Memory Task had fourteen items (see Appendix 1), of which twelve were included in the final score. The first (2 syllables) was often repeated incorrectly, potentially due to the acoustic properties of the recording or because it was the first item, and the eleventh (9 syllables) was determined to not conform to Spanish phonology due to several adjacent VV sequences; these were considered unreliable and excluded from the analysis. The remaining twelve items ranged from 3 syllables through 12 syllables, with two each for 7 and 8 syllables in length. Each item was scored based on the total number of syllables correctly repeated in left to right order; inserted, skipped or altered syllables were not counted against the score. A correctly produced syllable was produced entirely correctly for all segments; any incorrectly produced segment invalidated the whole syllable. The scorers did not attempt to determine whether the correct repetition was by chance, scoring all correctly produced syllables as correct.

The final score for the Swahili Repetition Memory Task was calculated as an average number of syllables that the participant could repeat correctly. This was done by finding the mean syllable length of all of the items with at least one error and those items repeated perfectly that were longer than the average length when there was an error.\(^5\) In other words, the final score was an average of all items, excluding those that were repeated perfectly and would have lowered the final score because the items themselves were short and easy to repeat, and therefore not an indication of the limit of the participant's memory. This final score represented the estimated maximum number of syllables that could be reliably repeated by the participant. All

---

\(^5\) This was done under the logic that any item produced perfectly that was longer than the average of the items with at least one error meant that the participant could produce more than that error length sometimes; items that were shorter than the average error length were excluded so as to not lower the score when the evidence from the other items showed that they could in fact repeat longer sequences, and that they would have likely repeated those items correctly if they were longer, although they were not.
participants were scored by two scorers, with an average taken as the final score. There was a high correlation between the two scorers ($r=0.96; \sigma=0.47$).

### 3.1.2 Artificial Dialect Repetition scoring (declarative memory measures)

All three scorers rated the semantic content measure from the Artificial Dialect Repetition. Each sentence was rated on a 0-4 scale, and the score for that participant for that round was an average of the ratings for each sentence in that round. The scale was based on the criteria shown in Table 7. Form was not the focus on this scale, with no preference given for novel artificial dialect forms or standard Spanish forms, although paraphrasing was considered worse than repeating the same words used in the input. Half-point values were allowed as needed.

**Table 7. Criteria for scoring the repetition of semantic content.**

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Everything was repeated, or everythong but one insignificant word was repeated.</td>
</tr>
<tr>
<td>3</td>
<td>Approximately the whole idea was repeated (no significant parts missing).</td>
</tr>
<tr>
<td>2</td>
<td>About half of the content was repeated.</td>
</tr>
<tr>
<td>1</td>
<td>At least one word was repeated.</td>
</tr>
<tr>
<td>0</td>
<td>No content from the input was repeated at all.</td>
</tr>
</tbody>
</table>

When participants deviated from the order of events in the input, the content was still considered repeated; this happened often during the full story repetition in Round 2B. As a common sense rule, the scores given for this measure reflected how much content was available to be scored in the manipulations and novel lexical items measure.

The performance on the novel lexical items was measured by two scorers; a value of 0 was assigned for either a skipped item, a completely inaccurate form, or a standard Spanish form.
that was similar to the novel lexical item. A value of 1 was assigned for a correctly repeated lexical item. For items produced close to the target, a partial value was given based on the Levenshtein distance from the target. Levenshtein distance is an algorithm used often in research in Historical Linguistics and in measuring dialect distance (cf. Heeringa 2004). The algorithm computes the minimal number of insertions, deletions and substitutions to transform one string to another. For example, the string *linguistics* has a Levenshtein distance of 7 from *language*, because there is are four substitutions (*a...a...ge* for *i...i...st*) and three deletions (*-ics*). Although this is only broadly accurate at a phonemic level, more precise methods are not yet reliable (see Kondrak 2003 & 2009 for some potential approaches and discussion). These partial scores were computed based on the relative Levenshtein distance for the form that was produced to the target compared to the most plausible standard Spanish alternative\(^6\) if one was available or compared to the total length of the lexical item in segments. For example, for the target item *despaz*, standard Spanish *despacio* was considered a plausible alternative; therefore a partially correct response would need to be more similar to *despaz* [despaθ] than *despacio* [despaθio] is, a Levenshtein distance of 2. Therefore, a partially correct response had to be within 1 phoneme of *despaz*, for a value of 0.5. The final score for the novel lexical items measure reflected the participant's ability to identify, remember and repeat new lexical items, on a scale from 0 to 1.

### 3.1.3 Scoring the manipulations in the Artificial Dialect Repetition

The ten manipulations were scored in similar ways. A value of 1 represented an instance produced in the novel artificial dialect form, and a value of 0 represent an instance produced in standard Spanish. Tokens that were skipped or produced with an unrelated substitute form (for

---

\(^6\) Based on either the standard Spanish translation of the lexical item, or on frequently occurring standard Spanish substitutes produced by several participants.
example, a word not containing a target phoneme) were excluded from the analysis, with only those instances repeated in one form or another considered for the final score calculation. The final score for a given participant in a given round was calculated by averaging the values of all of the instances repeated; due to the fact that older participants repeated more content than younger participants, the amount of instances from which this measurement was based varied significantly, but this was a necessary effect of the methodology and due to differences in performance based on age. The result is that all participants are compared on what they actually did produce and how well that conformed to the novel artificial dialect forms.

Additionally, participants often produced a form that was somewhere in between the novel artificial dialect form and that of standard Spanish; the details of how these were scored are explained below, as well as what was considered a valid instance for assigning a score to the form.

**Manipulation #1.** A valid instance of Manipulation #1 required that the word produced was recognizably similar to the target word, regardless of how the target phoneme was articulated. Both of the substitutions in this manipulation involved two phonological features. From [x] to [ʒ] there was a change in place of articulation and voicing; from [j] to [ʃ] there was a change in manner of articulation and voicing. Each feature was weighted as 0.5, and within that 0.25 was given for changing the feature and another 0.25 for changing it in the correct way. For example, if the participant produced [ʒ] for the change [j] to [ʃ], a score of 0.5 was assigned because the manner of articulation was changed, but not the voicing.

**Manipulation #2.** A valid instance of Manipulation #2 likewise required that the word produced was recognizably similar to the target word. No partial values were assigned because
this was based on the impression of the scorers on a categorical change from a diphthong to a monophthong. Detailed acoustic analysis would be possible, but has not been performed.

**Manipulation #3.** Because this manipulation was based on the pragmatic effect of the word, any lexical item with the intended meaning was considered valid for this rating. The native Spanish speaker research assistant determined whether or not the form produced sounded unusual in the way that the input did. This was also a categorical decision, so no partial values were assigned.

**Manipulation #4.** This manipulation required a word ending in -o in standard Spanish, and therefore any lexical item that was semantically and morphologically similar to the intended target and ended in -o was scored. The native Spanish speaker research assistant determined whether or not the form produced sounded unusual in the way that the input did. This was also a categorical decision, so no partial values were assigned.

**Manipulation #5.** This manipulation required a form with the <c,z> [s] phoneme in standard Ecuadorian Spanish, which restricted the scoring to, essentially, the same morphological root as the target word, regardless of inflectional or derivational morphology. Like with Manipulation #1, a 0.5 value was assigned for a partially correct sound, in this case for [ʃ], which altered the place of articulation of the [s] but not in the right way to reach [θ].

**Manipulation #6.** This manipulation required an adverb ending in *-mente* in standard Spanish, and any adverb was considered acceptable due to the high productivity and transparency of the suffix. A frequent pronunciation was *-mént*, which was assigned a 0.5 value as it was halfway between *-mente* and *-mén*.

**Manipulation #7.** This manipulation required producing the same lexical item as in the input because the regularized morphology was potentially unique to that verb. Participants often
produced alternative conjugations that sounded more like the novel form, such as *decía* (say.IMPF) instead of *deciú* for *dijo* (say.PRET). Because these forms were misinterpretations of the input and not regularizing the verbal forms, these were not considered instances of repetition and thus were ignored in the calculations. Legitimately intermediate forms were given partial scores using Levenshtein distance in the same way that it was used for the novel lexical items: if a form was produced that was closer to the novel target than the standard Spanish equivalent, a relative partial score was assigned. For example, *diciú* received a partial score because it is between standard Spanish form *dijo* and novel target *deciú*. For these items, overlapping manipulations were ignored, so the production of *deciú* or *deció* was considered equivalent for Manipulation #7.

**Manipulation #8.** This manipulation required the production of the same singular masculine article. Substitutions of definiteness were therefore ignored. In scoring these items, like in Manipulation #7, the overlapping effect of Manipulation #4 was ignored, so *uno* and *unu* were equivalent, as were *lo* and *lu*. No partial values were assigned for this manipulation because the articulation of articles is often very fast with adjacent vowels blending with those in the article, and therefore the score for this item was determined by a categorical judgment by the native Spanish speaker research assistant.

**Manipulation #9.** As this was a word order change involving two lexical items, both items were required to be repeated for this to be scored; semantically similar substitutions of lexical items were allowed. As a word order manipulation, there was no need for partial values because the modifier either was or was not after the noun.

**Manipulation #10.** The scoring of this manipulation was similar to that of Manipulation #9, but slightly more complex. A two verb phrase was required as well as an object; without the
object, the word order was ambiguous. The first verb was not strictly required (an OV production still showed influence from the artificial dialect's novel form. Due to the long distance movement in these sentences, participants sometimes mixed constructions. In the event that the participant produced both VO and OV word orders (either a self-correction or repeating part of the sentence twice\(^7\)), a 0.5 value was assigned. Avoidance, such as through fronting (e.g., *I the book have read*) was ignored as structurally ambiguous.

3.2 Verbal working memory results

The results for the Digit Span Task and the Swahili Repetition Memory Task are shown in Figure 1 and Figure 2 respectively plotted against age. As predicted, there was a strong correlation between both verbal working memory tasks and age, and the two measures have very similar distributions, although the Swahili scores were on average slightly higher.\(^8\)

**Figure 1. Age vs. Digit Span**

---

7 Participants sometimes produced phrases like *I have read the book read*.
8 The results from the Digit Span task were lower than those found by Gathercole et al. (2004) but had a similar correlation with age as shown in the graph to the right. There are several possible explanations. 1) Spanish numbers have more syllables, which correlates with lower performance on memory tasks (cf. Naveh-Benjamin & Ayres 1984; and Lovatt, Avons & Masterson 2000: 16-17). For example, Naveh-Benjamin & Ayres (1984: 745) found a 0.84 digit difference in Spanish and English memory scores. 2) The results from Gathercole et al. may be unusually high, with almost 9 digits for 15 year olds (cf. 7 for native English university students in Naveh-Benjamin & Ayres), which could be due to a practice effect from the other memory tasks included in the Gathercole et al. study. 3) The recordings in the current experiment did not have natural prosodic queues for grouping the numbers or the end of the list, which would likely be helpful for listeners to remember the numbers. 4) It is also possible that there could be cultural or educational differences in the use of numbers, although that would be surprising with single digits used for everyday communication.
As predicted, the limits on verbatim repetition demonstrated by the Swahili Repetition show that it is impossible for the participants to have repeated all of the content in the stories.
verbatim because the sentences (see Table 3 above) were much longer than the Swahili items; it is likely that some of the content, especially when repeating sentence-by-sentence, was repeated verbatim, but not all of it.

The verbal working memory measures correspond well to the predictions from the DP model, in which they should correlate with declarative memory, which increases during childhood and may remain steady through most of adulthood, as shown in the graphs. Interestingly, there is no decline in later adulthood (cf. Ullman 2005: 153), but there is also a minimal amount of data for older participants in this study.

Due to the similarities in distributions of these two verbal working memory measures ($r=0.73; p<0.0001$), it seems reasonable to consider their combined effects. In fact, Figure 3 shows cleaner data revealed by averaging Digit Span (digits) and Swahili Repetition (syllables).

**Figure 3. Age vs. Verbal Working Memory.**
The “units” represented by this measure of verbal working memory are unclear, but they do seem to relate to linguistic units, perhaps lexical items. In fact, if we imagine that some sequences of syllables might have been remembered as a group (potentially due to similarity with a Spanish word), then this label appears justified.9

Because there is a clear linear relationship from five years through adulthood, but no change throughout adulthood, all adults will be considered 16 years old in the statistical models. In addition to the verbal working memory task results, the DP model predicts that there will be increasing dominance of declarative over procedural memory through the beginning of adulthood, but not beyond (Ullman 2005: 153). Beyond this, for practical statistical reasons the significant numerical difference between 16 and 60 would be overwhelming in the model, despite its relative lack of significance in predictive power. The final significant linear correlation between verbal working memory and age is shown in Figure 4 (r=0.71; p<0.0001).

Figure 4. Correlation between verbal working memory and age.

---

9 What is actually represented by verbal working memory may be the number of individual items that can be stored at one time. For adults, Miller (1956) proposed “seven plus or minus two,” while later research has suggested the number may be lower, around 4, with the ability to remember more items based on higher level grouping or “chunking” (for discussion, see Ericsson, Chase & Faloon 1980; Naveh-Benjamin & Ayres 1986: 740; Baddeley 1994; Shiffrin & Nosofsky 1994; Cowan 2000; and Cowan, Chen & Rounder 2004). For children, the lower verbal working memory scores suggest either a lower number of items, or less chunking.
3.3 Additional declarative memory measures

With the clear correspondence between age and verbal working memory established, based on improved declarative memory with age, we can now compare those aspects of the Artificial Dialect Repetition task which were intended to measure declarative memory. Figure 5 shows the average of the semantic content repetition scores for all three rounds (Round 1, Round 2A and Round 2B), hinting at a ceiling effect. Figure 6 compares the verbal working memory score and semantic content repetition directly.

Figure 5. Semantic Content Repetition (Age). Figure 6. Semantic Content Repetition (Memory).

The positive linear relationship with verbal working memory is clear despite some variation and a ceiling effect. This relationship is statistically significant ($r=0.71; p<0.0001$).

A similar relationship is found for the repetition of novel lexical items in the Artificial Dialect Repetition across all rounds. A positive relationship with age is shown in Figure 7.

---

10 The pattern was similar for all three rounds for the repetition of semantic content, as well as for the repetition of novel lexical items discussed below.
Figure 8 shows that the performance on novel lexical item repetition is very strongly predicted by verbal working memory. This relationship is statistically significant ($r=0.75$; $p<0.0001$).

As predicted, these results show that adults outperformed children in all four tasks directly related to verbal working memory. This confirms hypotheses 1 and 2, showing that declarative memory improves as age increases.

### 3.4 Flexibility results from manipulations in Artificial Dialect Repetition

The statistical analysis required to test hypothesis 3 for procedural memory will not be as straightforward as that for testing the hypotheses about declarative memory. This section will present the results for Flexibility in the ten manipulations, which is to say the ability of the participants to adjust their linguistic systems to the input and repeat the novel forms accurately as a percentage. The results from the ten manipulations do not show an obvious advantage for children over adults, with the adults in fact often outperforming the children, and the different manipulations show mixed results. This is not completely unexpected, however, due to previous
research showing adults often learning faster than children and the prediction of the DP model that this may be due to adults compensating for weaker procedural memory by using their improved declarative memory to perform well on short term tasks but not to reach high levels of ultimate attainment.

At the same time, even a lack of obvious positive correlations between age and flexibility would indicate that young children are using some factor other than verbal working memory (or declarative memory in the DP model) to complete the task. If that were not the case, then the expectation would be to find distributions similar to those of the verbal working memory tasks, novel lexical items repetition and semantic content repetition. Given the clear advantage that adults have in those contexts, the lack of a clear advantage in others is potentially informative. Furthermore, a linear regression analysis can reveal a negative correlation for age on flexibility on some manipulations, with young children given an advantage over older children and adults.

Here, only the results from Round 2A are analyzed because Round 1 had significant variation due to it being the first round and varied interpretations of the instructions, and Round 2B had more missing data than the other rounds due to the harder task of repeating the whole story. Furthermore, a preliminary analysis of the results from the other rounds showed that they do not appear incompatible with the conclusions drawn from Round 2A here. The results from all 67 participants are reported here for eight manipulations; for Manipulation #7 and Manipulation #10 the results are tentative due to reporting the currently available subset of 21 participants.

Table 8 summarizes the descriptive statistics for all manipulations with the average Flexibility scores for three age groups. In Manipulation #2 and Manipulation #9, the young children actually outperformed the older children and adults. In the other manipulations, the adults tended to perform the best. The rest of this section will discuss multiple linear regression
analyses of these effects. As hinted at by the contour of the curves in the table, there are two general patterns: positive linear correlations (such as in Manipulation #8 and Manipulation #10) and U-shaped curves with young children and adults outperforming the intermediate ages (as in Manipulation #2 and Manipulation #9).

Table 8. Descriptive statistics for Flexibility on manipulations in Artificial Dialect Repetition.

<table>
<thead>
<tr>
<th>Manipulation</th>
<th>Ages 5-7</th>
<th>Ages 8-12</th>
<th>Ages 13+</th>
<th>Contour</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: palatalized fricatives [ʃ], [ʒ]</td>
<td>0.13</td>
<td>0.18</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>2: &lt;ue&gt;, &lt;ie&gt; monophthongized</td>
<td>0.66</td>
<td>0.45</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>3: rising pitch for emphasis</td>
<td>0.18</td>
<td>0.22</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>4: word-final [o] &gt; [u]</td>
<td>0.15</td>
<td>0.13</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>5: &lt;c,z&gt; produced as [θ]</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6: adverbs end in -mén</td>
<td>0.33</td>
<td>0.54</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>7: regularized verbs</td>
<td>0.37</td>
<td>0.45</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>8: regularized articles</td>
<td>0.06</td>
<td>0.15</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>9: post-nominal modifiers</td>
<td>0.78</td>
<td>0.56</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>10: OV for secondary verbs</td>
<td>0.36</td>
<td>0.52</td>
<td>0.78</td>
<td></td>
</tr>
</tbody>
</table>

As in the declarative memory results shown earlier, the graphs in rest of this section all will show the results from individual participants. Each data point is the average Flexibility score for a given participant plotted against either their memory score or age in years. Due to the difficulty of the experiment, participants did not repeat all target items. The average is based on those items that the participant did repeat. In certain cases, some participants repeated no target items for a certain manipulation and were thus excluded from the results; this was rare.

Manipulation #1 (palatalized fricatives [ʃ], [ʒ]). The results from this manipulation show an advantage for adults in Figure 9 and for higher verbal working memory in Figure 10. These distributions at first seem to not support an advantage in children over adults; upon closer
inspection, however, there does appear to be an advantage for the youngest children over the other young children. In Figure 9 this is shown from ages 5 through 7, where there is a negative effect for age; from 8 through adulthood, the effect for age is positive. Surprisingly, there is also a similar low-end advantage effect visible in Figure 10, with those with the lowest verbal working memory scores outperforming those with midrange scores. At the same time, from midrange verbal working memory scores through the highest scores, a clear positive correlation is shown. The distributions in the low end of these graphs are our first indication of an underlying factor giving an advantage to young children.

**Figure 9. Age vs. Manipulation #1.**

**Figure 10. Memory vs. Manipulation #1.**

Given that there are two underlying factors that explain the data, there are two ways to approach it statistically. The first is to manually split the data at the midpoint and analyze the effects of both factors separately. The second is to use a single statistical model that will predict high endpoints and a low midpoints; this is characteristic of a quadratic equation, meaning that a model can be simulated by including an $x^2$ term in multiple linear regression.\(^\text{11}\) The former is

\(^{11}\) All quadratic equations can be represented using the formula $ax^2+bx+c$, and a multiple linear regression model can include those three terms: a constant $c$, $x$ with a coefficient of $b$, and $x^2$ with a coefficient of $a$, where $a$, $b$, $c$, $x$, and $y$ are variables.
theoretically sound, given that there are two separate explanations for the effects at either end of the data set, but it is also based on an arbitrary decision by the researcher about where to locate the midpoint. The latter is more convenient because it is a single model and the statistical process will determine all of the necessary steps including where to set the midpoint. Additionally, if there may be interaction or even competition between the two memory systems (Ullman 2005: 147-151), then a single model, with the understanding that it is representing interaction between the memory systems, may be theoretically valid as well. The latter approach will be adopted here as a matter of statistical convenience in comparing alternative models; the former split model approach could be applied and would likely lead to similar results. The split model approach would be useful in the situation where the single model failed to return significant results, but also could potentially lead to a Type I error (false positive) if the data is split in an unfounded, statistically insignificant way. Another advantage for the single model is that if it involves a curve that can show a gradual shift in dominance between the two underlying memory systems, which is more likely than an immediate cutoff with no transition. Figure 11 shows the resulting model compared to the raw distribution and includes the linear regression data.

---

and $c$ are determined within the model. As an example, using $memory$ as a predictor, the first step would be to create a new variable $memory^2$ based on the values from $memory$ itself (simple multiplication), then standard linear regression can compute the coefficients for both $memory$ and $memory^2$ without needing to use nonlinear regression; the result will be nonlinear (quadratic) when interpreted as an equation of $memory$ rather than $memory$ and an artificial variable $memory^2$. 

44
In the case of Manipulation #1, a quadratic model based on the verbal working memory score was found to be more explanatory than one based on age; although both generated statistically significant models, the one based on verbal working memory explained more variance and each of its component parts was also significant, unlike in the model based on age. In general, verbal working memory was found to be a better predictor of flexibility when used to generate a quadratic model. This method will be used for the other manipulations when it fits.

**Manipulation #2 (<ue>, <ie> monophthongized).** A similar distribution is observed for the second manipulation. The model generated from verbal working memory is compared to the original distribution in Figure 12, along with the details of the linear regression data.
In this case, the data also support a significant model based on age, shown in Figure 13. Comparing the models reveals higher performance at the extremes, with lower performance in the middle. This similar distribution is not surprising because age correlates closely with verbal working memory. At the same time, there is a stronger correlation for memory, confirming the decision to use memory as a predictor instead of age.
Manipulation #3 (rising pitch for emphasis). Unlike the first two manipulations, the third manipulation is not explained by a quadratic model based on verbal working memory. However, a qualitatively similar model based on a positive effect for verbal working memory and a negative effect for age is significant, as shown in Figure 14. The line representing the model is based on the change in age as well as the projected corresponding change in verbal working memory, obtained through a secondary linear regressing generating a statistically significant model from age to projected verbal working memory score.

Figure 14. Age vs. Manipulation #3.

Although the model is significant, the results are not very substantial; all ages were essentially the same in performance on this manipulation. Revisiting the domain of prosody would be worthwhile in future research to find out the lack of substantial effect is generalizable.

Manipulation #4 (word-final [o] > [u]). The results for Manipulation #4 followed the same pattern as Manipulation #1 and Manipulation #2 and the results are shown in Figure 15.
Manipulation #5 (<c,z> produced as [θ]). Overall, the participants were either unaware of the pronunciation of <c,z> as [θ], or they were unable to produce this sound. It may have also been difficult to hear in the original recording on the headphones because [s] and [θ] are acoustically similar and not easy to differentiate in ideal conditions. Whatever the explanation, it was apparently too hard, and the vast majority of participants never produced any non-standard forms. This would be an interesting area for further research to investigate how much more exposure would be needed for participants to be able to reproduce it, as well as to determine whether adults or children have an advantage with such difficult sounds.

Manipulation #6 (adverbs end in -mén). This was the first manipulation that showed only a positive correlation with verbal working memory as shown in Figure 16. This relationship

In total, only 5 participants were not at floor: an 11 year old girl who consistently produced all instances of deciú with [θ], potentially because this lexical item was frequent, a 10 year old boy who repeated one instance of decidú with [θ] correctly, a 41 year old man who produced Martínez with a [θ]; a 21 year old woman who produced [ʃ] several times instead of [θ], and a 31 year old man who did the same once. The only additional evidence of even noticing this form was shown by several participants who produced a [t], such as [tinco] for cinco, which may suggest that the form was acoustically salient but too difficult to produce for the participants.
was statistically significant but weak (r=0.26; p=0.03). There were two outliers: five year olds who performed perfectly in this task.

**Figure 16. Memory vs. Manipulation #6.**

**Figure 17. Memory vs. Manipulation #7.**

**Manipulation #7 (regularized verbs).** There was a similar positive correlation with verbal working memory for Manipulation #7 as shown in Figure 17. This relationship was statistically significant and stronger than that for Manipulation #6 (r=0.54; p<0.01).

**Manipulation #8 (regularized articles).** Like the previous two manipulations, Manipulation #8 showed a strong correlation between memory and performance shown in Figure 18, and the relationship was statistically significant (r=0.64; p<0.001). A quadratic model fit slightly better (r=0.69; p<0.0001), but the fit of its components was not as strong; regardless, there is no indication of an advantage for young children even if the path is slightly curved rather than completely linear.
Manipulation 9 (post-nominal modifiers). This manipulation was the only one in the morphosyntactic domain to show the same distribution as the earlier phonological manipulations with an advantage for young children, as shown in Figure 19. Overall, the performance on this manipulation was high, with a number of participants at ceiling and none at floor; still, there was a strong tendency for those with midrange verbal working memories to not perform very well.
Manipulation #10 (OV for secondary verbs). Like Manipulation #8, Manipulation #10 shows a strong and statistically significant correlation between age and performance ($r=0.86$; $p<0.0002$), shown in Figure 20. There is also a possibility of young outliers; in Round 2B, a five year old with low verbal working memory performed very well, but he did not repeat any applicable instances in for Round 2A or Round 1.

Figure 20. Memory vs. Manipulation #10.

In many ways, Manipulation #9 and Manipulation #10 are very similar; both are in the domain of syntax and involve inversions of standard Spanish word order, and their frequencies in the experiment were comparable (9 instances for Manipulation #9, and 8 for Manipulation #10). Despite this, the results are as distinct as those for any of the manipulations. There are two plausible explanations. The first is related to typology. Given that Spanish is generally a head-initial language and there are instances of post-nominal modifiers (adjectives) but no instances of OV verb phrases, the introduction of additional post-nominal modifiers might have been more natural than OV word order. The second is directly related to language acquisition. Because

---

13 An exception might be clitic object pronouns as in *Lo veo* (CL see.1S 'I see it').
adults performed well on both tasks, it is plausible that they relied on the same mechanisms; this suggests then that the mechanism that allowed the young children to perform well on Manipulation #9 was not sufficient for Manipulation #10. It may simply be that Manipulation #9 is easier than Manipulation #10, but there is a more theoretically interesting possibility – it may be that the local word order shift in the Manipulation #9 noun phrases was possible to acquire naturally in procedural memory while the longer distance movement of in the verb phrase of Manipulation #10 was not processable in the linguistic system; only the adults could perform this well because of their stronger declarative memory systems, which allowed either memorization of the tokens or formulation and application of a metalinguistic rule.

3.5 Summary of flexibility results

Adults performed well throughout the experiment, and there was a strong tendency for performance to increase with increasing verbal working memory, for older children and adults. However, there was also an advantage for young children with poor verbal working memories shown on some of the manipulations. Although the results generally are expected from the predictions, one significant difference is that verbal working memory was a better predictor of performance than age. Manipulation #3 did not show significant trends for age or verbal working memory, although when considered together the contribution of age was negative (giving younger children an advantage) and the contribution of verbal working memory was positive (giving adults an advantage); in the end, these effects appear to have balanced out. Manipulation #5 generated no usable data because it was too difficult for everyone. In four of the remaining eight manipulations, participants with the poorest verbal working memories had an advantage, and in the other four they did not. The results are summarized in Table 9.
Table 9. Summary of statistical analysis of Flexibility results for Artificial Dialect Repetition.

<table>
<thead>
<tr>
<th>Manipulation</th>
<th>Domain</th>
<th>Verbal Working Memory Effects</th>
<th>Low advantage?</th>
<th>High advantage?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: palatalized fricatives [ʃ], [ʒ]</td>
<td>Phonetics</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2: &lt;ue&gt;, &lt;ie&gt; monophthongized</td>
<td>Phonetics</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3: rising pitch for emphasis</td>
<td>Phonology</td>
<td>marginal</td>
<td>marginal</td>
<td></td>
</tr>
<tr>
<td>4: word-final [o] &gt; [u]</td>
<td>Phonology</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>5: &lt;c,z&gt; produced as [θ]</td>
<td>Phonology</td>
<td>too difficult for everyone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6: adverbs end in -mén</td>
<td>Morphology</td>
<td>X</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>7: regularized verbs</td>
<td>Morphology</td>
<td>X</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>8: regularized articles</td>
<td>Morphology</td>
<td>X</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>9: post-nominal modifiers</td>
<td>Syntax</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>10: OV for secondary verbs</td>
<td>Syntax</td>
<td>X</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

The advantage for high verbal working memory was ubiquitous, and as it is associated with age the generalization that adults will tend to perform faster on experimental tasks is confirmed. The advantage for low verbal working memory, which correlates with young age, was shown in the domains of phonetics and phonology and in one of the syntax manipulations but for none of the morphology manipulations.

4. Discussion

The results of this study are consistent with the hypotheses and support the DP model. The performance on the verbal working memory tasks was shown to correlate strongly with age, explained in the DP model as increased use of the declarative memory system. In the Artificial Dialect Repetition, similar effects were found for the repetition of semantic content and new lexical items, two types of information known to be stored in declarative memory (Ullman 2012: 53).
Children showed no advantages in any of these tasks, and as age increased from 5 years through adulthood, performance on the tasks increased linearly until adulthood where performance leveled out.

If adulthood corresponds to increasing use of the declarative memory system, then childhood must correspond to use of the procedural memory system. Given that procedural memory corresponds to natural processing of language as suggested by the DP model, age effects in language can be explained by the use of these different memory systems and increased reliance on the declarative memory system as a child matures. Evidence from this experiment shows that the transition is gradual and that both children and adults have both systems available, although not at the same levels of accessibility.

Distinctions in use of these two memory systems can be shown in the measures of flexibility in ability to accurately repeat the manipulations in the Artificial Dialect Repetition. The results from four of the tasks patterned like the declarative memory based tasks discussed above, which suggests that adults are relying on declarative memory to outperform children. On the other four tasks with interpretable results, the adults still did very well and older children did increasingly well as age increased. For these participants, there is no reason then to believe different memory systems or other resources are in use in any of the tasks. Thus it is argued that adults, and older children in proportion to their age, used declarative memory for all of the tasks in the experiment. Therefore, it is also predicted that their apparent advantage would not lead to the full nativelike acquisition of the artificial dialect. Although the adults did appear to be learning as well as (or better than) the children, evidence from previous ultimate attainment studies show that they would likely not maintain that advantage long-term. Instead, they used a declarative memory strategy to perform well on the immediate task using resources like
metalinguistic understanding, memorization and analogy to simulate learning the language as a child would.

The failure of the younger participants to outperform the older participants in four of the manipulations is not unexpected; it just means that they were using their weaker declarative memory systems. Their advantage appeared in the other four manipulations, which must have been based on their use of procedural memory. The specific manipulations included systematic sound changes and a syntactic word order inversion, yet they failed to perform well on other similar manipulations in the domains of morphology and syntax; if the DP model is to be supported by this evidence, there must be something about these manipulations that is incompatible with the reliance on procedural memory.

Two of the sound-based manipulations involved systematic vowel changes. Manipulation #2 changed all <ue> or <ie> diphthongs into monophthongs, and Manipulation #4 involved raising word-final [o] to [u]. The participants did understand the material (at least most of the time), which means that already they were being flexible in some sense. Maye, Aslin & Tanenhaus (2008) describe an experiment in which vowels from a novel dialect are presented to participants for 20 minutes, after which a lexical decision task showed that their perceptual systems had been affected by the input. That, to some extent, must have been the case for both children and adults in the present study. While adults relied on declarative memory to produce these forms, an explanation for the young children performing well despite lower declarative memory is that they were more easily able to extend the perceptual accommodations for the accent of the artificial dialect into production through strong procedural memory. In that way, the perceptual and productive aspects of the child's linguistic system are more closely linked and more interactive than in adults and older children. Younger children are more flexible.
Manipulation #1 and Manipulation #6 are compositionally similar, but young children only showed an advantage for Manipulation #1. Both manipulations involve the acquisition of a new form (the sounds [ʃ] and [ʒ] in Manipulation #1, and the truncated adverbial suffix -mén in Manipulation #6) that is then substituted for an existing form in the speaker's linguistic system (phonemes in Manipulation #1, and the regular productive adverb suffix -mente in Manipulation #6). The first part of this task relies on declarative memory for learning the required form; the second part of the task requires systematically applying that form, which could be accomplished either through analogy and declarative memory (likely in adults) or with procedural memory.

Given that, there is no obvious reason within the DP model that children would display an advantage in Manipulation #1 but not Manipulation #6.

One possibility is that repeating adverbs was more difficult for younger children because they tend to be used in formal, adult speech; however, if they repeat adverbs at all, then it should actually be easier to repeat the new suffix if there is no existing form that might be used instead, or if that form is relatively unfamiliar. Another possibility is that young children had an advantage in perceiving or producing a new sound. The novel adverb suffix -mén is new and must be remembered, but it otherwise relies on existing components of the grammar (that is, the phonemes /m/, /e/ and /n/). Production of [ʃ] and [ʒ] requires new movements of the tongue and other articulators. The DP model may be able to explain this after all: if young children have strong procedural memory systems, and if procedural memory is responsible for motor control (Ullman 2001c: 106), then it follows directly that young children would be better at learning to physically articulate new sounds. This may explain more findings of phonology being subject to the strongest or earliest “sensitive period” age effects (e.g., Flege, Yeni-Komshian & Liu 1999; see also Baker, Trofimovich, Flege, Mack & Halter 2008).
As already discussed, it is surprising to find such a categorical difference between Manipulation #9 and Manipulation #10 when they both involve inverting two common word orders in Spanish. The distinction is likely due to the ease with which a particular manipulation can be acquired based on processing considerations; moving a verb to the end of a sentence is, apparently, more work than moving a pre-nominal modifier to the end of that noun phrase. This particular type of finding would be worth investigating with a longer version of this experiment to see how difficult manipulations are eventually learned by young children and how adults perform after the initial boost from stronger declarative memory systems.

Finally, the lack of a young child advantage in Manipulation #7 and Manipulation #8 may be surprising considering these both involved the regularization of irregular forms, which is found in both early stages of first language acquisition (cf. Ervin & Miller 1963; Pinker 1991: 533; Marcus, Pinker, Ullman, Hollander, Rosen, Xu & Clahsen 1992; Rogers, Rakison & McClelland 2004) and part of the natural generation of grammar performed by children in the development of creoles and sign languages (cf. McWhorter 2007; Senghas & Coppola 2001; Senghas, Kita & Özyürek 2004; inter alia). The explanation may be that the children do not perceive an underlying pattern of regularization in the input, but rather perceive a new irregular form (which just happens to be logically regular). If this is the case, then there should be no advantage for young children and the results should be similar to those of repeating novel lexical items. Likewise, adults are not systematically regularizing the forms in their speech; they have used declarative memory to memorize the specific tokens from the input, which Ullman has already suggested as part of adult second language learning (2001c: 105).

Those manipulations that were easily learned by younger children (Manipulation #1, Manipulation #2, Manipulation #4, and Manipulation #9) were learned as systematic processes;
the others were either too complex to be learned in procedural memory by young children in such a short time or featured aspects that needed to be learned in declarative memory before being generalized systematically, giving adults the only advantage. In this way, the study has contributed evidence in support of the DP model. Children did not always have an advantage, as they do in ultimate attainment outcomes. This is interpreted here to suggest that some systematic grammatical properties take some time to learn procedurally, even in young children with strong declarative memories; with more exposure it is predicted that the children would begin to learn procedurally, and patterns from ultimate attainment (for all types of grammatical properties) show that children do eventually do this. In this way, the finding that adults are faster than children was partially supported, but only for certain grammatical properties, and not in a way that necessarily suggests higher ultimate attainment. Several outliers – young children who performed very well on tasks without a procedural memory advantage (as in Manipulation #6) – hinted that young age might, with more time, correlate with increased Flexibility for all types of manipulations. Future research could address the time needed to learn different structures procedurally (see also discussion of the relative difficulty of linguistic features in Kerswill 1996).

This study has indicated that it is possible to measure age effects non-longitudinally and show the advantage that young children have; it is complicated by an unrelated advantage in adults that appears to suggest that adults are learning faster than children, but previous research on ultimate attainment shows that this is not, in any serious sense, the case. Adults are learning differently and performing certain tasks faster, but children are actually on their way to becoming fluent, effortless speakers of the language. Therefore, the important measure in this experiment is that of the young children outperforming older children and in some cases as well as adults; this occurs despite or even in negative correlation with poorer declarative memories.
The ability to measure the advantage young children have without looking at ultimate attainment is both practically and theoretically interesting. Because so-called “age effects” have been shown to more closely correlate with changes in declarative memory than age, it is suggested that these age effects are only indirect. Age correlates closely with changes in the procedural and declarative memory systems, a process that is mediated by competitive interaction between the memory systems (Ullman 2005: 151). Linguistic performance is then explained by the balance between the two memory systems rather than by age directly. This is shown in Figure 21. There may also be additional causes of age effects such as motivation.

**Figure 21. Age effects on language.**

This study has measured only declarative memory directly and shown indirect evidence for changes in procedural memory. Based on the schematic in Figure 21, we can imagine that the combined effects of procedural and declarative memory could cause the U-shaped Flexibility
scores in the experiment. A theoretical graph, extended from the data gathered here, of how the balance between these systems might change over time is shown in Figure 22.

Figure 22. Theoretical representation of the effect of shifting memory systems on language.

The data in this study do not reveal anything definite, but it might be possible for continued research with this methodology to reveal a way to measure the “critical period” (or sensitive period) as children are approaching or passing it. What would actually be measured is the balance between declarative and procedural memory, and the point at which dominance shifts between the two systems would indicate a substantial change in the way that the individual processes linguistic input by relying on one memory system or the other. Based on the four manipulations for which young children showed an advantage, we have four approximations of the point at which dominance shifts to declarative memory from procedural memory (the point at which performance is lowest): 3.2, 4.6, 3.8 and 4.5 for verbal working memory. Taken together, these suggest that the shift in dominance may fall around the time at which an individual would score 4 on the memory tasks involved in this study. Based on the correlation between age and
verbal working memory, this projects that point to be around 8-9 years of age, which is in the same range as previous estimates for the “critical period,” often just before puberty (Lenneberg 1967: 158, 180-181; DeKeyser, Alfi-Shabtay & Ravid 2010: 414).

The possibility of measuring the “critical period” in this way is far from conclusive, but it is suggestive. Further research would be required to test whether the age at which reliance on declarative memory overtakes procedural memory is the same in different contexts; it may be the case that, for example, children as old as 12 or 15 can, in the right circumstances, use procedural memory to learn as the younger children are shown to here. This may relate to length of exposure to input and practice effects. One potential difficulty in this methodology is that if age effects really do need to be studied focusing on ages under 5 years, alternative methodologies will need to be developed that will work with two to four year olds, because this experimental method was challenging to complete with five year olds and not possible with four year olds. Here, it has been shown that five and six year olds are capable of learning immediately based on novel input, without relying on declarative memory, with a gradual shift in balance between the two memory systems until adulthood where adults rely heavily on declarative memory.

4.1 Further research directions

A pilot study14 for extending this research to adult L2 acquisition has shown that there may be similar effects for different proficiency levels: advanced learners and native speakers may outperform novice learners in repetition of semantic content and learning new vocabulary, while novice learners may outperform more advanced learners when learning an artificial dialect through repetition. Unlike the participants in the current study of various ages, all adult L2

---

14 This pilot study was approved by the UIUC SLCL Human Subjects Review Committee, Fall 2011.
learners have similar memory systems. However, the use of procedural memory correlates with proficiency because “practice should lead to procedural learning and improved performance” (2005: 152). Therefore, the use of declarative memory to process novel input would more easily allow transfer to low proficiency L2 knowledge, which also relies on declarative memory. The more advanced learners would have more difficulty integrating the declarative knowledge with their procedural grammatical system that has built up over a long time. In short, experience should correlate negatively with flexibility. This research is in progress.

Other directions for future research possibilities include working with Child L2 learners to compare proficiency and age effects in the same population, and working with heritage learners to separate age of learning and age of testing. The present study used a broadly targeted stimulus, an artificial dialect with ten systematic manipulations; a natural line of inquiry is what would happen if fewer (or more) changes were included. It may or may not be the case that the advantage of young children appears specifically because this was a difficult task with many potential foci for limited cognitive resources for adults. It would also be useful to connect this research, directly or indirectly, with real-world cases of fossilization such as Japanese learners of English who have difficulty with the L/R contrast (cf. Miyawaki et al. 1975; Bradlow, Akahane-Yamada, Pisoni & Tohkura 1999; Ingvalson, McClelland & Holt 2011; inter alia) and whether the specific causes of fossilization could be identified. This could help learners either to attempt to avoid such circumstances (such as by being exposed to native speaker pronunciations starting early in language learning, or by being exposed to the language early in life) or to be aware of when they may be faced with extremely challenging forms to acquire and to know to seek extra help and linguistic input. For example, correcting the pronunciation if $L$ and $R$ in a Japanese learner's L2 English would be different depending on age of initial exposure to the contrast (as
opposed to learning from other L2 speakers), their current age, and how much interference they would face from the amount of experience they have not producing the contrast.

Another line of research at the moment is exploring how learners, especially children, restructure input (Fedzechkina, Jaeger & Newport 2012; see also Hudson Kam & Newport 2009; Wonnacott et al. 2012); that is, instead of expecting perfect performance, the goal of the experiment is actually to expect the learners to change the language to suit communicative need and learning constraints. Although there are many directions in which that research should be pursued (a particularly fascinating one would be to see a large-scale repeated filtering process from raw artificial pseudo-language to a complete natural grammar, along the lines of what MacCallum, Mauchm Burt & Leroi (2012) have done for music), the method used in this experiment would allow for precise tracking as the new linguistic system evolved, as shown in Figure 23. The horizontal axis represents time, and the vertical axis is the score on each individual token (in blue); the red line tracks the cumulative score for the participant in each round, in this case as a seven year old boy improves his repetition of Manipulation #2. Although there is not room to include any detail in this paper, this possibility for tracking individual learning is promising.

Figure 23. Cumulative performance for Participant #14 for Manipulation #2.
4.3 Conclusion

In summary, the findings of the present research study support Ullman's Declarative/Procedural model and have shown that the effects of using these two memory systems can be measured across-subjects with as little as 7 minutes of novel input and a repetition task. This is a promising direction for further research that seeks to contribute empirical evidence about the nature of age effects and their underlying causes, in line with Ullman's proposal for testing the DP model and other theories (2005: 163; 2012: 160). More broadly, age effects in language acquisition may not be specific to language or due to age; the effects may be part of a greater shifting balance between declarative and procedural memory as children mature, and the apparent relationship with age is only indirect, because age correlates with declarative memory and in turn the increasing dominance of declarative memory correlates with different language learning mechanisms. The DP model may not be the only theoretical approach that could explain these results, but it provides an immediate explanation for the correlation with verbal working memory and performance in adults, and the inverse correlation with young children performing well despite low verbal working memory. Regardless, whatever theory is used would need to address the gradual shift found here from an early age advantage, to the disadvantage found in the intermediate ages, and to the advantage found again for adults and the fact that adults do not reach the high levels of ultimate attainment that children do. This research has not addressed an absolute rule, but a statistical tendency; there may be exceptional adults who, with enough effort or exposure, may in fact be able to overcome the limited availability of their procedural memory system and reach high levels of ultimate attainment, but extra-linguistic factors such as time make this unlikely.
REFERENCES


Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. PNAS.


Gathercole et al., S. E. (2004). The Structure of working Memory from 4 to 15 Years of Age. Developmental Psychology 40(2), 177-190.


APPENDIX 1: EXPERIMENT MATERIALS

New (Artificial) Dialect  
(Bold denotes rising pitch for emphasis)

#1 (Personal Introduction)  
Hola, me jamu Adilu/a Martínez. Gustu muchu.  
Ju su de Kitale unu poblu pequeñu en África.  
Actualmén podemos idioma este practicar.

#2 (The Tortoise and the Hare)  
Había una vez unu coneʒu veloz y una tortuga despaz.  
Lu coneʒu quería contra tortuga esa competir.  
Lu coneʒu andú rá pidamén y logu decidiú hojas de nanasi comer.  
Peru la tortuga veniú lentamén y finalmén ganú la carrera.

#3 (Treasure Hunters)  
Cinco tesoradores cincu estaban unu tesoru buscandu.  
Eʃos poderon lu encontrar, pero ningunu lu sabú.  
"Tene una luz azul briʃante!" gritú Isa felizmén.  
"Es unu escudu de pedra forte!" decidiú niʃu otro certamén.  
"Miro a una chica beʃa!" pensú Samira seriamente.  
Lu niʃu sabiu explicú, "nu, es solamén unu espeʃu!"

#4 (The Giraffe and the Elephant)  
Unu día la ziraʃa estaba a lu elefante hablandu.  
La ziraʃa deciú: "Salu para la festa nacera de lu ʃeʃon."  
Lu elefante deciú: "¡Me vu tambén! Es amigu meʃor nostru."  
Eʃos habían a lu Ríu Mamba caminadu, peru lu ʃeʃon vivía a lu ladu otru.  
La ziraʃa deciú, "¡Ha cocodrilos catorce aʃi! Tenu ʃedu."  
Lu elefante deciú: "Pediremos ajuda a unu dios."  
"Usena, ¿cómo podemos lu ríu cruzar?" preguntú la ziraʃa.  
Rápidamén unu ŋuʃ de fugu caʃu de lu celu hacia los cocodrilos.  
Lu elefante y la ziraʃa cruzaron lu ríu tranquilamén.

Original Spanish (for comparison only; not heard)  
(Bold denotes new vocabulary)

Había una vez un conejo veloz y una tortuga despaz.  
El conejo quería competir contra esa tortuga.  
El conejo anduvo rápidamente y luego decidió comer hojas de nanasi.  
Pero la tortuga vino lentamente y finalmente ganó la carrera.

Cinco tesoradores estaban buscando un tesoro.  
Ellos pudieron encontrarlo, pero ninguno lo supo.  
"Tiene una luz azul brillante!" gritó Isa felizmente.  
"Es un escudo de piedra fuerte!" decidió otro niño ciertamente.  
"Miro a una chica bella!" pensó Samira seriamente.  
El niño sabio explicó, "No, es solamente un espejo!"

Un día la girafa estaba hablando al elefante.  
La girafa dijo: "Salgo para la fiesta nacera del león."  
El elefante dijo: "¡Me voy también! Es nuestro mejor amigo."  
Ellos habían caminado al Río Mamba, pero el león vivía al otro lado.  
La girafa dijo, "¡Hay catorce cocodrilos allí! Tengo miedo."  
El elefante dijo: "Pediremos ayuda a un dios."  
"Usena, ¿cómo podemos cruzar el río?" preguntó la girafa.  
Rápidamente un rayo de fuego cayó del cielo hacia los cocodrilos.  
El elefante y la girafa cruzaron el río tranquilamente.
**English Translations:**

#1: Hello, my name is Adilu/a Martínez. Nice to meet you. I'm from Kitale, a small town in Africa. Now we can practice this language.

#2: There once was a fast rabbit and a slow turtle. The rabbit wanted to race against the turtle. The rabbit went quickly and later decided to eat some nanasi leaves. But the turtle came slowly and finally won the race.

#3: Five treasure-hunters were searching for a treasure. They managed to find it, but no one knew what it was. “It has a brilliant blue light!” screamed Isa happily. “It's a shield of strong stone!” decided another child certainly. “I see a beautiful girl!” thought Samira seriously. The wise child explained, “No, it's only a mirror!”

#4: One day the giraffe was talking to the elephant. The giraffe said: “I’m leaving for the lion's birthday party.” The elephant said: “I’ll go too! He's our best friend.” They had walked to the Mamba River, but the lion lived on the other side. The giraffe said: “There are fourteen crocodiles there! I'm scared.” The elephant said: “We'll ask a god for help.” “Usena, how can we cross the river?” asked the giraffe. Quickly a bolt of lightening fell from the sky toward the crocodiles. The elephant and the giraffe crossed the river safely.

---

**Swahili:**

#. (syllables)
1. (2) paka
2. (3) samaki
3. (4) ninakula
4. (5) tutasafiri
5. (6) akili ni mali
6. (7) peña nia pana cha
7. (7) machu asana pasa
8. (8) dawa ta motu ni motu
9. (8) baba wa kambu si baba
10. (9) tisa ni karibu na kumi
11. (9) fuata nuki ula isali
12. (10) dalili ta vua ni mawingu
13. (11) kikulachu kiku guoni maku
14. (12) kila langu kuwa ufungulu waka

---

**Digits:**

**Numbers Set 1**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>45</td>
<td>148</td>
<td>912</td>
<td>3518</td>
<td>43678</td>
<td>749632</td>
<td>37158469</td>
<td>791354628</td>
<td>3871694254</td>
</tr>
<tr>
<td>7</td>
<td>89</td>
<td>625</td>
<td>4187</td>
<td>74953</td>
<td>892175</td>
<td>2817964</td>
<td>62137985</td>
<td>1365786289</td>
<td>5932861473</td>
</tr>
<tr>
<td>4</td>
<td>41</td>
<td>489</td>
<td>8145</td>
<td>46572</td>
<td>769932</td>
<td>4361782</td>
<td>722340961</td>
<td>936534179</td>
<td>3562948179</td>
</tr>
<tr>
<td>96</td>
<td>219</td>
<td>3982</td>
<td>43267</td>
<td>179325</td>
<td>8479263</td>
<td>97261354</td>
<td>367942581</td>
<td>9286547318</td>
<td>9632148756</td>
</tr>
<tr>
<td>5</td>
<td>58</td>
<td>932</td>
<td>1453</td>
<td>58279</td>
<td>538761</td>
<td>1743836</td>
<td>42318963</td>
<td>4659832179</td>
<td>8619245372</td>
</tr>
<tr>
<td>8</td>
<td>29</td>
<td>627</td>
<td>3812</td>
<td>12647</td>
<td>796135</td>
<td>8147365</td>
<td>19275843</td>
<td>4659832179</td>
<td>8619245372</td>
</tr>
</tbody>
</table>

**Numbers Set 2**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>52</td>
<td>638</td>
<td>8323</td>
<td>49728</td>
<td>534289</td>
<td>1745628</td>
<td>46972835</td>
<td>863197254</td>
<td>675482313</td>
</tr>
<tr>
<td>8</td>
<td>18</td>
<td>274</td>
<td>7693</td>
<td>75341</td>
<td>954723</td>
<td>6479318</td>
<td>26748153</td>
<td>752946813</td>
<td>1852476935</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>496</td>
<td>8467</td>
<td>47298</td>
<td>237458</td>
<td>2816453</td>
<td>79810432</td>
<td>376891254</td>
<td>1283545791</td>
</tr>
<tr>
<td>9</td>
<td>71</td>
<td>348</td>
<td>4891</td>
<td>96753</td>
<td>987264</td>
<td>4765812</td>
<td>43915876</td>
<td>698314257</td>
<td>7925343868</td>
</tr>
<tr>
<td>6</td>
<td>97</td>
<td>216</td>
<td>9325</td>
<td>65189</td>
<td>167598</td>
<td>6843927</td>
<td>17234958</td>
<td>967382415</td>
<td>8623971454</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>351</td>
<td>4185</td>
<td>19486</td>
<td>382149</td>
<td>3679425</td>
<td>15239867</td>
<td>753246159</td>
<td>5497926315</td>
</tr>
</tbody>
</table>