Aging Deficits in Naturalistic Speech Production and Monitoring Revealed Through Reading Aloud

Tamar H. Gollan and
Department of Psychiatry, University of California, San Diego

Matthew Goldrick
Northwestern University

Abstract

The current study investigated how aging affects production and self-correction of errors in connected speech elicited via a read aloud task. Thirty-five cognitively healthy older and 56 younger participants read aloud 6 paragraphs in each of three conditions increasing in difficulty: (a) Normal, (b) Nouns-Swapped (in which nouns were shuffled across pairs of sentences in each paragraph), and c) Exchange (in which adjacent words in every two sentences were reversed in order). Reading times and errors increased with task difficulty, but self-correction rates were lowest in the Noun-Swapped condition. Older participants read aloud more slowly, and after controlling for aging-related advantages in vocabulary knowledge, produced more speech errors (especially in the Normal condition), and self-corrected errors less often than younger participants. Exploratory analysis of error types revealed that aging increased the rate of function word substitution errors (saying *the* instead of *a*), whereas younger participants omitted content words more often than older participants. This pattern of aging deficits reveals powerful effects of vocabulary knowledge on speech production, and suggests aging speakers can compensate for aging-related decline in control over speech production with their higher vocabulary knowledge and careful attention to speech planning in more difficult speaking conditions. These results suggest a model of speech production in which planning of speech is relatively automatic, while monitoring and self-correction are more attention-demanding, in turn leaving speech production relatively intact in aging.

Speech patterns and behavior change in aging in numerous noticeable ways (for a recent review, see Kavé & Goral, 2017). Young adults speak more quickly, insert lexical fillers more often (e.g., *like, you know*), and produce a greater number of grammatically complex sentence structures than older speakers (Cooper, 1990; Kemper et al., 1987; 1992; 2003a; 2004; Rabagalia & Salthouse, 2011). While some age differences likely reflect cognitive

Correspondence concerning this article should be addressed to Tamar H. Gollan, 9500 Gilman Drive, La Jolla, CA 92093-0948, tgollan@ucsd.edu.

decline, others might reflect changes in communicative goals, differences in daily routine that change the nature of recent experience, differences in total amount of lifetime experience (Griffin & Spieler, 2006), and even some cognitive advantages. For example, older speakers use a greater variety of words when they speak (lexical diversity increases; Kemper et al., 2010), a result that might reflect the increase in vocabulary size that comes with longer lifetime experience (Bowles & Salthouse, 2008; Kavé & Yafé, 2014; Keuleers, Stevens, Mandera, & Brysbaert, 2015; Verhaegen, 2003). Older speakers may also be more talkative, producing more "off topic" speech, but that can be equally effective in achieving communicative goals (Arbuckle, 2000) and also rated as more interesting (James, 1998; Trunk & Abrams, 2009), than speech produced by younger speakers.

A general consensus in the field characterizes language production abilities as declining relatively more than language comprehension in aging (especially in tasks that do not place significant demands on working memory: Braver & West, 2008; Burke, MacKay, & James, 2000; Burke & Shafto, 2008; Carpenter, Miyake, & Just, 1994; Tun, Wingfield, & Stein, 1999; Waters & Caplan, 2004). Though changes in speech production are thought to be more salient than in comprehension, these too are often subtle and more difficult to observe when compared with other more apparent deficits found in aging e.g., in explicit memory, executive functioning, and processing speed (Dodson, 2017; Kemper, et al., 2001; 2003; Mayr, 2001; Park, Lautenschlager, Hedden, Davidson, Smith, & Smith, 2002; Salthouse, 1996; 2010; Verhaeghen, 2011) and because older speakers can compensate for some production difficulties because of their expanded knowledge base and vocabulary size (e.g., Dahlgren, 1988; Laver & Burke, 1988; Gollan & Brown, 2006; Kavé, Knafo, & Gilboa, 2010; Rabaglia & Salthouse, 2011; Shafto, 2015; Stine-Morrow, Miller, & Herzog, 2006). However, both anecdotally, and in controlled experimental studies, older adults report word finding problems (Goral, Spiro, Albert, Obler, & Connor, 2007; Nicholas, Obler, Albert, & Goodglass, 1985), and tip-of-the-tongue (TOT) retrieval failures significantly more often than younger speakers (Burke, MacKay, Worthley, & Wade, 1991; Maylor, 1990). This aging effect is especially robust with proper name targets (Burke et al., 1991; Burke, Locantore, Austin, & Chae, 2004; Evrard, 2002; James, 2006; Rastle & Burke, 1996; Maylor, 1990), but is also found with object naming especially in advanced aging (over 70–75, or in the eighth decade of life; Kavé et al., 2010; but see Dahlgren, 1998; Gollan & Brown, 2006; Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014).

While it might seem obvious that word-retrieval deficits should also be associated with a higher incidence of word substitution errors, much less is known as to whether or not aging increases other types of speech errors. A challenge in this domain is that speech errors are relatively rare; using a recorded corpus of speech, Garnham, Shillcock, Brown, Mill, and Cutler (1981) estimated that errors occur only once per every thousand words. Even if errors were to increase significantly in aging, this might still be difficult to observe (e.g., Connelly, Hasher, & Zacks, 1991). Investigators of speech production have developed a number of methods to induce speech errors at higher rates, but these typically involve relatively difficult production tasks, which could induce aging effects not found in more natural connected speech, or could even introduce speech production difficulties for reasons not directly related to speech processing.

Vousden and Maylor (2006) used a tongue twister paradigm to contrast phonological errors in younger and older speakers. In a slow speech condition, older speakers did not produce more tongue twister errors than young speakers. In contrast, in a fast rate condition, younger speakers produced many more speech errors while older speakers were simply unable to produce the tongue twisters. This implies a kind of aging deficit in tongue twister production, but leaves open the question as to whether aging changes the types of errors speakers make – and if aging effects are found only in particularly difficult speech production tasks (speech does not typically require rapid production of tongue twisters). MacKay and James (2004) revealed significant aging effects in a speech error elicitation task that required speakers to replace all /p/ sounds in single written words with /b/ sounds and vice versa (e.g., given *punk* participants would respond by saying "*bunk"*, or if given *ribbed* they would say "*ripped"*; and when given control words with neither target sound such as *dune* they were instructed to respond by saying "*neither"*). Interpretation of this work suffers from the same concern as Vousden and Maylor's (2006) study of tongue twisters: While it appears that older speakers do produce more speech errors in some testing conditions, it remains to be determined if aging similarly affects speech production in more naturalistic speaking tasks.

Identifying aging deficits in production of spontaneous connected speech has proved to be challenging (Kavé & Goral, 2017). Spontaneous speech affords considerable flexibility, which might easily allow older speakers to compensate for any processing declines that emerge in aging (Burke et al., 2000; Burke & Shafto, 2008; Caplan & Waters, 2005; Rabaglia & Salthouse, 2011) by using different words, and multiple syntactic structures to express the same concepts. Moreover, it is not always clear if speakers are in fact producing what they intended to produce or not. In some cases, deviations from intention become apparent when speakers interrupt themselves to change or correct what they initially produced. Aging-related difficulties in spontaneous speech might be apparent if older adults stopped to self-correct their own speech errors more frequently than younger adults. However, the ability to self-correct could itself be impaired in aging, and very little research has addressed this question. Two studies found that older speakers corrected their own speech errors at the same rate as younger speakers (Cooper, 1990; Macnamara, Obler, & Albert, 1992), and another study reported no aging effects on the primary task (/p/ and /b/ substitution; as explained above), but a sharp decline in ability to detect error production on control trials ("neither" responses to words without /p/ or /b/ sounds; MacKay & James, 2004). Possibly related evidence comes from studies of aging effects on proof-reading in which detection but not correction of spelling errors was intact in aging, while both detection and correction of grammar and meaning-based errors (which were more difficult for both young and old readers to identify than spelling errors) declined in aging (Shafto, 2015; for a slightly different pattern of aging effects see Abrams, Farrel, & Margolin, 2010).

Some suggestive evidence comes from a recent study designed to investigate how aging affects bilinguals' ability to switch languages in their speech (Gollan & Goldrick, 2016). In this study, young and older bilinguals read aloud mixed-language paragraphs. When they encountered a switch word, bilinguals sometimes spontaneously produced translations of the written words by mistake, or language intrusion errors (e.g., replacing *that* with *que* in a sentence such as *Estaba seguro that he had placed the shoe con la punta pointed upward para*

sostener la ventana open). Older bilinguals produced significantly more such intrusion errors, and also a significantly lower rate of very fast self-corrections (in which they began to produce an intrusion error and stopped themselves mid-error to self-correct before fully producing the error). Importantly, older bilinguals also produced significantly more within-language errors than young bilinguals (e.g., saying *these* instead of *those*, or *could* instead of *would*; errors that monolinguals might also produce). Although self-corrections of these errors were not examined in this previous study (which focused on bilingual language control), the aging deficit in monitoring these errors might be even stronger given that language switches are generally easier to identify than within-language errors (both in auditory comprehension and in reading aloud; Ivanova, Ferreira, & Gollan, 2017). To examine this question, we reanalyzed data from Gollan and Goldrick (2016; including all self-corrections,[1] not just mid-error interruptions). Older bilinguals self-corrected just 22% ($SD = 16$) of within-language errors, while young bilinguals self-corrected 42% ($SD = 19$), a mean age difference of 20%. This effect was four times as large as the age deficit in correction of intrusions (older bilinguals: 42% (SD=18%); younger bilinguals: 47% ($SD =$ 24%); mean age-difference 5%). While this might imply an aging effect on speech monitoring, it could reflect processing mechanisms specific to bilingualism or the relative difficulty of reading paragraphs with frequent language switches in them. If the effect is indeed not specific to bilinguals, it should be found when comparing young to older monolinguals in the reading aloud task.

In the present study, we examined aging effects on production of connected speech in monolingual English speakers using the paragraph reading task, a relatively naturalistic task that is also transparent with respect to the targets of intended speech. Importantly, though we relied on reading to elicit speech production, speech errors produced in the read aloud task arise primarily during planning and execution of speech production, and do not reflect errors in visual word processing and/or sentence comprehension; Gollan et al., 2014; Gollan & Goldrick, 2016). Perhaps most convincing in this regard is that bilinguals produce intrusion errors even with translation equivalents that do not resemble each other in form (e.g., see example above), and even Chinese-English bilinguals produce intrusions in this task though visually the language switches are quite obvious when the two languages have such distinct orthographies (Li & Gollan, in press). Because aging effects on speech production might be magnified in the absence of semantic or syntactic support (or more generally in difficult tasks), we included three different types of paragraphs including: a) Normal paragraphs, b) Nouns-Swapped paragraphs, in which all the nouns in each sentence were swapped randomly across positions in consecutive pairs of sentences (eliminating semantic support), and c) Exchange paragraphs, in which adjacent words were exchanged (eliminating both semantic and syntactic support). Paragraphs in the Normal condition elicited production of a coherent story, while those in the Nouns-Swapped condition required production of grammatical but nonsensical sentences (like Madlibs or Jabberwocky; see Methods section and the Appendix for examples), and the Exchange condition required producing strings of function and content words without any support from grammatical structure or meaning. If aging affects production of connected speech only in difficult processing conditions, we

---

[1](partial intrusions + self-corrected full intrusions)/(partial + full intrusions)

might expect to find significant age group differences only (or to a larger extent) in the Nouns-Swapped or Exchange conditions and no age differences in the Normal condition. Furthermore, if syntactic (rather than semantic) support plays a key role in processing, any aging differences might be magnified in the Exchange and Nouns-Swapped conditions. Conversely, more difficult conditions might be more likely to elicit compensatory strategies, which as reviewed above, could mask aging effects.

Our primary interest was whether older adults would produce speech errors more often, and self-correct those errors less often, than younger adults specifically without requiring them to read aloud at a faster pace than they would naturally choose. Importantly, as in our previous aging study (Gollan & Goldrick, 2016), we did not impose any speed requirements, and did not include any specific instructions about self-correction (self-corrections were spontaneous). As such, we expected older adults would likely read aloud more slowly than younger adults, a factor that might work against finding aging-related changes in speech error rate – but that might be best suited for observing ecologically valid aging effects on self-correction (participants might self-correct less often if they are pressured to produce speech quickly). Similarly, aging effects might be obscured by normative aging-related growth in vocabulary knowledge (which increases along with other crystallized abilities; Baltes, 1997; Hartshore & Germine, 2016) as a consequence of increased experience and exposure to print over time (e.g., Stanovich et al. 1995). Thus, in all our analyses we also considered the possible effects of vocabulary knowledge on production and self-correction of speech errors. Finally, further clues as to the possible cognitive mechanisms underlying possible aging effects on production of speech errors could be found if some types of errors appeared to be more affected by aging than others. We examined this via exploratory analysis of different types of speech errors.

Method

Participants

Thirty-five cognitively healthy English-speaking older and 56 younger monolinguals participated in the study. Older participants were screened extensively for cognitive status in annual neurological and neuropsychological test batteries as part of their participation as healthy controls at the University of California San Diego's (UCSD) Shiley-Marcos Alzheimer's Disease Research Center (ADRC), which follows a cohort of about 500 participants with equal numbers of patients versus cognitively healthy controls. All participants were diagnosed as cognitively healthy (i.e., "Normal controls" not "Probable Alzheimer's Disease" and not "Mild Cognitive Impairment" or any other neurological diagnosis) based on criteria developed by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's Disease and Related Disorders Association (ADRDA; McKhann et al., 1984). Participants with a history of alcoholism, drug abuse, severe psychiatric disturbances, severe head injury, and learning disabilities were also excluded. Young participants were recruited from the UCSD Psychology subject pool, and received course credit for their participation.

Table 1 shows participant demographics, neuropsychological test scores, and self-reported language history questionnaire responses for all young and older speakers tested, and for a

vocabulary-matched subset of participants. The vocabulary-matched subset was selected by one-to-one matching as many older with young participants as possible on a standardized measure of single word oral-reading ability (American National Adult Reading Test or ANART scores; Grober, Sliwinski, & Korey, 1991; Kreutzer, DeLuca, & Caplan, 2011), with some adjustments made as needed to match overall averages, and to produce group means that are also not significantly different on the multiple-choice vocabulary test (Shipley, 1940), and our measure of productive vocabulary (picture naming test scores; the Multilingual Naming Test or MINT; MINT; Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2012; Ivanova, Salmon, & Gollan, 2013). Vocabulary tests were administered just after the paragraph reading task in a fixed order (ANART, MINT, and Shipley; see below). Additional neuropsychological test scores were available for all older participants from their annual evaluations at the ADRC (e.g., Dementia Rating Scale; Mini Mental Status Exam; testing on these measures took place just before (n=12) or just after (n=23) they were tested on the paragraph reading task, with a lag between testing sessions of about 4.4 (SD=2.3) months on average. All participants reported being exposed to English from birth, regular and current use of English throughout their lifetime, and limited exposure to and proficiency in other languages. All but 3 older and 5 young participants reported being right-handed. Informed consent was obtained from all individuals prior to their participation in the research study. Study procedures were approved by the UCSD Institutional Review Board.

Materials

Paragraphs (n=18) were English-only versions of those used by Gollan and Goldrick (2018) and were adapted into three versions so that each paragraph could be presented in each of three conditions: (a) Normal, (b) Nouns-swapped, (c) Exchange. Paragraphs were 121.94 words long on average (SD = 14.68; range 100–148), and each participant read aloud six paragraphs of each type. Paragraphs were randomly divided into 3 groups that were blocked for presentation. Across 9 stimulus lists, the condition (Normal, Nouns-swapped, Exchange) and order of each group of paragraphs were counterbalanced. An example of each type of paragraph is presented in the Appendix In the Nouns-swapped condition, all nouns in the sentence were swapped randomly across positions in consecutive pairs of sentences, or across the final three sentences. For sentences without a noun, one personal singular or plural pronoun was selected and changed into a noun. Care was taken to ensure that modifiers not designated as nouns were not swapped (e.g., only *meters* in *a hundred meters ahead* would be swapped because *hundred* is a modifier in this sentence). Additionally, swapped nouns were changed as needed to fit the new context, preserving grammaticality (e.g., if the noun *woman* was swapped into the *meters* position in *a hundred meters ahead* it would be modified to *women*). Finally, capitalization and punctuation was left in its original position (so that the first word in a "sentence" would be capitalized, and commas and periods did not move with their originally adjacent nouns). In the Exchange condition, pairs of consecutive words were exchanged. If there were three words at the end of a sentence the last word was left alone, and additionally sometimes a word was left in its original place to avoid creating repetitions.

The **American National Adult Reading Test** (ANART; Grober et al., 1991; Kreutzer et al., 2011) consists of 50 written words with irregular spelling and the participants were asked to pronounce each word.

The **Shipley-Hartford vocabulary test** is part of the Shipley Institute of Living Scale (Shipley, 1940) and it measures receptive vocabulary. It has 40 written target words, presented in ascending order of difficulty, and participants choose the closest synonym to the target word from four presented options (Shipley reported a split-half reliability for this test as .87).

The **Multilingual Naming Test** (MINT; Gollan et al., 2012) consists of 68 black-and-white line drawings, administered in order of progressing difficulty. This test was designed to assess picture-naming ability in four languages (English, Spanish, Mandarin, Hebrew), and provides a measure of productive vocabulary in bilingual and monolingual speakers alike (Ivanova, Salmon, & Gollan, 2013).

Procedure

Participants were tested individually in a quiet, well-lit room. Paragraphs were presented in PowerPoint as words printed in Calibri font, size 20, double-spaced. Each paragraph was presented on a single slide. Participants were instructed to read each paragraph aloud as accurately as possible at a comfortable pace, and were audio-recorded and timed with a stop-watch. Prior to reading the first paragraph of each condition, participants completed a shorter practice paragraph. The experimenter corrected participants if they produced any errors during these practice trials. Participants were not instructed to self-correct errors they noticed, and if they self-corrected they were neither prompted to stop self-correcting nor encouraged to do so. If participants commented on, or asked about, the experimental manipulations they were prompted with "Just try to read what's written on the page as accurately as possible at a comfortable pace." Participants first completed the read-aloud task, and then the ANART, the MINT, and the Shipley (in that order). Errors were marked on a coding sheet during testing and were later checked against audio recordings.

Results

Two research assistants coded errors for all participants with each assistant coding half of the younger and half of the older participants. Errors were defined as any word produced differently from what was written on the page. Examples of error types are shown in Table 2.

A small number of speech productions could have been classified as more than one type of error, either as two attempted productions including one correct and one erroneous production, or as a specific error type and a repetition. In these cases, preference was given to 1) classifying them as errors (rather than marking them as correct), and 2) classifying them as specific error types rather than as repetitions. Disfluencies and hesitations were not marked as errors; repetitions were marked as errors only when they were not produced along with a more specific error type. We based this assumption on the observation that some of these instances appeared to reflect instances of doubled self-correction for emphasis.

Inter-rater reliability was assessed using 30 paragraphs, taken from 10 randomly selected participants (5 older and 5 younger from the vocabulary-matched subset). From each of these participants' data 3 paragraphs (one per condition) were selected at random, yielding 3,712 data points. Of these the two coders agreed perfectly on classifications in 99.5% of cases; both coders classified 98.8% of the responses as correct (n=3648), and 0.9% as the same error type (7 function-function substitutions, 10 omissions, 8 repetitions, 8 swap, and 1 inflection error). The two coders disagreed on 0.5% of cases (n=19; in 11 cases one coder classified the response as correct while the other coded an error, in 5 cases this situation was reversed, and in 3 cases there was disagreement about error type).

We analyzed the data using linear (for reading times) and logistic (for error and self-correction rates) mixed-effects regressions (Dixon, 2008; Jaeger, 2008; models were fit using the R package lme4, v. 1.1–12; Bates, Maechler, Bolker, & Walker, 2015). Analyses examined the effects of age group (older, younger), paragraph type (Normal, Nouns-Swapped, Exchange), ANART score, and all interactions. Age group was contrast coded (i.e., .5 for older, −0.5 for younger). Paragraph type was coded by three centered contrasts (Nouns-Swapped vs. Normal, Exchange vs. Normal, and Exchange vs. Nouns-Swapped; the third contrast was calculated by re-leveling paragraph type and re-fitting the regression model). ANART score was coded as a centered continuous variable to control for differences in vocabulary knowledge. ANART scores were used because (collapsing all young and older participants together) they showed the strongest simple correlation with participant accuracy in paragraph reading ($r = 0.32$, vs. 0.21 for the Shipley-Hartford vocabulary test and 0.02 for the MINT test) as well as with self-correction rate ($r = 0.14$, vs. 0.01 for the Shipley-Hartford and −0.10 for the MINT). ANART score was coded by (centered) proportion correct (logit-transformed), as raw scores blocked model convergence. Note that our figures use raw proportion correct in ANART; plotting our regression model fits in each figure will therefore yield non-linear curves (reflecting the non-linear transform of ANART score). Furthermore, models predicting binomial variables (accuracy, self-correction rate) predict log-odds of response; plotting these fits in raw proportion correction or proportion self-corrected will also lead to non-linear curves.

For each measure, we used the maximal random effects structure that would converge and was not over-parameterized (Bates, Kliegl, Vasishth, & Baayen, 2015). After excluding outlier paragraphs with reading times exceeding 100 seconds ($N = 8$, distributed across 3 participants), the models of reading times converged with the maximal random effects structure. For models of accuracy, subjects were entered as random intercepts, as more complex models failed to converge. For models of self-corrections, the full model was over-parameterized, so the random by-subject slope for the Exchange vs. Normal contrast was excluded (it had the smallest variance); this yielded a random effects structure with a random intercept by subject and a correlated, by-subject random slope for the Nouns-Swapped vs. Normal contrast. The significance of each fixed effect was assessed via likelihood ratio tests (Barr, Levy, Scheepers, & Tily, 2013). Tables in Supplementary Materials summarize the estimates of random effects, their correlations (for models with multiple random effects), and fixed effects.

We considered including years of education as a covariate. Vocabulary scores and education level were significantly correlated ($r = 0.49$) and this was driven primarily by the older group (older $r = .31$; younger $r = .04$). Given this collinearity, we preferred to use the objective standardized measure of vocabulary. Additionally, regression models including years of education explained less variance than those including vocabulary[2].

Reading Times

Although our primary measure of interest in the read aloud task was production and self-correction of speech errors, it was important to also consider possible differences between groups in reading times to determine if this could be influencing any of the observed error effects. As shown in Figure 1, participants read Exchange paragraphs most slowly, followed by Nouns-Swapped paragraphs, and Normal paragraphs were read fastest. This was confirmed by pair-wise comparisons in the regression (Exchange vs. Nouns-Swapped: $\beta = 19.165$, $SE\,\beta = 0.64$, $\chi^2(1) = 217.10$, $p < 0.0001$; Exchange vs. Normal: $\beta = 22.514$, $SE\,\beta = 0.719$, $\chi^2(1) = 225.61$, $p < 0.0001$; Nouns-Swapped vs. Normal: $\beta = 3.889$, $SE\,\beta = 0.418$, $\chi^2(1) = 71.44$, $p < 0.0001$). Additionally, older adults read aloud more slowly than younger adults ($\beta = 15.705$, $SE\,\beta = 1.665$, $\chi^2(1) = 62.37$, $p < 0.0001$), and individuals with higher vocabulary scores (i.e., ANART) read aloud more quickly than those with lower scores ($\beta = -5.699$, $SE\,\beta = 1.115$, $\chi^2(1) = 23.08$, $p < 0.0001$).

In absolute reading times, age-related slowing was biggest in more difficult paragraphs, as shown by significant interactions of paragraph type and age (Exchange vs. Nouns-Swapped: $\beta = 6.72$, $SE\,\beta = 1.488$, $\chi^2(1) = 18.39$, $p < 0.0001$; Exchange vs. Normal: $\beta = 7.237$, $SE\,\beta = 1.684$, $t = 4.30$[3], $p < 0.0002$; Nouns-Swapped vs. Normal: $\beta = 2.32$, $SE\,\beta = 0.972$, $\chi^2(1) = 5.65$, $p < 0.02$). However, these interactions were not significant when relative reading times were considered. We log-transformed reading times so that each regression coefficient expressed proportional changes in reading times. This revealed that older adults read each type of paragraph aloud approximately 30% more slowly than younger adults (as shown by a significant main effect: $\beta = 0.297$, $SE\,\beta = 0.032$, $\chi^2(1) = 62.14$, $p < 0.0001$ with no significant interactions: Exchange vs. Nouns-Swapped: $\beta = 0.012$, $SE\,\beta = 0.022$, $\chi^2(1) = 0.32$, $p < 0.57$; Exchange vs. Normal: $\beta = 0.025$, $SE\,\beta = 0.029$, $\chi^2(1) = 0.78$, $p < 0.38$; Nouns-Swapped vs. Normal: $\beta = 0.03$, $SE\,\beta = 0.021$, $\chi^2(1) = 1.98$, $p < 0.16$).

In absolute reading times, vocabulary effects were stronger in more difficult paragraphs, as shown by significant interactions of paragraph type and ANART (Exchange vs. Nouns-Swapped: $\beta = -2.48$, $SE\,\beta = 1.0$, $\chi^2(1) = 5.99$, $p < 0.02$; Exchange vs. Normal: $\beta = -3.374$, $SE\,\beta = 1.128$, $\chi^2(1) = 8.57$, $p < 0.005$; Nouns-Swapped vs. Normal: $\beta = -2.094$, $SE\,\beta = 0.651$, $\chi^2(1) = 10.19$, $p < 0.002$). However, when considering relative reading times, the Nouns-Swapped vs. Normal contrast remained significant, Exchange vs. Normal was

---

[2]Our a priori intention was to control for vocabulary in the analyses of reading times, accuracy, and self-corrections (hence we obtained ANART scores for both young and older participants). However, additional measures of individual differences in processing speed, working memory, attention, and list memory, were available for older adults (from annual testing sessions at the ARDC). As shown in the Supplemental Materials, comparison of these with vocabulary knowledge suggest that vocabulary is a stronger predictor of performance in the read-aloud task (see Table S1).
[3]The subset model excluding the Exchange vs. Normal paragraphs by age group interaction failed to converge, so p values were estimated by assuming t statistics approximated that of Z.

marginal, and Exchanges vs. Nouns-Swapped was not significant (Exchange vs. Nouns-Swapped: $\beta = -0.004$, $SE\,\beta = 0.014$, $\chi^2(1) = 0.08$, $p < 0.78$; Exchange vs. Normal: $\beta = -0.033$, $SE\,\beta = 0.019$, $\chi^2(1) = 2.97$, $p < 0.09$; Nouns-Swapped vs. Normal: $\beta = -0.041$, $SE\,\beta = 0.014$, $\chi^2(1) = 7.81$, $p < 0.006$).

Finally, aging-related slowing was significant even without controlling for age-group differences in vocabulary. Repeating the analysis of raw RT above and excluding the ANART control factor yielded a slightly smaller (but still significant) estimate of the aging effect ($\beta = 11.236$, $SE\,\beta = 1.603$, $\chi^2(1) = 39.30$, $p < 0.0001$). Similar results were found for the proportional analysis ($\beta = 0.216$, $SE\,\beta = 0.030$, $\chi^2(1) = 41.16$, $p < 0.0001$).

Accuracy

As shown in Figure 2[4], parallel to the reading times, accuracy was lowest in the Exchange paragraphs, higher in Nouns-Swapped paragraphs, and highest in Normal paragraphs. This was confirmed by pair-wise comparisons in the regression (Exchange vs. Nouns-Swapped: $\beta = -1.027$, $SE\,\beta = 0.042$, $\chi^2(1) = 674.38$, $p < .0001$; Exchange vs. Normal: $\beta = -1.212$, $SE\,\beta = 0.045$; $Z = -27.01$[5], $p < .0001$; Nouns-Swapped vs. Normal: $\beta = -0.185$, $SE\,\beta = 0.053$, $\chi^2(1) = 11.92$, $p < .001$). Higher vocabulary (ANART) scores were associated with higher accuracy ($\beta = 0.322$, $SE\,\beta = 0.102$, $\chi^2(1) = 9.31$, $p < .005$).

Interestingly, paragraphs that were read most quickly, and overall most accurately – Normal paragraphs – were also the only condition to show significant aging effects on accuracy. While there was no main effect of age ($\beta = -0.235$, $SE\,\beta = 0.152$, $\chi^2(1) = 2.32$, $p < .13$), there was a significant interaction of age and paragraph type (Exchange vs. Normal paragraphs: $\beta = 0.246$, $SE\,\beta = 0.095$, $\chi^2(1) = 6.45$, $p < .02$). Follow-up tests showed significantly lower accuracy for older vs. younger adults in the Normal paragraphs ($\beta = -0.427$, $SE\,\beta = 0.178$, $\chi^2(1) = 5.50$, $p < .02$), but no significant age effect in the Exchange paragraphs ($\beta = -0.157$, $SE\,\beta = 0.154$, $\chi^2(1) = 1.0$, $p < .32$). While not significant, the age effect within the Nouns-Swapped paragraphs ($\beta = -0.295$, $SE\,\beta = 0.192$, $\chi^2(1) = 2.32$, $p < .13$) fell in between the effect size in the Normal and Exchange paragraphs; interactions between these paragraph types and age were not significant (age by Nouns-Swapped vs. Exchange: ($\beta = 0.109$, $SE\,\beta = 0.091$, $\chi^2(1) = 1.40$, $p < .24$; age by Nouns-Swapped vs. Normal: $\beta = 0.138$, $SE\,\beta = 0.110$, $\chi^2(1) = 1.49$, $p < .23$). Thus, the Normal condition was fastest and least error prone for all participants (young and old), and elicited the smallest aging effect in terms of speed, but the largest aging effect with respect to decline in accuracy. This suggests that aging effects on accuracy were not an artifact of response speed.

Vocabulary effects were significant in Normal and Nouns-Swapped, but not in Exchange paragraphs. Pairwise interactions showed that the vocabulary effect was significantly weaker in Exchange paragraphs relative to the other two paragraph types (Exchange vs. Nouns-Swapped interaction: $\beta = -0.306$, $SE\,\beta = 0.062$, $\chi^2(1) = 25.03$, $p < .0001$; Exchange vs.

---

[4]There is a clear outlier in the young adult group, who produced many more errors than any participant young or old in the Nouns-Swapped and Exchange conditions. Exclusion of this participant from analyses does not change the significance of aging effects reported.
[5]The subset model excluding the Nouns-Swapped vs. Normal paragraphs contrast failed to converge, so p values were estimated by the Wald Z statistic.

Normal interaction: $\beta = -0.370$, $SE\,\beta = 0.067$, $\chi^2(1) = 32.15$, $p < .0001$). Follow-up tests within each paragraph type showed no significant vocabulary effect in Exchange paragraphs ($\beta = 0.163$, $SE\,\beta = 0.103$, $\chi^2(1) = 2.43$, $p < .12$); but ANART scores were significantly correlated with accuracy in both the Normal ($\beta = 0.568$, $SE\,\beta = 0.124$, $\chi^2(1) = 18.87$, $p < .0001$) and the Nouns-Swapped conditions ($\beta = 0.454$, $SE\,\beta = 0.130$, $\chi^2(1) = 11.41$, $p < .001$; there was no significant difference in ANART effect sizes across these two paragraph types: $\beta = -0.063$, $SE\,\beta = 0.079$, $\chi^2(1) = 0.62$, $p < .43$).

Note that if we fail to control for differences in vocabulary knowledge, the aging effect on accuracy is no longer reliable. Repeating the regression in the Normal condition (i.e., the condition that exhibited a significant age-related decrease in accuracy), but excluding the ANART control, yielded a non-significant effect of age ($\beta = -0.029$, $SE\,\beta = 0.174$, $\chi^2(1) = 0.03$, $p < .87$). This underscores how important this control can be in analyses of aging effects.

A potential concern, given our design, is that the aging effect on accuracy in naturalistic Normal paragraphs might have occurred because these were intermixed with unusual (Nouns-Swapped and Exchange) paragraphs. To consider this possibility, we re-ran the follow-up regression from within the Normal condition (i.e., including the ANART control), adding age, order (first, second, or third block, centered), and their interaction to the model. The effect of age was still significant ($\beta = -0.438$, $SE\,\beta = 0.178$, $\chi^2(1) = 5.82$, $p < .02$), but there was no main effect of order ($\beta = -0.072$, $SE\,\beta = 0.095$, $\chi^2(1) = 0.56$, $p < .46$) and no age by order interaction ($\beta = -0.260$, $SE\,\beta = 0.192$, $\chi^2(1) = 1.82$, $p < .18$). This suggests that the aging effect on accuracy is not an artifact of mixing normal and atypical text within the experiment.

Self-corrections

While reading times and accuracy exhibited similar main effects of condition, self-corrections produced a different pattern of results. As shown in Figure 3, Nouns-Swapped paragraphs exhibited the lowest rate of self-corrections, suggesting that the disruption of semantic information (while syntactic information remained intact) impaired error monitoring. Self-correction rates were significantly lower in Nouns-Swapped vs. Normal paragraphs ($\beta = -0.434$, $SE\,\beta = 0.148$, $\chi^2(1) = 8.89$, $p < .003$), and Noun-Swapped vs. Exchange ($\beta = 0.547$, $SE\,\beta = 0.119$, $\chi^2(1) = 21.21$, $p < .0001$), but there was no significant difference in self-correction rates between Exchange vs. Normal paragraphs ($\beta = 0.130$, $SE\,\beta = 0.106$, $\chi^2(1) = 1.51$, $p < .22$). Across conditions, individuals with higher vocabulary (ANART) score were more likely to self-correct ($\beta = 0.302$, $SE\,\beta = 0.126$, $\chi^2(1) = 5.50$, $p < .02$). Interactions of ANART and paragraph type were not significant (Exchange vs. Nouns-Swapped: $\beta = -0.072$, $SE\,\beta = 0.182$, $\chi^2(1) = 0.16$, $p < .70$; Noun-Swapped vs. Normal: $\beta = -0.005$, $SE\,\beta = 0.23$, $\chi^2(1) = 0$, $p < 0.99$; Exchange vs. Normal: $\beta = -0.07$, $SE\,\beta = 0.166$, $\chi^2(1) = 0.06$, $p < .81$).

Critically, across conditions, older adults were less likely to self-correct than younger adults ($\beta = -0.537$, $SE\,\beta = 0.185$, $\chi^2(1) = 8.11$, $p < .005$). This effect did not significantly differ across conditions; the interactions of age and paragraph type were not significant (Exchange vs. Nouns-Swapped: $\beta = 0.475$, $SE\,\beta = 0.263$, $\chi^2(1) = 3.33$, $p < .07$; Exchange vs. Normal:

$\beta = 0.056$, $SE \beta = 0.229$, $\chi^2(1) = 0.0$, $p < 0.99$; Noun-Swapped vs. Normal: $\beta = -0.436$, SE $\beta = 0.318$, $\chi^2(1) = 1.89$, $p < .17$). The marginal interaction comparing Exchange to Nouns-Swapped paragraphs reveals the possibility of a larger aging effect on self-corrections in the condition with the lowest overall rate of self-corrections; but note the high variability in self-correction rates between participants (in all conditions).

Across conditions, individuals with higher vocabulary (ANART) score were more likely to self-correct ($\beta = 0.302$, $SE \beta = 0.126$, $\chi^2(1) = 5.50$, $p < .02$). Interactions of ANART and paragraph type were not significant (Exchange vs. Nouns-Swapped: $\beta = -0.072$, $SE \beta = 0.182$, $\chi^2(1) = 0.16$, $p < .70$; Noun-Swapped vs. Normal: $\beta = -0.005$, $SE \beta = 0.23$, $\chi^2(1) = 0$, $p < 0.99$; Exchange vs. Normal: $\beta = -0.07$, $SE \beta = 0.166$, $\chi^2(1) = 0.06$, $p < .81$).

Note that if we fail to control for differences in vocabulary knowledge, the aging effect on self-corrections is no longer significant. Repeating the regression excluding the ANART control, yielded a marginal effect of age ($\beta = -0.322$, $SE \beta = 0.168$, $\chi^2(1) = 3.64$, $p < .057$). This again suggests that controlling this type of individual difference is crucial for analyses of between-participant differences in age.

Post-hoc analyses: Error types

In post-hoc analyses, we examined whether the accuracy or self-correction effects were driven by particular error types. We considered conducting analyses parallel to those above, using a logistic regression that was multinomial (one category per response type) instead of the standard binomial (correct vs. incorrect). However, this was not possible given the very small number of errors within each type. We therefore conducted a nonparametric bootstrap analysis to estimate the 95% confidence interval of the mean within each age group, limiting ourselves to a subset of participants matched for vocabulary size (see Table 1). Additionally, to maximize power, we collapsed across paragraph types.

Figure 4 shows the results of this analysis for number of errors. Swap errors (for examples see Table 2) were overwhelmingly confined to the Exchange paragraphs (97.86%, $N = 280$), and thus, we do not discuss these errors further.

Error types that young and older participants produced most often were largely but not exactly rank ordered in the same way across age-groups; while older participants produced function word substitution errors most often, followed by omissions, content word substitution errors, and inflection errors (with very few repetition and nonword errors), for younger participants, omission errors were most common followed by function and then content word substitution errors (and very few inflection, repetition, and nonword errors). Of great interest to our central question concerning aging effects, the 95% confidence intervals for function word substitution errors (e.g., *and* → *or*) did not overlap across age groups, suggesting that aging may specifically impact the processing of this word class. This aging effect was consistent across conditions (mean ratio, older/young errors across conditions: 2.355; broken down by condition it was 2.118 in Normal paragraphs, 2.987 in Nouns-Swapped, and 2.142 in Exchange). The vast majority of these substitutions (in both age groups) yielded alternative grammatical sentences[6] (Normal: older adults 74.29% ($N = 70$),

younger adults 82.86% ($N = 35$); Nouns-Swapped: older adults 83.54% ($N = 79$), younger adults 78.57% ($N = 28$); in Exchange, productions are designed to be ungrammatical).

The results for self-corrections of different error subtypes are shown in Table 3. The 95% confidence intervals for corrections of function word substitution errors exhibited the least degree of overlap across age groups, providing converging evidence that aging may specifically impact the processing of this word class[7].

Omission errors were the only type that trended in the opposite direction; younger participants omitted a greater number of words than older participants. We repeated the regression for accuracy, with omission error rate as the dependent measure, adding grammatical category as a main effect and as an interaction with age. This analysis, shown in Figure 5, revealed a significant main effect of grammatical class ($\beta = 1.331$, $SE \beta = 0.096$, $\chi^2 (1) = 261.15$, $p < .0001$) and an interaction ($\beta = 0.531$, $SE \beta = 0.191$, $\chi^2 (1) = 7.83$, $p < .006$). Follow-up regressions within each grammatical class showed a significant age effect for content words (with significantly *fewer* omissions for older than for younger adults $\beta = -0.773$, $SE \beta = 0.328$, $\chi^2 (1) = 6.18$, $p < 0.02$), and no significant age-group difference in omission of function words ($\beta = -0.103$, $SE \beta = 0.192$, $\chi^2 (1) = 0.28$, $p < 0.60$).

Though confidence intervals overlapped across age groups for inflection errors, given that MacKay and James (2004) reported older adults were significantly more likely to drop suffixes than young adults, we further subdivided these into inflection drop (e.g., skipping → skip), addition (e.g., skip → skipping) and substitution (e.g., skipping → skipped) errors (note that errors classified here as "inflection (drop)" errors are analogous to "omission errors" in MacKay & James, 2004). These showed some suggestive effects in the same direction. Inflection additions occurred similarly often in older adults' $M = 0.94$ ($SD = 1.00$, $N = 31$) as in younger adults' speech $M = 1.09$ ($SD = 0.83$, $N = 23$), while inflection drops seemed to occur more often in older adults' $M = 2.44$ ($SD = 4.15$, $N = 55$) than in younger adults' speech $M = 0.82$ ($SD = 0.75$, $N = 20$), and possibly inflection substitutions also occurred more often in older adults' $M = 1.56$ ($SD = 2.00$, $N = 41$) as in younger adults'

---

[6]To judge these for grammaticality, the first author rated each sentence outcome as grammatical or not, flagging ones she was unsure of. The second author checked all the flagged sentences. A small number of disagreements was settled with discussion. The second author then checked 10 additional sentences within each condition; there was only one disagreement across all 20 examples, suggesting the coding was reliable.

[7]Additionally, re-analysis of within-language error subtypes from Gollan and Goldrick (2016) also revealed a large number of function word substitution errors, especially in older bilinguals. Collapsing all errors produced with dominant and non-dominant language targets, there were four error subtypes for which both older and proficiency-matched younger bilinguals produced at least 20 errors (across participants):

- function word substitution errors (e.g., *a* → *the*; older bilinguals $M = 16.0$, $SD = 15.1$, $N=320$, younger bilinguals $M = 6.7$, $SD = 4.8$, $N=134$)

- nonword errors (e.g., applauding → applouding; older bilinguals $M = 11.1$, $SD = 14.5$, $N=221$, young bilinguals $M = 4.5$, $SD = 5.0$, $N=90$)

- inflection errors (e.g., abandoned → abandon; older bilinguals $M = 6.7$, $SD = 8.8$, $N=134$, younger bilinguals $M = 1.6$, $SD = 2.1$, $N=31$)

- form related (e.g., dared → darted; older bilinguals $M = 4.3$, $SD = 4.5$, $N=85$, younger bilinguals $M = 3.6$, $SD = 2.6$, $N=71$).

When combined with the data in the present study, it appears that aging increases function word substitution errors and inflection errors (whereas the nonword errors difference in this previous study might have more to do with reading aloud in a nondominant language).

speech $M = 0.45$ ($SD = 0.69$, $N = 16$). To examine if older adults were more likely to delete inflections than younger adults (given similarity to an effect previously reported by MacKay & James, 2004), we examined the ratio of drop to substitution errors, limiting ourselves to the 7 older adults and 5 younger adults who produced both types of errors. Within these very small groups, older adults did exhibit a higher ratio ($M = 3.05$, $SD = 2.63$) than younger adults ($M = 1.03$, $SD = 0.58$). However, inflection errors were very few in number, especially for young adults, reducing confidence in the reliability of these differences, we do not discuss these further.

Discussion

The present study aimed to determine if aging leads speakers to produce more speech errors, and fewer spontaneous self-corrections of errors, in connected speech. The results provided clear answers to these questions, and some clues as to the underlying cognitive mechanisms. Summarizing the key results, older participants read paragraphs aloud significantly more slowly than younger participants, and also produced more errors and self-corrected less often, but only after controlling for an aging-related advantage in vocabulary knowledge. The latter factor exerted robust effects, speeding reading times, reducing error rates, and increasing self-corrections. Importantly, aging effects were not restricted to difficult or unusual speech conditions. In fact, we obtained some evidence in the opposite direction; e.g., the aging-related increase in speech errors was significant on its own only in the Normal condition, in which speech was most naturalistic, produced most quickly, least error prone, and least variable (see Figure 2). Additionally, self-corrections, but not aging effects, varied across condition (though this might be interpreted with caution given high variability in self-correction rates (see Figure 3). Finally, error rates and self-corrections seemed to some extent to be driven by different underlying cognitive mechanisms; condition effects patterned differently in some measures (e.g., the exchange condition elicited the most errors, whereas nouns swap elicited lowest self-correct rates), but similarities were apparent in exploratory analysis of other measures. Specifically, in both accuracy and self-corrections, aging effects were most robust for whole-word function substitution errors (e.g., saying *the* instead of *a*; see Table 2), which older adults produced more often, and self-corrected less often, than young adults (see Figure 4 and Table 3).

Aging Effects on Accuracy

Our finding of a significant age-related increase in production of speech errors in the Normal condition establishes, in much more certain terms than was possible in previous experimental studies, that aging in fact leads speech production to be more error prone – once the vocabulary advantage often found in older age is controlled. Reading aloud is a fairly easy and naturalistic task that elicits rapid production of connected speech, and as explained in the Introduction, relies heavily on normal mechanisms of speech production (Ferreira, 1991; Gollan et al., 2014; Gollan & Goldrick, 2016; 2018; Kemper, Bontempo, Herman, McKedy, Schmalzried, Tagliaferri, & Kieweg, 2014). Importantly in the present context, even though intended speech was initially planned via reading, this ability remains relatively intact in aging (Burke et al., 2000; Burke & Shafto, 2008; especially in non-demanding conditions, e.g., Smiler, Gagne, & Stine-Morrow, 2003b; Stine-Morrow, Gagne,

Morrow, & DeWall, 2004; Stine-Morrow, Miller, Gagne, & Hertzog, 2008; Stine-Morrow et al., 2006; Waters & Caplan, 2004), and therefore input processing is unlikely to explain the aging-related increase in speech errors.

Instead, the aging-related decline in accuracy in the Normal condition may reveal difficulties with speech production that can be offset by high vocabulary knowledge and/or compensatory or strategic processing that is be triggered by more difficult speaking tasks. On this view, older speakers exerted more effort, especially in the Nouns-Swapped and Exchange conditions, to minimize errors in their speech – similar to previous studies in which compensatory processing and strategies recruited by older participants offset aging effects (Rabaglia & Salthouse, 2011; Stine-Morrow et al., 2008 in behavioral data; while in others the brain revealed evidence of compensatory processing given no aging effects in behavioral data: Tyler, Shafto, Randall, Wright, Marslen-Wilson, & Stamatakis, 2010). In contrast, the Normal condition would have been the only one in which older participants did not have their guard up, leading them to produce speech more naturally – and therefore including a greater rate of speech errors in older relative to young speakers.

How can we be sure that the errors did not in fact reflect age-group differences in reading ability or strategy? For example, one study silent reading of single sentences found that older adults skipped words more often than young adults (a "risky readers" strategy; Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006; but see Choi, Lowder, Ferreira, Swaab, & Henderson, 2017). Such a strategy could elicit more speech errors, perhaps particularly with function word targets, which are skipped more often than content words (O'Regan, 1979; Saint-Aubin & Klein, 2001). However, a recent study of aging effects on silent paragraph reading reported longer fixations for older than younger adults, and equivalent skipping rates across age groups (Whitford & Titone, 2017). If this is correct, aging should not have increased errors – it might have been expected to result in fewer errors relative to young adults. An alternative strategic difference across older and younger participants is that older participants might have relied more heavily on predictive processing in reading aloud (e.g., see Kliegl, Grabner, Rolfs, & Engbert, 2004, perhaps reflecting the need to maintain a larger eye voice span so as to provide more time for planning speech; Salthouse et al., 1984). Because function words tend to be more predictable than content words (Bell, Brenier, Gregory, Girand, & Jurafsky, 2009), this strategy might have older speakers to produce more function word substitution errors. Consistent with such a possibility, there was a significant difference between young and older participants in the rate of function word substitution errors (see Figure 4). However, aging does not consistently increase predictability effects in reading (Federmeier, Kutas, & Schul, 2010; Moers, Meyer, & Janse, 2017; Rayner et al., 2006; Whitford & Titone, 2017). Moreover, our exploratory analyses of errors did not provide any evidence that aging effects on function word substitution errors varied by condition. This result also argues against a strategic difference where older adults rely more heavily on semantic processing (so as to facilitate rapidly coordination of input and output processing). This strategy would be most accessible in the Normal condition, and would predict much stronger aging effects on function word substitution errors in that condition.

Finally, if age differences in function word substitution errors were driven by any strategy that increased skipping, one might have also expected older participants to completely omit

words relatively more than younger participants (but see Paulson, 2002). Instead, if anything there were trends in the opposite direction (see Figure 5; young adults omitted content words more often than older adults), perhaps implying a less careful approach to the task in the faster and younger participants. Given these results, and assuming an analogy between omission errors and skipping during paragraph reading (and that the error subtype analyses which we conducted on vocabulary matched groups would replicate with a larger group and statistical control), it seems less likely that between group differences in reading elicited aging-related decrease in accuracy in the present study. Instead, it is more likely that older participants expended considerable effort to avoid producing errors in the more difficult Nouns-Swapped and Exchange conditions (but not in the Normal condition); further work is clearly needed to determine the locus of aging effects on this type of speech error.

Aging Effects on Self-Correction

Our finding of lower self-correction rates in the present study suggests that the self-correction deficit identified in aging bilinguals in stopping intrusion errors in mid-utterance (Gollan & Goldrick, 2016), reflects general aging effects on speech production mechanisms rather than something specific to bilingual language switching. A question that arises is why, if older participants were especially focused on producing error-free speech in more difficult processing conditions, they nevertheless self-corrected their speech errors significantly less often than younger participants. One possibility is that there may be a distinction between planning and monitoring of speech wherein a more attentive and careful approach to planning does not necessarily also entail increased monitoring. Aging may impose more significant limitations on monitoring than on planning speech so that once a speech error was planned for production, older speakers were less effective at noticing and stopping to correct the error before continuing to read. On this view, the aging deficit in self-correction would also appear to be less amenable to compensatory processing or speakers' attempts to be more careful in their planning of speech for production (leading to differences between errors and self-correction variables in which conditions exhibited aging effects).

Why would self-correction be less amenable to strategic or compensatory processing in aging than the planning of speech itself (i.e., accuracy)? This apparent difference across measures is easiest to explain by assuming that monitoring (needed for self-correction) relies more heavily on cognitive mechanisms that decline in aging. Perhaps the most widely cited account of self-monitoring of speech is the Perceptual-Loop Theory, in which both planned and produced speech are monitored through language comprehension (Hartsuiker & Kolk, 2001; Levelt, 1983, 1989). The inner loop detects errors by comparing intended speech to comprehension of formulated word forms planned in inner speech, while the outer loop detects errors by comparing originally intended speech to comprehension of the produced utterance (in overtly produced speech). Though comprehension remains relatively intact in aging (see Introduction), monitoring is effectively a secondary task (with production as primary). Such dual-tasks are known to be impaired in aging (e.g., Kemper et al., 2010; 2014); this model therefore predicts an aging deficit in self-correction rate.

Another prominent account is the Conflict Detection Theory (Nozari, Dell, & Schwartz, 2011; Nozari & Novick, in press), in which monitoring of inner speech is triggered by

competition between response options with greater conflict initiating stronger monitoring. If conflict monitoring relies on executive control abilities – which are known to be impaired in aging (e.g., Mayr, Spieler, & Kliegl, 2001; Raz, 2000; West, 1996; but see Salthouse, 2005; Verhaeghen & Cerella, 2002; Verhaeghen, 2011) then this model too would predict that self-correction should decline in aging. However, according to this model, speech errors themselves (not only self-correction thereof) are driven by conflict, thus aging should similarly affect accuracy and self-correction. Indeed at the other end of the life span (i.e., in children), self-correction and production abilities appear to develop in close concert (Hanley, Cortis, Budd, & Nozari, 2016). A priori, it might also have seemed that if the speech production system is less effective at resolving conflict (leading to more errors), there should be more conflict, which should in turn trigger *greater* (not fewer) self-corrections. Thus, this model needs to make an additional assumption, which is that older adults cannot compensate for their speech production difficulties with increased monitoring – an assumption that seems a bit inconsistent with evidence of compensatory processing both in this study and in previous work (see above). Alternatively, this might reflect older adults emphasizing fluency in their productions. Given their already slower speech rate (relative to younger adults), older adults might simply have been less inclined to self-correct any errors they did detect, as corrections would render their speech even slower.

Finally, our finding that the Nouns-Swapped condition elicited the lowest rate of self-corrections for all participants – including young and old (and even marginally more so for older than for young) – is difficult to explain. Here it is likely important to consider that we did not place any requirements on speed or self-correction, and though speech was slowed in the Nouns-Swapped condition it was considerably more slowed in the Exchange condition (see Figure 1). Setting aside the Exchange condition as an outlier of sorts, we might conclude then that semantic processing – which was disrupted in the Nouns-Swapped condition – facilitates effective monitoring of speech for errors, a possibility that implies a high level of processing as a primary target of monitoring (i.e., meaning is being monitored, not just lexical and phonological form; Hartsuiker, Corley, & Martensen, 2005; Hartsuiker, Pickering, & De Jong, 2005; but see Slevc & Ferriera, 2006, and Postma, 2000. for arguments that speech is likely is monitored at multiple processing levels). Self-correction rates might be lower in older adults if they simply fail to remember exactly what they read before planning it for speech; maintaining only some vague information about the word, and eliciting no conflict to be detected. Consistent with this view, function word substitution errors exhibited strong aging effects on self-correction, and largely produced alternative grammatical constructions. As noted above, function words have relatively impoverished in semantic content, and therefore might be easier for older adults to forget – which in turn would make detection of a mismatch between written and spoken content impossible, and there would be no conflict to be detected during production.

Implications for Models of Speech Production and Cognitive Aging

The launch-point for the present study was the finding of an aging deficit in mid-utterance self-correction of errors produced during reading aloud of bilingual language switches. The present results indicate a broader usefulness of the read aloud task for studying group differences in production of speech errors (in bilinguals and monolinguals alike), and joins

other work in the field demonstrating the importance of controlling for between group differences in vocabulary and processing speed when looking to identify aging effects (though exactly how this should be done is far from simple; Bowles & Salthouse, 2008; Gray & Hills, 2014; Verhaegen, 2003; West, Crook, & Barron, 1992). With aging effects on vocabulary controlled, and without compensatory strategies triggered by unusually difficult speaking tasks, the reading aloud task revealed aging deficits – especially in the Normal condition – that are unique for their appearance in a simple naturalistic task that elicited largely error free connected speech, and construction of relatively simple sentences strung together to produce a meaningful narrative.

To what extent can the effects we observed be explained by commonly discussed cognitive mechanisms possibly underlying aging effects on speech production? Aging-related declines in accuracy might reflect a process internal to lexical access, e.g., disruptions to the flow of activation from concepts to lexical and especially to phonological representations, i.e., the *Transmission Deficit Hypothesis* (Burke et al., 1991), while reduced self-corrections might be more affected by declines in executive function abilities. However, the finding of a possibly greatest aging effect on function word substitution errors is very much unexpected on a transmission deficits account. Function words are very high frequency words, which makes them highly accessible – perhaps even retrieved relatively automatically (Bock & Levelt, 1994; Garrett, 1975; 1982; but see Ayora, Janssen, Dell 'Acqua, & Alario, 2009). This presumably, should also have sheltered function words considerably from competition for selection relative to content words. Another mechanism commonly invoked to explain aging effects is the Inhibitory Deficit Hypothesis, in which aging deficits reflect difficulty with managing competition within the language system (and nonlinguistic processing as well; Hasher, Zacks, & May, 1999; Zacks & Hasher, 1994). This account runs into similar problems given the relative automaticity of function word retrieval. Additional work will be needed to determine the source of this type of speech error, whether it is more common in aging when speech is elicited in other tasks, or perhaps in reading aloud of paragraphs designed explicitly to elicit such errors.

A broader possible implication of the current results is that more than a single cognitive mechanism may be needed to explain aging effects on accuracy versus self-correction – in turn implying at least partial dissociation in models of speech production between how much monitoring of speech versus speech production itself, rely on common cognitive mechanisms (as assumed by the Conflict Detection Theory). Thus, though executive control may be needed to resolve competition between alternative candidate lexical representations in speech, such control may be less critical for planning than for monitoring speech. If production and monitoring were similarly reliant on executive functions, and if aging-related changes in speech production primarily reflected decline this cognitive mechanism, then errors and self-corrections should have exhibited parallel aging effects. It will be important to confirm that any differences found between measures and across conditions in aging effects in fact reflect differences in speech planning versus monitoring – as opposed to differences in approach, compensatory strategies, or effects of higher vocabulary knowledge. Nonetheless, the apparent differences (comparing Figures 2 and 3) would be easiest to explain by assuming at least some dissociation between cognitive mechanisms underlying planning and monitoring of speech, in models of speech production.

A final, key question to consider is what are the implications of our finding that – other than a slower pace of speech – aging deficits in the present study would have been missed entirely had we not controlled for the concomitant aging-related increase in vocabulary knowledge. For obvious reasons, it is standard practice in individual differences research to control for the influence of experience-based confounding variables known to affect cognition (e.g., socioeconomic status). While vocabulary knowledge is necessarily based on experience with a large variety of lexical items, the ability to retain and utilize this information might reflect other individual differences in cognitive processes. Young adults with unexpectedly large vocabularies might therefore form an artificially high baseline – a young adult advantage rather than an older adult disadvantage. A key question for future work is to better understand the cognitive processes through which vocabulary knowledge exerts such powerful effects on language processing; this will clarify the extent to which the aging effects observed here reflect processes internal or external to the language production system.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Appendix

An example paragraph and its variants presented between subjects across different conditions.

Normal

With the light of the two oil lamps, the evil went away. The animals crossed and we continued walking. More or less a hundred meters ahead, we heard the cry of a sad woman. We stopped and then continued walking and she cried again in an even sadder voice. We asked the two gentlemen not to leave us in the darkness because we were very scared. They were a great help with their oil lamps. That time, we had intended to arrive to our village but it was not possible with all that happened to us on the road. We had to stay to sleep in another village. It is because of that reason that I say that the legend of the Weeping Woman is true. One can hear the voice of a young woman crying sadly.

Nouns-Swapped

With the lamps of the two oil animals, the light went away. The evil crossed and we continued walking. More or less a hundred cries ahead, we heard the voice of a sad meter. We stopped and then continued walking and she cried again in an even sadder woman. We asked the two lamps not to leave us in the gentlemen because we were very scared. They were a great darkness with their oil help. That village, we had intended to arrive to our road but it was not possible with all that happened to us on the village. We had to stay to sleep in another time. It is because of that woman that I say that the voice of the Weeping Reason is true. Women can hear the legend of a young one crying sadly.

Exchange

The with of light two the lamps oil, evil the away went. Animals the and crossed continued we walking. Or more a less meters hundred we, ahead the heard of cry sad a woman. Stopped we then and walking continued she and again cried an in sadder even voice. Asked we two the not gentlemen leave to in us darkness the we because very were scared. Were they great a with help oil their lamps. Time that, had we to intended to arrive village our it but not was with possible that all to happened on us road the. Had we stay to sleep to another in village. Is it of because reason that I that say the that of legend weeping the is woman true. Can one the hear of voice young a crying woman sadly.

References

Abrams L, Farrell MT, & Margolin SJ (2010). Older adults' detection of misspellings during reading. The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 65, 680–683.

Arbuckle TY, Nohara-LeClaire M, & Pushkar D (2000). Effect of off-target verbosity on communication efficiency in a referential communication task. Psychology and Aging, 15, 65–77. [PubMed: 10755290]

Ayora P, Janssen N, Dell'Acqua R, & Alario FX (2009). Attentional requirements for the selection of words from different grammatical categories. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35, 1344–1351.

Baltes PB (1997). On the incomplete architecture of human ontogeny: Selection, optimization, and compensation as foundation of developmental theory. American Psychologist, 52, 366–380 [PubMed: 9109347]

Barr DJ, Levy R, Scheepers C, Tily HJ (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68, 255–278.

Bates D, Kliegl R, Vasishth S, & Baayen H (2015). Parsimonious mixed models. arXiv preprint arXiv: 1506.04967.

Bates D, Maechler M, Bolker B, & Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67, 1–48.

Bell A, Brenier JM, Gregory M, Girand C, & Jurafsky D (2009). Predictability effects on durations of content and function words in conversational English. Journal of Memory and Language, 60, 92–111.

Bock K, & Levelt W (1994). Language production: Grammatical encoding In Gernsbacher MA (Ed.) Handbook of psycholinguistics, pp. 945–984. San Diego: Academic Press.

Burke DM, Locantore JK, Austin AA, & Chae B (2004). Cherry pit primes Brad Pitt: Homophone priming effects on young and older adults' production of proper names. Psychological Science, 15, 164–170. [PubMed: 15016287]

Burke DM, MacKay DG, Worthley JS, & Wade E (1991). On the tip of the tongue: What causes word finding failures in young and older adults. Journal of Memory and Language, 30, 542–579.

Burke DM, MacKay DG, James LE (2000). Theoretical approaches to language and aging In: Perfect TJ, & Maylor EA,(Eds.). Models of cognitive aging. Oxford: Oxford University Press p. 204–237.

Burke DM, & Shafto MA (2008). Language and aging In Salthouse TA (Ed.), The handbook of aging and cognition (pp. 373–443). New York, NY: Psychology Press.

Bowles RP, & Salthouse TA (2008). Vocabulary test format and differential relations to age. Psychology and Aging, 23, 366–376. [PubMed: 18573010]

Braver TS, & West RF (2008). Working memory, executive control, and aging In: Craik FIM, & Salthouse TA, (Ed.). The handbook of aging and cognition. New York: Psychology Press p. 311–372.

Carpenter PA, Miyake A, & Just MA (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging In Gernsbacher M (Ed.), Handbook of psycholinguistics. San Diego, CA: Academic Press.

Caplan D, & Waters G (2005). The relationship between age, processing speed, working memory capacity, and language comprehension. Memory, 13, 403–413. [PubMed: 15952262]

Choi W, Lowder MW, Ferreira F, Swaab TY, & Henderson JM (2017). Effects of word predictability and preview lexicality on eye movements during reading: A comparison between young and older adults. Psychology and Aging, 32, 232–242. [PubMed: 28333501]

Connelly SL, Hasher L, & Zacks RT (1991). Age and reading: The impact of distraction. Psychology & Aging, 6, 533–541. [PubMed: 1777141]

Cooper PV (1990). Discourse production and normal aging: Performance on oral picture description tasks. Journal of Gerontology: Psychological Sciences, 45, 210–214.

Dahlgren DJ (1998). Impact of knowledge and age on tip-of-the-tongue rates. Experimental Aging Research, 24, 139–153. [PubMed: 9555567]

Davidson DJ, Zacks RT, & Ferreira F (2003). Age preservation of the syntactic processor in production. Journal of Psycholinguistic Research, 32, 541–566. [PubMed: 14564992]

Dixon P (2008). Models of accuracy in repeated-measures designs. Journal of Memory and Language, 59, 447–456.

Dodson CS (2017). Aging and Memory In: Wixted JT (Ed..), Cognitive Psychology of Memory, Vol. 2 of Learning and Memory: A Comprehensive Reference, 2nd edition, Byrne JH (Ed.). pp. 403–421. Oxford: Academic Press.

Evrard M (2002). Ageing and lexical access to common and proper names in picture naming. Brain and Language, 81, 174–179. [PubMed: 12081390]

Federmeier KD, Kutas M, & Schul R (2010). Age-related and individual differences in the use of prediction during language comprehension. Brain and Language, 115, 149–161. [PubMed: 20728207]

Ferreira F (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. Journal of Memory and Language, 30, 210–233.

Garnham A, Shillcock RC, Brown GDA, Mill AID, & Cutler A (1981). Slips of the tongue in the London-Lund corpus of spontaneous conversation. Linguistics, 19, 805–817.

Garrett MF (1975). The analysis of sentence production In Bower GH (Ed.) The psychology of learning and motivation (vol. 9, pp. 133–177). New York: Academic Press.

Garrett MF (1982). Production of speech: Observations from normal and pathological language use In Ellis A (Ed.), Normality and pathology in cognitive functions (pp. 19–76). London, England: Academic Press.

Gollan TH, & Goldrick M (2018). A Switch is Not a Switch: Syntactically-Driven Bilingual Language Control. Journal of Experimental Psychology: Learning, Memory, & Cognition, 44, 143–156.

Gollan TH, & Goldrick M (2016). Grammatical constraints on language switching: Language control is not just executive control. Journal of Memory and Language, 90, 177–199. [PubMed: 27667899]

Gollan TH, Schotter ER, Gomez J, Murillo M, & Rayner K (2014). Multiple Levels of Bilingual Language Control: Evidence From Language Intrusions in Reading Aloud. Psychological Science, 25, 585–595. [PubMed: 24367061]

Gollan TH, Weissberger G, Runnqvist E, Montoya RI, & Cera CM (2012). Self-ratings of spoken language dominance: A multi-lingual naming test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. Bilingualism: Language and Cognition, 15, 594–615.

Goral M, Spiro A, Albert ML, Obler LK, & Connor LT (2007). Change in lexical retrieval skills in adulthood. The Mental Lexicon, 2, 215–240.

Gray WD & Hills T (2014). Does cognition deteriorate with age or is it enhanced by experience? Topics in Cognitive Science, 6(1), 2–4. [PubMed: 24425680]

Griffin ZM, & Spieler DH (2006). Observing the what and when of language production for different age groups by monitoring speakers eye movements. Brain and Language, 99, 272–288. [PubMed: 16290041]

Grober E, Sliwinsk M, & Korey SR (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. Journal of Clinical and Experimental Neuropsychology, 13, 933–949. [PubMed: 1779032]

Hanley RJ, Cortis C, Budd MJ, Nozari N (2016). Did I say dog or cat? A study of semantic error detection and correction in children. Journal of Experimental Child Psychology, 142, 36–47. [PubMed: 26476973]

Hartshorne JK & Germine LT (2015). When does cognitive functioning peak? the asynchronous rise and fall of different cognitive abilities across the life span. Psychological Science, 26, 433–443. [PubMed: 25770099]

Hartsuiker RJ, Corley M, & Martensen H (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al. 1975. Journal of Memory & Language, 52, 58–70.

Hartsuiker RJ, Pickering MJ, & De Jong NH (2005). Semantic and phonological context effects in speech error repair. Journal of Experimental Psychology: Learning, Memory, & Cognition, 31, 921–932.

Hasher H, Zacks RT, & May CP (1999). Inhibitory control, circadian arousal, and age In Gopher D & Koriat A (Eds.), Attention and performance XVII. Cognitive regulation and performance: Interaction of theory and application (pp. 653–675). Cambridge, MA: MIT Press.

Ivanova I, Ferreira VS, & Gollan TH (2017). Form overrides meaning when bilinguals monitor for errors. Journal of Memory and Language, 94, 75–102. [PubMed: 28649169]

Ivanova I, Salmon DP, & Gollan TH, (2013). The Multilingual Naming Test in Alzheimer's disease: Clues to the origin of naming impairments. The Journal of the International Neuropsychological Society, 19, 272–283. [PubMed: 23298442]

Jaeger TF (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of Memory and Language, 59, 434–446. [PubMed: 19884961]

James LE, Burke DM, Austin A, & Hulme E (1998). Production and perception of verbosity in younger and older adults. Psychology and Aging, 13, 355–367. [PubMed: 9793112]

James L (2006). Specific effects of aging on proper name retrieval: Now you see them, now you don't. Journal of Gerontology: Psychological Sciences, 61B, 180–183.

Kavé G, & Goral M (2017). Do age-related word retrieval difficulties appear (or disappear) in connected speech? Aging, Neuropsychology, and Cognition, 24, 508–527.

Kavé G, Knafo A, & Gilboa A (2010). The rise and fall of word retrieval across the lifespan. Psychology and Aging, 25, 719–724. [PubMed: 20853975]

Kavé G, & Yafé R (2014). Performance of younger and older adults on tests of word knowledge and word retrieval: Independence or interdependence of skills? American Journal of Speech-Language Pathology, 23, 36–45. [PubMed: 23831710]

Kemper S (1987). Life-span changes in syntactic complexity. Journal Gerontology Series B: Psychological Sciences and Social Sciences, 42, 323–328.

Kemper S, Bontempo D, Herman R, McKedy W, Schmalzried R, Tagliaferri B, & Kieweg D (2014). Tracking reading: Dual task costs of oral reading for young versus older adults. Journal of Psycholinguistic Research, 43, 59–80. [PubMed: 23463405]

Kemper S, Schmalzried R Hoffman L, Herman R (2010). Aging and the Vulnerability of Speech to Dual Task Demands. Psychology and Aging, 25, 949–962. [PubMed: 21186917]

Kemper S (1992). Language and aging In Craik FIM & Salthouse TA (Eds.), The handbook of aging and cognition (pp. 213–270). Hillsdale, NJ: Erlbaum.

Kemper S, Thompson M, & Marquis J (2001). Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content. Psychology and Aging, 16, 600–614. [PubMed: 11766915]

Kemper S, Herman R, & Lian C (2003a). Age differences in sentence production. Journals of Gerontology: Psychological Science, 58, 260–269.

Kemper S, Herman RE, & Lian CHT (2003b). The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech or noise. Psychology and Aging, 18, 181–192. [PubMed: 12825768]

Kemper S, Herman RE, & Liu CJ (2004). Sentence production by younger and older adults in controlled contexts. Journals of Gerontology: Psychological Sciences, 58B, P220–P224.

Keuleers E, Stevens M, Mandera P, & Brysbaert M (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. Quarterly Journal of Experimental Psychology, 68, 1665–1692.

Kliegl R, Grabner E, Rolfs M, & Engbert R (2004). Length, frequency, and predictability effects of words on eye movements in reading. European Journal of Cognitive Psychology, 16, 262–284.

Kreutzer JS, DeLuca J, & Caplan B (Eds.). (2011). American National Adult Reading Test (ANART) In Encyclopedia of clinical neuropsychology. New York, NY: Springer 10.1007/978-0-387-79948-3

Laver GD & Burke DM (1993). Why do semantic priming effects increase in old age? A meta-analysis. Psychology and Aging, 8, 34–43. [PubMed: 8461113]

Levelt WJM (1983). Monitoring and self-repair in speech. Cognition, 14, 41–104. [PubMed: 6685011]

Levelt WJM (1989). Speaking: From intention to articulation. Cambridge, MA: MIT Press.

Li C, & Gollan TH (in press). Cognates interfere with language selection but enhance monitoring in connected speech. Memory & Cognition.

MacKay DG, & James LE (2004). Sequencing, speech production, and selective effects of aging on phonological and morphological speech errors. Psychology and Aging, 19, 93–107. [PubMed: 15065934]

Mattis S (1988). Dementia rating scale: Professional manual. Odessa, FL: Psychological Assessment Resources.

Maylor E (1990). Recognizing and naming faces: Aging, memory retrieval, and the tip of the tongue state. Journal of Gerontology: Psychological Sciences, 45, 215–226.

Mayr U (2001). Age differences in the selection of mental sets: The role of inhibition, stimulus ambiguity, and response-set overlap. Psychology and Aging, 16, 96–109. [PubMed: 11302371]

Mayr U, Spieler DH, & Kliegl R (2001). Aging and executive control. New York: Routledge.

McKhann G, Drachman D, Folstein M, Katzman R, Price D, & Stadlan EM (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group. Neurology, 34, 939–944. [PubMed: 6610841]

McNamara P, Obler LK, Au R, Durso R and Albert M (1992) Speech monitoring skills in Alzheimer's disease, Parkinson's disease, and normal aging. Brain and Language, 42, 38–51. [PubMed: 1547468]

Moers C, Meyer AS, & Janse E (2017). Effects of word frequency and transitional probability on word reading durations of younger and older speakers. Language and Speech, 60, 289–317. [PubMed: 28697699]

Morris JCA, Heyman RC, Mohs et al., (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease, Neurology, 39, 1159–1165. [PubMed: 2771064]

Nicholas M, Obler LK, Albert ML, & Goodglass H (1985). Lexical retrieval in healthy aging. Cortex, 21, 595–606. [PubMed: 4092486]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Nozari N & Novick J (in press). Monitoring and control in language production. Current Directions in Psychological Science.

O'Regan JK (1979). Eye guidance in reading: Evidence for the linguistic control hypothesis. Perception & Psychophysics, 25, 501–509. [PubMed: 492916]

Park DC, Lautenschlager G, Hedden T, Davidson N, Smith AD, & Smith P (2002). Models of visuospatial and verbal memory across the adult life span. Psychology and Aging, 17, 299–320. [PubMed: 12061414]

Paulson EJ (2002). Are oral reading word omissions and substitutions caused by careless eye movements? Reading Psychology, 23, 45–66.

Postma A (2000). Detection of errors during speech production: a review of speech monitoring models. Cognition, 77, 97–131. [PubMed: 10986364]

Rabaglia CD & Salthouse TA (2011) Natural and constrained language production as a function of age and cognitive abilities, Language and Cognitive Processes, 26, 1505–1531.

Ramscar M, Hendrix P, Shaoul C, Milin P, & Baayen H (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. Topics in Cognitive Science, 6, 5–42. [PubMed: 24421073]

Rastle KG, & Burke DM (1996). Priming the tip of the tongue: Effects of prior processing on word retrieval in young and older adults. Journal of Memory and Language, 35, 585–605.

Rayner K, Reichle ED, Stroud MJ, Williams CC, & Pollatsek A (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. Psychology and Aging, 21, 448–465. [PubMed: 16953709]

Raz N (2000). Aging of the brain and its impact on cognitive performance: integration of structural and functional findings In: Craik FIM & Salthouse TA, (Eds.). The handbook of aging and cognition, pp.1–90. Mahwah: Erlbaum

Salthouse TA (1984). Effects of age and skill in typing. Journal of Experimental Psychology: General, 113, 345–371. [PubMed: 6237168]

Salthouse TA (1996). The processing speed theory of adult age differences in cognition. Psychological Review, 103, 403–428. [PubMed: 8759042]

Salthouse TA (2005). Relations between cognitive abilities and measures of executive functioning. Neuropsychology, 19 (4), 532–545. [PubMed: 16060828]

Salthouse TA, (2010). Is flanker-based inhibition related to age? Identifying specific influences of individual differences on neurocognitive variables. Brain & Cognition, 73, 51–61. [PubMed: 20303636]

Schotter ER, Li C, & Gollan TH (submitted). Eye movements reveal monitoring of uncorrected intrusion errors in reading aloud: evidence from Chinese-English bilinguals

Shafto MA (2015). Proofreading in young and older adults: The effect of error category and comprehension difficulty. International Journal of Environmental Research and Public Health, 12, 14445–14460. [PubMed: 26580634]

Shipley WC (1940). A self-administered scale for measuring intellectual impairment and deterioration. Journal of Psychology, 9, 371–377.

Smiler A, Gagne DD, & Stine-Morrow EAL (2003). Aging, memory load, and resource allocation during reading. Psychology and Aging, 18, 203–209. [PubMed: 12825770]

Saint-Aubin J, & Klein RM (2001). Influence of parafoveal processing on the missing-letter effect. Journal of Experimental Psychology: Human Perception and Performance, 27, 318–334. [PubMed: 11318050]

Stanovich KE, West RF, & Harrison MR (1995). Knowledge growth and maintenance across the life span: The role of print exposure. Developmental Psychology, 31, 811–826.

Stine-Morrow EAL, Miller LMS, & Herzog H (2006). Aging and self-regulated language processing. Psychological Bulletin, 132, 560–582. [PubMed: 16822167]

Stine-Morrow EAL, Miller LMS, Gagne DD, & Hertzog, (2008). Self-regulated reading in adulthood. Psychology and Aging, 23, 131–153. [PubMed: 18361662]

Stine-Morrow EAL, Gagne DD, Morrow DG, & DeWall BH (2004). Age differences in rereading. Memory and Cognition, 32, 696–710. [PubMed: 15552347]

Tun PA, Wingfield A, & Stine EAL (1991). Speech-processing capacity in younger and older adults: A dual-task study. Psychology and Aging, 6, 3–9. [PubMed: 2029365]

Trunk DL, & Abrams L (2009). Do younger and older adults' communicative goals influence off topic speech in autobiographical narratives? Psychology and Aging, 24, 324–337. [PubMed: 19485651]

Tyler LK, Shafto MA, Randall B, Wright P, Marslen-Wilson WD, & Stamatakis EA (2010). Preserving syntactic processing across the adult life span: the modulation of the frontotemporal language system in the context of age-related atrophy. Cerebral Cortex, 20, 352–364. [PubMed: 19505991]

Verhaeghen P (2003). Aging and vocabulary scores: A meta-analysis. Psychology and Aging, 18, 332–339. [PubMed: 12825780]

Verhaeghen P (2011). Aging and executive control: Reports of a demise greatly exaggerated. Current Directions in Psychological Sciences, 20, 174–180.

Verhaghen P, & Cerella J (2002). Aging, executive control, and attention: a review of meta-analyses. Neuroscience and Biobehavioral Reviews, 26, 849–857. [PubMed: 12470697]

Vousden JI & Maylor EA (2006) Speech errors across the lifespan. Language and Cognitive Processes, 21, 48–77.

Waters GS, & Caplan D (2004). Verbal working memory and on-line syntactic processing: Evidence from self-paced listening. Quarterly Journal of Experimental Psychology, 57A, 129–163.

Wechsler D (1981). WAIS-R : Wechsler adult intelligence scale-revised. New York, N.Y., Psychological Corporation.

Weintraub S, Besser L, Dodge HH, Teylan M, Ferris S, Goldstein FC, … Morris JC (2018). Version 3 of the Alzheimer Disease Centers' Neuropsychological Test Battery in the Uniform Data Set (UDS). Alzheimer Disease and Associated Disorders, 32, 10–17. 10.1097/WAD. 0000000000000223 [PubMed: 29240561]

West R (1996). An application of prefrontal cortex function theory to cognitive aging. Psychological Bulletin, 120, 272–292. [PubMed: 8831298]

West RL, Crook TH, & Barron KL (1992). Everyday memory performance across the life span: Effects of age and noncognitive individual differences. Psychology and Aging, 7, 72–82. [PubMed: 1558708]

Whitford V, & Titone D (2017). The effects of word frequency and word predictability during first-and second-language paragraph reading in bilingual older and younger adults. Psychology and Aging, 32, 158–177. [PubMed: 28287786]

Zacks RT, & Hasher H (1994). Directed ignoring: Inhibitory regulation of working memory In Dagenbach D & Carr TH (Eds.), Inhibitory processes in attention, memory, and language (pp. 241–264). San Diego, CA: Academic Press.
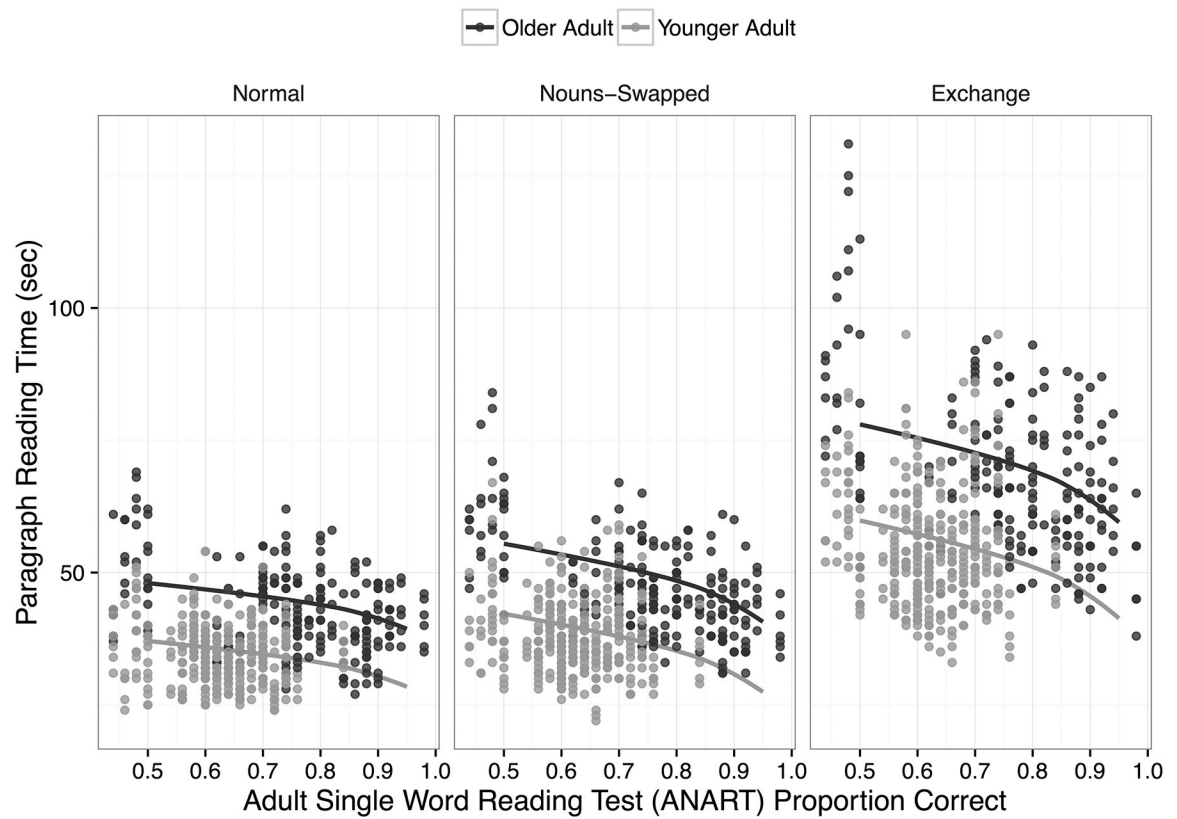
Figure 1.
Paragraph reading time (seconds) by ANART proportion correct, separated by paragraph type and subject age group, with fitted regression lines. Note that the regression uses a log-odds transformation of ANART score, yielding non-linear fits (see Results section for further discussion).
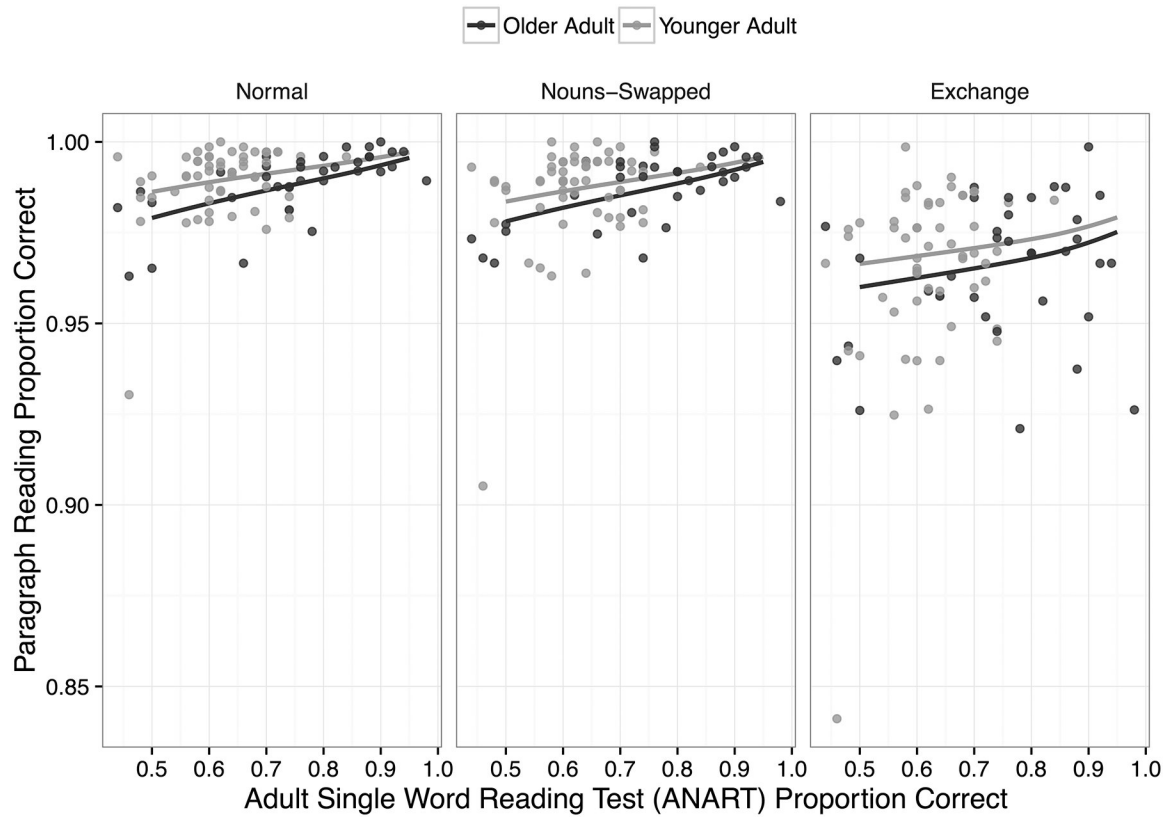
Figure 2.
Proportion of words correctly produced by ANART proportion correct, separated by paragraph type and subject age group, with fitted regression lines. Note this is a logistic regression, which yields non-linear curves when plotted as proportions (see Results section for further discussion).

Figure 3.
Proportion of errors self-corrected by ANART proportion correct, separated by paragraph type and subject age group, with fitted regression lines. Note this is a logistic regression, which yields non-linear curves when plotted as proportions (see Results section for further discussion).
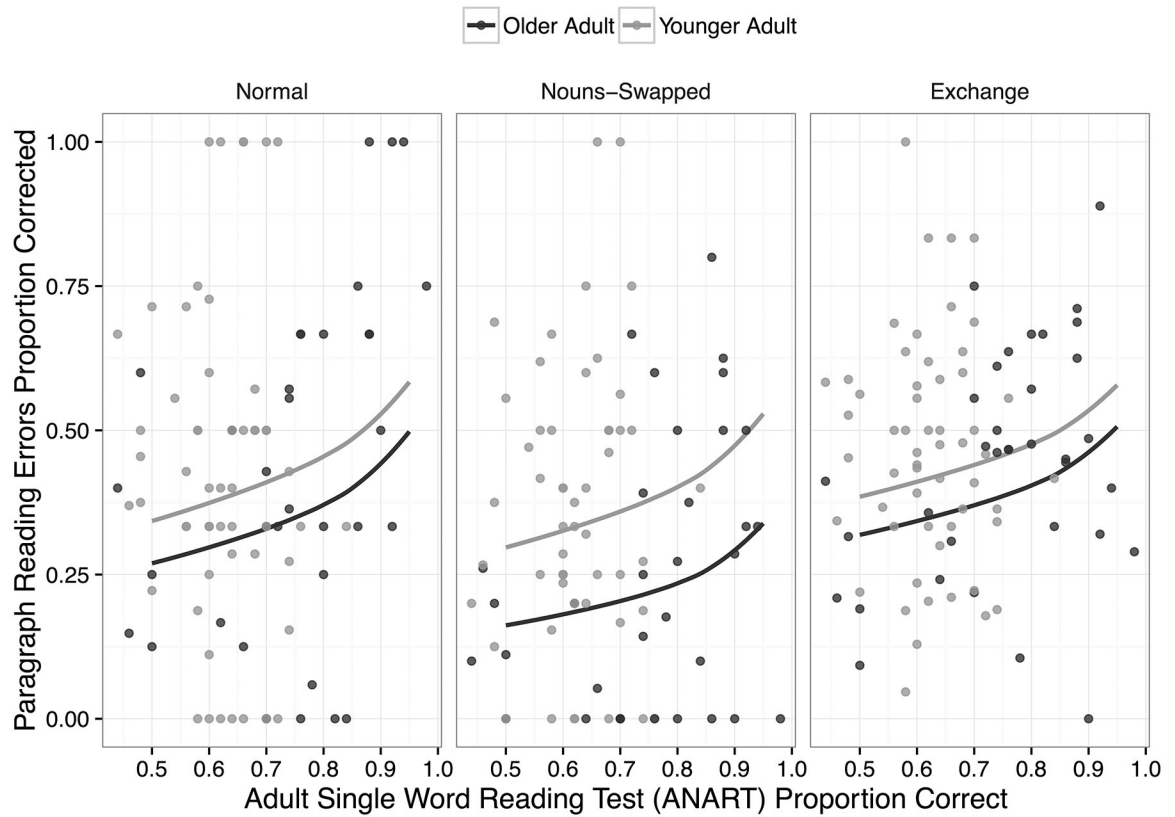
Figure 4.
Mean number of errors per participants within the matched groups of older and younger participants (for errors occurring at least 20 times in each participant group). Error bars show bootstrapped 95% confidence intervals for means (estimated using 1,000 re-samplings of the participant means with replacement).

Figure 5.
Proportion of words omitted by ANART proportion correct, separated by subject age group and word class, with fitted regression lines. Note this is a logistic regression, which yields non-linear curves when plotted as proportions (see Results section for further discussion).
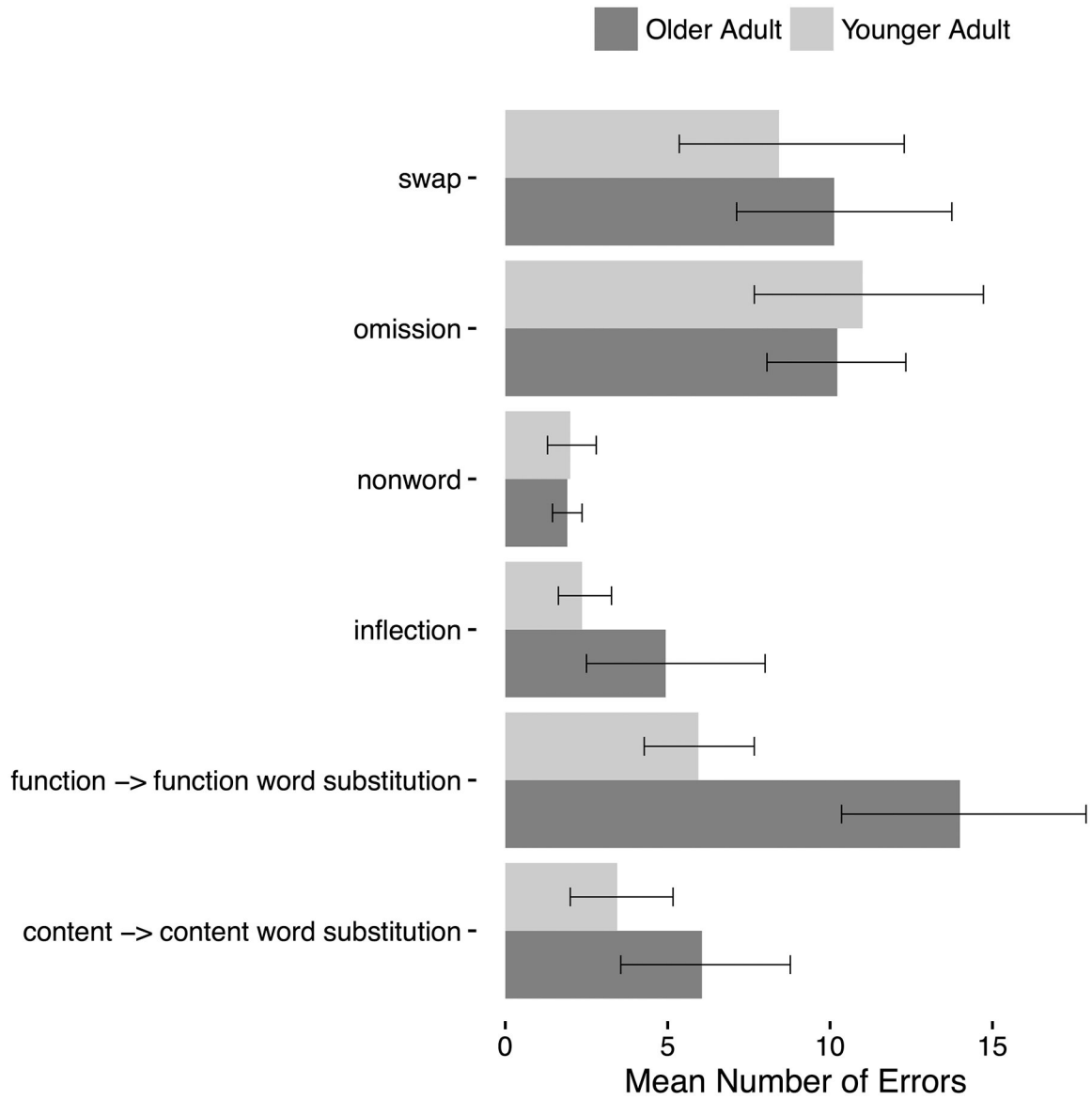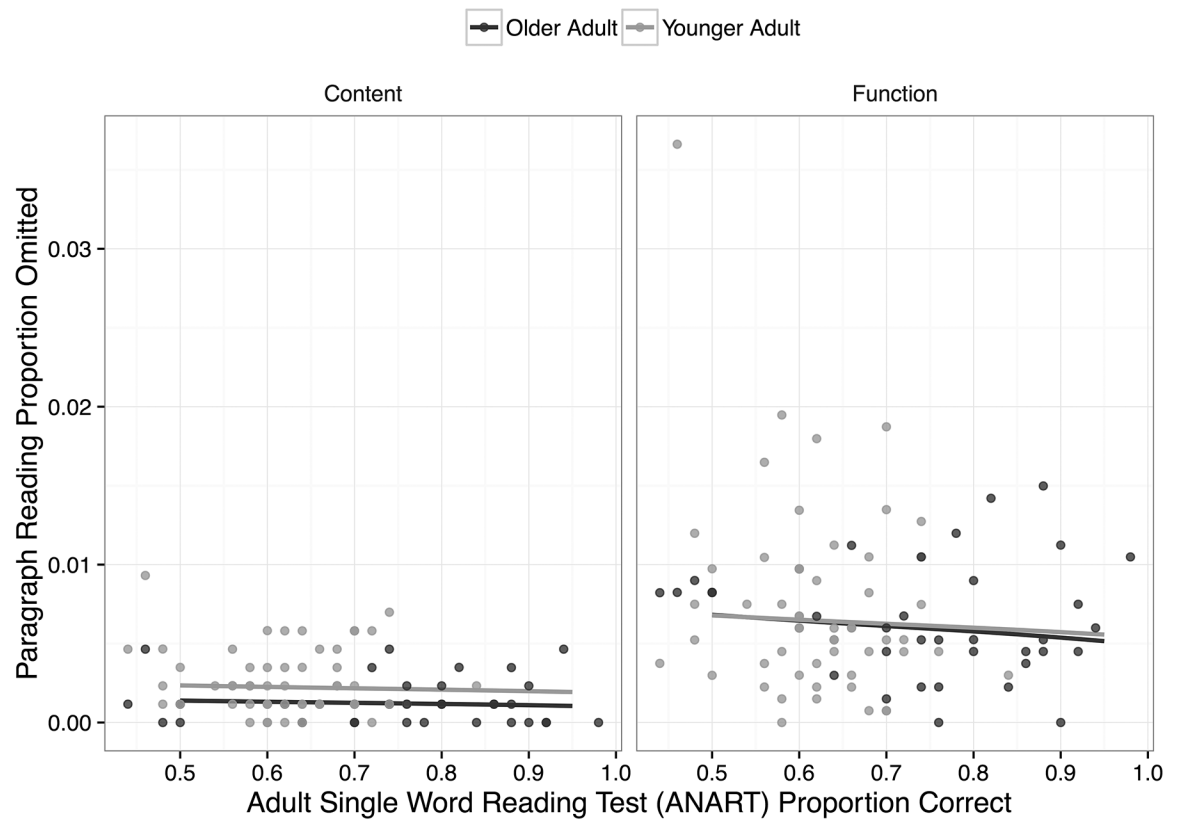
Table 1.

Means, standard deviations and group comparisons for demographic characteristics and performance on neuropsychological tests.

Participant Characteristics	All Participants					Vocabulary Matched				
	Older (n = 35)		Younger (n = 56)		P	Older (n = 18)		Younger (n = 18)		p
	M	SD	M	SD		M	SD	M	SD	
Age	72.3	3.2	20.4	2.5	<.001	71.6	2.9	21.1	3.3	<.001
Years of Education	17.5	2.4	13.9	1.4	<.001	17.3	1.8	14.2	1.5	<.001
Current % English use	99.8	0.9	99.5	1.6	.395	99.6	1.2	99.7	1.2	.785
% English use during childhood	99.4	2.0	98.4	3.4	.097	99.2	2.6	98.3	3.3	.402
Dementia Rating Scale (DRS)[a]	140.4	2.4	--	--	--	139.8	2.3	--	--	--
Mini Mental Status Exam (MMSE)[a]	29.1	1.1	--	--	--	28.9	1.1	--	--	--
ANART[b]	12.3	7.2	18.9	4.1	<.001	15.9	7.8	16.7	3.6	.703
Shipley Vocabulary	36.4	2.4	31.5	2.9	<.001	34.7	1.9	34.5	1.4	.766
English Multilingual Naming Test[c]	67.1	1.2	64.6	2.3	<.001	66.4	1.4	65.7	2.2	.208

[a] Younger participants did not complete these tests. Older participants completed the tests as part of their annual neuropsychological evaluation at the ADRC.

[b] Number of errors produced during standardized measure of single word reading aloud.

[c] Maximum possible picture-naming test score is 68.

Table 2.

Error types, counts, definitions, and examples for all error types that vocabulary-matched young and older speakers produced during the read aloud task (collapsed across condition). *Italicized text* = error word the speaker produced; [ ] = the written target word (that the speaker should have but did not produce).

Error Type	% of Total Errors that were Produced by Older Participants	Definition	Example(s)	Context Example
Word Substitution (n=547)	67.1	Speaker substituted a target word with a different whole word	at → in (n=345) county → country (n=171) alone → along (n=18) more → move (n=13)	This was odd because almost everyone *in* [at] the dance was poor.
Omission[a] (n=382)	48.2	Speaker skipped the target word	a →	Were they great [a] with help oil their lamps.
Swap[b] (n=280)	57.9	Speaker exchanged adjacent words	cups the → the cups	Since and *the cups* [cups the] and plate
Inflection (n=105)	75.2	Speaker produced a syntactic error, either adding, replacing, or omitting an affix	asked → ask sneaked → sneaking growling → growl	When his grandmother *ask* [asked] him to do his chores.
Repetition (n=91)	67.0	Speaker produced the same word twice or multiple adjacent words	to → to to (n=44) and a → and a and a (n=15)	If it was it seemed like his heart would jump *to to* [to] his throat
Insertion (n=50)	48.0	Speaker inserted a word that was not written in the paragraph	→ the	The rest of the people in *the* [ ] town lived well.
Nonsense word (n = 41)	51.2	Speaker produced a nonword	poverty → proverty malign → malyn	At that time there was a lady with twelve children who lived in extreme *proverty* [poverty].
Multi-error (n=14)	50.0	Speaker produced more than one error (e.g., in this example an insertion and repetition)	[ ] → the bottom of of	If it wasn't his hard sank to *the bottom of of* [ ] his feet from disappointment

[a]Mostly function word targets;

[b]Swap errors were produced almost exclusively in the Exchange condition

Table 3.

Mean percentage of errors self-corrected per participant within the matched older and younger groups (for errors that occurred at least 20 times in each participant group). Error bars show bootstrapped 95% confidence intervals for means (estimated using 1,000 re-samplings of the participant means with replacement).

Error Type	Proportion Corrected	
	Older Adult	Younger Adult
Swap	29% (19%, 41%)	36% (21%, 51%)
Omission	57% (46%, 68%)	61% (49%, 72%)
Nonword	36% (14%, 64%)	35% (14%, 60%)
Inflection	21% (3%, 40%)	20% (0%, 45%)
Function-> Function Word Substitution	25% (17%, 32%)	41% (32%, 51%)
Content-> Content Word Substitution	29% (18%, 40%)	33% (17%, 52%)