International Journal of Language & Communication Disorders

INT J LANG COMMUN DISORD, JANUARY–FEBRUARY 2019,
VOL. 54, NO. 1, 62–78

Research Report

Measuring word retrieval in narrative discourse: core lexicon in aphasia

Hana Kim†, Stephen Kintz‡, Kristen Zelnosky§ and Heather Harris Wright†

†Department of Communication Sciences and Disorders, East Carolina University, Greenville, NC, USA
‡Department of Audiology & Speech Pathology, University of Arkansas—Little Rock, AR, USA
§Department of Communication Sciences and Disorders, Radford University, Radford, VA, USA

(*Received March 2018; accepted September 2018*)

Abstract

Background: Discourse analysis procedures are time consuming and impractical in a clinical setting. Critical to clinicians are simple and informative discourse measures that require minimal time and labour to complete. Many studies, however, have overlooked difficulties that clinicians face. We recently developed core lexicon lists for nouns, verbs, adjectives and adverbs for two narrative discourse tasks with healthy control groups. Core lexicon lists consist of important lexical items required to produce coherently meaningful discourse in response to discourse tasks. Measuring core lexicon is useful for quantifying word retrieval impairments at the discourse level in clinical populations.
Aims: To apply an age-based core lexicon list for nouns, verbs, adjectives and adverbs for the wordless picture books *Good Dog Carl* (1985) and *Picnic* (1984) and to determine how well the lists measured linguistic impairments in persons with aphasia (PWA).
Materials & Methods: Lemma forms were extracted from 470 control participants who were divided into seven age groups. Twenty-five core lexicons were identified for four word classes (nouns, verbs, adjectives and adverbs) among the seven age groups. The nouns, verbs, adjectives and adverbs for each PWA (N = 11) were then compared with the core lexicon for their respective age group. Per cent agreement was computed by comparing the number of total items within each list to the number of items that PWA produced. A Spearman's correlation coefficient was computed between the WAB-R AQ and the per cent agreement for each word type for PWA.
Outcomes & Results: The percentage of agreement for each word type among the age cohorts ranged between 56% and 96%. Of the four word types, core verbs significantly correlated with the WAB AQs for both discourse tasks. A post-hoc analysis found significant differences between fluent and non-fluent aphasia for core verbs.
Conclusions & Implications: Core lexicon analysis appears to be a practical way to capture impairments in word retrieval at the discourse level. Core verbs may be a better indicator to understand holistic language performances for PWA. Use of the core lexicon checklist can serve as an option to reconcile ecological validity with clinical usability.

What this paper adds

What is already known on the subject

Discourse abilities in aphasia have garnered attention because discourse represents the process by which thinking is converted into language. Aphasiologists have used various methods for quantifying discourse production that are not easily detectable by standardized test batteries. Recently, a group of researchers has attempted to develop a core lexicon analysis that requires less time and effort. Their findings concluded that the core lexicon list can capture PWA's ability to access target words required to produce coherently meaningful discourse.

What this paper adds to existing knowledge

The study considered age and word class effects (nouns, verbs, adjectives and adverbs) in developing core lexicon lists. There is evidence that older adults' ability to access target words may be reduced compared with younger adults,

Address correspondence to: Heather Harris Wright, College of Allied Health Science, East Carolina University, 3206 F Allied Health Sciences, MS 668, Greenville, NC 27834, USA; e-mail: wrighth@ecu.edu

International Journal of Language & Communication Disorders
ISSN 1368-2822 print/ISSN 1460-6984 online © 2018 Royal College of Speech and Language Therapists
DOI: 10.1111/1460-6984.12432

and PWA's word retrieval ability is selectively disrupted depending on word class. In this study, multiple core lexicon lists were developed for different age groups and word classes.

What are the potential or actual clinical implications of this work?
The results have potential clinical implications as the core lexicon lists are easily quantifiable and time-saving for assessment. The multiple core lexicon lists by word class were created based on normative data, and can be used for contrasting word retrieval ability of cognitively healthy individuals with that of PWA.

Introduction

Persons with aphasia (PWA) have been defined as having an acquired language impairment which presents with deficits in word retrieval (Goodglass and Winfield 1997). Traditionally, speech–language pathologists have focused on THE PWA's ability to retrieve words because it indicates disruptions in lexico-semantic and/or phonological representations (Dell *et al.* 1999). In this sense, commonly used measures of language difficulties in clinical settings capture word retrieval impairments at the single-word level.

The PWA's word retrieval impairment is thought to be different between word classes, such as nouns and verbs, depending on the aphasia language profile. In research involving comparisons between nouns and verbs, verb deficits have been reported in individuals with agrammatic aphasia on single-word naming tasks and discourse tasks, whereas other aphasia subtypes, such as fluent aphasia, represent relative deficits in retrieving nouns (Bates *et al.* 1991, Chen and Bates 1998, Kim and Thompson 2000, Luzzatti and Chierchia 2002, Schwartz *et al.* 1980). However, some researchers have argued that there are no clear dissociations between retrieving nouns and verbs across different aphasia types (Berndt *et al.* 1997, Jonkers and Bastiaanse 1998, Mätzig *et al.* 2009, Williams and Canter 1982, Zingerser and Berndt 1990).

During discourse production, word-retrieval problems in PWA have proven to be more dynamic because contextual effects may influence retrieval processes at the discourse level (Basso *et al.* 1990, Williams and Canter 1982, Wilshire and McCarthy 2002). Relatively few studies have investigated the PWA's ability to retrieve words by word class beyond the word level (Berndt and Haendiges 2000, Kambanaros 2010, Mayer and Murray 2003, Pashek and Tompkins 2002, Zingeser and Berndt 1988). Contrasting findings have been reported, where some studies have shown that persons with anomic aphasia performed better on retrieving nouns than verbs (Pashek and Tompkins 2002, Zingeser and Berndt 1988), and the others found an opposite pattern (Berndt and Haendiges 2000). These conflicting results highlight the differences in lexical retrieval at the word and discourse levels, indicating that lexical retrieval at the word level may not inform or predict lexical retrieval at discourse level. Therefore, a goal of the current study is to develop a quantitative measure of word retrieval ability in discourse production that is clinically practicable. In the following sections, we briefly summarize existing discourse measures and the challenges with these measures that led to the current study. We then review core lexicon measures developed in previous literatures.

Discourse analysis

Discourse is any natural form of language comprising utterances or phrases (Wright and Capilouto 2012) and may be 'the most elaborative linguistic activity' (Ska *et al.* 2004: 302). Owing to the complexity of discourse processing, quantifying discourse production in clinical settings is a challenging task (Armstrong 2000, Prins and Bastiaanse 2004).

To date, researchers have suggested a great deal of outcome measures to examine the amount of information provided in discourse such as correct information unit (CIU; Nicholas and Brookshire 1993) and main concept (Nicholas and Brookshire 1995), which are rule-based scoring measures. In keeping with Nicholas and Brookshire's (1995) idea, Wright and colleagues developed a main event analysis, which is operationally defined as essential elements within the discourse (Capilouto *et al.* 2005) and is discourse-task specific. Recently, multilevel approaches that include micro- and macro-linguistic assessments have received experimental attention from researchers because they provide a breadth of information on discourse ability (Marini *et al.* 2011, Sherratt 2007, Wright and Capilouto 2012).

Although such analyses have been applied to investigate empirically discourse abilities in PWA, application and usability in clinical settings have not been readily investigated to our knowledge. Maddy *et al.* (2015) examined the extent to which clinicians have used discourse analysis to evaluate PWA in clinical settings. In semi-structured interviews with nine clinicians, they found that external influences such as time constraints and lack of training obstruct application and use of discourse analysis. For example, discourse analysis requires collecting, transcribing and analyzing language samples. A trained clinician generally requires more than four times the actual length of the discourse sample just to complete the transcription process alone (Armstrong *et al.* 2007, Boles and Bombard 1998, Elia *et al.* 1994). This time frame excludes the time required for training and analysis, thus making many analyses impractical for use in clinical settings.

In recent discussions on the topic of discourse outcome measures, several groups of researchers agree that discourse analysis requires arduous processes (de Riesthal and Diehl 2018, Dietz and Boyle 2018a, 2018b, Kintz and Wright 2018, Kurland and Stokes 2018, Wallace *et al.* 2018, Whitworth 2018). Commenting on roadblocks of discourse analysis, they have raised their voices in pursuing clinical feasibility to extend the use of discourse outcome measures by reducing time and effort. For many years, there has been an increasing emphasis on evaluating discourse without transcribing (McNeil *et al.* 2001, Olness *et al.* 2012, but see de Riesthal and Diehl 2018). Along with the advantage of lessening the burden on clinicians, non-transcription discourse analysis may also permit clinicians to achieve reliability (McNeil *et al.* 2001).

Core lexicon in aphasia

In acknowledgement of these clinical barriers for discourse analysis, recently researchers have developed a lexicon-based analysis that does not require an arduous transcription process (Dalton and Richardson 2015, Dillow 2013, Fromm *et al.* 2013, MacWhinney *et al.* 2010). Lexicon is not only a critical aspect of communication but the building block of discourse (Kintz *et al.* 2016). Without access to the intended word, the ability to deliver a message may be reduced. Moreover, core lexicon, which is one such analysis currently in development, refers to the pivotal lexical items required to produce a semantically meaningful and coherent narrative (MacWhinney *et al.* 2010). As such, it can be expected that core lexicon production reflects the ability to access the target word (MacWhinney *et al.* 2010), and further, informational discourse performance (Andreetta *et al.* 2012).

MacWhinney *et al.* (2010) introduced a core lexicon analysis for the Cinderella story by analyzing the discourse samples from 25 healthy participants and 24 PWA. They collected the discourse samples from AphasiaBank (MacWhinney 2000), a collaborative project whose goal is to develop a database of language samples from PWA. Participants told the Cinderella story after looking through a 25-page wordless picture book. The researchers used the Computerized Language Analysis program (CLAN; MacWhinney 2000) to extract the core lexicons from the language samples. They found that the PWA's discourse abilities were characterized with reduced lexical diversity and greater use of light verbs (i.e., frequently occurring verbs in language samples such as *be, have, come* etc.) compared with the control group. However, the core lexicon lists only included nouns and verbs, and the researchers did not consider other word classes, such as adjectives and adverbs, which may contribute to increased lexical diversity in discourse production (Sarno *et al.* 2005).

Dalton and Richardson (2015) reported that a 24-item core lexicon list, independent of word class (i.e., verbs, nouns, adverbs, adjectives), discriminated between neurologically healthy controls and PWA. To develop the core lexicon list, the researchers accessed the transcripts of 92 healthy controls from AphasiaBank. They extracted all the lemmas produced within one of the discourse tasks, a sequential picture description task. The lemmas were extracted by using the CLAN command, where 24 lemmas produced by 50% or more of the control participants were included within the core lexicon. To determine if core lexicon could distinguish between the two groups, the researchers examined the transcripts of 166 healthy controls and 235 PWA. The researchers found a significantly different number of core lexicon items between PWA and healthy controls, and Broca's aphasia and other aphasia subtypes. They also concluded that the core lexicon list can reflect the participants' ability to convey the gist of a narration. However, the relative influence of lexical processing, known to be susceptible to aging, was not considered during development of the core lexicon lists.

Fromm *et al.* (2013) compared the core lexicon lists for a different type of discourse: procedural discourse (how to make a peanut butter and jelly sandwich). No differences were found between healthy controls (n = 145) and PWA (n = 141). They included additional measures as well the number of words, the number of utterances, time on task and mean length of utterance. The healthy control group produced significantly more words and utterances, and also had longer utterance durations compared with the aphasia group. Fromm and colleagues suggested these measures reflect quantitative differences among the groups. Further, they suggested that core lexicon is a qualitative assessment, in turn suggesting the groups' procedural discourse samples differed quantitatively but not qualitatively. Results from these studies demonstrate potential pitfalls to using procedural discourse tasks for developing core lexicon measures such as fewer lexical items produced (Fergadiotis *et al.* 2011).

Dillow (2013) analyzed the core lexicon lists for the Cinderella story. Dillow created core verb and noun lists following MacWhinney *et al.*'s (2010) procedures. In contrast to earlier studies, they attempted to add an adjective core lexicon list, but they did not include it due to their criterion to establish the lexicon. The scores for core verbs, core nouns and the entire core lexicon lists differentiated aphasia subtypes from the control group. They also analyzed how different word types of core lexicon affected the ability to differentiate the aphasia subtype groups. Core verbs differed for the following groupings (Anomic > Conduction > Wernicke > Broca): adults with anomic aphasia and adults with Broca's aphasia; adults with anomic aphasia and adults with conduction aphasia; and adults with Broca's aphasia and adults with Wernicke's aphasia. For core nouns, participants with anomic aphasia produced significantly more core nouns than those with Broca's and Wernicke's aphasia. Likewise, participants with conduction aphasia also produced significantly more core nouns than those with Broca's aphasia and Wernicke's aphasia. When considering the complete lexicon, adults with Broca's aphasia differed significantly from the adults with anomic aphasia and conduction aphasia. Compared with studies using an aggregated core lexicon list, this study demonstrates that separate core lexicon lists by word class differentiate each subtype from one another.

These findings are promising in that core lexicon analysis provides an alternative approach to more time-intensive, lexical-level discourse analyses. Whereas the transcription process and training are necessary for existing measures, clinicians can simply check if the words are present or not while listening to the recorded language samples once the core lexicon lists are established. However, limitations of previous research exist that need to be addressed. In generating the core lexicon, many researchers disregarded some word types (e.g., adjectives and adverbs) (Dillow 2013, Fromm *et al.* 2013, MacWhinney *et al.* 2010) or combined words types to create a single core lexicon list (Dalton and Richardson 2015). Different words types, such as nouns, verbs, adjectives, and adverbs, carry important and unique semantic information that differentiate them (Neville 2014). Based on previous research that production of modifiers manifested qualitative changes in language usage for PWA (Sarno *et al.* 2005), it would be worth developing core lexicon lists for different word types as an exploratory purpose. Moreover, previous studies of core lexicon have not considered age differences. Lexical selection by someone in their 20s may differ from someone in their 80s. Age should be considered when creating a core lexicon for a stimulus.

The purpose of the current study was to apply an age-based core lexicon list for nouns, verbs, adjectives, and adverbs for the wordless picture books, *Good Dog Carl* (*GDC*; Day 1985) and *Picnic* (McCully 1984), to determine how well the lists measure linguistic impairment in PWA. The specific aims of the current study, then, were twofold: (1) to determine the per cent agreement between groups and their core lexicon; and (2) to examine the correlation among lexicon lists and aphasia impairment as determined by the aphasia quotient (AQ) from the Western Aphasia Battery—Revised (WAB-R; Kertesz 2006). Based on the well-documented word-retrieval deficits on verbs and nouns in PWA, we hypothesized that core nouns and verbs would positively correlate with the WAB-R AQs. If PWA demonstrate improved production of modifiers with better language performance, as shown by Sarno *et al.* (2005), then it would be hypothesized that core adjectives and adverbs positively correlate with aphasia severity.

Materials and methods

Participants

Language samples from 470 cognitively healthy participants (273 females, 197 males) and 11 PWA were included in the study. The normative data presented are a subset of data from a larger study examining discourse processing across the lifespan (Wright and Capilouto 2017) and was approved by the respective universities (Arizona State University and University of Kentucky). The database included discourse samples and cognitive measures collected from over 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 (Folstein *et al.* 2001); and (4) self-reported no history of stroke, head injury or progressive neurogenic disorders. Demographic information for the control participants can be found in table 1.

All PWA met the following criteria: (1) native English speaker; (2) aided or unaided visual acuity as indicated by Beukelman and Mirenda's (1998) vision screening form; (3) aided or unaided hearing acuity within normal limits as measured by the ability to hear pure tones at 25 dB HL for the frequencies of 500, 1000 and

Table 1. Neurologically healthy adult demographic information

Age group (years)	*N* (female:male)	Age (SD)	Education (SD)
20s	66 (35:31)	23.93 (3.69)	15.76 (1.93)
30s	63 (39:24)	34.12 (3.11)	16.15 (3.28)
40s	67 (41:26)	44.34 (3.01)	15.36 (2.54)
50s	68 (43:25)	55.57 (2.65)	15.85 (2.54)
60s	67 (38:29)	64.78 (4.93)	15.45 (2.49)
70s	76 (43:33)	73.85 (2.88)	15.32 (2.32)
80s	63 (34:29)	83.29 (2.71)	14.76 (2.70)
Total	470 (273:197)		

Note: SD, standard deviation.

Table 2. Participants with aphasia demographic information

	Age (years)	Gender	Education	WAB-R AQ	Post-onset (months)	Aphasia type (fluent/non-fluent)
P1	65	Male	18	76.3	67	Conduction (fluent)
P3	73	Male	12	85.2	25	Anomic (fluent)
P4	84	Female	12	62.6	26	Conduction (fluent)
P5	55	Male	14	57.6	26	Broca's (non-fluent)
P6	66	Female	14	56.3	171	Broca's (non-fluent)
P7	34	Female	14	90.7	21	Anomic (fluent)
P9	38	Female	14	57.7	151	Broca's (non-fluent)
P10	62	Female	20	61.3	96	Broca's (non-fluent)
P11	72	Male	12	64.9	57	Transcortical motor (non-fluent)
P12	65	Female	11	89.4	120	Anomic (fluent)
P13	65	Male	14	54.4	n.a.	Broca's (non-fluent)
Average (SD)	61.7 (14.7)		14.1 (2.7)	68.8 (14.0)		

Note: WAB-R AQ, Western Aphasia Battery—Revised (Kertesz 2006). Maximum WAB-R AQ raw score = 100.

2000 Hz; (4) no reported history of other neurological disorders; (5) presented with aphasia as determined by performance on the WAB-R AQ subtests (Kertesz 2006); (6) chronic aphasia (at least 6 months post-onset); and (7) left hemisphere damage. Initially, 13 PWA were recruited, and then two aphasia participants (P2 and P8) were disqualified from the study due to other neurological disorders. Thus, 11 right-handed participants with present or past evidence of stroke participated in this study. Demographic information for the PWA can be found in table 2.

Experimental procedures

All participants were tested individually in a laboratory setting. Since the normative data were collected for a large study, the cognitively healthy participants attended two sessions, lasting no more than 2 h for each session. Before study participation, they completed consent forms and then completed screening measures to confirm that they met the inclusion criteria. Next, a cognitive test battery and a set of discourse tasks were administered. The order of test administration was randomized across participants. The cognitive test and discourse task results irrelevant to this study are not reported here.

For participants in the PWA group, the WAB-R was administered first and then cognitive and discourse tasks were randomized across participants. During the experimental procedures, they were allowed to take breaks as needed. This study is focused solely on some of those discourse measures (described below).

Discourse task

Two wordless picture books were used to collect narrative discourse samples from participants. They included *Good Dog Carl* (*GDC*; Day 1985) and *Picnic* (McCully 1984). Because limited to no text is included in the books, the task is a storytelling or story generation task, rather than a story-retelling task. Storytelling tasks are 'more representative of spontaneous communication' (Liles 1993, cited in Hughes *et al.* 1997: 19). Additionally, because participants are telling stories from books rather than from shorter pictured stimuli (e.g., single pictures), participants provide longer samples and use a more diverse vocabulary (Fergadiotis and Wright 2011, 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 the following 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.' The examiner then provided an example to participants with another story, for *The Great Ape* (Krahn 1978). Finally, participants were presented with the book and allowed to look through it for as long as they needed to tell the whole story by themselves. While telling the story, the books were still viewable by the participant.

Language sample preparation

All samples were either audio or video recorded, and then orthographically transcribed by trained research assistants using a set of programs called CLAN.

A total of 10% of the control participants were randomly selected for inter- and intra-rater reliability for the entire transcription. The inter- and intra-rater agreements were 95% and 98% respectively. For the aphasia group, two PWA were selected due to the small number of participants, and inter- and intra-rater agreements for the entire transcription were 91% and 93% respectively.

Core lexicon

The core lexicons for *GDC* and *Picnic* were created by determining the 25 most frequently used lemmas produced for nouns, verbs, adjectives and adverbs within each age group. The researchers accomplished this goal by assigning the proper syntactic category for each word within the narrative transcripts and extracting the lemma forms and their respective frequencies within each age group. To assign proper syntactic categories, the researchers used CLAN (MacWhinney 2000) and the methods outlined by MacWhinney (2000) and MacWhinney *et al.* (2010). CLAN uses the programs MOR and POST, which are respectively tied to a dictionary of lexical items and English grammar rules, to categorize words automatically into their respective syntactic category with an accuracy of 95% (for a review, see MacWhinney 2000). Once the words within each transcript were categorized, the researchers automatically extracted *GDC* and *Picnic* into separate files using the GEM program of CLAN. MOR, POST and GEM are terms for CLAN commands. The MOR command is used primarily for morphosyntactic analysis for each word. The POST command following the MOR command automatically resolves grammatical ambiguity. The GEM command is to sort different discourse tasks in the transcripts (see appendix A for the CLAN commands). This step is necessary to create two independent lists for each story. For each story and age group, the lemma forms were extracted for all the participants into a single list that included their frequency information. For example, if 20 participants used the lemma *go* once and a single participant used *go* five times, the lemma list would indicate that *go* was produced 25 times for that age group. The top 25 most frequent lemmas were collected for each word class within each story for every age group. While the top 25 most frequent lemmas is an arbitrary cut-off, previous researchers used similar numbers (Dalton and Richardson 2015). A complete list of the top 25 core lexicon for each age group for the two stimuli is presented in appendix B.

Per cent agreement was determined by comparing the 25 core lexical items within each list among seven neurologically healthy groups. Per cent agreement was calculated by dividing the number of agreements by the total number of core lexical items on each list (the number of agreements/25)∗100. For example, an aphasia speaker (P1) who is in his 60s was evaluated by using the 60s age group core lexicon lists. If P1 produces four items from the 60s core verb list, the numerator is four and the denominator is 25 in the fraction.

Core lexicon production in aphasia

The PWA's transcripts were prepared for analysis in a similar manner as described above. Counting of how many core lexical items were produced in PWA was based on PWA's transcripts. These lists were compared with the age-matched core lexicon list for each story. For this study, we chose not to count synonyms, to maintain consistency with Dalton and Richardson's (2015) procedures, which acknowledges the importance of producing the target words (e.g., Andreetta *et al.* 2012, Verhaegen and Poncelet 2013). If a PWA produced any lemmas on any of the core lexicon lists, they would receive a point. If the PWA did not produce the lemma form, they did not receive a point. Only one point was provided regardless of how many times the lemma form may have been used by the participant. The number of lemmas produced was divided by the total number of lemmas on the core lexicon list for each syntactic category type resulting in a per cent agreement between the PWA and age-matched cohorts for the core lexicon lists.

Results

The purpose of the study was to apply a core lexicon list for nouns, verbs, adjectives and adverbs within the narrative discourse, *GDC* and *Picnic*, for different age groups and compare them with core lexicon productions by PWA. These age-based core lexicon lists were used to address the aims of the current study.

The percentage of agreement for each word type was calculated across the seven age groups. Adverbs appeared to have the best agreement among the age groups with the lowest agreement only at 72%. Verbs had the next best agreement among the age cohorts, ranging between 64% and 92%. The per cent agreement for adjectives ranged from 56% to 92%. The per cent agreement for nouns ranged from 56% to 98%. See tables 3–6 for agreements among age groups for syntactic category types for each narrative discourse task (*GDC* and *Picnic*).

Table 3. Per cent agreement between cognitively healthy age cohorts for nouns

	20s	30s	40s	50s	60s	70s	80s
20s		80%	72%	80%	60%	60%	60%
30s	88%		80%	68%	80%	72%	68%
40s	92%	92%		76%	72%	64%	60%
50s	84%	92%	92%		80%	60%	56%
60s	80%	84%	88%	98%		68%	64%
70s	72%	80%	80%	84%	88%		72%
80s	68%	72%	72%	80%	80%	84%	

Note: *Good Dog Carl* appears in the section at bottom left. *Picnic* with shading appears in the section at upper right.

Table 4. Per cent agreement between cognitively healthy age cohorts for verbs

	20s	30s	40s	50s	60s	70s	80s
20s		80%	88%	88%	80%	76%	72%
30s	92%		72%	76%	68%	68%	64%
40s	80%	84%		84%	92%	84%	76%
50s	88%	84%	80%		80%	80%	68%
60s	88%	80%	80%	84%		88%	84%
70s	88%	80%	76%	80%	88%		80%
80s	76%	72%	80%	76%	80%	76%	

Note: *Good Dog Carl* appears in the section at bottom left. *Picnic* with shading appears in the section at upper right.

Table 5. Per cent agreement between cognitively healthy age cohorts for adjectives

	20s	30s	40s	50s	60s	70s	80s
20s		84%	80%	76%	72%	80%	76%
30s	72%		84%	80%	76%	80%	72%
40s	56%	72%		88%	88%	92%	84%
50s	64%	76%	68%		88%	88%	84%
60s	60%	80%	68%	72%		92%	84%
70s	60%	76%	64%	80%	76%		92%
80s	64%	76%	64%	80%	76%	76%	

Note: *Good Dog Carl* in the section at bottom left. *Picnic* with shading appears in the section at upper right.

Table 6. Per cent agreement between cognitively healthy age cohorts for adverbs

	20s	30s	40s	50s	60s	70s	80s
20s		84%	80%	76%	72%	80%	76%
30s	80%		84%	80%	76%	80%	72%
40s	76%	84%		88%	88%	92%	84%
50s	80%	88%	84%		88%	88%	84%
60s	80%	88%	84%	84%		92%	84%
70s	80%	84%	88%	96%	84%		92%
80s	80%	88%	88%	88%	88%	92%	

Note: *Good Dog Carl* appears in the section at bottom left. *Picnic* with shading appears in the section at upper right.

Table 8. Correlations (Spearman's rho) between AQs and core lexicon by word class

Comparison	Spearman's Rho	Significance
GDC		
Nouns	.146	.687
Verbs	.869**	.001
Adjectives	–.307	.388
Adverbs	–.171	.636
Picnic		
Nouns	.338	.309
Verbs	892**	.000
Adjectives	.266	.429
Adverbs	.574	.065

Note: $^{*}p < .05$, $^{**}p < .01$.
Sources: *Good Dog Carl* (Day 1985) and *Picnic* (McCully 1984).

To investigate the relationship between the core lexicon and aphasia impairment, the per cent agreement for each word class was obtained between the PWA and age-matched cohorts for the core lexicon lists. Spearman's correlation coefficients were computed between WAB-R AQs and core lexicon agreements for nouns, verbs, adjectives and adverbs for each narrative task. For both *GDC* and *Picnic*, significant correlations were found between core verbs and WAB AQs, $r(9) = .869$, $p < .001$, $r(9) = .892$ $p < .001$. PWA with better AQs had greater core lexicon agreements for verbs. Significant correlations were not found among AQs and other word classes (nouns, adjectives and adverbs) (tables 7 and 8).

Post-hoc analysis: aphasia type

A post-hoc analysis was conducted to determine if production of different word types obtained by the core lexicon measure differed between individuals with fluent ($N = 5$) and individuals with non-fluent ($N = 6$) types of aphasia. To conduct this analysis, the PWA were divided into two groups (fluent versus non-fluent) based

Table 7. Per cent agreement for the participants with aphasia with their respective age group for the core lexicon

		Good Dog Carl				*Picnic*			
Participant ID	Age group (years)	Nouns	Verbs	Adjectives	Adverbs	Nouns	Verbs	Adjectives	Adverbs
P1	60s	36	16	8	4	24	28	28	28
P3	70s	24	44	16	16	16	28	28	24
P4	80s	44	32	36	40	24	16	44	36
P5	50s	36	8	20	16	16	20	8	16
P6	60s	n.a.	n.a.	n.a.	n.a.	16	12	8	8
P7	30s	56	48	16	20	36	52	24	16
P9	30s	48	20	16	24	16	16	16	20
P10	60s	40	16	16	12	24	8	12	4
P11	70s	8	28	28	32	8	28	28	32
P12	60s	52	48	20	20	52	48	28	40
P13	60s	48	8	24	24	28	4	32	12

Sources: *Good Dog Carl* (Day 1985) and *Picnic* (McCully 1984).

Table 9. Mann–Whitney *U*-test of the difference in the core lexicon between two aphasia types (fluent versus non-fluent)

Task	Nouns	*n*	Mean rank	*p*
GDC	Fluent	5	6.50	.310
	Non-fluent	5	4.50	
	Verbs			
	Fluent	5	7.50	.032*
	Non-fluent	5	3.50	
	Adjectives			
	Fluent	5	5.20	.841
	Non-fluent	5	5.80	
	Adverbs			
	Fluent	5	5.10	.690
	Non-fluent	5	5.90	
Picnic	Nouns			
	Fluent	5	8.00	.082
	Non-fluent	6	4.33	
	Verbs			
	Fluent	5	8.30	.030*
	Non-fluent	6	4.08	
	Adjectives			
	Fluent	5	7.70	.126
	Non-fluent	6	4.58	
	Adverbs			
	Fluent	5	8.10	.052
	Non-fluent	6	4.25	

Notes: $^{*}p < .05$, $^{**}p < .001$.
Sources: *Good Dog Carl* (Day 1985) and *Picnic* (McCully 1984).

on the WAB-R aphasia classification. A Mann–Whitney *U*-test indicated that for *GDC* production of core verbs was significantly greater for fluent aphasia (mean rank = 7.50) than for non-fluent aphasia (mean rank = 3.50), $U = 2.50$, $z = -2.11$, $p < .05$. For *Picnic*, fluent aphasia (mean rank = 8.30) also produced more core verbs than non-fluent aphasia (mean rank = 4.08), $U = 000$, $z = -2.124$, $p < .05$ (table 9).

Discussion

The purpose of the study was to apply age-developed core lexicon lists for the narrative discourse tasks *GDC* and *Picnic* to determine if core lexicon lists for word type would correlate with aphasia severity. For the normative data, while comparatively high agreement across age groups was observed for adjectives and verbs, adverb and noun use had considerable variability across the age cohorts, suggesting a need to develop and use core lexicon lists that account for age with clinical populations. Further, only verbs significantly correlated with WAB AQs for both narrative tasks for the PWA. These findings suggest that the core lexicon comparisons between age-matched controls and PWA may be useful for determining atypical patterns of lexical usage in discourse production, which in turn is reflective of aphasia severity.

Core lexicon and aphasia

Core verbs for both tasks significantly correlated with overall aphasia severity as measured by the WAB-R AQ, providing partial support for our hypothesis that core verbs and nouns correlate with AQs. These findings agree with findings by other researchers who were able to differentiate aphasia subtypes (Dillow 2013). Whereas some researchers have created a single core lexicon list (Dalton and Richardson 2015, MacWhinney *et al.* 2010), Dillow (2013) demonstrated that a single list is not sufficiently able to discern between aphasia types and thus created core lexicon lists for nouns and verbs separately. The study extended these results by adding lists for adverbs and adjectives as an exploratory investigation. In the current study, we did not have a large enough sample to consider different subtypes of aphasia and determine if each core lexicon list differed across aphasia subtypes. However, findings of the post-hoc analysis lend weight to our results in that the only difference identified was that individuals with fluent aphasia produced significantly more core verbs than individuals with non-fluent aphasia for both tasks.

The results support and extend Dillow's (2013) results, wherein verbs are important in differentiating aphasia subtypes. As overall aphasia severity increases, fewer verbs are produced. This finding is unsurprising, since verbs are often considered the building blocks or central themes of utterances (Healy and Miller 1970). Additionally, these findings have critical implications in terms of how researchers and clinicians should assess and treat verbs in discourse production of PWA. However, it was somewhat surprising that no significant correlations were found between core noun production and overall language severity obtained from the standardized, norm-referenced measure (i.e., WAB-R AQ), considering the substantial impact of noun production in clinical decisions. For language assessment in PWA, the WAB-R and the Boston Naming Test (BNT; Kaplan *et al.* 1976) require naming noun objects and are the most frequently used tests in clinical settings (Guo *et al.* 2014, Verna *et al.* 2009). Based on these findings, it may be insufficient for clinicians to rely on performance of noun production alone when drawing clinical decisions regarding word retrieval abilities of their patients with aphasia.

Further, the initial analyses based on 11 PWAs did not demonstrate significant correlations between aphasia severity and production of adjectives and adverbs. However, the subsequent statistical analyses (Mann–Whitney *U*-test) of the fluent and non-fluent aphasia groups detected lower statistical power ($p = .052$) for capturing significant differences in adverb production (table 9). Because of the small number of aphasia participants and relatively restricted range of aphasia severity included in the separate group analyses, these results should be interpreted with caution (i.e., fluent aphasia groups presented with more mild aphasia compared with non-fluent aphasia). Future studies should consider potential joint effects of different word types to capture the level of aphasia severity.

In contrast to the current study, previous investigators have employed different elicitation techniques such as story retelling of Cinderella (Dillow 2013, MacWhinney, 2010), and procedural discourse (Dalton and Richardson 2015, Fromm *et al.* 2013). Considering that the core lexicon measure is developed based on the entire spoken lexicon, a sufficient number of words should be produced to create a reliable and sensitive measure for capturing unusual lexicon patterns of PWA. Although narrative discourse obtained from wordless picture books has not been used in clinical settings frequently, it does provide lexically diverse language samples (Fergadiotis and Wright 2011), thereby increasing the probability of capturing the severity of aphasia using this measure. Additionally, the existence of pictorial stimuli may be an important factor to elicit discourse samples with high quality and quantity (Grosjean 1980). A task that provides picture stimuli having frame-by-frame presentation may function as cognitive schema, which leads to more episodes (Coelho 2002). As such, narrative discourse tasks with pictorial support may be appropriate for collecting language samples, as well as for developing core lexicon measures.

There is no converging evidence from previous research with respect to criterion for the lemmas. For example, MacWhinney *et al.* (2010) did not stipulate a criterion and generated 10 core nouns and 10 core verbs. Fromm *et al.* (2013) did not specify an inclusionary criterion and included 10 core lexicon items as well, though with comparatively short language samples obtained from procedural discourse. Other studies required that at least 50% of the core lexical items be produced by the control participants to be included in the core lexicon list (Dalton and Richardson 2015, Dillow 2013). Given that the core lexicon measure is a relatively novel method, an important next step is to determine the impact of different inclusionary criteria for lexical items, and then investigate the sensitivity and specificity of measuring language impairments. For best practice and usability of the core lexicon measure, a systematic approach to the criterion should be considered in future investigations.

Clinical implications

Discourse outcome measures are evolving in response to clinical utility. Such changes can enhance our understanding of discourse impairment of PWA, as well as aid in alleviating some difficulties that clinicians face. This study is a step forward in addressing the issue of clinical feasibility for discourse analysis in clinical settings. Researchers investigating discourse ability in PWA claim that the transcription process is an obstacle that prevents clinicians from using discourse analysis in clinical settings (de Riesthal and Diehl 2018, Kintz and Wright 2018).

In this sense, the core lexicon measure is a meaningful outcome because it is easily quantifiable and time-saving for assessment without transcription. Additionally, eliciting sufficient quality and quantity of language samples in a limited period of time is important. The explicit task instructions (identified in the method) are distinct from traditional instructions (i.e., 'tell me everything you see going on in this picture') and induce individuals to provide the core event line of pictures depicted in narrative discourse (Olness 2006, Wright and Capilouto 2009). In this study, we did not ask participants to describe every scene, but instead to build the story. This led participants to focus on temporal and/or causal information, not simply list all objects viewed in each scene. Most participants in this study took between 5 and 15 min to complete both tasks, whereas one PWA with the longest language sample for our participants took 21 min to finish them. In turn, we were able to elicit a language sample in a very appropriate time frame despite comparatively more picture stimuli included in the task.

Another clinical contribution of the current study is that the core lexicon measure was created based on the performance of cognitively healthy controls. Discourse disruptions featured in PWA lie on the continuum of normal discourse performance. By contrasting PWA's lexical usage to typical lexical usage produced by cognitively healthy controls within similar age cohorts, we can gain some insight into the nature of PWA's language profiles and to what extent they are preserved or impaired. Finally, though separate core lexicon lists by word class may be useful for evaluating overall changes in lexical use before and after treatment, they are restricted to providing clinical information about lexical retrieval. Core lexicon does not inform, clinically, about syntactic structure, rate of speech or fluency.

Conclusions and future directions

Multiple core lexicon lists were developed in this study for two discourse elicitation tasks and seven, 10-year age-cohort groups and compared with narratives elicited from PWA to determine the suitability of core lexicons for predicting aphasia severity and potential clinical use. Results of the study are promising, as they broaden our understanding of how meaningful the verb core lexicon is for PWA and also have clinical implications. The core verbs were verified as a comparatively simple means for predicting the language function of PWA, while other core lexicon lists were not. These findings have potential clinical implications in that verb counts (i.e., using a core verb list or counting verbs produced) might be a discourse measure that is sensitive to capturing comprehensive language ability.

However, there are several limitations to the study that need to be considered in future investigations. A major limitation is the construction of the core lexicon lists. The lists included the 25 most common lemmas produced by cognitively healthy participants for each stimulus. While there is a precedent for defining text from a frequency list (Gottron 2009), the cut-off was mostly arbitrary with ease of use being the most important factor in that decision. A combined frequency list may be plagued with outliers if an individual uses a single lemma significantly more than others. For example, a discourse sample that includes the word 'no' a thousand times might place 'no' at the top of the frequency list, but it would not be descriptive of the text. Future research should address this issue.

We were not able to find noun, adjective, or adverb core lexicon lists that were sensitive to severity of aphasia, perhaps due to the small sample size. Reviewing the demographic information of these participants with aphasia, nearly half presented with Broca's aphasia. Possibly, the expected, limited verb retrieval of individuals with Broca's aphasia drove the statistically significant results. In the same vein, more fluent types of aphasia need to be included to ensure the necessity of expanding grammatical category for both research and clinical judgment similar to how Sarno *et al.* (2005) were able to demonstrate the predictive value of the production of modifiers. Future studies should include a larger number of participants as well as a sufficient number of participants with different aphasia types so as to determine whether the findings are specific to type of aphasia.

Lastly, core lexicon lists should be applied to different discourse elicitation tasks such as picture descriptions, procedural discourse tasks, and storytelling, to explore discourse adequacy by measuring the core lexicon that is most useful for clinical populations. Finally, to establish ecological validity and utility of the core lexicon measure, it is essential that researchers investigate its correlations to other linguistic measures as well as to the standardized tools. Along with the acceptable validity, it could be expected that such an effort will acquire a more useful clinical prediction by requiring less time, training, and efforts when completing these key evaluations in clinical settings.

Acknowledgements

The research was partially supported by a National Institute on Aging Grant (number R01AG029476). The authors are especially grateful to the study participants. ***Declaration of interest:*** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

ANDREETTA, S., CANTAGALLO, A. and MARINI, A., 2012, Narrative discourse in anomic aphasia. *Neuropsychologia*, **50**(**8**), 1787–1793.

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

ARMSTRONG, L., BRADY, M., MACKENZIE, C. and NORRIE, J., 2007, Transcription-less analysis of aphasic discourse: a clinician's dream or a possibility? *Aphasiology*, **21**, 355–374.

BASSO, A., RAZZANO, C., FAGLIONI, P. and ZANOBIO, M. E., 1990, Confrontation naming, picture description and action naming in aphasic patients. *Aphasiology*, **4**, 185–195.

BATES, E., CHEN, S., TZENG, O., LI, P. and OPIE, M., 1991, The noun–verb problem in Chinese aphasia. *Brain and Language*, **41**(**2**), 203–233.

BERNDT, R. S. and HAENDIGES, A. N., 2000, Grammatical class in word and sentence production: evidence from an aphasic patient. *Journal of Memory and Language*, **43**, 249–273.

BERNDT, R. S., MITCHUM, C. C., HAENDIGES, A. N. and SANDSON, J., 1997, Verb retrieval in aphasia. 1. Characterizing single word impairments. *Brain and Language*, **56**(**1**), 68–106.

BEUKELMAN, D. R. and MIRENDA, P., 1998, *Augmentative and Alternative Communication: Management of Severe Communication Disorders in Children and Adults* (Baltimore, MD: Brookes).

BOLES, L. and BOMBARD, T., 1998, Conversational discourse analysis: appropriate and useful sample sizes. *Aphasiology*, **12**, 547–560.

CAPILOUTO, G., WRIGHT, H. H. and WAGOVICH, S. A., 2005, CIU and main event analyses of the structured discourse of older and younger adults. *Journal of Communication Disorders*, **38**, 431–444.

COELHO, C. A., 2002, Story narratives of adults with closed head injury and non-brain-injured adults: Influence of socio-economic status, elicitation task, and executive functioning. *Journal of Speech, Language, and Hearing Research*, **45**(**6**), 1232–1248.

CHEN, S. and BATES, E., 1998, The dissociation between nouns and verbs in Broca's and Wernicke's aphasia: findings from Chinese. *Aphasiology*, **12**(**1**), 5–36.

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*, **24**(**4**), S923–S938.

DAVIS, H. and SILVERMAN, S. R., 1978, *Hearing and Deafness* (New York: NY: Holt, Rinehart & Winston of Canada).

DAY, A., 1985, *Good Dog Carl* (New York, NY: Simon & Schuster).

DELL, G. S., CHANG, F. and GRIFFIN, Z. M., 1999, Connectionist models of language production: lexical access and grammatical encoding. *Cognitive Science*, **23**, 517–542.

DE RIESTHAL, M. and DIEHL, S. K., 2018, Conceptual, methodological, and clinical considerations for a core outcome set for discourse. *Aphasiology*, **32**, 469–471.

DIETZ, A. and BOYLE, M., 2018a, Discourse measurement in aphasia: consensus and caveats. *Aphasiology*, **32**, 487–492.

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

DILLOW, E., 2013, Narrative discourse in aphasia: main concept and core lexicon analyses of the Cinderella story. Doctoral dissertation, University of South Carolina.

ELIA, D., LILES, B. Z., DUFFY, R. J., COELHO, C. A. and BELANGER, S. A., 1994, An investigation of sample size in conversational analysis. Paper presented at the ASHA Convention, New Orleans, LA, USA.

FERGADIOTIS, G. and WRIGHT, H. H., 2011, Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*, **25**(**11**), 1414–1430.

FERGADIOTIS, G., WRIGHT, H. H. and CAPILOUTO, G. J., 2011, Productive vocabulary across discourse types. *Aphasiology*, **25**(**10**), 1261–1278.

FOLSTEIN, M., FOLSTEIN, S. and FANJIANG, G., 2001, *Mini-Mental State Examination*, 2nd edn (Lutz, FL: PAR).

FROMM, D. A., FORBES, M., HOLLAND, A. and MACWHINNEY, B., 2013, 'PWAs and PBJs: Language for describing a simple procedure', paper presented at *Clinical Aphasiology Conference, Tucson, Arizona, 28 May–2 June*.

GOODGLASS, H. and WINGFIELD, A., 1997, Word-finding deficits in aphasia: Brain—behavior relations and clinical symptomatology. In *Anomia* (New York, NY: Elsevier), pp. 3–27.

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* 2009 (Berlin, Heidelburg: Springer), pp. 94–105.

GROSJEAN, F., 1980, Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, **28**, 267–283.

GUO, Y. E., TOGHER, L. and POWER, E., 2014, Speech pathology services for people with aphasia: what is the current practice in Singapore? *Disability and Rehabilitation*, **36**, 691–704.

HEALY, A. F. and MILLER, G. A., 1970, The verb as the main determinant of sentence meaning. *Psychonomic Science*, **20**(**6**), 372–372.

HUGHES, D. L., MCGILLIVRAY, L. and SCHMIDEK, M., 1997, *Guide to Narrative Language: Procedures for Assessment* (Eau Claire, WI: Thinking).

JONKERS, R. and BASTIAANSE, R., 1998, How selective are selective word class deficits? Two case studies of action and object naming. *Aphasiology*, **12**(**3**), 245–256.

KAMBANAROS, M., 2010, Action and object naming versus verb and noun retrieval in connected speech: comparisons in late bilingual Greek–English anomic speakers. *Aphasiology*, **24**, 210–230.

KAPLAN, E., GOODGLASS, H. and WEINTRAUB, S., 1976, *Boston Naming Test, Experimental Edition* (Boston, MA: Boston University).

KERTESZ, A., 2006, *The Western Aphasia Battery—Revised* (San Antonio, TX: Pearson).

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**), 1–25.

KINTZ, S., FERGADIOTIS, G. and WRIGHT, H. H., 2016, Aging effects on discourse production. In H. H. Wright (ed.), *Cognition, Language, and Aging* (Philadelphia, PA: John Benjamins), pp. 81–106.

KRAHN, F., 1978, *The Great Ape* (New York: The Viking Press).

KINTZ, S. and WRIGHT, H. H., 2018, Discourse measurement in aphasia research. *Aphasiology*, **32**, 472–474.

KURLAND, J. and STOKES, P., 2018, Let's talk real talk: an argument to include conversation in a D-COS for aphasia research with an acknowledgment of the challenges ahead. *Aphasiology*, **32**, 475–478.

LILES, B. Z., 1993, Narrative discourse in children with language disorders and children with normal language: a critical review of the literature. *Journal of Speech, Language, and Hearing Research*, **36**(**5**), 868–882.

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 Analysing Talk* (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**(**6–8**), 856–868.

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**(**2**), 211–236.

MARINI, A., ANDREETTA, S., DEL TIN, S. and CARLOMAGNO, S., 2011, A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology*, **25**(**11**), 1372–1392.

MÄTZIG, S., DRUKS, J., MASTERSON, J. and VIGLIOCCO, G., 2009, Noun and verb differences in picture naming: past studies and new evidence. *Cortex*, **45**(**6**), 738–758.

MAYER, J. and MURRAY, L., 2003, Functional measures of naming in aphasia: word retrieval in confrontation naming versus connected speech. *Aphasiology*, **17**(**5**), 481–497.

MCCULLY, E. A., 1984, *Picnic* (New York, NY: HarperCollins).

MCNEIL, M. R., DOYLE, P. J., FOSSETT, T. R. D., PARK, G. H. and GODA, A. J., 2001, Reliability and concurrent validity of the information unit scoring metric for the story retelling procedure. *Aphasiology*, **15**, 991–1006.

NEVILLE, H. J., 2014, Language Hemisphere? In *Modularity and the Motor Theory of Speech Perception: Proceedings of a Conference to Honor Alvin M. Liberman* (New York, NY: Psychology Press), pp. 269.

NICHOLAS, L. E. and BROOKSHIRE, R. H., 1993, A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research*, **36**(**2**), 338–350.

NICHOLAS, L. E. and BROOKSHIRE, R. H., 1995, Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech and Hearing Research*, **38**(**1**), 145–156.

OLNESS, G. S., 2006, Genre, verb, and coherence in picture-elicited discourse of adults with aphasia. *Aphasiology*, **20**(**02–04**), 175–187.

OLNESS, G. S., GYGER, J. and THOMAS, K., 2012, Analysis of narrative functionality: toward evidence-based approaches in managed care settings. *Seminars in Speech and Language*, **33**, 55–67.

PASHEK, G. V. and TOMPKINS, C. A., 2002, Context and word class influences on lexical retrieval in aphasia. *Aphasiology*, **16**(**3**), 261–286.

PRINS, R, and BASTIAANSE, R., 2004, Review. *Aphasiology*, **18**, 1075–1091.

SARNO, M. T., POSTMAN, W. A., CHO, Y. S. and NORMAN, R. G., 2005, Evolution of phonemic word fluency performance in post-stroke aphasia. *Journal of Communication Disorders*, **38**(**2**), 83–107.

SCHWARTZ, M. F., SAFFRAN, E. M. and MARIN, O. S., 1980, The word order problem in agrammatism: I. Comprehension. *Brain and Language*, **10**(**2**), 249–262.

SHERRATT, S., 2007, Multi-level discourse analysis: a feasible approach. *Aphasiology*, **21**(**3–4**), 375–393.

Ska, B., Duong, A. and Joanette, Y., 2004, Discourse impairments. In R. D. Kent (ed.), *The MIT Encyclopedia of Communication Disorder* (Cambridge, MA: MIT Press), pp. 302–304.

Verhaegen, C. and Poncelet, M., 2013, Changes in naming and semantic abilities with aging from 50 to 90 years. *Journal of the International Neuropsychological Society*, **19**(**2**), 119–126.

Verna, A., Davidson, B. and Rose, T., 2009, Speech–language pathology services for people with aphasia: a survey of current practice in Australia. *International Journal of Speech–Language Pathology*, **11**, 191–205.

Wallace, S. J., Worrall, L. E., Rose, T. and Le Dorze, G., 2018, Discourse measurement in aphasia research: have we reached the tipping point? A core outcome set or greater standardisation of discourse measures? *Aphasiology*, **32**, 479–482.

Whitworth, A., 2018, The tipping point: are we nearly there yet? *Aphasiology*, **32**, 483–486.

Williams, S. E. and Canter, G. J., 1982, Action-naming performance in four syndromes of aphasia. *Brain and Language*, **32**(**1**), 124–136.

Wilshire, C. E. and McCarthy, R. A., 2002, Evidence for a context-sensitive word retrieval disorder in a case of nonfluent aphasia. *Cognitive Neuropsychology*, **19**, 165–186.

Wright, H. H. and Capilouto, G. J., 2009, Manipulating task instructions to change narrative discourse performance. *Aphasiology*, **23**(**10**), 1295–1308.

Wright, H. H. and Capilouto, G. J., 2012, Considering a multilevel approach to understanding maintenance of global coherence in adults with aphasia. *Aphasiology*, **26**(**5**), 656–672.

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**(**3**), 900–917.

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) (available at: http://doi.org/10.3886/ICPSR36634.v1).

Zingerser, L. B. and Berndt, R. S., 1990, Retrieval of nouns and verbs in agrammatism and anomia. *Brain and Language*, **39**(**1**), 14–32.

Appendix A: CLAN commands

(1) Generate a morphological Analysis: mor + t* SUB*.gem.cex
(2) Generate Syntactic Categories: post + t* SUB*.gem.mor.cex
(3) Extract the two stories: gem + t%mor + t* SUB + sStory + d + f_n*.cha
(4) Extract all lemma forms with frequencies information: freq + t%mor + s@"r*,|*,o%" + u + d2*.gem.mor.pst.cex

Appendix B: Top 25 core lexicon produced by the control group

Good Dog Carl

20s				30s			
Nouns	Verbs	Adjectives	Adverbs	Nouns	Verbs	Adjectives	Adverbs
baby	go	good	then	baby	go	little	then
dog	get	little	back	Carl	get	good	back
Carl	put	back	there	dog	put	clean	all
mother	look	hungry	just	crib	take	back	there
mom	take	big	all	mom	make	hungry	up
crib	come	messy	where	bed	run	big	upstairs
bed	have	open	upstairs	back	have	open	now
back	play	sure	up	laundry	come	great	away
room	make	own	out	room	decide	sure	where
laundry	run	left	down	mother	play	nice	out
mess	see	dirty	away	milk	see	happy	just
chute	decide	happy	next	time	watch	ready	so
milk	leave	nice	very	mess	clean	messy	next
bread	say	huge	in	window	leave	dirty	on
home	clean	next	now	bread	dance	tired	down
window	eat	right	again	cookie	jump	left	off
time	let	able	also	chute	turn	next	very
child	turn	hot	around	home	ride	smart	in
butter	dry	dangerous	so	butter	let	first	shortly
cookie	do	first	shortly	kitchen	dry	right	sure
bath	know	great	on	house	say	whole	again
kitchen	watch	long	more	chocolate	know	awesome	over
grape	find	ready	how	grape	eat	huge	really
fish tank	ride	young	once	bath	wash	old	around
soap	dance	early	as	floor	find	pretty	soon

40s				50s			
Nouns	Verbs	Adjectives	Adverbs	Nouns	Verbs	Adjectives	Adverbs
baby	go	little	then	baby	go	little	then
Carl	get	good	back	Carl	get	good	back
dog	put	hungry	there	dog	look	big	there
crib	look	back	up	bed	put	hungry	up
back	have	big	all	mother	take	sure	all
bed	take	clean	just	crib	have	messy	just
mom	come	dry	now	back	clean	open	now
laundry	play	messy	upstairs	mom	come	next	upstairs
mother	run	great	out	laundry	play	fun	out
milk	make	sure	very	milk	see	dirty	where
room	see	dirty	where	room	make	right	on
chute	let	happy	away	window	ride	left	here
window	know	nice	down	mess	say	ready	in
mess	say	fun	in	chute	decide	smart	so
bread	clean	old	on	cookie	run	own	away
chocolate	watch	next	around	time	turn	great	down
time	ride	ready	here	bread	dry	happy	next
cookie	decide	right	over	powder	let	pretty	shortly
floor	leave	first	off	kitchen	leave	nice	very
kitchen	do	safe	next	floor	watch	whole	again
soap	open	whole	again	butter	know	wonderful	how
butter	climb	wonderful	right	chocolate	eat	close	off
home	find	bad	home	swim	excited	yet	
grape	lay	full	shortly	grape	give	fine	over
fish tank	jump	gentle	together	fish	climb	first	sure

60s				70s			
Nouns	Verbs	Adjectives	Adverbs	Nouns	Verbs	Adjectives	Adverbs
baby	go	little	then	baby	go	good	then
dog	get	good	back	dog	get	little	back
Carl	look	clean	there	Carl	look	big	there
bed	put	big	all	back	put	next	all
back	take	dry	up	crib	have	dirty	just
crib	have	hungry	just	bed	take	open	now
mother	come	open	now	mother	come	nice	out
mom	run	happy	out	laundry	see	hungry	up
laundry	play	great	upstairs	window	clean	first	again
chute	make	next	where	milk	say	old	upstairs
milk	decide	dirty	very	powder	do	sure	where
mess	say	left	in	floor	know	wonderful	over
powder	see	nice	so	chute	turn	happy	very
window	ride	old	down	bread	find	great	off
time	do	sure	probably	boy	sit	same	here
bread	find	first	over	mess	make	messy	on
floor	watch	messy	shortly	mom	climb	right	in
butter	know	pretty	off	kitchen	watch	pretty	shortly
cookie	climb	smart	again	cookie	play	cute	so
chocolate	clean	soft	away	butter	decide	tired	down
aquarium	let	tired	on	chocolate	dance	whole	next
home	dance	able	around	thing	let	close	away
room	sit	close	also	time	dry	smart	yet
soap	swim	full	here	aquarium	eat	innocent	how
thing	eat	whole	soon	puff	open	own	apparently

80s			
Nouns	Verbs	Adjectives	Adverbs
baby	go	little	then
dog	get	good	back
Carl	look	big	there
back	put	next	all
bed	take	happy	up
boy	have	dirty	just
mother	come	nice	now
crib	play	great	out
window	see	open	where
laundry	do	old	on
bread	say	messy	down
milk	clean	smart	again
powder	know	sure	here
floor	watch	right	very
butter	find	ready	upstairs
chute	straighten	wonderful	in
chocolate	make	left	so
puff	ride	small	apparent
child	wash	wet	off
cookie	give	able	over
fish	open	whole	how
head	let	close	shortly
time	climb	first	around
room	guess	pretty	really
thing	run	different	away

Picnic

20s				30s			
Nouns	Verbs	Adjectives	Adverbs	Nouns	Verbs	Adjectives	Adverbs
mouse	go	little	all	mouse	go	little	back
family	get	stuffed	back	picnic	get	stuffed	all
picnic	have	happy	there	family	see	missing	there
truck	look	missing	then	truck	have	happy	then
baby	start	lost	just	road	find	pink	out
road	eat	red	meanwhile	baby	look	sad	where
back	find	sad	out	back	eat	good	up
time	play	pink	where	child	start	big	meanwhile
child	see	good	very	time	realize	lost	just
animal	decide	big	still	girl	decide	hungry	so
kid	come	hungry	so	kid	begin	ready	still
berry	call	left	up	berry	come	bumpy	around
flower	know	bumpy	around	flower	run	whole	very
girl	notice	ready	finally	animal	hear	great	now
mom	pick	excited	together	doll	know	left	together
rest	do	small	alone	raspberry	fall	excited	here
grass	begin	scared	now	grass	take	scared	finally
day	drive	old	really	food	do	old	behind
rock	cry	same	off	baseball	pick	own	alone
raspberry	run	beautiful	maybe	rock	continue	young	down
dad	take	own	again	water	sit	small	again
food	hear	whole	behind	toy	give	beautiful	once
baseball	hug	glad	even	lunch	hit	high	even
way	fall	long	here	mom	forget	nearby	really
lake	sit	nice	away	bush	hug	nice	on

40s			
Nouns	Verbs	Adjectives	Adverbs
mouse	go	little	all
picnic	look	stuffed	back
truck	get	missing	there
family	have	red	then
road	play	pink	just
baby	eat	happy	where
back	find	lost	out
time	start	good	very
berry	see	big	up
flower	realize	sad	still
rat	come	hungry	so
child	sit	small	around
animal	know	left	here
rock	run	ready	now
food	call	old	together
blanket	cry	beautiful	again
bush	do	scared	finally
doll	take	nice	alone
kid	drive	whole	behind
day	fall	young	off
lake	decide	great	suddenly
grass	pick	excite	away
toy	lay	bumpy	down
lunch	begin	poor	on
area	notice	full	probably

50s			
Nouns	Verbs	Adjectives	Adverbs
mouse	go	little	all
picnic	look	stuffed	back
truck	get	happy	there
family	have	missing	then
road	play	big	where
baby	eat	pink	out
back	see	good	meanwhile
kid	find	sad	just
child	start	lost	so
grandpa	come	hungry	very
mom	sit	ready	up
flower	call	whole	around
raspberry	do	full	still
time	decide	bumpy	now
animal	pick	left	here
berry	realize	beautiful	together
toy	know	wonderful	finally
dad	take	poor	maybe
doll	run	old	behind
blanket	cry	great	down
grass	drive	small	really
food	think	nice	even
grandma	notice	tall	again
day	hear	lonely	off
way	continue	scared	alone

60s			
Nouns	Verbs	Adjectives	Adverbs
mouse	go	little	all
picnic	look	happy	there
truck	have	good	back
family	get	lost	just
road	play	pink	so
baby	find	big	then
time	eat	missing	out
flower	start	stuffed	very
back	see	red	where
mom	sit	sad	up
doll	come	hungry	around
kid	do	whole	still
berry	take	great	meanwhile
dad	know	ready	maybe
toy	decide	poor	now
grandpa	cry	beautiful	here
child	pick	nice	together
raspberry	run	old	finally
food	fall	young	even
brother	realize	bumpy	down
water	drive	left	really
baseball	lay	full	off
bush	call	tall	again
blanket	swim	wonderful	on
area	jump	aware	probably

70s			
Nouns	Verbs	Adjectives	Adverbs
mouse	go	little	all
picnic	look	good	there
truck	have	happy	back
road	get	big	out
baby	play	lost	then
family	eat	missing	now
time	seem	pink	where
flower	find	red	up
doll	sit	stuffed	so
back	come	nice	still
toy	start	sad	very
berry	take	ready	just
child	pick	old	around
kid	know	poor	here
basket	do	great	meanwhile
water	cry	hungry	again
grandpa	decide	bumpy	maybe
bush	lay	whole	suddenly
tree	run	left	off
thing	call	full	down
rock	realize	young	meantime
rest	fall	beautiful	finally
place	think	small	even
middle	swim	wonderful	together
girl	hear	glad	on

80s			
Nouns	Verbs	Adjectives	Adverbs
mouse	go	little	there
picnic	have	lost	all
truck	look	big	then
road	get	good	back
baby	play	happy	out
family	eat	ready	just
flower	come	red	where
toy	see	missing	still
time	find	pink	so
back	sit	poor	up
mother	know	sad	here
food	take	old	now
basket	start	bumpy	around
picture	pick	great	very
thing	cry	nice	on
baseball	decide	wonderful	meanwhile
water	think	beautiful	down
berry	do	whole	even
rock	guess	hungry	maybe
place	hug	stuffed	off
watermelon	lay	glad	again
doll	fall	pretty	apparently
home	jump	next	along
banjo	swim	full	away
child	put	small	ever

Total number of different nouns, verbs, adjectives and adverbs produced in the controls

Good Dog Carl

	Nouns	Verbs	Adjectives	Adverbs
20s	555	467	421	128
30s	543	469	399	116
40s	570	483	413	119
50s	569	473	392	135
60s	585	484	431	129
70s	641	508	440	129
80s	650	510	432	127

Picnic

	Nouns	Verbs	Adjectives	Adverbs
20s	557	448	423	128
30s	483	391	392	110
40s	531	422	453	136
50s	494	407	421	116
60s	575	443	469	121
70s	587	442	461	116
80s	567	440	451	111

Total number of different nouns, verbs, adjectives and adverbs produced by participants with aphasia

Good Dog Carl

	Nouns	Verbs	Adjectives	Adverbs
P1	19	9	8	1
P3	30	32	6	7
P4	61	16	32	12
P5	27	7	10	8
P6	n.a.	n.a.	n.a.	n.a.
P7	27	18	7	7
P9	35	12	14	9
P10	18	12	7	3
P11	7	8	7	8
P12	42	31	7	6
P13	31	9	19	7

Picnic

	Nouns	Verbs	Adjectives	Adverbs
P1	13	13	11	7
P3	21	18	15	8
P4	28	10	19	10
P5	34	7	17	7
P6	11	7	4	2
P7	20	22	7	6
P9	18	11	12	5
P10	23	5	7	4
P11	13	7	8	8
P12	41	24	13	10
P13	24	5	26	3