Aging and the Statistical Learning of Grammatical Form Classes

Jessie Schwab

Faculty Advisor: Dr. Darlene Howard

Faculty Co-Advisor: Dr. Elissa Newport

Georgetown University
Abstract

When acquiring a new language, learners are forced to place unfamiliar words into categories, oftentimes with few explicit indicators about when and how that word can be used grammatically. Previous research (Reeder, Newport & Aslin, 2013) has demonstrated that college students are able to learn grammatical form classes from distributional information alone. Specifically, after brief exposure to an artificial language, participants are able to determine when to assign a word to a category and when to treat it as lexically distinct, given no cues other than the grammatical patterning of the input. In the following research, we examined the effect of healthy aging on this type of statistical learning. We found that adults over the age of 65 are in fact able to learn the grammatical form classes of the artificial language, but to a significantly lesser extent than young adults. This finding has implications for cognitive aging research, in terms of determining which aspects of learning decline in healthy aging, as well as for the field of language acquisition in general.
Aging and the Statistical Learning of Grammatical Form Classes

Previous research on language acquisition has focused on infants and young adults, particularly with regard to the idea of a critical or sensitive period of language development and the existence of maturational constraints on language learning (Penfield & Roberts, 1959; Lenneberg, 1967; Snow & Hoefnagel-Hohle, 1978; Hylenstam & Abrahamsson, 2000; Flege, Yeni-Komshian, & Liu, 1999). Nevertheless, researching language acquisition in aging could be useful in a variety of ways. First, knowing how language acquisition changes as people age could help researchers understand the cognitive and neural bases of language processing. Additionally, as our population continues to age, there will be a growing number of older adults trying to learn a new language or to relearn one after stroke. Therefore, it will be important to know which aspects of language processing and language-learning ability are spared and which decline in older adults. Finally, there is a great deal of overlap between language and other aspects of cognition, such as memory and attention, so delving deeper into the study of language acquisition and aging would help us to understand cognitive aging more broadly.

Findings from Cognitive Aging

Over the past several decades, research has examined which cognitive abilities decline in older adults, as well as what physical changes can be observed in the aging brain. For example, adults experience declines in processing speed, working memory, and the encoding of information as they age (Craik, 1994; Park, 2000). Older adults have been found to perform worse on working memory tasks as the load increases (i.e., more items are added) (Cappel, Gmeindl, & Reuter-Lorenz, 2010; Rypma, Eldreth, & Rebbechi, 2007; Rypma, Prabhakaran, Desmond, & Gabrieli, 2001). Researchers have also found age differences during tasks involving executive functions, such as memory updating or reordering, which are more complex
and require more than simply rote rehearsal (Reuter-Lorenz & Park, 2010). Age-related declines in processing speed are associated with decreasing white matter, a measure of neural transmission in the brain (Sullivan & Pfefferbaum, 2006; Kennedy & Raz, 2009). There is also a lower volume of gray matter in older adults, although these declines are not equal throughout the brain (Terry, 2000; Resnick, Pham, Kraut, Zonderman & Daratzikos, 2003). In terms of regions of the brain, it is the structures of the prefrontal cortex (PFC) that undergo the most notable functional changes and volumetric declines during aging (Hedden & Gabrieli, 2004). There are also major age-related declines in the striatum, an area of the brain that has many connections with the PFC (Hedden & Gabrieli, 2004).

Since the amount of available attentional resources declines in older age, older adults might not be as likely to carry out the effortful or strategic memory processes that are associated with better performance (Craik, 2006.) Older adults may therefore make use of their preserved knowledge during tasks, while younger adults rely on more fluid processing functions (Hedden & Gabrieli, 2004). Thus, while there are a variety of declines in the brain during aging, semantic knowledge remains relatively stable, and vocabulary can actually increase over the lifespan (Park et. al, 2002).

Aging and Language Acquisition

Previous second-language (L2) acquisition research examining aging includes studies correlating immigrants’ L2 ability with their age of arrival in the United States. Johnson and Newport (1989) tested 46 Korean and Chinese immigrants to the U.S. on English grammaticality judgment tasks and measured how their performance was related to their age of arrival. The age of arrival ranged from 3 to 39 years, with each immigrant having lived in the U.S. at least 3 years. Johnson and Newport found a linear decline in performance beginning with immigrants
who had an age of arrival of 8 and ending with those who had an age of arrival of 16, and age of arrival was no longer predictive for those who arrived post maturation. These results provided evidence for a critical period of language learning, with age having no effect on L2 development after individuals reached maturity.

Yet the results of subsequent studies have led to conflicting interpretations of the effect of aging on L2 acquisition. Birdsong and Molis (2001) replicated Johnson and Newport’s 1989 study with a larger sample (61 participants) of Spanish-speaking immigrants and obtained different results. In this study, there was a significant age-related deficit in English L2 learning for immigrants who arrived after the age of maturation (16 years old.) Hakuta, Bialystok, & Wiley (2003) performed a similar study examining second language acquisition for Spanish-speaking and Chinese-speaking immigrants to the U.S. They found, based on a large sample of U.S. census data, that age of immigration (up to about 95 years of age) had a negative linear correlation with English proficiency. Overall, these studies provide some evidence against a critical period of language learning, which views learning as declining throughout maturation (between ages 7 and 15) and then reaching a plateau. Instead, they suggest that age effects on L2 learning continue throughout the lifespan.

While these correlational language and aging studies are important for understanding real-world L2 development, it is difficult to separate out the various biological, environmental, or socioemotional factors that may be contributing to age differences in immigrants’ language learning. Therefore, some researchers have sought to bring aging and L2 acquisition research into the laboratory. For example, attempts have been made to distinguish between declines in inherent language-learning ability and general cognitive declines, since, as described above, it is
well known that certain cognitive functions decline in older age (Park, 2000). In *Child Language Acquisition and Development*, Matthew Saxton (2010) wrote the following:

> Hearing, vision, general problem-solving skills and memory (both long-term and short-term) get progressively weaker with age (Seifert, Hoffnung & Hoffnung, 2000). It is conceivable, therefore, that receptivity to language remains strong throughout life, but that deterioration in these other, non-linguistic functions mediates increasingly poor language learning. (p 70)

In order to address this question, Mackey and Sachs (2011) performed a laboratory study with older native Spanish speakers (ages 65-79) who were learning English. They had the participants perform communication tasks with native English speakers and correlated this with their working memory abilities. The results showed that only older adults who performed highly on working memory tasks (in their native language) showed L2 development, which was coded as the learner producing at least two more advanced question forms in each of the communicative posttest tasks than they had in the pretest.

Midford and Kirsner (2005) used an artificial grammar paradigm developed by Reber (1967) to compare age differences in language learning based on teaching method (or the explicitness of the instruction.) During initial exposure, all participants encountered letter strings that had been generated by a finite-state grammar. Younger (mean age 20.6) and older (mean age 65.9) participants were assigned to one of four groups: complex-without-rules, complex-with-rules, simple-without-rules, or simple-with-rules. The “complex” grammar contained a possible 31 transitions between letters and up to 9 letters in each string. The “simple” grammar contained only 24 possible transitions between letters with up to 8 letters in each string. During an initial study phase, all groups were exposed to twenty-seven different grammatical letter
strings (each heard three times) and told to remember the letter strings. In the “with-rules” (but not the “without-rules”) groups, participants were first shown a diagram that explained to them either the “simple” or “complex” grammar, and they then worked through six example strings, before having to provide three examples themselves. In the test phase, all groups were shown four sets of grammar strings and asked to determine how “correct” each item was. The first set contained ten grammatical and ten non-grammatical items, and the following sets each contained ten new grammatical and ten new non-grammatical items, in addition to a repetition of the initial 20 items. To measure learning, researchers looked at both reaction time and accuracy. Results showed longer reaction time for older than younger adults, and this difference did not vary across conditions. In terms of accuracy, the greatest age differences occurred when the grammar was simple and participants were given an explanation of the grammar rules. In contrast, accuracy was not significantly different for older and younger adults when the grammar was complex and the rules were not explained. Midford and Kirsner (2005) concluded that the complex grammar condition encouraged the use of implicit processes and that these results suggested that implicit learning is spared to a large extent in older adults.

Alison Lenet et al. (2011) extended Midford and Kirsner’s results by using a real language (Latin) and varying the degree of explicitness in instruction, explicit versus less-explicit. Prior to treatment all participants were given an initial vocabulary lesson of 35 Latin nouns and 11 verbs. Then, all participants received a treatment consisting of two Latin grammar lessons, each lasting about 30 minutes, in which participants read or listened to sentences in Latin and chose pictures or English translations that most closely corresponded with what they saw or heard. Immediately after each response, both groups were told whether they were correct or incorrect, but only the explicit treatment group was also given a grammatical explanation as to
why they were wrong. To measure learning, the researchers had participants complete a grammar pre-test prior to treatment and a post-test following treatment. The grammar tests each had four sections: aural interpretation, written interpretation, written grammaticality judgment, and written production. Results showed that the overall learning in the older group (mean age 72.3) was not significantly less than the younger group (mean age 18.7). However, while the younger group learned better in the explicit condition, the older group only showed significant learning in the less explicit feedback condition. These results were consistent with Midford and Kirsner’s (2005) study and provide further evidence that older adults perform better when relying on implicit methods of learning.

**Implicit Learning and Aging**

The results of Midford and Kirsner (2005) and Lenet et al. (2011) suggest that implicit instruction improves performance in older adults (but not in the younger groups.) Thus, it seems that implicit learning is a factor in successful language acquisition for older adults in particular. Yet previous cognitive aging literature focusing on implicit learning, or the ability to pick up on regularities in our environment outside of conscious awareness, has demonstrated that while some aspects of implicit learning are spared in healthy aging, other aspects decline (e.g., Howard & Howard, 2012).

The Spatial Contextual Cueing task (developed by Chun & Jiang, 1998) exemplifies one type of implicit learning (Howard & Howard, 2012). In this task, participants view a screen with 11 distracter items (“Ls” in different directions) in search of a letter “T,” positioned on its side. Participants are supposed to respond to the “T” as quickly as possible by pressing keys on the keyboard corresponding to its direction (either facing right or left.) Although the participant is unaware, the screen sometimes portrays a repeated pattern of items (that occurs multiple times
over the session), while other times it portrays a novel pattern, in which the letters are randomly arranged. Over time, participants respond more quickly and more accurately to the repeated trials, indicating implicit learning of spatial contexts. Howard et al. (2004b) tested a group of young and older adults on this task and found no significant age differences in learning. Thus, the implicit learning of spatial contexts seems to be relatively spared in healthy aging.

On the other hand, the implicit learning of some kinds of sequential patterns does seem to decline in healthy aging. Nissen & Bullemer (1987) introduced a serial reaction time (SRT) task, in which, over several blocks of trials, participants view a computer screen and respond to an event, which consists of the filling in of one of four circles in a row, by pressing a corresponding key. Although participants are not aware of it, on most blocks, the event occurs in a predictable sequence. Implicit sequence learning is thus demonstrated when participants have faster reaction times and are more accurate on blocks that contain the pattern compared to blocks that do not.

Howard & Howard (1997) developed a modification, called the Alternating Serial Response Time (ASRT) task. This task differs from the SRT task in that predictable events alternate with random ones. Results show that people implicitly learn that some groups of three events (“triplets”) occur more frequently than others, and they gradually come to respond faster and more accurately to the frequent as opposed to the infrequent events. Yet when younger (mean age 19.83) and older adults (mean age 70.96) completed 10 sessions of the ASRT task over the course of several days, significant age-related deficits in learning were found (Howard et al., 2004b). Age-related differences in implicit sequence learning are also apparent in a Triplets Learning Task (TLT) which takes out the motor sequencing component of the ASRT (Howard, Howard, Dennis, & Kelly, 2008).
In terms of the neural basis of implicit learning, neuroimaging data has shown that the mediotemporal lobe, which is largely responsible for explicit and declarative memory, is also involved in the type of implicit spatial context learning that takes place during the SCCT (Preston & Gabrieli, 2008; Howard & Howard, 2012). Yet a fronto-striatal system has been implicated in implicit sequence learning, which occurs during ASRT or TLT, particularly after practice (Rauch et al., 1997; Howard & Howard, 2012). Interestingly, aspects of language processing have also been found to involve the structures of the frontal lobe and basal ganglia (Ullman, 2001; Newman, Supalla, Hauser, Newport, & Bavelier, 2008).

Research has shown that there may also be age differences in the recruitment of various neural substrates during learning. In their fMRI study, Rieckmann, Fischer & Bäckman (2010) examined learning in an SRT task and found that learning in younger adults was associated with increased activation in the striatum and decreased activation in the medial temporal lobe. On the other hand, learning for older adults was associated with increased activation in both the striatum and the medial temporal lobe. fMRI data have also measured brain activity during TLT “learning” (i.e., responses to higher probability triplets) (Simon, Vaidya, Howard & Howard, 2011). Results showed that younger and older adults both recruited the hippocampus early in learning, but with training, younger adults began to rely on the basal ganglia while older adults continued to rely on the hippocampus.

Statistical Learning and Language Acquisition

The fact that some aspects of implicit learning (specifically, the ability to learn sequential patterns) have been found to decline over the lifespan invites the question of whether the more unconscious aspects of language acquisition also decline in older age. Specifically, was Saxton (2010) right in thinking that “receptivity to language remains strong throughout life” or does our
ability to pick up on the grammatical sequences and patterns in a language also decline in older age?

According to Conway & Pisoni (2008), language at each of its levels of organization (phonemes, morphemes, etc.) is made up sequential structures that involve statistical or probabilistic patterning. In this vein, a great deal of recent language acquisition research has focused on the importance of statistical learning in language development (for reviews see Romberg & Saffran, 2010; Gomez & Gerken, 2000; Saffran, Werker, & Werner, 2006). For the purposes of this paper, statistical learning and implicit learning can be thought to reflect the same mechanisms, though Perruchet & Pacton (2006) argue that differences exist between the two.

Saffran, Aslin, and Newport’s (1996) influential study on infant word segmentation found that 8-month-old infants are able to segment a sequence of nonsense syllables into “words” from a mere two minutes of exposure. The infants were able to learn word boundaries based only on the transitional probabilities between syllables, which were higher within words of the artificial language (1.0) than across word boundaries (0.33) of the word strings they presented to the infants. This study indicates that language acquisition involves implicit (or statistical) learning, in which learners are sensitive to probabilistic regularities in the language input.

Yet language learning involves multiple levels of organization. In addition to figuring out the boundaries of words within speech, learners have to extend that knowledge to the acquisition of word meanings and grammatical categories (i.e., verb, noun, adjective). As they learn and acquire more of a language, learners often have to decide when to consider a word as belonging to a known category, generalizing from the properties of words that they have heard before, and when to treat it as novel and lexically distinct (Mintz, 2002; Reeder et al., 2013).
The idea of a Universal Grammar, in which individuals are theorized to possess an innate language learning mechanism (Chomsky, 1965), cannot fully explain this ability, since the languages of the world have a variety of grammatical categories and distinctions. Reeder et al. (2013) also argue that the theory of semantic “bootstrapping,” or associating semantic properties with grammatical classes (Pinker, 1987), is not a sufficient explanation, since semantic meaning and syntactic categories do not always match perfectly. Thus, a number of linguists have looked to the “distributional learning” hypothesis, which argues that learners use statistical learning in order to acquire grammatical categories based on the structure of the input (i.e., the distributional information surrounding a word) (Cartright & Brent, 1997; Mintz, Newport, & Bever, 1995, 2002; Mintz, 2002, 2003; Reeder et al., 2013).

Mintz (2002) first tested the ability of learners to organize words into grammatical form class categories based solely on the distributional information surrounding the words of an artificial language. Results showed that when college-aged participants were exposed to 6 minutes of the artificial language (with each language treatment consisting of 22 three-word training sentences and 10 three-word test sentences), they were able to learn and generalize properties of words. Reeder et al. (2013) expanded upon these results, attempting to pinpoint what type of distributional information leads learners to form categories, as well as the extent to which learners are able to generalize that category to a word seen in a minimal number of contexts. Results showed that as long as the language input contained a dense sampling of possible sentence strings (with relatively few strings withheld during initial exposure) and that the words were seen in overlapping contexts, learners were able to form categories. Additionally, learners were able to extend their knowledge of the grammatical categories to
novel words for which they had been given minimal distributional information, i.e. they had only
seen the word in one context.

The present experiment attempts to replicate and extend the results of Reeder, Newport,
& Aslin to determine whether there are age-related differences in the type of statistical learning
that is involved in form-class category formation. Based on the previous cognitive aging
literature, we expected that older adults would be able to learn the form classes, but because this
task involves picking up on regularities among sequences of word items, that there would be
significant age-related deficits in learning.

Method

Participants: Participants were 20 monolingual native English-speaking young adults (13
female, 7 male) and 20 monolingual native English-speaking older adults (13 female, 7 male).
The young adults ranged in age from 19 to 24 (21.3 ± 1.6 years old) and the older from 66 to 87
(73.6 ± 5.6 years old). The average number of years of education for the young adults was 14.9
(SD =1.07) and for the older adults 17.2 (SD = 2.17). For counterbalancing purposes,
participants were randomly assigned to the different language groups described below, with ten
young and ten older participants in each group. Young participants were recruited by flyers
around Georgetown University, and older participants were recruited by Washington Post
advertisements. All participants gave informed consent and were given monetary compensation
for their participation. The Georgetown University Institutional Review Board approved the
experimental procedures.

Stimulus Materials: The artificial language task used in this experiment was developed by
Reeder et al. (2013). The language exposure was presented in the form of an alien game;
participants were told to listen to an alien named Zooma practice sentences in a language called “SillySpeak,” which she had to learn before arriving on a new planet. The artificial grammar within the program had the structure (Q)AXB(R), where each letter represents a set of words. “X” represented the target category we were interested in, “A” and “B” were the “context” categories, or the distributional cues surrounding X, and “Q” and “R” were optional flanker categories used so that the length of the word strings could vary from three to five words. The Q and R categories consisted of two words each, the A and B categories had three words each, and the X category had four words. The words of the grammar were spad, klidum, flairb, daffin, glim, tomber, zub, lapal, fluggit, mawg, bleggin, gentif, frag, and sep in both languages, and the languages differed only in which nonsense words were assigned to each category, as shown below.

### Language 1

<table>
<thead>
<tr>
<th>Q</th>
<th>A</th>
<th>X</th>
<th>B</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>spad</td>
<td>flairb</td>
<td>tomber</td>
<td>fluggit</td>
<td>gentif</td>
</tr>
<tr>
<td>klidum</td>
<td>daffin</td>
<td>zub</td>
<td>mawg</td>
<td>frag</td>
</tr>
<tr>
<td>glim</td>
<td>lapal</td>
<td>bleggin</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Language 2

<table>
<thead>
<tr>
<th>Q</th>
<th>A</th>
<th>X</th>
<th>B</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>frag</td>
<td>gentif</td>
<td>spad</td>
<td>zub</td>
<td>lapal</td>
</tr>
<tr>
<td>daffin</td>
<td>mawg</td>
<td>fluggit</td>
<td>tomber</td>
<td>flairb</td>
</tr>
<tr>
<td>klidum</td>
<td>bleggin</td>
<td>glim</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Procedure:** Prior to testing, participants signed a consent form and filled out a biographical questionnaire. Participants were then tested individually during a session that lasted approximately 25 minutes. Participants were told, “Listen carefully while Zooma practices [a new language called Sillyspeak.] When she arrives on her new planet, she is going to say some more Sillyspeak sentences. Your job will be to decide if you have heard Zooma say that sentence before.” During the session, they listened to 12 minutes of one of the two artificial languages described above. The exposure set consisted of 10 of the 36 possible AXB combinations. Three of the “X” words (X1-X3) were heard in combination with every “A” and every “B” word. One of the “X” words (X4) was heard in only one context. With the optional Q and R flanker words, the exposure set consisted of 40 (Q)AXB(R) sentences, and each participant heard the exposure set four times through. The order of the sentences in the exposure set was randomized for each participant (by one of six possible orders) and presented in a Python program on a Macintosh OSX laptop.

Following exposure, participants heard a series of 70 three-word test strings (e.g., “glim zub mawg”). They were asked to rate each string on a scale from 1 to 5, with a rating of “1” being that the sentence definitely did not come from the language they were exposed to and “5” being that the sentence definitely did come from the exposure language. The test strings were of three types—grammatically familiar (20 AXB strings presented during exposure), grammatically novel (26 AXB strings that were withheld during exposure) or ungrammatical (24 strings in the form of AXA or BXB).

**Results**

*X123 Analyses*
Figure 1 displays the average ratings of younger and older adults on the grammatically familiar (GF), grammatically novel (GN) and ungrammatical (UG) sentence strings that contained the X1-X3 words (which were heard in combination with every “A” and every “B” word). We submitted these ratings to a mixed-design 2 x 2 x 3 Analysis of Variance (ANOVA), with age and language (1 or 2) as between-subjects factors and sentence string type as a within-subjects factor. The ANOVA yielded a significant main effect of string type, $F(2, 72)=60.686$, $p<.0001$, and a string type by age interaction, $F(2,72)=10.989$, $p<.0001$. No other main effects or interactions approached significance.

From Figure 1, it appears that while both younger and older adults learned the grammatical structure, the older adults did not do so to the same extent as younger adults. To examine this interaction more fully, we conducted separate repeated measures Analyses of Variance (ANOVA) on the younger and older groups, with sentence string type as the within-subjects factor.

The younger group showed a main effect of string type, $F(2,36)=47.4$, $p<.0001$. Paired samples t-tests showed no significant difference between grammatically familiar and grammatical novel ratings, $t(19)=-.167$, $p=.869$. However, the grammatically familiar strings were rated significantly higher than the ungrammatical strings, $t(19)=7.773$, $p<.001$, as were the grammatically novel strings, $t(19)=7.016$, $p<.001$.

The older group showed the same pattern of results, i.e., they also showed a main effect of string type, $F(2,36)=14.38$, $p<.0001$, no significant difference between grammatically familiar and grammatically novel strings, $t(19)=-.550$, $p=.59$, but the grammatically familiar strings were rated significantly higher than the ungrammatical strings, $t(19)=4.256$, $p<.001$, as were the grammatically novel strings, $t(19)=4.164$, $p<.01$. 
These analyses indicate that both the younger and older groups were able to distinguish grammatical from ungrammatical sentence string types. To determine the source of the significant age x sentence type interaction reported above, we collapsed the ratings for grammatically familiar and grammatically novel sentence string types into one “grammatical” category. The mean difference in ratings between grammatical and ungrammatical strings was significantly greater for young adults, .99 (SE=.13), than for older adults, .40 (SE=.09), \( t(38)=3.74, p=.001 \).

Figures 2a, 2b, and 2c display individual data, with each bar showing a difference score for a single young or older adult, with the individuals ordered from lowest to highest difference score. Figure 2a shows the difference in ratings for grammatically familiar minus ungrammatical sentence strings, Figure 2b for grammatically novel minus ungrammatical sentence strings, and Figure 2c for grammatically familiar minus grammatically novel sentence strings. The figures demonstrate that the general pattern of learning remains consistent at the individual level. Although there is some overlap between the age groups, the younger participants tend to be the ones who have the larger difference scores for the GF vs. UG and the GN vs. UG comparisons, whereas for the GF vs. GN comparison, the participants of both ages have scores near zero.

**X4 Analyses**

We separately analyzed the X4 sentence string types, in which the “X” words had been heard in only one context, as opposed to the X1-X3 words, which participants heard in all contexts (with every A word and every B word.) If participants follow the same pattern of learning for the X4 sentence string types as they do for the X1-X3 types, then that would show that they were able to generalize X4 to its entire range of possible grammatical contexts, even though they only heard it in one of the contexts.
Figure 3 displays the average ratings of younger and older adults on the grammatically familiar (GF), grammatically novel (GN) and ungrammatical (UG) sentence strings that contained the X4 words (which were heard in only one “A” or “B” context). An ANOVA identical to that performed on the X123 items yielded a significant main effect of sentence string type, $F(2, 72) = 16.782, p < .0001$, and a marginal sentence type by age interaction, $F(2,72) = 2.816, p = .066$. No other main effects or interactions approached significance.

From Figure 3, it appears that the pattern of learning for the X4 test string ratings is different from the X123 test strings. The GN X4 sentence types were rated lower than GF ones, though they are still rated higher than the UG sentences. It also appears that while both younger and older adults followed the same pattern of learning for the X4 strings, in that their UG ratings are in the direction of being lower than the grammatical ones, the older adults did not do so to the same extent as younger adults.

Separate ANOVAS on each group indicated that the younger group showed a main effect of sentence string type, $F(2,36) = 14.703, p < .0001$; GF strings were rated significantly higher than GN, $t(19) = 2.215, p < .04$, GF strings were rated significantly higher than the UG strings, $t(19) = 4.44, p < .001$, and GN strings were rated significantly higher than the UG strings, $t(19) = 3.794, p < .01$.

The older group also showed a main effect of sentence string type, $F(2,36) = 3.451, p = .043$, but the pattern was different from that of the younger group. There was no significant difference between GF and GN strings, $t(19) = 1.213, p = .24$, but the GF strings were rated significantly higher than the UG strings, $t(19) = 2.489, p = .02$, and there was a marginally significant difference between the GN and UG strings, $t(19) = 2.054, p = .054$. 
Figures 4a, 4b, and 4c display individual data for the X4 sentence string ratings, demonstrating that the general pattern of learning remains consistent at the individual level. Although there is some overlap between the age groups, the younger participants tend to be the ones who have the larger difference scores for the GF vs. UG comparison for X4 sentence strings. For the GN vs. UG comparison, the young participants again tend to have larger difference scores, but the difference scores overall are much smaller. For the GF vs. GN comparison, the participants of both age groups are varied in their individual difference scores with no visible pattern in rating.

**Discussion**

**Overview of Findings**

In the present study, we aimed to replicate the results of Reeder et al. (2013), who found that younger adults were able to learn the grammatical form classes of the artificial language for the X123 sentence strings, in which each “X” word was heard in context with every “A” and every “B” word. We also tested older adults in order to determine whether there were age-related differences in this type of learning. Our results for the X123 sentence strings replicated the results of the previous study, in that young participants rated both grammatically familiar sentence strings and grammatically novel sentence strings as significantly higher than ungrammatical ones, but not as significantly different from each other. This demonstrates that young participants learned the actual grammatical structure of the artificial language; they were not just memorizing sentence strings. Thus, young learners are clearly able to make use of the statistical patterning of the language input and are able to form grammatical categories even in the absence of other cues to word learning (such as prosodic or semantic cues.)
Our results demonstrated that older adults were also able to learn the grammatical form classes for the X123 strings, showing the same pattern in sentence string rating as the young adults. However, the difference in rating between grammatical and ungrammatical sentence strings was not as large for older adults, suggesting that older adults were not able to learn the grammar as well as younger adults.

While the results for the X123 sentence strings were as predicted, the results for the X4 sentence strings were surprising in that we did not replicate the results of Reeder et al. (2013). Although young participants rated grammatically novel sentence strings as significantly higher than ungrammatical ones, they also rated grammatically familiar strings as significantly different than grammatically novel ones. Thus, they did not extend their knowledge of a grammatical category to “X” words that they had only heard in one “A” or “B” context.

For older adults, there was no clear pattern of learning for the X4 sentence strings. In this case, while the grammatically familiar strings were rated higher than the ungrammatical strings, there was no significant difference in rating between the grammatically novel and grammatically familiar strings, and there was only a marginally significant difference in rating between the grammatically novel and ungrammatical strings. In terms of age differences between the groups, there was only a marginal sentence type by age interaction for the X4 sentence strings.

**Age-related Differences**

Overall, the results on the X123 portion of this experiment are consistent with the “distributional learning” hypothesis (Cartright & Brent, 1997; Mintz, Newport, & Bever, 1995, 2002; Mintz, 2002, 2003; Reeder et al., 2013). Both the young and older group were able to acquire grammatical categories based solely on the linguistic structure of the input (in this case, the distributional information surrounding the “X” words.)
Nevertheless, there were age differences apparent in grammatical form class learning. These differences are consistent with previous aging and implicit learning findings, in which older adults have been found to learn less than younger adults on tasks that involve picking up on probabilistic sequential patterns (Howard et al., 2004; Howard, Howard, Dennis, & Kelly, 2008). Such declines in implicit learning may be related to losses in the basal ganglia, in that there is evidence that the basal ganglia are involved in implicit sequence learning, and that this system declines in old age (Simon, Vaidya, Howard, & Howard, 2011; Bennett, Madden, Vaidya, Howard, & Howard, 2011; Doyon et al., 2009; Seger, 2006). As stated earlier, previous research has also shown the basal ganglia to be linked to language processing (Ullman, 2001; Newman, Supalla, Hauser, Newport, & Bavelier, 2008). This suggests that the statistical learning measured in probabilistic language acquisition tasks calls on similar brain mechanisms as non-verbal implicit sequence learning. Both types of learning are showing similar patterns of age differences.

X4 Results

Our results were surprising in that they were different from those of Reeder et al. (2013). In the present study, younger adults did not generalize properties of the grammatically familiar sentence strings to the grammatically novel ones, rating them instead as significantly different from each other. One possible explanation is the set-up of our experiment. We used the “alien game” form of the experiment, which was designed for use with children and which Reeder et al. (2013) had not used in their study. Reeder et al. (2013) had instead used a “listen only” version, in which the input came to participants via headphones and there was no visual component. We chose the alien version in order to make the task more engaging so that it would be easier for both groups to pay attention. However, it was a less naturalistic language learning
setting, in that learners were not just hearing the input but were also interacting visually with the program. Therefore, participants may have actually been too focused. Aslin and Newport (2012) have suggested that during rule learning, adult learners will be more or less likely to generalize a rule (i.e., the properties of a word) to a word heard in minimal contexts – versus treating it as an exception to the rule – depending on the degree of overlap among the contexts. In this case, younger learners’ increased attention may have made them more aware of gaps in the input and thus less likely to generalize the properties of words that they had heard in only one context.

Alternative Interpretations

We are interpreting our results to mean that older adults are poorer than young adults in their ability to learn grammatical categories from distributional information, but alternative interpretations are possible. For example, since this task measured learning after it occurred, one might argue that older adults are simply not retaining what they learned to the same degree as younger adults. Yet the present study demonstrates the same patterns in age-related differences as TLT and other sequence-learning tasks measuring learning on-line, while it occurs. Thus, we think it likely that the present study also demonstrates age differences in actual learning.

It is possible that older adults might be using the rating scale differently than the younger adults. Nevertheless, there was no main effect of age on the overall ratings collapsed across sentence type, suggesting that the older people did not tend to use extreme ratings more than the young. The older adults did, however, show significant differences in rating among sentence types, suggesting that they understood the rating task.

It is also possible that older adults were not able to hear the input as well, as speech perception declines with age (e.g., Seifert, Hoffnung, & Hoffnung, 2000). However, the fact that
older adults did reveal significant differences among some sentence types suggests that they were able to discriminate among the sounds.

*Future Directions*

One question for future research is whether the format of the instruction contributed to the age differences we observed. Previous research has shown that older adults perform worse than younger adults on language tasks when given explicit instructions, whereas there are fewer (if any) age-related differences in learning when instructions are more implicit (Midford & Kisner, 2005; Lenet et al., 2011). Thus, a possible follow-up study might look at whether age differences decrease if the task instructions are made less explicit (i.e. if participants are told to listen passively to the language and are not told that they will be tested on it until afterward.)

Another remaining question concerns how variations in the type of input that learners hear would influence age differences. In previous research using this task, Reeder et al. (2013) varied the levels of artificial language input that learners received (i.e., the proportion of distributional cues that occurred in the input and the amount of overlap of these contexts across words.) They found that learning in college-aged participants decreased as the distributional cues in the input became less dense and overlapping (Reeder et al., 2013). Therefore, future research should examine the extent to which variability in the artificial language input affects learning in older participants.

Participants in our experiment were also only exposed to the language input for a limited amount of time. The magnitude of age differences in learning with ASRT tasks has been shown to increase in older adults over extended practice (Howard et al., 2004a), so we would expect that longer training would increase age-related differences in form class learning as well. We
might try varying the *amount* of language input (in terms of time) that participants are exposed to in order to see if longer training increases or decreases age differences in learning.

In the Alien task, “Zooma” jumps around the screen, and a spaceship moves along the bottom to indicate the participant’s progress. These complex visual stimuli might have been distracting for older people, who typically have more difficulty ignoring irrelevant material. In particular, research has found that older adults experience declines both in their ability to attend to complex stimuli and to inhibit irrelevant information (Luo & Craik, 2008; Craik, 2006; Hasher & Zacks, 1988). Thus, in order to limit the amount of potentially distracting stimuli, this experiment could be repeated using the “listen only” version of the task mentioned before, in which participants heard the input through headphones and there was no visual component to the task.

**Conclusion**

The current study suggests that it is not just a decline in general cognitive functions, such as explicit memory and processing speed, that leads to language learning deficits in older adults. Rather, older adults also show a decreased ability to use statistical learning in order to distinguish the grammatical categories in a language. While this may seem to indicate bleak prospects for older adults attempting to learn a new language, it is important to recognize that older adults were still able to learn the grammatical form classes. In addition, there is significant potential for future research to point to specific contexts in which older adult language learning might be enhanced.

**Acknowledgments**

I would like to thank Dr. Howard for all of her thoughtful guidance and dedication to advising me on this project; Dr. Newport and Kathryn Schuler for providing me with the artificial
language task and instructions, lending technical support, and giving constant helpful feedback and suggestions; Chelsea Stillman for help with data analysis and overall support for my project; Halley Feldman, Alyssa Coffin, and Caroline Wambach for help with data collection; and Katherine Gamble for helpful comments and discussion. This research was supported by grant R01AG036863 from the National Institute on Aging, National Institutes of Health; NICHD grant R01HD037082; and the Georgetown Undergraduate Research Opportunities Program (GUROP) Summer Fellowship. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.
References


Mackey, A., & Sachs, R. (2011). Older learners in SLA research: A first look at working


Rieckmann, A., Fischer, H., & Bäckman, L. (2010). Activation in striatum and medial temporal lobe during sequence learning in younger and older adults: Relations to


Appendix

Figure 1.

Average ratings for old and young groups on X123 grammatically familiar, grammatically novel, and ungrammatical sentence strings. Error bars represent +/- 1 SE, and asterisks show significant differences (p<.05).
Each bar shows an individual participant’s rating difference between (a) grammatically familiar and ungrammatical strings (GF – UG), (b) grammatically novel and ungrammatical strings (GN – UG), (c) grammatically familiar and grammatically novel strings (GF-GN)

Figure 2a, b, c.
Figure 3.

Average ratings for old and young groups on X4 grammatically familiar, grammatically novel, and ungrammatical sentence strings. Error bars represent +/- 1 SE, and asterisks show significant differences (p<.05).
Each bar shows an individual participant’s rating difference for X4 between (a) grammatically familiar and ungrammatical strings (GF – UG), (b) grammatically novel and ungrammatical strings (GN – UG), (c) grammatically familiar and grammatically novel strings (GF-GN).