

Research Article

Lexical and Acoustic Characteristics of Young and Older Healthy Adults

Sunghye Cho,^a Naomi Nevler,^b Sanjana Shellikeri,^b Natalia Parjane,^b
David J. Irwin,^b Neville Ryant,^a Sharon Ash,^b Christopher Cieri,^a
Mark Liberman,^a and Murray Grossman^b

Purpose: This study examines the effect of age on language use with an automated analysis of digitized speech obtained from semistructured, narrative speech samples.

Method: We examined the Cookie Theft picture descriptions produced by 37 older and 76 young healthy participants. Using modern natural language processing and automatic speech recognition tools, we automatically annotated part-of-speech categories of all tokens, calculated the number of tense-inflected verbs, mean length of clause, and vocabulary diversity, and we rated nouns and verbs for five lexical features: word frequency, familiarity, concreteness, age of acquisition, and semantic ambiguity. We also segmented the speech signals into speech and silence and calculated acoustic features, such as total speech time, mean speech and pause segment durations, and pitch values.

Results: Older speakers produced significantly more fillers, pronouns, and verbs and fewer conjunctions, determiners, nouns, and prepositions than young participants. Older speakers' nouns and verbs were more familiar, more frequent (verbs only), and less ambiguous compared to those of young speakers. Older speakers produced shorter clauses with a lower vocabulary diversity than young participants. They also produced shorter speech segments and longer pauses with increased total speech time and total number of words. Lastly, we observed an interaction of age and sex in pitch ranges.

Conclusions: Our results suggest that older speakers' lexical content is less diverse, and these speakers produce shorter clauses than young participants in monologic, narrative speech. Our findings show that lexical and acoustic characteristics of semistructured speech samples can be examined with automated methods.

Not all people speak a language in the same way, even if they are native speakers of the same language. Language use is affected by factors such as an individual's age and sex. While these fundamental speaker characteristics have received considerable attention in the literature, the findings concerning the effects of these variables are mixed. In the case of age, for example, previous studies consistently observe that older speakers exhibit reduced fluency (Bortfeld et al., 2001; Heller & Dobbs, 1993; Kemper, 1992; Spieler & Griffin, 2006), increased pause duration (Bóna, 2014; Hartman & Danhauer, 1976),

and increased pause rate (Bóna, 2014; Martins & Andrade, 2011) when compared to young speakers. Some previous studies have found that vocabulary diversity in language use is maintained or even increases as people age (Horton et al., 2010; LaGrone & Spieler, 2006; Uttil, 2002; Verhaeghen, 2003), suggesting that older speakers use a greater variety of words compared to younger speakers. Moscoso del Prado Martín (2017) studied natural conversations and also found that vocabulary diversity increases throughout one's lifetime. On the other hand, Luo et al. (2019), who also examined language use in natural conversations, found no age effect on vocabulary diversity when interlocutors were not taken into account. When the interlocutors were considered, they found that older speakers used fewer unique words and more common words with children than young adult speakers did. Also, most previous studies have observed that older speakers speak more slowly than young speakers when reading isolated sentences (Bóna, 2014; Jacewicz

^aLinguistic Data Consortium, University of Pennsylvania, Philadelphia

^bPenn Frontotemporal Degeneration Center, University of Pennsylvania, Philadelphia

Correspondence to Sunghye Cho: csunghye@ldc.upenn.edu

Naomi Nevler and Sunghye Cho contributed equally to this article.

Editor-in-Chief: Frederick (Erick) Gallun

Editor: David A. Eddins

Received December 6, 2019

Revision received May 8, 2020

Accepted June 3, 2020

https://doi.org/10.1044/2020_JSLHR-19-00384

Publisher Note: This article is part of the Forum: Select Papers From the 8th Aging and Speech Communication Conference.

Disclosures: Murray Grossman participates in clinical trials sponsored by Alector, Eisai and Biogen that are unrelated to this study. He also receives research support from Biogen and Avid that is unrelated to this study and research support from National Institutes of Health. He receives financial support from Neurology for his work as an Associate Editor. Mark Liberman served on Scientific Advisory Board for Baidu Research, USA, and a co-editor of the Annual Review of Linguistics. All other authors have nothing to disclose.

et al., 2010; Spieler & Griffin, 2006) and during natural conversations (Horton et al., 2010; Jacewicz et al., 2010; Kemper et al., 2003). Yet, some studies, such as that of Cooper (1990), have not found a significant difference in the number of total words or total speech time as a function of age. In this study, we examined the effect of age and its interaction with language use by means of an automated analysis of digitized speech obtained from a semistructured speech sample.

The effect of sex on language has also been extensively studied, but again, previous studies have reported mixed results. For example, previous studies consistently have found that male speakers produce more filled pauses (e.g., *um* or *uh*) than female speakers (Bortfeld et al., 2001; Mulac et al., 1988; Shriberg, 1996). Similarly, Moscoso del Prado Martín (2017) has shown that male speakers' syntactic diversity decreases from 45 years of age onward with increased speech disfluency markers in contrast to female speakers, whose syntactic structures show increased diversity with fewer disfluency markers. In the case of the total number of words, Bortfeld et al. (2001) have found that male speakers do not necessarily produce more words than female speakers, whereas some studies (e.g., Mulac et al., 1986, 1988) have found that female speakers produce longer sentences than male speakers. Other studies (e.g., Dovidio et al., 1988; Mulac et al., 2000) report a higher total number of words and increased turn-taking during conversations between male compared to female speakers. Also, the interaction of age and sex on the total number of words has been studied with mixed findings. For example, Ardila and Rosselli (1996) found that the total number of words does not differ by sex in young and middle-age (16–50 years) groups but significantly differs in an older group (51–65 years), where older women produce significantly more words than their male counterparts (see also Mulac et al., 1986, 1988). The interaction of age and sex on pitch is also reported; previous studies (Ferrand, 2002; Linville, 1987; Mueller, 1997; Russell et al., 1995; Sataloff et al., 1997) have found that pitch, which is commonly measured by fundamental frequency (f_0), increases in older men but decreases in older women, suggesting that the sex difference in pitch is modulated by age. In contrast, a recent large-scale study by Nishio and Niimi (2008) did not find a significant correlation between age and pitch in male speakers but found a strong negative correlation in female speakers.

In the context of these inconclusive reports, it is not unreasonable to revisit the effect of age and sex on speakers' language production. Differences in results may come from different data types (reading vs. natural conversation), different definitions of similar concepts, and different methods of measuring lexical and acoustic features. In this study, we analyze 1-min picture description speech samples of the same picture, which allows all participants to express themselves in their own words with minimal constraints while controlling for potential confounding factors along with age and sex, such as topic or familiarity of the topic and interlocutors in conversations. This approach has been successfully applied in many previous studies (e.g., Ardila & Rosselli, 1996; Cooper, 1990; Cousins et al., 2018; Kavé et al., 2009;

Nevler et al., 2019, 2017), so we can assess the coherence and appropriateness of the content of the speech samples and compare our results with those from previous studies.

It is striking that very few studies have considered both lexical and acoustic features at the same time, which leaves a major gap in our understanding of the effect of age and sex on language use. This might partly be due to the fact that many previous studies rely on manual assessments of lexical and acoustic features, and the manual examination of both aspects of speech is extremely time-consuming. Because of the recent development of natural language processing (NLP) and automatic speech recognition tools, in this study, we were able to establish and illustrate objective, quick, replicable, and fully automated methods of analyzing the effects of age and sex on language use. This will allow us to clarify some of the previously observed mixed results using an objective and highly repeatable method. Thus, the goals of this study are to (a) examine and verify age- and sex-related properties of both lexical and acoustic characteristics of speech reported in previous studies with modern, fully automated methods; (2) further explore features of speech that have not previously been analyzed; and (3) establish normed linguistic data that are specific to picture description.

Method

Participants

We collected about 1-min-long picture descriptions from two age groups using the Cookie Theft picture from the Boston Diagnostic Aphasia Examination (Goodglass et al., 1983). A young age group consisting of 76 volunteers (18–22 years) were all undergraduate students, recruited at the University of Pennsylvania. This group volunteered to participate in a pilot study, where they performed three neuropsychological tests (F-letter fluency, judgment of line orientation, and symbol–digit substitution) and four different picture description tasks, including the Cookie Theft. We only examined the Cookie Theft picture descriptions in this report. The students received course credit for their participation in this study.

The other group consisted of 37 older adults, whose age ranged from 52 to 89 years at the time of recording. Most of these participants were caregivers of patients at the Frontotemporal Degeneration Center of the Hospital of the University of Pennsylvania. We used their Cookie Theft descriptions from the Boston Diagnostic Aphasia Examination to examine the effect of age on semistructured, narrative, natural speech samples. They contributed their speech samples on a voluntary basis. None of the young or older participants reported any hearing or speaking difficulties, and all were native speakers of English.

The two age groups were matched on sex ratio ($p = .11$) but significantly differed in education level ($p < .001$), since our young participants were all undergraduate students who had not yet completed their bachelor's degree, while our older participants were a highly educated group,

where most of them (29 out of 37) had received higher education. However, when considering the age of the participants, education levels were at ceiling. Also, we note that the variation in education level was small. For this reason, we did not covary for education level in statistical tests. All speakers participated in an informed consent procedure approved by the institutional review board at the University of Pennsylvania.

Six prospective participants were tested but were excluded from the analysis due to incomplete data. Two young speakers out of the six participants were excluded because of missing demographic information, either sex or age. One older control's sample and three young adults' samples were excluded due to poor audio quality, a low signal-to-noise ratio of the speech sample. The total number of participants after exclusions was 113. Age and education level did not significantly differ by sex within each group. Demographic characteristics of the participants are summarized in Table 1.

Text Data Processing and Measurements

We employed spaCy (Honnibal & Johnson, 2015; <https://spacy.io>), an NLP library in Python, to automatically tag part-of-speech (POS) information of all tokens in the speech samples. We used spaCy's basic language model ("en_core_web_sm") for English to process the data. There are two POS tagging schemes in spaCy: one is the Penn Treebank tag set (Marcus et al., 1993), and the other is the Universal POS tag set (Petrov et al., 2012), which was automatically mapped from the Penn Treebank tag set. We wrote a Python program to automatically tokenize the transcripts of speech samples and annotate the POS category (both the Universal set and the Penn Treebank set) of each token, along with its lemma.

We used the Universal set to report the general trend of POS production in the two age groups. We summed the token count of each POS category for each participant and calculated the number of tokens per 100 words for each POS category. The total number of words was also compared by group. We note that the interjection category spaCy tagged consisted of filler words, such as *um* or *uh*, over 90% of the time in our data.

The Penn Treebank tag set and word lemma were used to calculate three derived lexical measures: the number

of tense-inflected verbs, mean number of words per clause, and vocabulary diversity. The number of tense-inflected verbs per 100 words (number of modal auxiliaries per 100 words + number of present tense verbs per 100 words + number of past tense verbs per 100 words) approximated the number of clauses in a picture description. Conjoined verbs did not increase the number of clauses in our methods. The mean length of clause in words was measured as the number of all tokens / the number of tense-inflected verbs.

Vocabulary diversity or lexical diversity is a measure that shows how diverse one's vocabulary usage is, and it was previously measured with a type-token ratio (TTR) as the number of unique words / the number of total words. However, one problem of a simple TTR is that it is sensitive to the text length. Various methods have been proposed to cope with this challenge (e.g., Covington & McFall, 2010; Jarvis, 2002; McKee et al., 2000; Moscoso del Prado Martín, 2017; Tweedie & Baayen, 1998), and in this article, we reported the moving-average TTR (MATTR; Covington & McFall, 2010) to compare the group difference in lexical diversity. This method calculates TTR for a fixed length of window of tokens, moving one word at a time from the beginning to the end of a text, and it averages the measured TTRs of all windows. Since the shortest picture description in our data contained 47 words, we set a window of 45 words. We calculated TTR scores with the number of unique lemma counts within each window and averaged the TTRs of all windows from each picture description. We also tried the MATTR with the word order of each speech sample randomized as well as other measures, such as Guiraud's measure (Guiraud, 1954, as cited in Tweedie & Baayen, 1998), Summer's index ($= \log(\log(\text{type})) / \log(\log(\text{token}))$), and Uber index (Jarvis, 2002). All of them gave similar results, so we only reported the MATTR measure.

Even though the accuracy of spaCy's POS tagger is known to be very high (about 97% in spaCy's official release), we further validated the POS tags from spaCy with manual POS tags using a subset of our data. A professional linguist manually tagged POS categories of all words produced by six older speakers in our data and calculated the error rates of spaCy's POS tagger. The mean error rate was 5.4% (range: 2.7%–7.3%), with a standard deviation of 1.7%, which suggests that automatic POS tags were, on average, 94.6% correct. Since the accuracy of automatic POS tags was reasonably high, automatically generated POS tags were used for analysis without any modification.

We also rated five other lexical measures for noun and verb tokens using published norms. We used concreteness/abstractness ratings from the study of Brysbaert et al. (2014), which rated words' semantic concreteness/abstractness on a scale from 1 (*most abstract*) to 5 (*most concrete*). Additionally, semantic ambiguity (number of different meanings of a given word) from the study of Hoffman et al. (2013), word frequency (\log_{10} -scaled frequency per million words from the SUBTLEX_{US} corpus; Brysbaert & New, 2009), age of acquisition (the average age at which people acquire a given word; Brysbaert et al., 2018), and word

Table 1. Demographic characteristics of the participants.

Characteristic	Older (n = 37)	Young (n = 76)	p
Age			< .001
M (SD)	68.5 (8)	20.0 (0.9)	
Range	52–89.6	18–22	
Sex			.108
Female	23 (62.2%)	35 (46.1%)	
Male	14 (37.8%)	41 (53.9%)	
Education			< .001
M (SD)	15.9 (2.5)	13.5 (0.9)	
Range	12–20	11.5–15.5	

familiarity (how many people know a given word; Brysbaert et al., 2018) were rated for each noun and verb. After determining these measures, we calculated each individual's mean scores of the measures for nouns and verbs. The mean scores were used for group comparisons.

Acoustic Data Processing and Measurements

We used an in-house Gaussian mixture models–hidden Markov models based speech activity detector (SAD) developed at the University of Pennsylvania Linguistic Data Consortium to segment the speech samples into speech and silence segments. The minimum duration for speech was set at 250 ms, and the minimum duration for silent pauses was set at 150 ms. This method of speech segmentation relied purely on acoustic signal properties without the use of transcripts. We then validated the SAD output by visually reviewing the segments.

We pitch-tracked segments of continuous speech with the Praat (Boersma & Weenink, 2019) pitch-tracking algorithm and extracted the 10th to 90th percentile estimates of f_0 for each speech segment. The f_0 is the lowest (or longest) periodicity in a complex sound wave and is the physical measure that most closely represents the perceived pitch. Frequency limits for pitch tracking were set at 75–300 Hz. We also extracted the durations of speech and pause segments and the number of pauses. We converted f_0 estimates from hertz to semitones, using each subject's 10th percentile as the reference frequency in order to control for individual physiological differences in voice characteristics, such as height, weight, sex, and so forth. We calculated additional acoustic parameters: f_0 range, which is represented as the 90th percentile f_0 in the conversion just described; mean speech segment duration; total speech time, calculated by the summation of all speech segment durations in the sample; pause count; and pause rate, calculated as the number of pauses per minute over the total speech time. Detailed description and justification of SAD and pitch-tracking specifications as well as the methods for the acoustic measurement conversion and calculation have been published previously (Nevler et al., 2017).

Statistical Considerations

First, we performed Levene's test and visually plotted density and distribution of the data to see if the data met the requirements for parametric tests. Then, we performed Student's t tests to compare the two age groups (young vs. older) and reported t statistics and p values. When measures did not meet the requirements for parametric tests, we performed a Mann–Whitney U test and reported U and p values. To show the magnitude of the effect size, we also reported Cohen's d , assuming that a value of 0.2 is a small effect, 0.5 is a medium effect, and 0.8 is a large effect. We also built two-way analysis of variance models with an interaction term (Age Group \times Sex) to test the interaction of age group and sex on linguistic and acoustic measures. We confirmed that all variables that showed significant interactions

between age group and sex met the assumptions of analysis of variance by plotting the models' residuals.

Results

Lexical Measures

Word-Level Features

The results of all statistical analyses of the lexical measures are summarized in Table 2. Older participants produced significantly more pronouns, verbs, and so-called interjections, 90.23% of which were filler words such as *um* or *uh*, compared to young speakers (see Figure 1A). Also, older speakers produced significantly fewer prepositions, conjunctions, determiners, and nouns per 100 words compared to young speakers (see Figure 1B). Group variances, which are shown as standard deviation values in Table 2, were mostly similar for all POS categories except nouns, where older speakers showed a larger group variance than young speakers. The counts of adjectives and adverbs per 100 words did not differ by group (see Figure 1C).

Older participants produced more familiar nouns and verbs compared to young speakers (see Figures 2Aa and 2Ba). The group difference in word frequency was significant for verbs (see Figure 2Bb), but not for nouns (see Figure 2Ab). Semantic ambiguity for nouns (see Figure 2Ac) differed by group, and the same measure for verbs was marginally significant (see Figure 2Bc). Both concreteness and age of acquisition measures did not differ by group for nouns and verbs.

Global Lexical Features

The means and standard deviations of all global lexical measures are also summarized in Table 2. The number of tense-inflected verbs per 100 words significantly differed by group (see Figure 3A). Furthermore, lexical diversity significantly differed by group (see Figure 3B), in that older participants presented lower vocabulary diversity than young speakers. Young speakers showed a larger group variance than older speakers with several outliers (see Figure 3B), but the group difference was still significant after outliers were removed. The group difference in mean length of clause was also significant (see Figure 3C), with older speakers producing shorter clauses than young speakers. Lastly, the total number of words also significantly differed by group (see Figure 3D); the older group generally produced more words than the young group.

Acoustic Features

Table 3 summarizes the statistical results of the acoustic measures. The 90th f_0 percentile, which represents the f_0 range, was similar in the young and older groups (see Table 3). The younger speakers had, on average, longer speech and shorter pause segments (see Figures 4A and 4B). The number of pauses seems higher in the older speaker's samples (see Table 3); however, after controlling for the lengthier samples by calculating the pause rate over the duration of the speech sample, pause rate did not differ significantly

Table 2. Mean (standard deviation) of all lexical measures and group comparisons of young and older participants.

Variable	Older	Young	<i>t</i> or <i>U</i>	<i>p</i>	Cohen's <i>d</i>
Filler words	5.49 (2.56)	4.32 (2.42)	<i>t</i> = 2.33	.023	0.48
Pronoun	7.28 (2.41)	4.64 (2.24)	<i>t</i> = 5.57	< .001	1.14
Verb	22.52 (3.47)	20.48 (3.41)	<i>t</i> = 2.96	.004	0.6
Preposition	10.03 (1.97)	11.85 (2.89)	<i>U</i> = 902	.002	0.69
Conjunction	4.34 (1.84)	5.3 (1.95)	<i>t</i> = -2.55	.013	0.5
Determiner	14.27 (2.5)	15.7 (3.07)	<i>t</i> = -2.65	.009	0.49
Noun	20.36 (4.38)	21.59 (2.91)	<i>U</i> = 1083.5	.049	0.36
Adjective	5.61 (1.83)	5.62 (2.5)	<i>t</i> = 0.02	.98	0
Adverb	5.63 (2.12)	5.56 (2.67)	<i>t</i> = 0.37	.71	0.07
Familiarity (noun)	2.36 (0.03)	2.34 (0.03)	<i>t</i> = 2.73	.008	0.55
Familiarity (verb)	2.29 (0.05)	2.25 (0.05)	<i>t</i> = 4.1	< .001	0.8
Frequency (noun)	3.57 (0.17)	3.6 (0.15)	<i>t</i> = -0.9	.37	0.19
Frequency (verb)	4.54 (0.25)	4.38 (0.23)	<i>t</i> = 3.19	.002	0.66
Ambiguity (noun)	1.69 (0.06)	1.71 (0.06)	<i>t</i> = -2	.049	0.39
Ambiguity (verb)	2.11 (0.05)	2.13 (0.05)	<i>t</i> = -1.93	.057	0.37
Concreteness (noun)	4.49 (0.23)	4.43 (0.21)	<i>t</i> = 1.43	.16	0.3
Concreteness (verb)	2.6 (0.18)	2.65 (0.21)	<i>t</i> = -1.2	.23	0.23
AoA (noun)	4.42 (0.32)	4.53 (0.37)	<i>t</i> = -1.59	.12	0.3
AoA (verb)	4.7 (0.24)	4.75 (0.2)	<i>t</i> = -0.97	.34	0.2
Tense-inflected verb	12.39 (1.86)	11.06 (1.82)	<i>t</i> = 3.59	< .001	0.73
Vocabulary diversity	0.68 (0.00)	0.69 (0.01)	<i>U</i> = 968.5	.008	0.40
MLC	8.26 (1.32)	9.33 (1.85)	<i>t</i> = -3.52	< .001	0.63
Total words	176.57 (64.98)	136.39 (48.98)	<i>t</i> = 3.33	.002	0.73

Note. Part-of-speech counts and the number of tense-inflected verbs are per 100 words. AoA = age of acquisition; MLC = mean length of clause.

between the two age groups (see Figure 4C and Table 3). Total speech duration (see Figure 4D and Table 3) was longer in the older age group.

Interaction of Age Group and Sex

We examined the effect of age group and sex on the three variables that previous studies have explored: pitch, number of filler words, and total number of words. A linear regression model revealed a significant effect for the interaction of age group and sex on pitch range, $F(1, 109) = 4.37, p = .039$ (see Figure 5A), where the model predicts a gradual decrease in pitch differentiation between the sexes with increasing age. The number of filler words (interjections in spaCy) per 100 words significantly varied by age group, $F(1, 109) = 5.81, p = .018$, and sex, $F(1, 109) = 5.41, p = .022$, but the interaction of the two factors was not significant (see Figure 5B). The total number of words only differed by age group, $F(1, 109) = 13.32, p < .001$, but not by sex or the interaction of sex and age group ($p > .05$; see Figure 5C).

Correlation of Lexical and Acoustic Measures

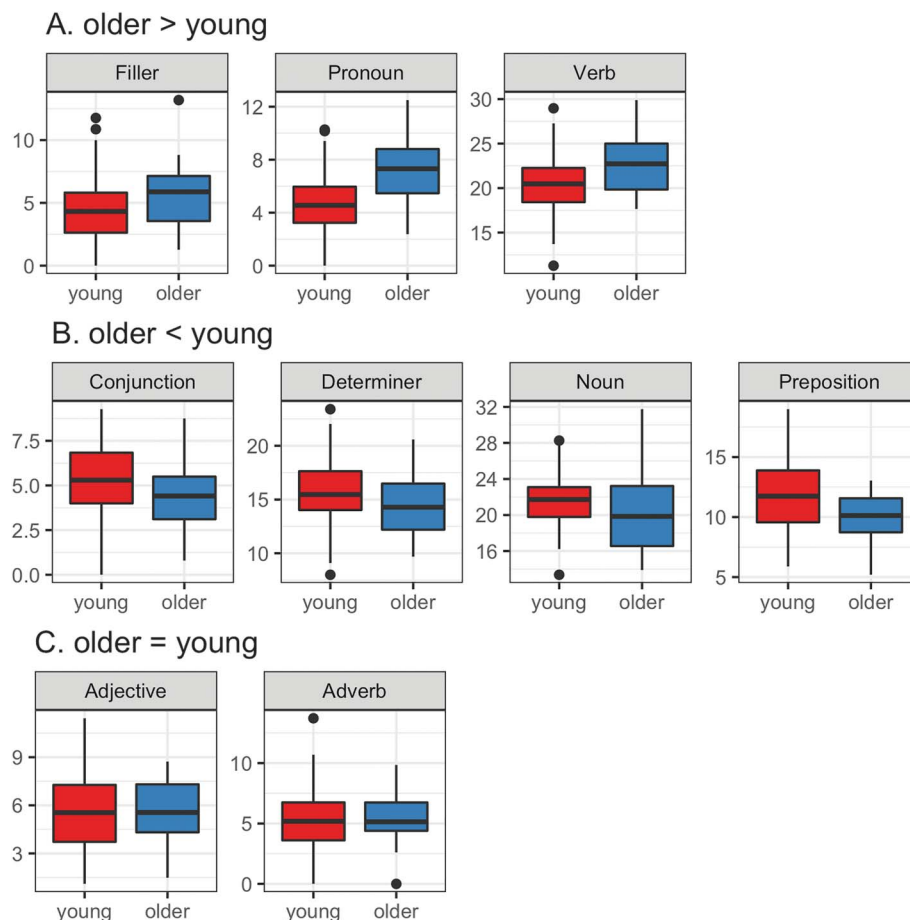
Correlations of the lexical and acoustic measures are summarized in Table 4. We find that total speech time shows a strong positive correlation with the total number of words, which is an expected pattern. Noun familiarity is strongly correlated with total number of words and total speech time.

Pause rate per minute is negatively correlated with total speech time and total word counts but positively correlated with filler words per 100 words. This suggests that speakers who produce more pauses also produce more filler words (i.e., filled pauses), whereas speakers who produce fewer pauses tend to speak longer with more words. Interestingly, speech rate (words per minute) in our study is only correlated with the total number of words and noun frequency, but not with other measures. Participants who speak fast produce more words and more frequent nouns. Lastly, the mean length of clause is negatively correlated with noun frequency, suggesting that speakers who produce longer clauses tend to use less frequent nouns.

Discussion

In this study, we employed automated methods to investigate the effect of age and sex on both lexical and acoustic features in a digitized, semistructured speech sample. Our results, in general, report reduced fluency and shorter clauses in older speakers in narrative, monologic, natural speech. We found that older speakers used more pronouns, filler words, and verbs when describing a picture, whereas young participants used more prepositions, determiners, nouns, and conjunctions. Also, older speakers produced more tense-inflected verbs (per 100 words), compared to young participants. At the same time, older speakers showed a lower lexical diversity score than young speakers in this task. Furthermore, older participants used nouns and verbs with higher familiarity, frequency (verbs only), and

Figure 1. Number of part-of-speech categories per 100 words by group. The y-axis of each panel is in a different range to optimize sensitivity to each feature.



less ambiguity than young speakers when describing a picture. These findings indicate that the lexical content of older speakers seems to be generally less diverse than young speakers in narrative, monologic speech. On the acoustic side, older speakers' speech contained longer pauses with increased total speech time compared to young participants. The increased total speech time and total number of words in older speakers were correlated with their frequent use of familiar nouns. Finally, we examined the effect of age and sex on some important aspects of speech. We discuss each of these themes below.

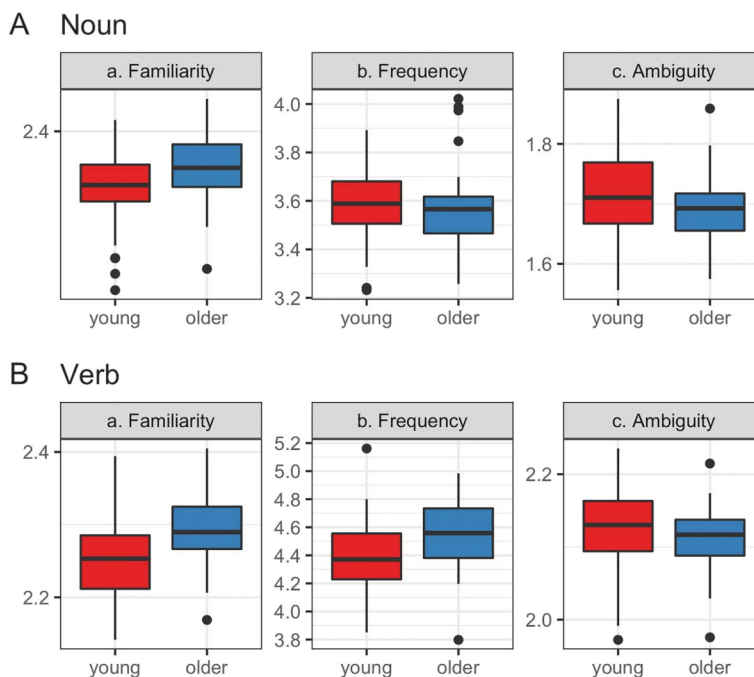
Older Speakers' Lexical Content Is Less Diverse

The automated methodology employed in this study enabled us to discover novel findings of the association of age with the counts of POS categories in narrative, monologic speech. The results that older speakers produce more filler words and pronouns have been previously described (Bortfeld et al., 2001; Heller & Dobbs, 1993; Kemper, 1992; Spieler & Griffin, 2006). However, no one,

to our knowledge, has examined the entire range of POS categories. A previous study by Ardila and Rosselli (1996) is the only study that has considered POS categories and age difference in depth, but these authors collapsed determiners, pronouns, adverbs, prepositions, and conjunctions together as grammatical connectors, making it hard to assess fine differences in these categories. Because of recent developments in NLP, we were able to examine all POS categories individually and found age differences not only for the POS categories that have been frequently discussed in the literature but also for other categories in narrative speech samples.

Our lexical analyses provided a clear result: Older speakers produced shorter clauses with more tense-inflected verbs and lower lexical diversity (more repetition) in the type of speech sample we examined. Furthermore, nouns and verbs that were produced by older participants were more familiar, more frequent (for verbs), and less ambiguous than those produced by young speakers. These results support a conclusion of decreased lexical agility with aging, which is in line with previous findings (Heller & Dobbs,

Figure 2. Lexical scores of nouns and verbs produced by the participants. The y-axis of Panel A shows the z-scored word familiarity scale—percentage of people who know a given word; that of Panel B displays log-scaled word frequency per million words; that of Panel C is for the number of different meanings of a given word.



1993; Kemper, 1992; Nicholas et al., 1985; Ramsay et al., 1999; Schmitter-Edgecombe et al., 2000).

Older Speakers Use Shorter Clauses

We found that older speakers produced fewer words per clause than young participants in narrative speech, which is in line with some previous studies (e.g., Ardila & Rosselli, 1996; Jacewicz et al., 2010), but not with others (e.g., Horton et al., 2010). Mean length of clause in our study was negatively correlated with noun frequency, which was, in turn, positively correlated with speech rate. This suggests that speakers who used more frequent nouns tended to produce shorter clauses and speak more slowly regardless of their age.

One potential reason that previous studies have presented mixed results for mean length of clause might be due to differences in speakers' education levels. Many studies that have investigated the effect of aging on speech (e.g., Horton et al., 2010; Jacewicz et al., 2010; Kavé et al., 2009) did not consider speakers' education level, even though previous studies (Ardila & Rosselli, 1989; Labov, 2001; Prichard, 2016) have shown that the education level of a speaker affects many aspects of speech. Most of our older participants had received higher education and about 16 years of education on average. Since our young and older participants were comparable in terms of their education level, the group difference in mean length of clause in this study seems to be a reflection of cognitive decline with aging. However,

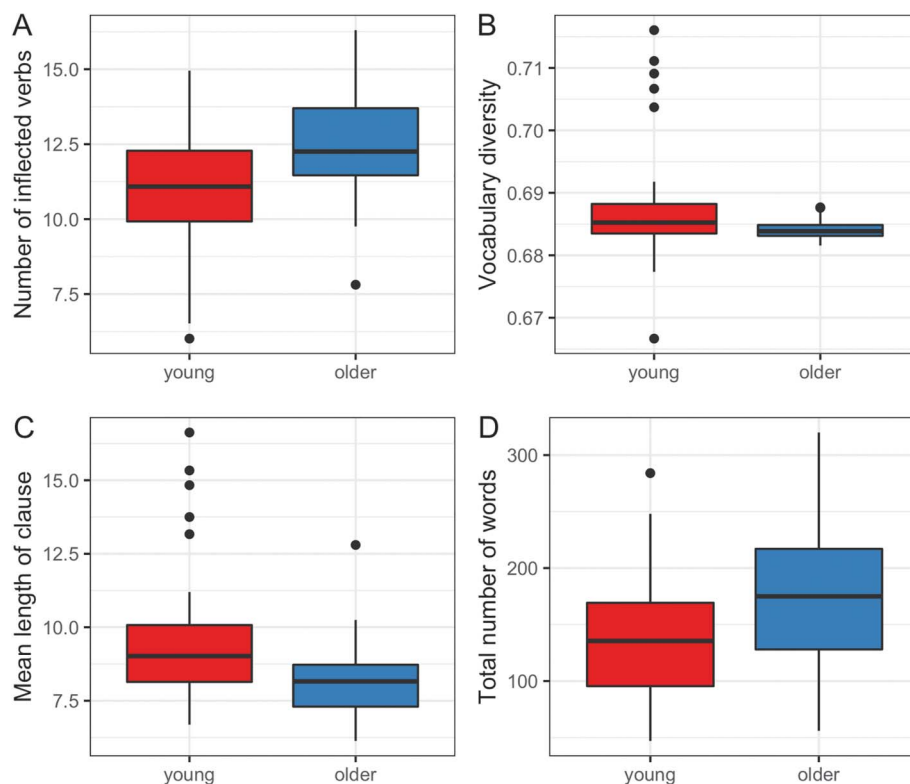
since we only looked at narrative, monologic speech samples, this relation of age, education level, and mean length of clause calls for further exploration with a larger data set of natural conversations in future research.

Age Differences Are Reflected in Part in the Acoustic Properties of Speech in Picture Descriptions

In our acoustic analysis, we found that the older speakers produced shorter speech segments, coinciding with our lexical analysis that suggested the production of shorter clauses with limited lexical content. This was in contrast to the younger speakers who produced longer speech segments with a greater mean length of clause. These differences did not result in an incomplete description of the picture, as the older participants simply spent more time speaking and produced more clauses and words. Their total speech time was longer on average than that of the young speakers, but this measure excluded pause time, which was also longer in the older speakers' samples. This could be regarded as a compensatory mechanism, implemented by older speakers to complete the cognitive task of describing the picture in detail. In our study, these findings seem to follow from the fact that the stimulus supports the production of highly natural narrative speech while controlling for the topic.

Speech rate in our older group did not differ significantly from their younger counterparts. This is in contrast to some previous reports (Bóna, 2014; Horton et al., 2010).

Figure 3. Derived measures from POS categories. Panel A shows the total number of tense-inflected verbs (= number of modal auxiliary verbs + number of present tense verbs + number of past tense verbs), and Panel B displays vocabulary diversity, which is estimated with entropy. Panel C presents mean length of clause (= total number of words / total number of inflected verbs). Panel D displays the group difference in total number of words.



Speech rate is used as an umbrella term, and different investigators calculate it in different ways. In our analysis, we calculated the number of words produced per minute of speech time, excluding pauses. Had we included the pauses, which were significantly longer in our older group, we may have gotten the impression that speech rate is reduced, such that the rate of word production is similar between age groups, but pause time is longer in older adults. Some researchers refer to the measurement of rate with an omission of pauses as “articulation rate” and still find it to be reduced in older speakers (Bóna, 2014); however, it is difficult to compare our findings as the studies differ in the speakers as well as the speech sample characteristics. A larger sample with a wider, fuller range of speaker ages and variable task stimuli, including natural dialogues, may shed light on this question.

Noun familiarity and pause rate were strongly correlated with total speech time and total number of words in the picture description task, which goes along with the finding that the elderly cohort in our study exhibited longer speech times and produced more words. Also, the interpretation of pause length and filled pauses in our corpus is consistent with the hypothesis that pause duration represents lexical retrieval time for speakers of any age, and this in turn is expected to be longer in an aging group as their cognitive processing speed declines. It is a limitation of our

current study that we cannot compare the speakers’ performance on nonspeech measures of cognitive processing as we do not have data from an appropriate task. In future studies, we plan to incorporate such neurocognitive tests to better address this issue.

The Interaction of Age and Sex

Our results showing a greater total number of words in older speakers compared to younger speakers and no sex effects in the picture description task are in line with the findings in Bortfeld et al. (2001). However, our result did not agree with the observed interaction of age and sex reported by Ardila and Rosselli (1996) or the effect of sex (either female or male speaking more than the other sex) in other studies (e.g., Dovidio et al., 1988; Mulac et al., 1986, 2000). These incongruent results might be due, in part, to differences in the types of speech samples that previous studies have examined (e.g., monologue vs. dialogue). The question of “who talks when and for how long” in conversations depends on the interlocutors’ perceived sociopolitical status compared to one another, as well as specific cultural norms (Dovidio et al., 1988; Mulac et al., 2000; Ng & Deng, 2017). We tried to eliminate such confounding factors by using a picture description task, providing a neutral and

Table 3. Young versus older age groups acoustics comparisons.

Variable	Older	Young	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
90th pitch quantile (ST)	6.26 (2.61)	6.29 (2.96)	-0.06	.951	0.01
Mean speech segment duration (s)	2.00 (0.57)	2.29 (0.60)	-2.0	.017	0.48
Total speech time (s)	50.94 (17.02)	38.25 (13.83)	4.0	< .001	0.85
Mean pause duration (s)	0.91 (0.37)	0.57 (0.12)	5.0	< .001	1.4
Total number of pauses	25.54 (8.29)	18.66 (7.60)	4.0	< .001	0.88
Pause rate per minute (ppm)	31.53 (9.07)	29.49 (6.25)	1.0	.166	0.28
Speech rate (wpm)	208.63 (31.66)	215.51 (27.64)	-1.0	.239	0.24

Note. ST = semitone; ppm = pauses per minute; wpm = words per minute.

uniform context for speakers' language use. However, since we had a relatively small number of speakers with homogeneous education level and our data were monologue speech samples, our findings will need to be tested against larger-cohort cross-sectional and longitudinal studies and with dialogue speech samples.

We also showed that the number of filler words significantly varied by both age and sex. The result of older speakers' showing reduced fluency with more filler words is consistent with previous studies (Bortfeld et al., 2001; Heller & Dobbs, 1993; Kemper, 1992; Spieler & Griffin, 2006). Also, our result of male speakers using more filler words than female speakers aligns with previous studies (Bortfeld et al., 2001; Mulac et al., 1988; Shriberg, 1996). Since we

found the same result of filler word usage in narrative, monologic speech samples and previous studies have evaluated a variety of sources of speech data, it appears that the pattern of old and/or male speakers producing more filler words than young and/or female speakers may be a general trend in natural speech.

In this work, we found that pitch range, as represented by our *f*₀ range (the 90th percentile *f*₀), was similar between the age groups. However, separating the groups by sex revealed an interaction, whereby the difference in *f*₀ range between male and female speakers was much larger in the younger age group than in the older age group. This phenomenon of diminished differentiation of pitch between the sexes with aging has been previously reported (Ferrand,

Figure 4. Group comparisons for duration measures. The units on the y-axis in Panels A, B, and D are seconds, and the units in Panel B are pauses per minute.

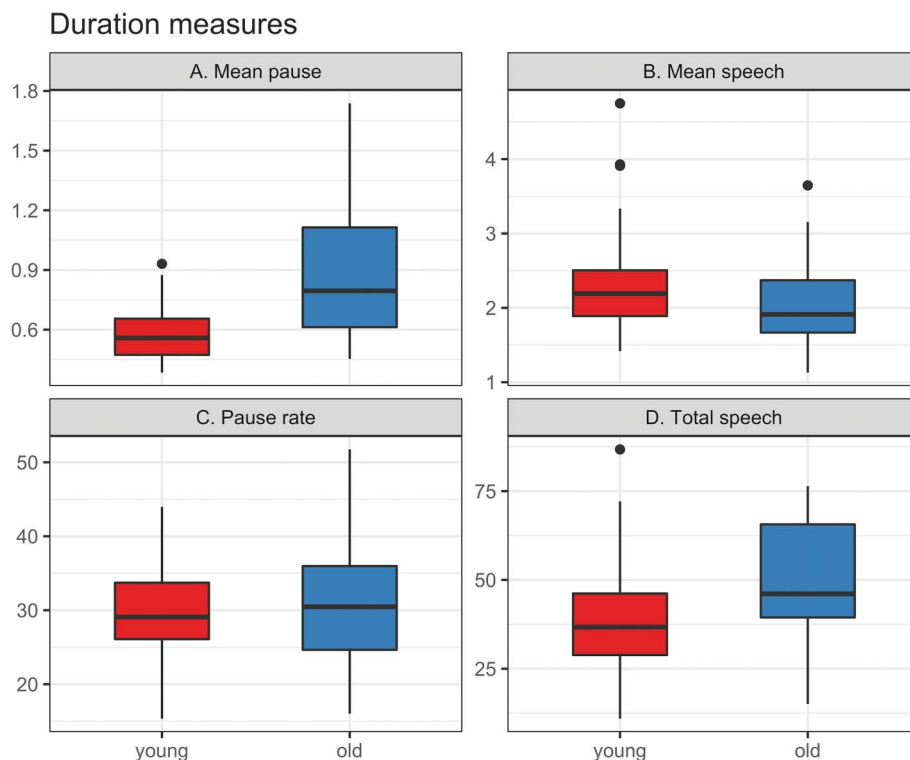


Figure 5. Pitch range (A), frequency of filler words (B), and total number of words (C) for participants grouped by age and sex. ST = semitone; F = female; M = male; f0 = fundamental frequency.

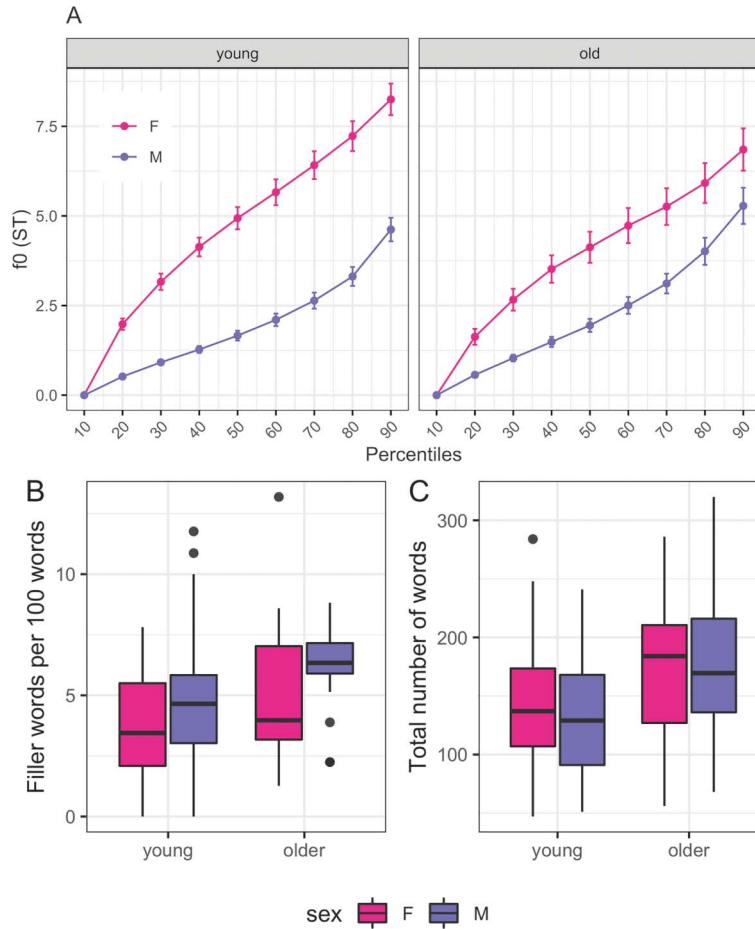


Table 4. Correlation matrix for the lexical and acoustic measures.

Variable	Total speech time	Total number of words	Lexical diversity	Familiarity (noun)	Number of filler words	Pause rate	Speech rate	Frequency (noun)
Total speech time								
Total number of words	.92***							
Lexical diversity	.01	.03						
Familiarity (noun)	.37***	.38***	.06					
Number of filler words	.12	.08	.13	.01				
Pause rate	-.21*	-.25**	.02	-.12	.32***			
Speech rate	-.15	.22*	.05	.00	-.15	-.08		
Frequency (noun)	-.05	.09	-.03	.11	.01	.01	.33***	
MLC	-.16	-.18	.06	-.07	.06	.11	-.09	-.27**

Note. MLC = mean length of clause.

* $p < .05$. ** $p < .01$. *** $p < .001$.

2002; Linville, 1987; Mueller, 1997; Russell et al., 1995; Sataloff et al., 1997). Several hypotheses can be suggested to explain this finding. One possibility may be related to a potential evolutionary or psychosocial need for the sexual vocal differences to be more distinct in the younger age group. Alternatively, we can consider physiological explanations that involve hormonal changes (e.g., Gugatschka et al., 2010) or reduced vocal fold muscular bulk and tone in older female speakers, causing their pitch to lower as they age (e.g., Xue & Hao, 2003). Regardless of the basis for this finding, our observations suggest that acoustic data should be adjusted by sex differently in different age groups. With the current study design, we were not able to fit a model to test the nature of this age and sex interaction in a more complete way; however, in future studies with a more complete data set, we hope to model this interaction in more depth.

Conclusions

This study compared lexical and acoustic features of semistructured, narrative speech samples between healthy older and young adults using automated methods. We discovered differences between age groups in the counts of POS categories and lexical characteristics of nouns and verbs. Our results show that older speakers use less diverse and more limited lexical content and produce shorter clauses and longer pauses than young speakers. We also confirmed previous findings, including the interaction of age and sex with respect to pitch and the more frequent use of pronouns and filler words by older speakers. Most importantly, this study shows that semistructured speech samples can be studied with automated methods.

Although our study provides novel methods and findings, there are limitations. First of all, since our data included monologue speech samples, some of our findings may or may not be applicable to natural dialogues. Examining natural dialogues that have been carefully and systematically controlled for interlocutors' sociopolitical status will further shed light on the effect of aging on natural speech. We plan to analyze a large-scale speech corpus with natural dialogues, such as the Switchboard corpus (Godfrey & Holliman, 1993) or the Fisher corpus (Cieri et al., 2004), to examine the effect of aging on both lexical and acoustic features in natural dialogues. Also, our methodology did not examine the effect of aging on syntactic aspects of language, which is an important area to investigate. We plan to explore this area further with a syntactic dependency parser in the near future. Lastly, since we only investigated one picture description from each individual, we were not able to assess individual variability in this study. Future research with multiple picture descriptions will be needed to investigate individual variability in narrative speech.

Acknowledgment

This study was supported by the National Institutes of Health (AG017586, AG053940, AG052943, NS088341, DC013063, AG054519, awarded to M. G.), the Institute on Aging at the

University of Pennsylvania (awarded to M. L.), the Alzheimer's Association (AACSF-18-567131, awarded to N. N.), an anonymous donor, and the Wyncote Foundation (awarded to M. G.).

References

- Ardila, A., & Rosselli, M.** (1989). Neuropsychological characteristics of normal aging. *Developmental Neuropsychology*, 5(4), 307–320. <https://doi.org/10.1080/87565648909540441>
- Ardila, A., & Rosselli, M.** (1996). Spontaneous language production and aging: Sex and educational effects. *International Journal of Neuroscience*, 87(1–2), 71–78. <https://doi.org/10.3109/00207459608990754>
- Boersma, P., & Weenink, D.** (2019). *Praat: Doing phonetics by computer* [Computer program]. <https://www.praat.org>
- Bóna, J.** (2014). Temporal characteristics of speech: The effect of age and speech style. *The Journal of the Acoustical Society of America*, 136(2), EL116–EL121. <https://doi.org/10.1121/1.4885482>
- Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E.** (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, 44(2), 123–147. <https://doi.org/10.1177/00238309010440020101>
- Brybaert, M., Mandera, P., & Keuleers, E.** (2018). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, July 2018, 467–479. <https://doi.org/10.3758/s13428-018-1077-9>
- Brybaert, M., & New, B.** (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990. <https://doi.org/10.3758/BRM.41.4.977>
- Brybaert, M., Warriner, A. B., & Kuperman, V.** (2014). Concrete-ness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904–911. <https://doi.org/10.3758/s13428-013-0403-5>
- Cieri, C., Graff, D., Kimball, O., Miller, D., & Walker, K.** (2004). *Fisher English Training Speech Corpus (LDC2004S13)*. Linguistic Data Consortium.
- Cooper, P. V.** (1990). Discourse production and normal aging: Performance on oral picture description tasks. *Journal of Gerontology*, 45(5), 210–214. <https://doi.org/10.1093/geronj/45.5.P210>
- Cousins, K. A., Ash, S., Olm, C. A., & Grossman, M.** (2018). Longitudinal changes in semantic concreteness in semantic variant primary progressive aphasia (svPPA). *eNeuro*, 5(6), 1–10. <https://doi.org/10.1523/ENEURO.0197-18.2018>
- Covington, M. A., & McFall, J. D.** (2010). Cutting the Gordian knot: The moving-average type-token ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94–100. <https://doi.org/10.1080/09296171003643098>
- Dovidio, J. F., Brown, C. E., Heltman, K., Ellyson, S. L., & Keating, C. F.** (1988). Power displays between women and men in discussions of gender-linked tasks: A multichannel study. *Journal of Personality and Social Psychology*, 55(4), 580–587. <https://doi.org/10.1037/0022-3514.55.4.580>
- Ferrand, C. T.** (2002). Harmonics-to-noise ratio: An index of vocal aging. *Journal of Voice*, 16(4), 480–487. [https://doi.org/10.1016/S0892-1997\(02\)00123-6](https://doi.org/10.1016/S0892-1997(02)00123-6)
- Godfrey, J. J., & Holliman, E.** (1993). *Switchboard-1 Release 2 (LDC97S62)*. Linguistic Data Consortium.
- Goodglass, H., Kaplan, E., & Weintraub, S.** (1983). *Boston Diagnostic Aphasia Examination*. Lea & Febiger.
- Gugatschka, M., Kiesler, K., Obermayer-Pietsch, B., Schoekler, B., Schmid, C., Groselj-Strele, A., & Friedrich, G.** (2010). Sex

- hormones and the elderly male voice. *Journal of Voice*, 24(3), 369–373. <https://doi.org/10.1016/j.jvoice.2008.07.004>
- Guiraud, H.** (1954). *Les Caractères Statistiques du Vocabulaire* [Statistical characteristics of vocabularies]. Presses Universitaires de France.
- Hartman, D. E., & Danhauer, J. L.** (1976). Perceptual features of speech for males in four perceived age decades. *The Journal of the Acoustical Society of America*, 59(3), 713–715. <https://doi.org/10.1121/1.380894>
- Heller, R. B., & Dobbs, A. R.** (1993). Age differences in word finding in discourse and nondiscourse situations. *Psychology and Aging*, 8(3), 443–450. <https://doi.org/10.1037/0882-7974.8.3.443>
- Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T.** (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, 45(3), 718–730. <https://doi.org/10.3758/s13428-012-0278-x>
- Honnibal, M., & Johnson, M.** (2015, September). An improved non-monotonic transition system for dependency parsing. In *Conference Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1373–1378). <https://doi.org/10.18653/v1/d15-1162>
- Horton, W. S., Spieler, D. H., & Shriberg, E.** (2010). A corpus analysis of patterns of age-related change in conversational speech. *Psychology and Aging*, 25(3), 708–713. <https://doi.org/10.1037/a0019424>
- Jacewicz, E., Fox, R. A., & Wei, L.** (2010). Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America*, 128(2), 839–850. <https://doi.org/10.1121/1.3459842>
- Jarvis, S.** (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing*, 19(1), 57–84. <https://doi.org/10.1191/0265532202lt220oa>
- Kavé, G., Samuel-Enoch, K., & Adiv, S.** (2009). The association between age and the frequency of nouns selected for production. *Psychology and Aging*, 24(1), 17–27. <https://doi.org/10.1037/a0014579>
- Kemper, S.** (1992). Adults' sentence fragments: Who, what, when, where, and why. *Communication Research*, 19(4), 444–458. <https://doi.org/10.1177/009365092019004003>
- Kemper, S., Herman, R. E., & Lian, C. H.** (2003). 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(2), 181–192. <https://doi.org/10.1037/0882-7974.18.2.181>
- Labov, W.** (2001). *Principles of linguistic change. Volume 2: Social factors*. Blackwell.
- LaGrone, S., & Spieler, D. H.** (2006). Lexical competition and phonological encoding in young and older speakers. *Psychology and Aging*, 21(4), 804–809. <https://doi.org/10.1037/0882-7974.21.4.804>
- Linville, S. E.** (1987). Maximum phonational frequency range capabilities of women's voices with advancing age. *Folia Phoniatrica et Logopaedica*, 39(6), 297–301. <https://doi.org/10.1159/000265873>
- Luo, M., Robbins, M. L., Martin, M., & Demiray, B.** (2019). Real-life language use across different interlocutors: A naturalistic observation study of adults varying in age. *Frontiers in Psychology*, 10, 1412. <https://doi.org/10.3389/fpsyg.2019.01412>
- Marcus, M., Santorini, B., & Marcinkiewicz, M. A.** (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2), 313–330. <https://doi.org/10.21236/ADA273556>
- Martins, V. d. O., & Andrade, C. R. F. d.** (2011). Study of pauses in elderly. *Revista Da Sociedade Brasileira de Fonoaudiologia*, 16(3), 344–349. <https://doi.org/10.1590/S1516-80342011000300017>
- McKee, G., Malvern, D., & Richards, B.** (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, 15(3), 323–337. <https://doi.org/10.1093/lilc/15.3.323>
- Moscoso del Prado Martín, F.** (2017). Vocabulary, grammar, sex, and aging. *Cognitive Science*, 41(4), 950–975. <https://doi.org/10.1111/cogs.12367>
- Mueller, P. B.** (1997). The aging voice. *Seminars in Speech and Language*, 18(2), 159–168. https://doi.org/10.5005/jfp/books/12711_51
- Mulac, A., Lundell, T. L., & Bradac, J. J.** (1986). Male/female language differences and attributional consequences in a public speaking situation: Toward an explanation of the gender-linked language effect. *Communication Monographs*, 53(2), 1150129. https://doi.org/10.5005/jfp/books/12711_51
- Mulac, A., Seibold, D. R., & Farris, J. L. E. E.** (2000). Female and male managers' and professionals' criticism giving differences in language. *Journal of Language and Social Psychology*, 19(4), 389–415. <https://doi.org/10.1177/0261927X00019004001>
- Mulac, A., Wiemann, J. M., Widenmann, S. J., & Gibson, T. W.** (1988). Male/female language differences and effects in same-sex and mixed-sex dyads: The gender-linked language effect. *Communication Monographs*, 55(4), 315–335. <https://doi.org/10.1080/03637758809376175>
- Nevler, N., Ash, S., Irwin, D. J., Liberman, M., & Grossman, M.** (2019). Validated automatic speech biomarkers in primary progressive aphasia. *Annals of Clinical and Translational Neurology*, 6(1), 4–14. <https://doi.org/10.1002/acn.3.653>
- Nevler, N., Ash, S., Jester, C., Irwin, D. J., Liberman, M., & Grossman, M.** (2017). Automatic measurement of prosody in behavioral variant FTD. *Neurology*, 89, 1–8.
- Ng, S. H., & Deng, F.** (2017). Language and power. In *Oxford research encyclopedia of communication* (pp. 1–22). <https://doi.org/10.1093/acrefore/9780190228613.013.436>
- Nicholas, M., Obler, L., Albert, M., & Goodglass, H.** (1985). Lexical retrieval in healthy aging. *Cortex*, 21(4), 595–606. [https://doi.org/10.1016/S0010-9452\(58\)80007-6](https://doi.org/10.1016/S0010-9452(58)80007-6)
- Nishio, M., & Niimi, S.** (2008). Changes in speaking fundamental frequency characteristics with aging. *Folia Phoniatrica et Logopaedica*, 60(3), 120–127. <https://doi.org/10.1159/000118510>
- Petrov, S., Das, D., & McDonald, R.** (2012). A universal part-of-speech tag set. In *Proceedings of the International Conference on Language Resources and Evaluation* (pp. 2089–2096).
- Prichard, H.** (2016). *The role of higher education in linguistic change (PhD thesis, University of Pennsylvania)*. <https://doi.org/10.1075/cilt.106.27rai>
- Ramsay, C. B., Nicholas, M., Au, R., Obler, L. K., & Albert, M. L.** (1999). Verb naming in normal aging. *Applied Neuropsychology*, 6(2), 57–67. https://doi.org/10.1207/s15324826an0602_1
- Russell, A., Penny, L., & Pemberton, C.** (1995). Speaking fundamental frequency changes over time in women: A longitudinal study. *Journal of Speech, Language, and Hearing Research*, 38(1), 101–109. <https://doi.org/10.1044/jshr.3801.101>
- Sataloff, R. T., Rosen, D. C., Hawkshaw, M., & Spiegel, J. R.** (1997). The aging adult voice. *Journal of Voice*, 11(2), 156–160. [https://doi.org/10.1016/S0892-1997\(97\)80072-0](https://doi.org/10.1016/S0892-1997(97)80072-0)
- Schmitter-Edgecombe, M., Vesneski, M., & Jones, D. W.** (2000). Aging and word-finding: A comparison of spontaneous and constrained naming tests. *Archives of Clinical Neuropsychology*, 15(6), 479–493. [https://doi.org/10.1016/S0887-6177\(99\)00039-6](https://doi.org/10.1016/S0887-6177(99)00039-6)

-
- Shriberg, E.** (1996). Disfluencies in Switchboard. In A. Cutler (Ed.), *ICSLP 96: Fourth International Conference on Spoken Language Processing* (pp. 11–14). A.I. duPont Hospital for Children and the University of Delaware.
- Spieler, D. H., & Griffin, Z. M.** (2006). The influence of age on the time course of word preparation in multiword utterances. *Language and Cognitive Processes, 21*(1–3), 291–321. <https://doi.org/10.1080/01690960400002133>
- Tweedie, F. J., & Baayen, R. H.** (1998). How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities, 32*, 323–352. <https://doi.org/10.1023/A:1001749303137>
- Uttil, B.** (2002). North American adult reading test: Age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology, 24*(8), 1123–1137. <https://doi.org/10.1076/jcen.24.8.1123.8375>
- Verhaeghen, P.** (2003). Aging and vocabulary scores: A meta-analysis. *Psychology and Aging, 18*(2), 332–339. <https://doi.org/10.1037/0882-7974.18.2.332>
- Xue, S. A., & Hao, G. J.** (2003). Changes in the human vocal tract due to aging and the acoustic correlates of speech production: A pilot study. *Journal of Speech, Language, and Hearing Research, 46*(3), 689–701. [https://doi.org/10.1044/1092-4388\(2003/054\)](https://doi.org/10.1044/1092-4388(2003/054))