International Journal of Language & Communication Disorders

INT J LANG COMMUN DISORD, JANUARY/FEBRUARY 2021, VOL. 56, NO. 1, 6–19

Research Report

Development of a measure of function word use in narrative discourse: core lexicon analysis in aphasia

Hana Kim†, Stephen Kintz‡ and Heather Harris Wright†

†Department of Communication Sciences and Disorders, East Carolina University, Greenville, NC, USA
‡Department of Audiology and Speech Pathology, University of Arkansas, Little Rock, AK, USA

(*Received June 2019; accepted July 2020*)

Abstract

Background: Although discourse-level assessments contribute to predicting real-world performance in persons with aphasia (PWA), the use of discourse measures is uncommon in clinical settings due to resource-heavy procedures. Moreover, assessing function word use in discourse requires the arduous procedure of defining grammatical categories for each word in language transcripts.
Aims: The purpose of this exploratory study was twofold: (1) to develop core function word lists as a clinician-friendly means of evaluating function word use in discourse; and (2) to examine the ability of the core function word measure to differentiate PWA from cognitively healthy adults and persons with fluent aphasia from non-fluent aphasia.
Methods & Procedures: The 25 most commonly used function words (core function words) were extracted from narrative language samples from 470 cognitively healthy adults, which were divided into seven age groups (20s, 30s, 40s, 50s, 60s, 70s and 80s). The percent agreement of core function words for 11 PWA (fluent aphasia = 5; non-fluent aphasia = 6) and 11 age- and education-matched controls were then calculated. Percent agreement for the core function words produced was compared between the controls and the PWA group, and between participants with fluent aphasia and non-fluent aphasia.
Outcomes & Results: The results indicated that PWA produced fewer core function words from the lists than the control group, and that core function word use was strongly correlated with aphasia severity. Persons with non-fluent aphasia produced fewer core function words than those with fluent aphasia, although this could be a confound of aphasia classification from the use of the Western Aphasia Battery (WAB)—Revised.
Conclusions & Implications: Core function word lists consisting of a limited number of items for quantifying function word use in discourse remain in a nascent stage of development. However, the findings are consistent with previous studies analysing the total production of function words in language samples produced by PWA. Therefore, core function words may potentially serve as a clinician-friendly manner of quantifying function words produced in discourse.

Keywords: aphasia, discourse analysis, function word, core lexicon.

What this paper adds

What is already known on the subject

- Function word analysis in discourse requires arduous processes of identifying the error production and grammatical category of function words in discourse. Previous studies have demonstrated that core lexicon measures are an efficient, simple means of quantifying discourse in PWA. However, function words have never been considered for generating an independent core lexicon list.

What this paper adds to existing knowledge

- As an exploratory study, we focused primarily on developing a clinician-friendly measure to evaluate function word production in discourse, motivated by the idea of an adaptation strategy within the core lexicon framework. Our findings demonstrated that by using a simple scoring system that the core lexicon

Address correspondence to: Hana Kim, Department of Communication Sciences and Disorders, East Carolina University, Greenville, NC, USA; email: only.hana.kim@gmail.com

ISSN 1368-2822 print/ISSN 1460-6984 online
DOI: 10.1111/1460-6984.12567

measure provides, we differentiated the control group from the PWA group, and persons with fluent aphasia from persons with non-fluent aphasia. Additionally, we found significant correlations between function word production and aphasia severity determined by WAB Aphasia Quotient (AQ).

What are the potential or actual clinical implications of this work?

- The results add empirical evidence for the utility of core function word lists for quantifying function word usage in discourse in PWA. Counting the presence and absence of function words in discourse will allow clinicians to avoid labour-intensive preparatory work, and to obtain useful diagnostic information in a less time-consuming way.

Introduction

Word retrieval is a common problem for persons with aphasia (PWA) and researchers have most commonly examined word-retrieval difficulties related to retrieval of content words (e.g., Bates *et al.* 1991, Caramazza and Hillis 1991, Chen and Bates 1998, Kim and Thompson 2000, Luzzatti and Chierchia 2002, Saffran *et al.* 1980) with relatively few studies focusing on the retrieval of function words in discourse (e.g., Chapman and Ulatowska 1989, Gleason *et al.* 1980, Manning and Franklin 2016, Nicholas *et al.* 1985). At the discourse level, function word production (e.g., referents, prepositions) contributes to elaborative phrasing or sentence structure in binding story elements (Halliday and Hasan 1976). However, function word production at the discourse level has received little attention to date.

Investigations that have addressed function word production in discourse have reported differences between subtypes of aphasia (Gordon 2008, 2006, Kolk and Heeschen 1992, Manning and Franklin 2016, Rochon *et al.* 2000, Saffran *et al.* 1989, Salis and Edwards 2004). For example, Saffran *et al.* (1989) compared grammatical production between participants with aphasia and agrammatism ($N = 5$), participants with aphasia and non-agrammatism ($N = 5$) and cognitively healthy adults ($N = 5$).[1] In their study, the participants told the Cinderella story and the proportion of closed class words was computed. Closed class words consisted of pronouns, determiners, prepositions, conjunctions, quantifiable adverbs, verb inflections, verb particles and auxiliary verbs, and they are considered function words. Significant differences for proportion of closed class words were found between the two aphasia groups ($p < 0.001$), with the agrammatism aphasia group producing a lower proportion of closed class words compared to the non-agrammatism aphasia group. The non-agrammatism aphasia group and the control group did not significantly differ in proportion of closed class words produced. In a later study, Gordon (2006) attempted to replicate Saffran *et al.*'s (1989) findings. Gordon included eight participants with fluent aphasia, eight participants with non-fluent aphasia and six participants without brain damage. Fluency and aphasia severity were determined by performance on the Boston Diagnostic Aphasia Exam (Goodglass *et al.* 2001). Similar to Saffran *et al.*'s (1989) findings, the non-fluent aphasia group produced a lower proportion of closed-class words and had less frequent use of obligatory determiners compared with the fluent aphasia group. Gordon concluded that such measures can facilitate clinical judgement in differentiating fluent and non-fluent aphasia groups.

Simplified utterances with the omission of function words, as observed in Broca's aphasia, are proposed to reflect adaptive mechanisms in PWA and are evidently manifested at the connected speech level (Kolk and Heeschen 1992). The abnormal behaviour changes, referred to as 'elliptical style', have been attributed to reduced processing capacity; thus, they attempt to reduce the computational overload by producing simplified utterances (Kolk and Heeschen 1990: 229). However, it should be noted that this theory is based on Dutch and German studies, and has been applied to interpret findings with English-speaking PWA.

Salis and Edwards (2004) replicated Kolk and Heeschen's (1996) findings and included two groups: persons with non-fluent aphasia with agrammatism ($N = 4$) and a control group ($N = 3$). They investigated the overuse of simplified utterances in discourse collected from spontaneous language and picture-description tasks. The picture-description task consisted of drawings of shapes and colours depicting different spatial positions. Outcome measures included omissions and substitutions of determiners, verbs, inflections and prepositions. Similar to Kolk and Heeschen's results, the non-fluent aphasia group differed from the controls and had more omissions of determiners and prepositions. Although Salis and Edwards did not provide further explanation regarding the underlying mechanisms that accounted for these findings beyond that of a limited processing capacity, their main finding was that function word production is associated with grammatical impairment in persons with non-fluent aphasia.

Narrative-based tasks are indispensable for language assessment in that they provide a window into understanding how PWA perform their daily communicative tasks. Undoubtedly, function word use should be evaluated at the discourse level because it demands contextual requirements. For example, the selection of pronouns is deeply linked with semantic information of thematic roles that have been previously introduced in discourse. However, infrequent use of discourse analysis in clinical settings due to resource-heavy preparatory work for discourse assessment has been reported (Bryant *et al.* 2017, Maddy *et al.* 2015, Pak-Hin and Law 2004). It is also unrealistic for clinicians to complete a full transcription of language samples and define grammatical categories and errors for measuring function word use at the discourse level.

For these reasons, the current study focused primarily on developing a function word measure to potentially address issues of clinical feasibility. Although only a few English-language studies have examined function word ellipsis in discourse produced by PWA within the same framework raised by Kolk and Heeschen (1996) (Salis and Edwards 2004), the notion that the omission of function words is a possible strategy to reduce the computational demands to produce connected speech is compelling. Further, a recent article revisiting adaptive mechanisms regarding PWA's language use reported that adaptive mechanisms on omissions of function words and disfluencies in utterances are pervasive (Cui and Zhong 2018). Thus, the scope of the current study is limited to the examination of the presence and absence of function words in discourse produced by PWA. In contrast to conventional procedures of discourse analysis, a novel approach was taken to quantify function words in discourse in which a checklist was provided, reducing clinicians' burden. A review of previous studies revealed differential performance of function word production at the discourse level, but the findings were mixed across different subtypes of aphasia (e.g., non-fluent versus fluent, agrammatism versus non-agrammatism). As a first step toward investigating the clinical utility of the approach, we began with a fluent–non-fluent comparison of core function word production because of the widespread use of the fluency classification in clinical settings (e.g., Clough and Gordon 2020). In the following sections, we review the novel, core lexicon research that has led to the computational analysis for this study.

Core lexicon measures

Recently, researchers have developed and applied core lexicon analysis to aphasia narratives (Dalton *et al.* 2020, Dalton and Richardson 2015, Kim *et al.* 2019, Kim and Wright 2020a, 2020b, MacWhinney *et al.* 2010). Core lexicon measures consist of critical lexical items that play a significant role in constructing a semantically coherent narrative (MacWhinney *et al.* 2010). It is intended to provide clinicians with a means to quantify word-retrieval ability at the discourse level. The use of core lexicon measures has many advantages. They have been created with computational language analysis programs, such as the Computerized Language Analysis (CLAN; MacWhinney 2000), which reduces analysis errors (Dalton *et al.* 2020, Dalton and Richardson 2015, Kim *et al.* 2019, Kim and Wright 2020a, 2020b, MacWhinney *et al.* 2010). In using the core lexicon measures providing a checklist, clinicians may potentially need to devote less time to completing narrative-based analysis. The simple binary scoring system of checking the presence and absence of core lexicon items does not require clinicians to undertake special training for use. Lastly, core lexicon lists that are created based on the performance of cognitively healthy adults provide a reference to understand the degree to which clinical populations deviate from normalcy.

MacWhinney *et al.* (2010) included core lexicon measures as a method for demonstrating the use of TalkBank tools with the AphasiaBank database. They included language samples from 25 healthy participants to extract the 10 most frequent nouns and verbs by using CLAN. They found group differences in the core lexicon measure: the aphasia group presented with reduced lexical diversity and greater use of light verbs. Function words were not considered in the study. In a more recent study using a core lexicon measure, Dalton and Richardson (2015) included function words in their core lexicon list. The study included 92 cognitively healthy adults to build core lexicon items, and then they examined if the core lexicon measure could discriminate a different group of cognitively healthy participants ($N = 166$) from PWA ($N = 235$), as well as among aphasia subtypes. The cognitively healthy group produced more items on the core lexicon list compared with the aphasia group. Within the aphasia group, participants with fluent aphasia performed better on the measure compared with participants with non-fluent aphasia. The researchers also rated the participants' ability to convey the gist of the story (main concept analysis) to investigate the relationship between core lexicon and main concept performance. Significant correlations between the two measures were found across the subtypes of aphasia (i.e., anomic, Broca's, conduction, Wernicke's). Dalton and Richardson concluded that the inclusion of function words as part of the core lexicon list may have contributed to the significant correlations; however, no attempts were made to generate an independent core lexicon list to test this hypothesis.

Table 1. Neurologically healthy adult demographic information

Age group	*N* (female:male)	Age (SD) (years)	Education (SD) (years)	MMSE (SD)[a]
20s	66 (35:31)	24.3 (2.7)	15.8 (2.0)	56.1 (6.2)
30s	63 (39:24)	34.1 (3.2)	16.0 (3.2)	52.1 (9.4)
40s	67 (41:26)	44.4 (3.1)	15.4 (2.5)	53.0 (6.1)
50s	68 (43:25)	53.5 (2.6)	15.8 (2.5)	51.9 (5.9)
60s	67 (38:29)	65.2 (4.5)	15.6 (2.8)	54.5 (8.3)
70s	76 (43:33)	73.5 (2.9)	15.4 (2.3)	56.7 (9.1)
80s	63 (34:29)	83.3 (2.5)	15.2 (3.0)	62.4 (12.6)
Total (female:male)	470 (273:197)			
Mean (SD)		54.4 (19.9)	15.6 (2.6)	55.2 (9.1)

Note: [a]Mini-Mental State Exam (MMSE; Folstein *et al.* 2001). Study inclusion criteria were an MMSE *t*-score ≥ 30.

The purpose of this exploratory study, then, was to examine the potential use of core function word lists as a means to evaluate function word production in narrative discourse. MacWhinney *et al.*'s (2010) core lexicon procedures served as the basis for extracting the function word items. Moreover, normative data and data from individuals with aphasia were provided for comparison, as the current study is exploratory in nature and seeks to provide a basis for clinical feasibility for clinicians and researchers. The research questions addressed were as follows:

- Does the production of core function words differ between the cognitively healthy adult group and PWA?
- Does the production of core function words differ between persons with fluent aphasia and persons with non-fluent aphasia?

Method

Participants

The study included language samples from 470 healthy participants (273 females, 197 males) and 11 PWA (fluent aphasia = 5; non-fluent aphasia = 6). The normative data presented are a subset of data from a larger study examining discourse processing across the lifespan (Wright and Capilouto 2017). The database included discourse samples and cognitive measures collected from 470 participants ranging in age from 20 to 89 years. Control participants were divided into seven age groups: 20s, 30s, 40s, 50s, 60s, 70s and 80s. All control participants (1) were native English speakers; (2) passed hearing (Davis and Silverman 1978) and vision screenings (Beukelman and Mirenda 1998); (3) presented with normal cognitive functioning, as indicated by the Mini-Mental State Exam (MMSE; Folstein *et al.* 2001); and (4) self-reported no history of stroke, head injury or progressive neurogenic disorders. For demographic information for the control participants, see table 1.

All PWA met the following criteria: (1) native English speaker; (2) passed hearing and vision screenings; (3) no reported history of other neurological disorders; (4) presented with aphasia as determined by performance on the Western Aphasia Battery (WAB)–Revised Aphasia Quotient (AQ) subtests; (5) has chronic aphasia (at least 6 months post-onset); and (6) has left hemisphere damage. A convenience sample of PWA was recruited from local support groups and university speech–language–hearing clinics. Two participants with aphasia (P2 and P8) were disqualified from the study due to other neurological disorders. Aphasia type for the PWA was determined by the taxonomic classification system of the WAB-R. For group comparisons, the participants were divided into fluent and non-fluent types of aphasia, as determined by the WAB-R criterion, using performance on the spontaneous speech subtest. According to the WAB-R classification, fluent types of aphasia (including anomic, conduction, transcortical sensory and Wernicke's aphasia) display comparatively intact syntactical form, utterances of moderate length and ease of articulation, whereas non-fluent types of aphasia (including transcortical motor, Broca's, mixed transcortical and global aphasia) use a limited grammatical form with short utterances and a reduced intonation. In the spontaneous speech subtest, individuals acquiring ratings from 0 to 4 are considered non-fluent, and those with ratings from 5 to 10 are considered fluent.[2] For demographic information for the PWA, see table 2.

Experimental procedures

All participants were tested individually in a laboratory setting. Participants provided informed consent for the study before beginning the experimental tasks. Since the normative data were collected for a large study, the cognitively healthy participants attended two sessions,

Table 2. Participants with aphasia demographic information

	Age (years)	Gender (male/female)	Education (years)	WAB-R AQ[a]	Aphasia type (fluent/non-fluent)[b]
P1	65	M	18	76.3	Conduction (fluent)
P3	73	M	12	85.2	Anomic (fluent)
P4	84	F	12	62.6	Conduction (fluent)
P5	55	M	14	57.6	Broca's (non-fluent)
P6	66	F	14	56.3	Broca's (non-fluent)
P7	34	F	14	90.7	Anomic (fluent)
P9	38	F	14	57.7	Broca's (non-fluent)
P10	62	F	20	61.3	Broca's (non-fluent)
P11	72	M	12	64.9	Transcortical motor (non-fluent)
P12	65	F	11	89.4	Anomic (fluent)
P13	65	M	14	54.4	Broca's (non-fluent)
Mean	61.7		14.1	68.8	
SD	14.7		2.7	14.0	

Notes: [a]Western Aphasia Battery (WAB)—Revised Aphasia Quotient (AQ) (Kertesz 2006).
[b]Fluency determined by the WAB-R classification.

lasting no more than 2 h for each session. The PWA participants attended one session, lasting approximately 2 h. All participants who were PWA were administered the WAB-R first, followed by cognitive and discourse tasks. The order of the experimental tasks and discourse tasks was randomized across participants.

Discourse elicitation tasks

Two wordless picture books were used to collect discourse samples from participants. They included *Good Dog Carl* (*GDC*; Day 1985) and *Picnic* (McCully 1984). This type of story-generation task involves various aspects of natural communication (Fergadiotis 2011). In story books following the schema of a typical Western traditional story, unpredictable events occur and the main characters arrive at the highest peak of tension as the story proceeds. These story structures also trigger internal responses in speakers. Moreover, during the task, speakers produce diverse lexical items that need to be retrieved for the characters, settings and events described in the book (Fergadiotis *et al.* 2011, Wright *et al.* 2011). *GDC* is a 30-page book that follows a temporally driven story structure conveying the events that unfold as a dog is left to take care of a baby. *Picnic* is a 31-page story that represents a spatially and temporally driven story structure conveying the adventures of a family of mice going on a picnic.

For the discourse task, the examiner provided an explicit task instruction to elicit the core event line of the story books, given the potential impact of elicitation instructions on participants' narrations (Olness 2006, Wright and Capilouto 2009). The examiner read the following script as instructions: 'These are wordless picture books that allow an individual to make up their own story. First, I'll look through the book to get an idea of the story. Then, I will start at the beginning and tell you the story that goes with the pictures.' The examiner provided an example of how to tell a story using a wordless picture book entitled *The Great Ape* (Krahn 1978), which is comparable in length and includes 36 pages. Similar to experimental tasks (*GDC* and *Picnic*), this story includes an introduction of the main characters and the setting (a group of people, including a little girl, travel to an island). The main events are placed along a temporal succession with detailed descriptions (while people are watching a great ape swinging between mountains, a little girl falls off a cliff and the ape catches her). The story ends with the resolution of the problem (people attempt to rescue the girl). (For the scripted story of *The Great Ape*, see appendix A.) Participants were then presented with the book and allowed to look through it for as long as they needed. The stimuli were viewable until participants finish storytelling. If a participant stopped speaking after $\leq$ 15 s, the examiner prompted him/her by asking if they had more information to provide. The order of the picture books (*GDC*, *Picnic*) was randomized across participants.

Language sample preparation

All samples were either audio or video recorded, and then orthographically transcribed by trained research assistants in CHAT of the CLAN program. Inter- and intra-rater reliability for word-by-word transcription were determined for 10% of each participant group (i.e., the control group and the aphasia group). For the control group, inter- and intra-rater agreements were 95% and 98%, respectively. For the aphasia group, two PWA were randomly selected for interrater reliability, and two different PWA were selected for intra-rater reliability, due to the small number of participants. Inter- and intra-rater agreements were 91% and 93%, respectively.

To generate the core lexicon list for function words based on the normative data, we used the GEM, MOR and FREQ programs associated with CLAN. GEM extracts the specific stories under analysis from the larger discourse sample and creates a separate file for each participant containing the specific stories (i.e., *GDC* and *Picnic*) under review. To automate the process of finding function words, MOR was used to assign automatically a syntactic category to each word (for reviews, see MacWhinney 2000 and MacWhinney *et al.* 2010). The MOR program uses a dictionary of lexical items and English grammar rules to assign each word to its respective syntactic category. It has an accuracy of 95% (for a review, see MacWhinney *et al.* 2010). The FREQ program extracts words, word classes or other coded items from the discourse samples and generates a frequency list. FREQ was used to generate a list of all the function words produced within the transcripts along with their frequency information. For this study, the function word lists included pronouns, determiners, prepositions, conjunctions, coordinators, quantifiers, negatives and copula verbs (e.g., be, look). For the narrative task, the top 25 most frequent function words were used to create the core function word lists for each age group. While the top 25 most frequent function words is an arbitrary cut-off, previous researchers have used similar numbers (Dalton and Richardson 2015).

Core function word agreement

The core function word lists were used to calculate the percent agreement for the function words produced by the PWA using the CLAN code. A FREQ code was generated that would automatically count whether a participant produced any of the core function words from the appropriate list (i.e., the matching age and narrative task list), and the number of times the core function word was used. Percent agreement was determined by giving 1 point for each function word that was part of the core function word list produced by participants, regardless of the number of times the word was produced. For example, if a PWA produced '*the*' 26 times and '*and*" 12 times in language samples of story A, and both are included in the core lexicon list, the participant would receive 2 points out of 25 (the total number of function word items on the list). Percent agreement of the PWA was calculated by dividing 2 (numerator) by 25 (denominator) in the fraction, then multiplying by 100.

Results

Preliminary analyses

The current study was motivated by the core lexicon framework, as well as by the adaptive mechanisms of PWA. As such, following previous work, preliminary analyses were conducted to determine: (1) if age cohorts significantly differed between the percentage of core function words produced for *GDC* and *Picnic*; and (2) if the percentage of core function words produced by PWA significantly correlated with overall language severity as determined by the AQ.

To address the first preliminary analysis, two one-way analyses of variance (ANOVAs) were conducted with age cohort as the independent variable and the percentage of core function words produced as the dependent variable for *GDC* and *Picnic*. Following the procedures outlined by Tabachnick and Fidell (1996), the data for *GDC* contained four outliers < –3.0 SD (standard deviations) from the mean as determined by converting the percentages of core function words produced to *z*-scores. *Picnic* contained one outlier < –3.0 SD. Levene's test of equality of variance was not significant for *GDC*, $p = 0.90$, or *Picnic*, $p = 0.71$, indicating the error variance was equal across groups. Normality was assessed by visual inspection of the histograms and dividing skewness scores with skewness error and dividing the kurtosis scores by the kurtosis error to create a *z*-score where any scores between –3.3 and 3.3 is considered normal. For *Picnic*, the 50s cohort had a slight negative skew with a *z*-score of –3.34. The kurtosis scores were within an acceptable range. For *GDC*, there was moderate skewing for the 50s ($Z = -4.11$), 60s ($Z = -4.06$), 70s ($Z = -4.47$) and 80s ($Z = -3.91$) cohorts. The kurtosis scores also had a moderate deviation for the 50s ($Z = 4.29$) and 70s ($Z = 4.43$) cohorts. Removing the four outliers corrected both kurtosis and *z*-scores, and corrected all but the 80s cohort ($Z = -3.39$) for *GDC*. While ANOVA tests have recently demonstrated to be robust for moderate to severe violations of normality (Blanca *et al.* 2017), we conducted ANOVAs for *Picnic* and *GDC* with outliers included and outliers removed, as well as transforming the data to meet the requirement of normality.

For *GDC* with outliers, there was a significant difference in the percentage of core function words produced across the age cohorts, $F(6, 463) = 2.13$, $p < 0.05$, $\eta_p^2 = 0.03$. Removing outliers produced a lower *p*-value, $F(6, 459) = 2.53$, $p = 2.53$, $\eta_p^2 = 0.03$. Post-hoc comparisons with a Tukey HSD correction found that only the 30s (mean = 86.54, SD = 1.11) and 80s (mean = 90.67, SD = 1.07) cohorts were significantly different when outliers were removed. There was no significant difference in our post-hoc comparisons with outliers. These results indicate that the older adults produced more core function words from the age invariant list compared with younger adults. A one-way ANOVA conducted on data transformed for mild to moderate negative skewing (x = SQRT [–6 – x]) did not change the results. For the age-invariant core function word lists for *GDC*, see appendix B.

For *Picnic*, there was a significant difference between age cohorts and the number of core function words produced, $F(6, 463) = 4.78$, $p < 0.001$, $\eta_p^2 = 0.06$. The removal of the single outlier did not change the results, $F(6, 462) = 4.83$, $p < 0.001$ $\eta_p^2 = 0.06$. Post-hoc comparisons with Tukey HSD correction demonstrated that the 20s (mean = 83.70, SD = 0.96) cohort differed significantly from the 60s (mean = 87.52, SD = 0.82), 70s (mean = 88.58, SD = 0.77) and 80s (mean = 88.00, SD = 0.87) cohorts. The 30s cohort (mean = 84.76, SD = 0.78) also differed significantly from the 70s cohort (mean = 88.58, SD = 0.77). Removing the outlier did not also change the post-hoc results. These results indicate that older adults produce more core function words compared with younger adults. A one-way ANOVA conducted on the data transformed for mild to moderate negative skewing (x = SQRT [–6 – x]) did not change the results. For the age-invariant core function word lists for *Picnic*, see appendix B. The results indicate that the percentage of core function words produced for *Picnic* and *GDC* differed by age. For this reason, the age-corrected list will be used to address the second preliminary question and main research questions. For the age-corrected lists for *GDC* and *Picnic*, see appendix C.

To address the second preliminary question of whether core function words produced by PWA significantly correlated with aphasia severity, a bivariate Spearman's correlation was conducted on the percentage of function words produced by 11 PWA from the age-corrected list for *GDC* and *Picnic*. Following Goodwin and Leech's (2006) guidelines, Spearman's correlations were used because: (1) visual inspection of the histograms, as well as Q-Q plots, determined a lack of normality; (2) large SDs suggested larger variability potentially inflating r-scores; and (3) the small n ($n < 30$) can exacerbate any problems associated with non-normality and larger variability. For *GDC*, AQ significantly correlated with the percentage of core function words produced, $r = 0.90$, $p < 0.001$. For *Picnic*, AQ also significantly correlated with the percentage of core function words produced, $r = 0.64$, $p < 0.05$.

Main analyses

To address the first research question of whether there would be a difference in the percentage of core lexicon produced for the PWA and control groups, 11 PWA and 11 age- and education-matched control participants were selected. Because of the low N and inability to determine normality, a Mann–Whitney test was conducted on age, education and number of core function words produced for *GDC* and *Picnic*. The test demonstrated that PWA and control participants were correctly matched for age, $Z < 0.001$, $p > 0.99$, and education, $Z = –0.24$, $p = 0.85$. There were significant differences between the two groups for *GDC*, $Z = –3.68$, $p < 0.001$, and for *Picnic*, $Z = –3.70$, $p < 0.001$. These results indicate that cognitively healthy participants produced significantly more core function words compared with PWA for both narratives. For the percent agreement for core function words by story for each group, see Table 3.

To address the second research question of whether participants with fluent ($N = 5$) and non-fluent aphasia ($N = 6$) significantly differed in the percentage of core function words produced, a Mann–Whitney test was conducted. There were significant differences between participants with fluent and non-fluent aphasia for *GDC*, $Z = –2.62$, $p < 0.01$, and *Picnic*, $Z = –2.75$, $p < 0.01$. For *GDC*, participants with fluent aphasia (mean = 74.40%, SD = 6.69) produced more core function words than participants with non-fluent aphasia (27.20%, SD = 5.93). For *Picnic*, the trend continued with participants with fluent aphasia (mean = 69.60%, SD = 8.76) produced more core function words than participants with non-fluent aphasia (mean = 15.20%, SD = 4.38).

Discussion

The purpose of this exploratory study was to develop core function word lists that could potentially be used for clinical settings, and to explore whether they capture differences between PWA and persons without aphasia, and between persons with fluent and non-fluent aphasia. Due to the exploratory nature of the study, we performed preliminary analyses before the main analyses. The results of the preliminary analyses indicated that age-related differences were found for core function words produced in both story narratives. As such with determining the relationship between function word production and overall aphasia severity, age-corrected core function word lists for each stimulus were used. The second preliminary analysis indicated that function word production using these core function word lists significantly correlated with aphasia severity. The results of the main analyses indicated that core function word lists discriminated PWA from age- and education-matched controls, and also discriminated persons identified by their performance on the WAB-R with non-fluent aphasia from persons with fluent aphasia. What follows is a discussion of the results and potential clinical implications.

Table 3. Percent agreement for *Good Dog Carl* (*GDC*) and *Picnic* in age-invariant core function word lists

	PWA[a]			Controls[b]	
ID	*GDC*[c]	*Picnic*[c]	ID	*GDC*[c]	*Picnic*[c]
P1	68	72	6050	84	92
P3	88	76	7076	88	88
P4	64	68	8040	100	100
P5	24	32	5036	92	84
P6	NA	24	6016	92	92
P7	80	68	3034	96	88
P9	20	16	3008	88	76
P10	44	20	6037	96	92
P11	32	8	7013	100	96
P12	72	88	6014	96	92
P13	24	20	6018	84	84
Mean	46.9	44.7		92.4	89.5
SD	28.9	29.4		5.8	6.5

Notes: [a]Persons with aphasia.
[b]Age- and education-matched PWA counterparts.
[c]Percent agreement; maximum = 100.

Function word production and aphasia

The age- and education-matched healthy controls performed significantly better on core function word production than PWA. These findings are not surprising as it is generally well known that PWA differ from cognitively healthy adults on measures of function words (Chapman and Ulatowska 1989, Gleason *et al.* 1980, Manning and Franklin 2016, Nicholas *et al.* 1985). Moreover, it is very promising that the small number of function word items provided by core function word lists detected abnormal performance in PWA from the control group. The three participants with the highest AQs on the WAB-R (P3, P7, P12) in our sample produced fewer core function word items compared with their control counterparts. These results suggest that performance of cognitively healthy speakers provides a firm baseline from which to examine the degree of language impairment in clinical populations, as suggested by Webster *et al.* (2007).

Worthy of consideration is that function word production using these core function word lists significantly correlated with overall aphasia severity in our analysis. The ability to produce core function words may be associated with overall aphasia severity as determined by the WAB-R AQ. Individuals with more severe aphasia produced fewer core lexicon items than individuals with less severe aphasia. Based on the results, it appears that PWA tend to use some strategy to compensate for their communication difficulties so that they achieve better performance under unfavourable cognitive and linguistic conditions (Isserlin 1922, Kolk 1998; for a review, see Cui and Zhong 2018). Within the adaptation theory framework suggested by Kolk and colleagues, the omission of function words, which are comparatively less informative for delivering content of the presented pictorial stimuli compared with content words, or which may be the most impaired for certain aphasia types, has been suggested to be the most typical feature of the strategy. Our findings suggest the possible use of an elliptical strategy with function words in PWA not exclusively confined to agrammatism or non-fluent output. As mentioned previously, the focus of the current study was to investigate the potential utility of core function word measures for time-conscious clinicians. Providing a definitive explanation of relationships between PWA's discourse production output and such adaptation strategy is beyond the scope of this investigation. Thus, testing this hypothesis should be considered in future investigations.

Consistent with previous studies, function word production was a clinical marker used to discriminate non-fluency (Gordon 2008 2006, Manning and Franklin 2016). Our six PWA who presented with a non-fluent type of aphasia produced fewer core function words compared with persons with fluent aphasia. The findings are not surprising, as previous studies reported similar patterns for persons with fluent and non-fluent aphasia during connected speech—participants with non-fluent aphasia (i.e., Broca's, transcortical motor aphasia) produced fewer closed class words and use of determiners compared with participants with fluent aphasia (i.e., anomic, Wernicke's, transcortical sensory aphasia) (Gordon 2008, 2006). These findings could be interpreted as evidence of a trade-off between semantic and syntactic processes occurring in fluent and non-fluent aphasia in opposite directions (Gordon 2008). PWA with relatively intact syntactic ability (i.e., persons with fluent aphasia types) tended to show a reduced semantic ability, indicating that these individuals produced a higher proportion of syntactically laden words, such as function words, in connected speech. PWA with decreased syntactic ability (i.e., persons with non-fluent aphasia types) tended to rely on semantic processing, which led to the production of a higher proportion of semantically laden words and a lower proportion of syntactically laden words. These theoretical findings and interpretations inform the dichotomous fluent/non-fluent aphasia distinction and their typical characterization in language skills. However, the generalizations of language profiles in PWA should be made cautiously, considering variability within individuals' symptoms and resources, particularly in clinical practices. As a final note, we acknowledge that there are some concerns about the circularity nature of these findings driven by the fact that fluency classification in the WAB-R depends partly on function word use.

Although the exploratory nature of the study does not permit determination of clinical use of the core function word lists, we speculate that how the core function word lists were constructed contributed to the current study's statistically significant findings. Earlier studies of function words provide evidence that both semantic and syntactic processes are activated when function words are processed in context (e.g., Colé and Segui 1994, Friederici 1982, 1985, Friederici *et al.* 2000, Hinojosa *et al.* 2001). Within an utterance, function words are presented between content words, and thus function word use requires adherence to syntactic and semantic constraints in response to communicative intentions. Moreover, both semantic and syntactic information come into play when prepositions and pronouns are produced (e.g., Bird *et al.* 2002, Friederici 1982, Grodzinsky 1984). For example, the pronoun *she* indicates a singular female image, requiring an active, semantic representation of the antecedent. Pronouns and prepositions account for a substantial proportion of the core function words listed (see appendix B). Together, by creating subsets of core function words depending on the relative weights of syntactic and semantic processing, future studies could investigate the utility of core function word production for quantifying differences among aphasia subtypes. This will help us to broaden our understanding of function word processing in PWA's connected speech.

Measurement issues

The core function word measure provides an advantage over other methods of quantifying function word production because it can potentially reduce the workload burden on clinicians by providing a limited number of critical lexical items in checklists. However, it can be argued that the use of core function word lists does not allow for the identification of error production of function words in PWA's language samples. As noted above, Dalton and Richardson (2015) demonstrated that counting the presence of lexical items is a valid means of quantifying word-retrieval ability at the discourse level. Moreover, our approach was motivated by the idea of the adaptation strategy that a handful of studies have investigated regarding the production of simplified utterances (i.e., omission of function words) in PWA's discourse, which supports this way of scoring (e.g., Kolk 1995, Kolk and Heeschen 1992, De Roo *et al.* 2003, Ruiter *et al.* 2010 2013, Salis and Edwards 2004). The concept of the adaptation strategy in which the omission of function words in PWA's discourse could be a manifestation of a reduced linguistic and/or cognitive capacity has been theoretically substantiated. Considering the limited time and resources in clinical settings, defining a grammatical category for each word in language samples in traditional ways may deter the use of discourse analyses. Identifying the error production and grammatical category of function words in discourse produced by PWA is admittedly important in quantifying function word production, it may come at the expense of clinical utility of the measure. Development of the core lexicon measure as a clinician-friendly method for discourse analysis is still in its early stage. Subsequent studies are needed to determine reliability and validity of the measure.

Since core lexicon analysis is a relatively new measure, no consensus exists among researchers for defining the predetermined core lexicon items. MacWhinney *et al.* (2010) generated 10 core nouns and 10 core verbs from a discourse task (approximately 20% of all lexicon items produced). Fromm *et al.* (2013) identified 10 nouns and 10 verbs based on how frequently the lexical items were produced by the control participants. Dalton and Richardson (2015) aggregated all word classes in a core lexicon list with lexical items produced by > 50% of the sampling cohort. Only the core lexicon list developed by Dalton and Richardson included function words. In developing a new language test, including items of varying difficulty enhances the sensitivity of indexing language impairments (Ivanova and Hallowell 2013). Further, while there is a precedent for defining text from a frequency list (Gottron 2009), the cut-off we used (25 most frequently produced function words) was arbitrarily determined with ease of use being the most important factor in that decision. Thus, the number of items in our core lexicon list was similar to the numbers identified in previous studies (Dalton and Richardson 2015). Future studies should consider a systematic approach to the establishing criterion for determining core lexical list length and items included.

Conclusions and limitations

The aim of this study was to provide a starting point for the development of a clinical tool for quantification of function word production in discourse produced by PWA. Clinical challenges with regards to language assessment at the discourse level have been well known (e.g., Bryant *et al.* 2016, Dietz and Boyle 2018, Elia *et al.* 1994, Maddy *et al.* 2015). Particularly, measuring function word production at the discourse level requires an additional process of determining grammatical categories and proper use of them (Pak-Hin and Law 2004); however, it is critical to identify clinical characteristics of PWA. Thus, motivated by the concept of adaptation, we extended previous work of core lexicon analysis by generating individual core function word lists. Our findings demonstrated different production of core function words between the control and PWA groups, and between persons with fluent and non-fluent aphasia. Although theoretical explanation of simplified utterances with omission of function words remains obscure, our findings serve as evidence that the presence and absence of core function words may be an additive approach for assessing fluency at the discourse level in clinical practice.

Despite the potential clinical usability of the core function word measure, there are several notable limitations with the current investigation. A first limitation is related to our participants and their clinical subgroupings determined by the WAB-R, which is influenced by function word use for the fluency classification. Because the core function word measure was designed for clinical utilization, the WAB-R was used for classification of subgroups (fluent and non-fluent types of aphasia) for the main analyses, as it is one of the most popular standardized tests in clinical settings. Since it has been suggested that there is a lack of reliability in clinical judgements on fluency according to WAB-R scoring (Clough and Gordon 2020, Gordon 1998, Holland *et al.* 1986, Nozari and Faroqi-Shah 2017), a comprehensive approach to determine subgroups of aphasia would improve our understanding of appropriateness of this measure as a clinical outcome measure. Another related limitation is that our participants with non-fluent types of aphasia fortuitously have more severe aphasia compared with their counterparts (fluent types of aphasia). It is likely that the nature of our study participants could be associated with our statistical findings. Further, it is quite possible that our participants with fluent types of aphasia, whose aphasia is also comparatively less severe, produced longer samples containing more core function word items, and earned higher scores on the core lexicon measure. Based on the property of core lexicon measures providing a closed set of function word items, it is likely to offset the effect of sample length. However, as is common with new measures, investigating potential confounds that may influence measurement will be necessary to improve the quality of measurement.

Future directions

In this investigation of the potential clinical utility of core lexicon measures, we did not discriminate proper use of function words for scoring, which is similar to previous research (Dalton and Richardson 2015) and should be considered in future studies. The benefit of this scoring system may have potential clinical utility for clinicians by providing a time-efficient tool. Moreover, the most fascinating feature of core lexicon measures is that it offers multiple subtests for one language sample. Core content word lists by word class (nouns, verbs, adjectives, adverbs) have been previously developed and validated based on the same narrative discourse tasks (Kim and Wright 2020a, Kim *et al.* 2019). Using core function word lists with content word lists will allow clinicians to obtain a fuller and more accurate picture in PWA's overall language ability. Simultaneously, time spent in testing can be reduced by using the explicit instructions (identified in the method) that make PWA describe core lines of event in multiple pictures provided, rather than describing all pictured objects and actions. In the current study, the time required for the discourse tasks to be finished ranged from 5 to 15 min, with the exception of P13 who has non-fluent aphasia with the lowest AQ out of all PWA. Nonetheless, the simplified way of scoring is viewed as exploratory and requires replication with a larger sample of participants with aphasia to strengthen conclusions regarding clinical feasibility.

Although core function word lists are devised to benefit clinical practices, it contains the most occurring 25 function words irrespective of subcategory. As mentioned above, function words possess different and unique properties, and certain function words are more involved in either semantic or syntactic processing than others (De Roo *et al.* 2003, Gordon 2006, Kemmerer 2005, Ruigendijk and Bastiaanse 2002, Salis and Edwards 2004). Kemmerer (2005) reported that within the preposition category, spatial and temporal prepositions are independently processed, and could be separately impaired. Thus, future studies should further determine core function word use by subcategory in PWA and to what extent subcategories are useful in clinical practice.

Finally, PWA's discourse performance has shown to be dissimilar in response to different discourse elicitation tasks that impose cognitive and linguistic demands on speakers (e.g., Armstrong 2000, Coelho 2002). Depending on the degree of contextual support in the illustrations provided, it is possible that PWA retrieve core lexicon items more easily in one task compared with others. At the same time, the inconclusive theoretical explanation on the uniqueness of scoring in the current method is involved in PWA's cognitive burden, which can result in simplifying their utterances without function words. A comprehensive approach to illustrate intertwined relationships among cognitive resources, discourse tasks, and function word production at the discourse level will enhance the validity of the core lexicon measure.

Acknowledgements

This research was partially supported by National Institute on Aging (grant number R01AG029476). The authors are especially grateful to the study participants. The data supporting the findings of this study are available from the corresponding author upon reasonable request. ***Declaration of interest***: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Notes

1. The aphasia types referred to represent the terms used in the reviewed manuscripts. In the traditional aphasia classification system, aphasia can be classified as either a fluent type or a non-fluent type. Broca's aphasia, one non-fluent type of aphasia, is the prototypical non-fluent aphasia. Agrammatism, impairment in syntactic processing (comprehending and/or producing), is a common feature of Broca's aphasia. However, it should be noted that Broca's aphasia is not always accompanied by agrammatism.
2. Using a standardized language battery is often a reasonable first step in determining subtypes of aphasia in research and clinical practices. Considering the widespread use of the WAB-R test in clinical assessment, we used it to identify fluency in grouping our participants. However, it is important to acknowledge that fluency classification using rating scales such as the WAB-R is not consistently reliable (Gordon 1998).

References

Armstrong, E., 2000, Aphasic discourse analysis: the story so far. *Aphasiology*, **14**, 875–892.

Bates, E., Chen, S., Tzeng, O., Li, P. and Opie, M., 1991, The noun–verb problem in Chinese aphasia. *Brain and Language*, **41**, 203–233. https://doi.org/10.1016/0093-934X(91)90153-R.

Beukelman, D. R. and Mirenda, P., 1998, *Augmentation and Alternative Communication* (Baltimore, MD: Paul H. Brookes Publication).

Bird, H., Franklin, S. and Howard, D., 2002, 'Little words'—not really: function and content words in normal and aphasic speech. *Journal of Neurolinguistics*, **15**, 209–237.

Blanca, M., Alarcón, R., Arnau, J., Bono, R. and Bendayan, R., 2017, Non-normal data: is ANOVA still a valid option? *Psicothema*, **29**, 552–557.

Bryant, L., Ferguson, A. and Spencer, E., 2016, Linguistic analysis of discourse in aphasia: a review of the literature. *Clinical Linguistics & Phonetics*, **30**, 489–518.

Bryant, L., Spencer, E. and Ferguson, A., 2017, Clinical use of linguistic discourse analysis for the assessment of language in aphasia. *Aphasiology*, **31**(**10**), 1105–1126.

Caramazza, A. and Hillis, A. E., 1991, Lexical organization of nouns and verbs in the brain. *Nature*, **349**, 788–790. https://doi.org/10.1038/349788a0.

Chapman, S. B. and Ulatowska, H. K., 1989, Discourse in aphasia: integration deficits in processing reference. *Brain and Language*, **36**, 651–668.

Chen, S. and Bates, E., 1998, The dissociation between nouns and verbs in Broca's and Wernicke's aphasia: findings from Chinese. *Aphasiology*, **12**, 5–36. https://doi.org/10.1080/02687039808249441.

Clough, S. and Gordon, J. K., 2020, Fluent or nonfluent? Part A. Underlying contributors to categorical classifications of fluency in aphasia. *Aphasiology*, **34**(**5**), 515–539. https://doi.org/10.1080/02687038.2020.1727709.

Coelho, C. A., 2002, Story narratives of adults with closed head injury and non-brain-injured adults: influence of socioeconomic status, elicitation task, and executive functioning. *Journal of Speech, Language, and Hearing Research: JSLHR*, **45**, 1232–1248. https://doi.org/10.1044/1092-4388(2002/099)

Colé, P. and Segui, J., 1994, Grammatical incongruency and vocabulary types. *Memory & Cognition*, **22**, 387–394. https://doi.org/10.3758/BF03200865.

Cui, G. and Zhong, X., 2018, Adaptation in aphasia: revisiting language evidence. *Aphasiology*, **32**, 855–875. https://doi.org/10.1080/02687038.2018.1458068.

Dalton, S. G. and Richardson, J. D., 2015, Core-lexicon and main-concept production during picture-sequence description in adults without brain damage and adults with aphasia. *American Journal of Speech–Language Pathology*, **39**, 1125–1137. https://doi.org/10.1044/2015_AJSLP-14-0161.

Dalton, S. G. H., Hubbard, H. I. and Richardson, J. D., 2020, Moving toward non-transcription based discourse analysis in stable and progressive aphasia. *Seminars in Speech and Language*, **41**(**1**), 32–44.

Davis, H. and Silverman, S., 1978, *Hearing and Deafness*. 4th ed. (New York: Holt, Rinehart, & Winston.).

Day, A., 1985, *Good Dog, Carl* (New York: Scholastic).

Dietz, A. and Boyle, M., 2018, Discourse measurement in aphasia research: have we reached the tipping point? *Aphasiology*, **32**, 459–464.

Elia, D., Liles, B. Z., Duffy, R. J., Coelho, C. A. and Belanger, S. A., 1994, An investigation of sample size in conversational analysis. In *ASHA Convention, New Orleans, LA*.

Fergadiotis, G., 2011, *Modeling Lexical Diversity Across Language Sampling and Estimation Techniques* (Tempe: Arizona State University).

Fergadiotis, G., Wright, H. H. and Capilouto, G. J., 2011, Productive vocabulary across discourse types. *Aphasiology*, **25**, 1261–1278. https://doi.org/10.1080/02687038.2011.606974.

Folstein, M. F., Folstein, S. E., McHugh, P. R. and Fanjiang, G., 2001, *Mini-Mental State Examination: MMSE-2* (Lutz, FL: Psychological Assessment Resources).

Friederici, A. D., 1982, Syntactic and semantic processes in aphasic deficits: the availability of prepositions. *Brain and Language*, **15**, 249–258.

Friederici, A. D., 1985, Levels of processing and vocabulary types: evidence from on-line comprehension in normals and agrammatics. *Cognition*, **19**, 133–166. https://doi.org/10.1016/0010-0277(85)90016-2.

Friederici, A. D., Opitz, B. and Von Cramon, D. Y., 2000, Segregating semantic and syntactic aspects of processing in the human brain: an fMRI investigation of different word types. *Cerebral Cortex*, **10**, 698–705.

Fromm, D. A., Forbes, M., Holland, A. and MacWhinney, B., 2013, PWAs and PBJs: language for describing a simple procedure. In: Paper presented at Clinical Aphasiology Conference, Tucson, Arizona, 28 May–2 June.

Gleason, J. B., Goodglass, H., Obler, L., Green, E., Hyde, M. R. and Weintraub, S., 1980, Narrative strategies of aphasic and normal-speaking subjects. *Journal of Speech, Language, and Hearing Research*, **23**, 370–382.

Goodglass, H., Kaplan, E. and Barresi, B., 2001. *BDAE-3: Boston Diagnostic Aphasia Examination—Third Edition* (Philadelphia: Lippincott Williams & Wilkins).

Goodwin, L. D. and Leech, N. L., 2006, Understanding correlation: factors that affect the size of r. *The Journal of Experimental Education*, **74**, 249–266.

Gordon, J. K., 1998, The fluency dimension in aphasia. *Aphasiology*, **12**, 673–688.

Gordon, J. K., 2008, Measuring the lexical semantics of picture description in aphasia. *Aphasiology*, **22**, 839–852. https://doi.org/10.1080/02687030701820063.

Gordon, J. K., 2006, A quantitative production analysis of picture description. *Aphasiology*, **20**, 188–204. https://doi.org/10.1080/02687030500472777.

Gottron, T., 2009, Document word clouds: visualising web documents as tag clouds to aid users in relevance decisions. In *International Conference on Theory and Practice of Digital Libraries* (Berlin, Heidelburg: Springer), pp. 94–105.

Grodzinsky, Y., 1984, The syntactic characterization of agrammatism. *Cognition*, **16**, 99–120.

Halliday, M. A. K. and Hasan, R., 1976, *Cohesion in English*. (London: Longman).

Hinojosa, J. A., Martin-Loeches, M., Casado, P., Munoz, F., Carretie, L., Fernandez-Frias, C. and Pozo, M. A., 2001, Semantic processing of open- and closed-class words: an event-related potentials study. *Cognitive Brain Research*, **11**, 397–407.

Holland, A. L., Fromm, D. and Swindell, C. S., 1986, The labeling problem in aphasia: An illustrative case. *Journal of Speech and Hearing Disorders*, **51**, 176–180.

Isserlin, M., 1922, Ueber Agrammatisms. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, **75**, 332–410.

Ivanova, M. V. and Hallowell, B., 2013, A tutorial on aphasia test development in any language: key substantive and psychometric considerations. *Aphasiology*, **27**, 891–920. https://doi.org/10.1080/02687038.2013.805728.

Kemmerer, D., 2005, The spatial and temporal meanings of English prepositions can be independently impaired. *Neuropsychologia*, **43**, 797–806.

Kertesz, A., 2006, *Western Aphasia Battery-Revised (WAB-R)* (San Antonio, TX: Pearson).

Kim, H. and Wright, H. H., 2020a, Concurrent validity and reliability of the core lexicon measure as a measure of word retrieval ability in aphasia narratives. *American Journal of Speech–Language Pathology*, **29**(**10**), 101–110.

Kim, H. and Wright, H. H., 2020b, A tutorial on core lexicon: development, use, and application. *Seminars in Speech and Language*, **41**(**10**), 20–31.

Kim, H., Kintz, S., Zelnosky, K. and Wright, H. H., 2019, Measuring word retrieval in narrative discourse: core lexicon in aphasia. *International Journal of Language & Communication Disorders*, **54**, 62–78.

Kim, M. and Thompson, C. K., 2000, Patterns of comprehension and production of nouns and verbs in agrammatism: Implications for lexical organization. *Brain and Language*, **74**, 1–25. https://doi.org/10.1006/brln.2000.2315.

Kolk, H., 1995, A time-based approach to agrammatic production. *Brain and Language*, **50**, 282–303. https://doi.org/10.1006/brln.1995.1049.

Kolk, H., 1998, Disorders of syntax in aphasia: linguistic-descriptive and processing approaches. In *Handbook of Neurolinguistics* (New York: Academic Press), pp. 249–260.

Kolk, H. and Heeschen, C., 1990, Adaptation symptoms and impairment symptoms in Broca's aphasia. *Aphasiology*, **4**, 221–231. https://doi.org/10.1080/02687039008249075.

Kolk, H. and Heeschen, C., 1992, Agrammatism, paragrammatism and the management of language. *Language and Cognitive Processes*, **7**, 89–129. https://doi.org/10.1080/01690969208409381.

Kolk, H. and Heeschen, G., 1996, The malleability of agrammatic symptoms: A reply to Hesketh and Bishop. *Aphasiolgy*, **10**(**1**), 81–96.

Krahn, F., 1978, *The great ape* (New York: The Viking Press).

Luzzatti, C. and Chierchia, G., 2002, On the nature of selective deficits involving nouns and verbs. *Italian Journal of Linguistics*, **14**, 43–72.

MacWhinney, B., 2000, *The CHILDES Project: Tools for Analyzing Talk: Volume I: Transcription Format and Programs (3rd ed.)* (Mahwah, NJ: Lawrence Erlbaum Associates Inc.).

MacWhinney, B., Fromm, D., Holland, A., Forbes, M. and Wright, H., 2010, Automated analysis of the Cinderella story. *Aphasiology*, **24**, 856–868. https://doi.org/10.1080/02687030903452632.

Maddy, K. M., Howell, D. M. and Capilouto, G. J., 2015, Current practices regarding discourse analysis and treatment following non-aphasic brain injury: a qualitative study. *Journal of Interactional Research in Communication Disorders*, **6**, 211.

Manning, M. and Franklin, S., 2016, Cognitive grammar and aphasic discourse. *Clinical Linguistics & Phonetics*, **30**, 417–432.

McCully, E. A., 1984, *Picnic* (New York: Harper & Row).

Nicholas, M., Obler, L. K., Albert, M. L. and Helm-Estabrooks, N., 1985, Empty speech in Alzheimer's disease and fluent aphasia. *Journal of Speech, Language, and Hearing Research*, **28**, 405–410.

Nozari, N. and Faroqi-Shah, Y., 2017, Investigating the origin of nonfluency in aphasia: A path modeling approach to neuropsychology. *Cortex*, **96**, 119–135.

Olness, G. S., 2006, Genre, verb, and coherence in picture-elicited discourse of adults with aphasia. *Aphasiology*, **20**, 175–187.

Pak-Hin, A. K. and Law, S. - P., 2004, A Cantonese linguistic communication measure for evaluating aphasic narrative production: normative and preliminary aphasic data. *Journal of Multilingual Communication Disorders*, **2**, 124–146.

Rochon, E., Saffran, E. M., Berndt, R. S. and Schwartz, M. F., 2000, Quantitative analysis of aphasic sentence production: further development and new data. *Brain and Language*, **72**, 193–218. https://doi.org/10.1006/brln.1999.2285.

De Roo, E., Kolk, H. and Hofstede, B., 2003, Structural properties of syntactically reduced speech: a comparison of normal speakers and Broca's aphasics. *Brain and Language*, **86**, 99–115. https://doi.org/10.1016/S0093-934X(02)00538-2.

Ruigendijk, E. and Bastiaanse, R., 2002, Two characteristics of agrammatic speech: omission of verbs and omission of determiners, is there a relation? *Aphasiology*, **16**, 383–395. https://doi.org/10.1080/02687030244000310.

Ruiter, M. B., Kolk, H. H. J. and Rietveld, T. C. M., 2010, Speaking in ellipses: the effect of a compensatory style of speech on functional communication in chronic agrammatism. *Neuropsychological Rehabilitation*, **20**(**3**), 423–458. https://doi.org/10.1080/09602010903399287.

Ruiter, M. B., Kolk, H. H. J., Rietveld, T. C. M. and Feddema, I., 2013, Combining possibly reciprocally dependent linguistic parameters in the quantitative assessment of aphasic speakers' grammatical output. *Aphasiology*, **27**(**3**), 293–308. https://doi.org/10.1080/02687038.2012.710319.

Saffran, E. M., Berndt, R. S. and Schwartz, M. F., 1989, The quantitative analysis of agrammatic production: procedure and data. *Brain and Language*, **37**, 440–479. https://doi.org/10.1016/0093-934X(89)90030-8.

Saffran, E. M., Schwartz, M. F. and Marin, O. S. M., 1980, Evidence from aphasia: isolating the components of a production model. *Language production*, **1**, 221–241.

Salis, C. and Edwards, S., 2004, Adaptation theory and non-fluent aphasia in English. *Aphasiology*, **18**, 1103–1120. https://doi.org/10.1080/02687030444000552.

Tabachnick, B. G. and Fidell, L. S., n.d., , 1996. *Using Multivariate Statistics* (Northridge: CA: Harper Collins).

Webster, J., Franklin, S. and Howard, D., 2007, An analysis of thematic and phrasal structure in people with aphasia: what more can we learn from the story of Cinderella? *Journal of Neurolinguistics*, **20**, 363–394.

Wright, H. H. and Capilouto, G. J., 2009, Manipulating task instructions to change narrative discourse performance. *Aphasiology*, **23**, 1295–1308. https://doi.org/10.1080/02687030902826844.

Wright, H. H., Capilouto, G. J., Srinivasan, C. and Fergadiotis, G., 2011, Story processing ability in cognitively healthy younger and older adults. *Journal of Speech Language and Hearing Research*, **54**, 911–917. https://doi.org/10.1044/1092-4388(2010/09-0253).

Wright, H. H. and Capilouto, G. J. (2017). *Discourse Processing in Healthy Aging in the United States*. ICPSR36634-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2017-03-02. https://doi.org/10.3886/ICPSR36634.v1.

Appendix A: Scripted story of *The Great Ape* (Krahn 1978)

A ship captain and his first mate have sighted something in the water. A father and daughter are also on board the ship. The crew along with the father and little girl left the ship in a small boat and travelled to an island they spotted. Now they are on foot and have a great deal of camera equipment with them. They come across a group of natives watching a turtle race. The captain taps one of the natives on the shoulder and asks a question. The native points to the top of a mountain. The crew begins to climb the mountain. They climb and climb until the captain calls out to them as he points to something in the distance. He is pointing to a great ape swinging on a swing that is held up by a huge tree between two mountains. The crew begins to climb the mountain looking at the ape and the ape looks back in time to see the little girl fall. The ape catches her. He smiles at her and puts her on top of his head and starts to swing some more. The crew opens a chest they have been carrying and pull out a pump and something else. Oh, it is a giant banana. They blow it up. The ape reaches for it. The crew starts to run down the mountain with the banana hoping the ape will follow them. And he does. He follows them into the water as they head back to their ship. Once they get to the ship, the ape gets the banana and turns to look at it. When he does, he accidentally sits on the ship and the little girl falls off his head into a ship mate's arms. The ape continues back to shore, pleased with his banana. He stops about half-way back and feels the top of his head. He realized that the little girl is gone, and he is sad. The little girl is on the deck of the ship, waving goodbye to the ape and crying. The ship enters New York Harbor. The father takes a picture of the little girl with the Empire State Building in the background. Meanwhile, the ape is in the mountains looking very sad. A plane flies over his head and drops something out. He catches it. It is the picture of the little girl! The ape is very happy and hugs the picture. The End.

Appendix B

Table B1. Age-invariant core function word lists

	***Good Dog Carl* (*GDC*)**	***Picnic***
1	a	a
2	and	all
3	be	and
4	for	be
5	have	for
6	he	he
7	her	her
8	him	him
9	his	his
10	I	in
11	in	it
12	into	look
13	it	of
14	look	on
15	of	one
16	on	she
17	she	so
18	so	some
19	some	that
20	that	the
21	the	their
22	they	them
23	to	they
24	up	to
25	with	with

Appendix C

Table C1. Age-corrected core function word lists for *Good Dog Carl* (*GDC*) and *Picnic*

	20s	30s	40s	50s	60s	70s	80s
GDC							
1	the	the	the	the	the	the	the
2	and	and	and	and	and	and	and
3	to	be	be	be	be	be	be
4	be	to	to	to	to	to	he
5	they	a	a	a	a	a	a
6	a	they	he	he	he	he	to
7	he	he	they	they	they	in	in
8	in	in	on	on	on	they	on
9	on	on	in	in	in	on	they
10	of	of	of	of	of	of	of
11	some	so	some	so	him	him	him
12	that	that	into	into	into	his	it
13	so	with	so	that	that	it	his
14	him	she	him	it	his	into	that
15	into	some	with	some	it	that	so
16	with	into	her	with	so	with	into
17	she	him	his	him	with	I	I
18	it	for	she	up	some	so	she
19	his	it	it	she	she	some	with
20	her	his	that	his	her	up	some
21	for	her	for	her	up	she	her
22	up	up	up	I	I	out	out
23	as	I	I	for	for	for	but
24	after	as	as	like	have	have	for
25	have	down	out	out	like	her	at
Picnic							
1	the	the	the	the	the	the	and
2	and	and	and	and	and	and	the
3	be	be	be	be	be	be	be
4	they	they	they	they	they	they	they
5	to	to	to	to	a	a	a
6	she	a	of	a	to	to	to
7	of	she	a	of	of	of	of
8	a	of	he	she	he	he	he
9	her	her	she	he	in	in	in
10	in	in	in	in	all	all	all
11	that	he	all	all	she	she	one
12	their	that	their	their	that	that	that
13	all	for	that	her	one	their	it
14	for	all	her	that	their	it	with
15	he	their	one	for	his	one	their
16	so	so	for	with	it	with	them
17	one	on	so	so	on	his	him
18	them	it	his	it	with	on	on
19	it	some	with	one	so	her	I
20	on	with	it	on	for	so	his
21	some	have	some	some	her	for	so
22	have	one	on	his	some	him	some
23	with	his	at	at	them	I	for
24	as	him	them	them	him	them	she
25	at	where	him	but	have	some	but