AJSLP

Supplement Article

Core-Lexicon and Main-Concept Production During Picture-Sequence Description in Adults Without Brain Damage and Adults With Aphasia

Sarah Grace Dalton[a] and Jessica D. Richardson[a]

Purpose: We sought to identify the core lexicon of a picture-description task using transcripts from the AphasiaBank database and to determine differences in core-lexicon usage between control speakers and persons with aphasia (PWAs). We also investigated the relationship between core lexicon and an established discourse measure, main-concept analysis.

Method: A core lexicon was developed by identifying lemmas produced by 92 control speakers. Transcripts were scored—165 control transcripts and 238 PWA transcripts—using the core lexicon and a recently developed main-concept list. Median tests examined differences between controls, PWAs, and aphasia subtypes. Spearman's correlations assessed the relationship between core-lexicon and main-concept performance.

Results: A 24-item core lexicon was identified. Significant differences were found between control speakers and PWAs, and between aphasia subtypes, for core-lexicon and main-concept scores. Core-lexicon and main-concept performance was significantly and positively correlated for all groups.

Conclusions: We report the development of a core lexicon, differences in core-lexicon usage between speakers, and the relationship between core-lexicon and main-concept scores. Research is needed to determine the clinical utility and psychometric properties of these discourse measures and their potential contribution to multilevel discourse analysis of functional communication.

In clinical settings, assessment of language difficulties in a person with aphasia (PWA) generally consists of measuring impairments in discrete language domains, most often relying on commonly used diagnostic instruments or on assessment batteries developed inhouse (Simmons-Mackie, Threats, & Kagan, 2005; Verna, Davidson, & Rose, 2009). Derived scores may or may not predict real-world performance and thus have low ecological validity; indeed, it is not uncommon or unexpected for psychometrically sound diagnostic instruments to lack correspondence to real-world abilities (Chaytor & Schmitter-Edgecombe, 2003). Further, time constraints and resource limitations may prevent full examination of how deficits in multiple domains interact and affect functional communication (Simmons-Mackie et al., 2005). There has recently been a focus on improving the ecological validity of assessment measures by ensuring that they predict communication abilities in everyday situations (Marini, Andreetta, del Tin, & Carlomagno, 2011; Mayer & Murray, 2003; Ross & Wertz, 1999). As part of this initiative, a great deal of research has focused on the discourse of individuals with aphasia and how it differs from that of individuals with normal language abilities (for a review, see E. Armstrong, 2000). *Discourse* is the highly individualized and complex process humans use to send information to and receive information from others. The motivations behind discourse vary widely and include cooperation, survival, and societal engagement (Carey, 1988; Dimbleby & Burton, 1998). Of particular importance, discourse may be a better predictor of functional communication abilities and treatment outcomes than standard assessment measures (Larfeuil & Le Dorze, 1997; Ross & Wertz, 1999), highlighting the need for valid and reliable discourse measures.

Discourse in typical and clinical populations has been investigated with a variety of structuralist and functionalist techniques (E. Armstrong, 2000), yielding a plethora of

[a]University of South Carolina

Correspondence to Sarah Grace Dalton, who is now at the University of New Mexico: hudspeth@unm.edu

Jessica D. Richardson is now at the University of New Mexico.

Editor: Anastasia Raymer
Associate Editor: Melissa Duff

Received September 16, 2014
Revision received January 26, 2015
Accepted May 18, 2015
DOI: 10.1044/2015_AJSLP-14-0161

Disclosure: The authors have declared that no competing interests existed at the time of publication.

measures for clinicians and researchers to use. Structuralist measures, such as type–token ratio, T-units, and mean length of utterance, examine an individual's microlinguistic use of language. Functionalist measures, such as rating scales of communicative effectiveness, measures of conversational repair and turn-taking, and ratings of coherence, examine the speaker's ability to successfully communicate a message and are macrolinguistic in nature. There are other discourse-measurement approaches that do not fit neatly into either the structuralist or functionalist divisions (E. Armstrong, 2000). One approach examines the informativeness of discourse by assessing the quantity of correct or relevant information provided. This approach includes analysis of main concepts (MCs; Kong, 2009; Nicholas & Brookshire, 1993b, 1995), content units (Yorkston & Beukelman, 1980), and/or correct information units (Nicholas & Brookshire, 1993a). Main-event analysis (Capilouto, Wright, & Wagovich, 2005) attempts to combine aspects of the informational and functional measures already discussed by including informational units that also contain relational concepts, thus providing information about discourse cohesion.

A multilevel approach that includes assessment of functional, structural, informational, and efficiency domains of discourse may be what is needed to best predict real-world discourse and conversational performance (E. Armstrong, 2000; Marini et al., 2011; Wright & Capilouto, 2012). The depth and breadth of information yielded by such assessments would be highly beneficial for prognosis, treatment planning, and outcomes assessment. However, clinicians often lack the time (Duncan & Murray, 2012; Simmons-Mackie et al., 2005; Verna et al., 2009) and specialized training (Duncan & Murray, 2012) needed to transcribe and analyze a discourse sample with a single measure, much less the various measures needed for a multilevel approach. What is needed are standardized and norm-referenced measures that can provide a similar array of information while reducing the time and training required to make discourse assessment clinically feasible. In this supplement article we discuss the potential utility of a quick and easy-to-derive discourse measure that may be a useful addition to clinician-friendly multilevel assessment.

An important consideration when developing measures is the type of discourse to be analyzed. Frequently elicited discourse tasks include personal narrative, storytelling (or retelling), procedural description, and picture (scene or sequence) description. Significant differences have been found among the discourse tasks. For example, single pictures elicit more descriptive than narrative statements compared to picture sequences (Olness, Ulatowska, Wertz, Thompson, & Auther, 2002), and procedural tasks elicit simpler language than picture sequences or story retelling (Ulatowska, North, & Macaluso-Haynes, 1981). In addition, story retellings and picture-sequence descriptions often elicit greater lexical diversity than picture-scene description, even when participants are instructed to tell a story with a beginning, middle, and end (Fergadiotis & Wright, 2011). Personal narratives probably best approximate typical communicative interactions and therefore may have the highest ecological validity; however, precisely because they are highly individualized, they are the most difficult to standardize and compare across PWAs (e.g., Nicholas & Brookshire, 1993b). Research involving a picture-sequence description was decided upon for this study because picture-sequence description consistently elicits a narrative similar to story retelling (Fergadiotis & Wright, 2011; Ulatowska et al., 1981) while requiring less time for collection and analysis—important factors to consider when attempting to create clinically useful measures.

MacWhinney, Fromm, Holland, Forbes, and Wright (2010) suggested that analysis of the typical vocabulary, the "target lexicon," used during structured narrative tasks could provide a time-efficient and informative index of functional communication abilities. For example, clinicians could bypass lengthy transcriptions, instead generating a list of words spoken during narration for later comparison to the target lexicon. Such a measure could also potentially be scored online (or, more reasonably, with an audio recording) using a simple checklist that the clinician monitors while the speaker is producing the narrative. Task-specific lexicons have so far been investigated for a storytelling task (Cinderella; MacWhinney et al., 2010) and a procedural task in which clients describe how to make a peanut butter-and-jelly sandwich (Fromm, Forbes, Holland, & MacWhinney, 2013). Using a subset of data from the AphasiaBank, a database of videos and discourse transcripts produced by control speakers and PWAs, MacWhinney et al. (2010) identified 306 words (including nouns, verbs, adjectives, and adverbs) produced by control speakers that seemed to capture the gist of the Cinderella story. From this list, nouns and verbs were separated out and considered to be the target lexicon for comparison or reference. The researchers then determined the 10 most frequently used nouns and verbs from that target lexicon, separately for PWAs and control speakers. They found significantly reduced lexical diversity for PWAs, characterized by less informative words used in place of the highly informative words produced by control speakers (e.g., saying "girl" rather than "stepsister"), and fewer words produced by multiple speakers. Transcripts of PWAs contained fewer abstract nouns and a larger proportion of "light" verbs (*be, have, come, go,* etc.; MacWhinney et al., 2010). Using similar methods, Fromm et al. (2013) examined the lexicon for the peanut butter-and-jelly task, finding no differences in the top nouns and only two differences in the top verbs used in each group.

Although there are many different lexical options that can be selected during storytelling or picture description, aphasia often interferes with an individual's ability to retrieve the most typical and familiar words for the telling, likely placing a strain on working-memory capacity. Because increased working-memory load affects word retrieval and sentence production even in individuals with normal language (Ferreira & Pashler, 2002; Hartsuiker & Barkhuysen, 2006), it is even more likely that it would negatively affect discourse in PWAs. There is some evidence to suggest that performance on a target lexicon may predict other structural or informational discourse performance (Andreetta, Cantagallo, & Marini, 2012), but direct evidence of this relationship is lacking. We would intuitively expect that if there are shared events or concepts for a community narrative, then that narrative might also share a common vocabulary. If individuals experience difficulty retrieving this common vocabulary, it could impede their ability to produce those shared events or concepts. An examination of the relationship between shared lexicon and concepts is needed.

MC analysis (Kong, 2009, 2011; Nicholas & Brookshire, 1993b, 1995; Richardson & Dalton, 2015) is a rule-based system of scoring how accurately and completely speakers produce concepts considered to be essential for conveying the gist of a shared narrative. Significant differences in MC production between control speakers and PWAs, as well as between speakers with fluent and nonfluent aphasia, have been documented (Kong, 2009; Nicholas & Brookshire, 1993b, 1995). MC analysis is easy to perform (L. Armstrong, Brady, Mackenzie, & Norrie, 2007; Kong, 2009), reliable across raters (Kong, 2009, 2011; Nicholas & Brookshire, 1993a, 1995; Richardson & Dalton, 2015), and stable across sessions for control speakers and PWAs (Boyle, 2014; Kong, 2011; Nicholas & Brookshire, 1993b, 1995). It is important to note that MC measures, specifically percentage of accurate and complete MCs, have outperformed standardized assessment measures of impairment and activity limitations in predicting socially valid change in response to treatment (Ross & Wertz, 1999). Given these strengths and the recent introduction of MC checklists for several discourse tasks (Richardson & Dalton, 2015), MC analysis is uniquely suited for the aforementioned investigation.

Given that lexical retrieval difficulties are characteristic in even the mildest of aphasia subtypes, and that these difficulties appear to negatively affect functional communication, our aims were threefold. First, we sought to develop a core lexicon (CoreLex) list for the Broken Window picture sequence (Menn et al., 1998) from a large sample of control speakers selected from the AphasiaBank database. Second, we sought to report on CoreLex performance for both control speakers and PWAs, investigating differences between the two groups and between aphasia subtypes. We hypothesized that control speakers and PWAs would perform differently on CoreLex production and that differences would exist between aphasia subtypes. Third, we wanted to characterize the relationship between CoreLex and MC performance in all speakers. We expected that control speakers would produce more of the CoreLex and more MCs than all other comparison groups. We hypothesized that there would be a strong positive correlation between CoreLex and MCs —that is, as CoreLex decreases, MC production decreases.

Method

Transcripts

Transcripts from 92 control speakers (55 female, 37 male) were selected from the 198 control participants in the AphasiaBank database at the time of the study to create the core lexicon. The mean age was 58.3 (±21.6) years, with a range of 20–89.5 years (see Table 1). The sample included 23 speakers from four age groups (20–39, 40–59, 60–79, and 80–99 years), with age groups relatively matched for gender and years of education (M = 15.6, SD = 2.5). This same speaker sample was used in previous research to develop the MC checklist used in this study. For a more detailed discussion of the rationale and distribution characteristics of the normative sample, please see Richardson and Dalton (2015).

To investigate CoreLex and MC production in this study, transcripts of 166 control speakers (85 female, 81 male) and 235 PWAs (97 female, 138 male) were retrieved from the AphasiaBank database. The average age of the control speakers was 62.6 (±19.4) years, whereas the average age of PWAs was 60.7 (±12.8) years. Of the 235 PWAs, 79 were diagnosed with anomic aphasia (Aphasia Quotient [AQ] M = 85.2, SD = 7.3), 63 with Broca's aphasia (AQ M = 49.7, SD =15.9), 46 with conduction aphasia (AQ M = 69.9, SD = 9.8), and 22 with Wernicke's aphasia (AQ M = 52.6, SD = 14.2; see Table 2). In addition, there were 25 "recovered" PWAs who did not receive a clinical diagnosis of aphasia on the basis of Western Aphasia Battery–Revised (WAB or WAB-R; Kertesz, 2006) scores (AQ ≥ 93.8) and who are subsequently referred to in the AphasiaBank database as "not aphasic by WAB" (NABW; AQ M = 96.4, SD = 1.7). (Additional transcripts for subtypes of global, transcortical motor, and transcortical sensory are available, but with numbers so small they were excluded from this analysis.)

The Broken Window picture-description narrative was extracted from transcripts using the Computerized Language Analysis (CLAN) command `gem +sWindow +n +fWindow +d1 +t*PAR +t%mor *.cha`. This command outputs a separate file (+fWindow) for each participant (*.cha) containing only the selected gem (gem +sWindow +n), with utterances produced by the participant only (+t*PAR). It also includes morphological information such as parts of speech (+t%mor; see the Appendix for examples of control and PWA transcripts). On occasion, the gem within an individual's transcript was labeled with a different name; to ensure that all gems had been located, a manual search was completed for each missing transcript. When a missing gem was identified, the command was edited to correspond to the name in that particular file (i.e., +swindow or +sBreakingWindow) and the command was rerun.

CoreLex

We first identified the entire spoken lexicon used by our normative sample during Broken Window narration. We then used the CLAN command `freq +t*PAR +s"@r-*,|-*,o-%" +o *.gem.cex +d2 -s"[+exc]"` to identify the unique lemmas (for example, the lemma "run" would include "run," "runs," "running," "ran," etc.) in Broken Window transcripts. Lemmas produced by 50% or more of the normative sample were included in the CoreLex, with this cutoff selected because it yields a reasonably sized (clinically manageable) lexicon and has served as a cutoff or threshold criterion in previous language research (e.g., Brown's stages of development; Owens, 2008).

Table 1. Demographic information for all control speakers, divided by age and inclusion in the normative sample.

Speaker group	Age (yrs)	Gender		Education (yrs)	Race/ethnicity		
		Female	Male		White	African American	Hispanic/Latino
All control speakers (n = 166)	62.6 (±19.4)	85	81	15.4 (±2.5)	160	3	3
Normative sample (n = 92)	58.3 (±21.6)	55	37	15.6 (±2.5)	88	2	2
20–39 years (n = 23)	29.6 (±5.8)	14	9	15.9 (±2.5)	20	2	1
40–59 years (n = 23)	48.4 (±6.3)	15	8	15.7 (±2.5)	22	1	0
60–79 years (n = 23)	71.6 (±4.7)	13	10	15.7 (±2.4)	23	0	0
80–99 years (n = 23)	83.9 (±2.9)	13	10	15.3 (±2.8)	23	0	0

In addition, because the purpose was to establish a "core" lexicon, a less stringent cutoff would include words used by fewer than half of control speakers when telling the Broken Window story. Twenty-four lemmas were identified as being produced by 50% or more of the control participants (see Figure 1). (We also present an additional 26 lemmas, so that the top 50 lemmas are displayed.) Previous research has established a target lexicon using only nouns and verbs from participant transcripts. For this study, however, we chose to include the most commonly used words, regardless of word type (i.e., nouns, verbs, adjectives, adverbs, pronouns, determinants, etc.), to ensure that any differences in word-type production across the aphasia subtypes would be preserved (Bates & Goodman, 1997; Caramazza & Berndt, 1985; Goodglass & Menn, 1985). In this way, the clinical utility of the CoreLex may be maximized by preserving differences among the aphasia subtypes. In addition, production of coordinating and subordinating junctions, referents, and/or prepositions often indicates that speakers are using more elaborated phrase and sentence structures in attempts to tie story elements together (Halliday & Hasan, 1976); inclusion of these structure words may therefore strengthen the relationship between CoreLex and other discourse measures.

After identifying the CoreLex for Broken Window, the command was run with all transcripts of control speakers and PWAs to identify the lemmas produced. The output was subsequently scored; speakers (row) received a score of 1 if the lemma (column) was present in the output and a score of 0 if it was absent. The number of CoreLex lemmas produced by a given speaker served as the CoreLex score for that speaker. (The number of times a lemma was produced within a speaker transcript was not considered; instead, we focused on the presence or absence of the lemma in the transcript).

To provide additional information about the lexical retrieval of persons with different types of aphasia, we ran a CLAN command (`freq +t*PAR +s"@r-*,o-%" +o *.Window.cex +d2 -s"[+exc]"`) that generated all lemmas produced by each aphasia subtype. The 50 most frequent lemmas (as measured by the number of individuals producing the lemma at least once) produced in each aphasia subtype were visually compared to the 50 most frequent lemmas produced by control speakers for presence of CoreLex items, the top 50 lemmas (i.e., lemmas commonly produced by control speakers that did not reach criterion for CoreLex), and unique lemmas that were not present in the top 50 list for control speakers (see Figures 2, 3, 4, 5, and 6).

Main Concepts

Transcripts in the current study were examined for the previously identified list of eight Broken Window MCs (Richardson & Dalton, 2015; see Table 3) and were scored using a multidimensional scoring system modified from Kong (2009) and used by Richardson and Dalton (2015). A numeric value was assigned to each MC attempt on the basis of the accuracy and completeness of the concept. A score of 0 was assigned for MCs that were missing (absent; AB) from the transcript. A score of 3 was assigned when the concept produced was accurate and contained all of the essential elements (accurate and complete; AC). Between those two extremes, a score of 1 was assigned to an inaccurate and incomplete (II) concept and a score of 2 was assigned to an inaccurate and complete (IC) concept and also to an accurate and incomplete concept (AI). The scoring formula used to yield an overall MC score for each speaker was then (3 × AC) + (2 × AI) + (2 × IC) + (1 × II). To provide additional information regarding the MC production of persons with different types of aphasia, we present information about the percentage of speakers in each group that attempted production of each MC. Attempts include AC, AI, IC, and II codes.

Data Analysis

Omnibus median tests were conducted to confirm hypothesized differences between CoreLex and MC scores for control speakers and PWAs. Nonparametric median tests were selected to evaluate differences between groups because the groups exhibited skewed and heterogeneous distributions, indicating that parametric statistics were not appropriate. Planned comparisons using median tests with Holm–Bonferroni corrections were used to identify differences between aphasia subtypes. Spearman correlations, the nonparametric test corresponding to Pearson correlations, were performed to investigate the relationship between CoreLex and MC scores for all participants, and for the control participants and the aphasia subtypes separately.

Results

Core Lexicon

The speakers comprising the normative sample (n = 92) produced 8,255 words, with 767 unique lemmas. All control speakers (n = 166) produced 15,018 words, with 1,040 unique

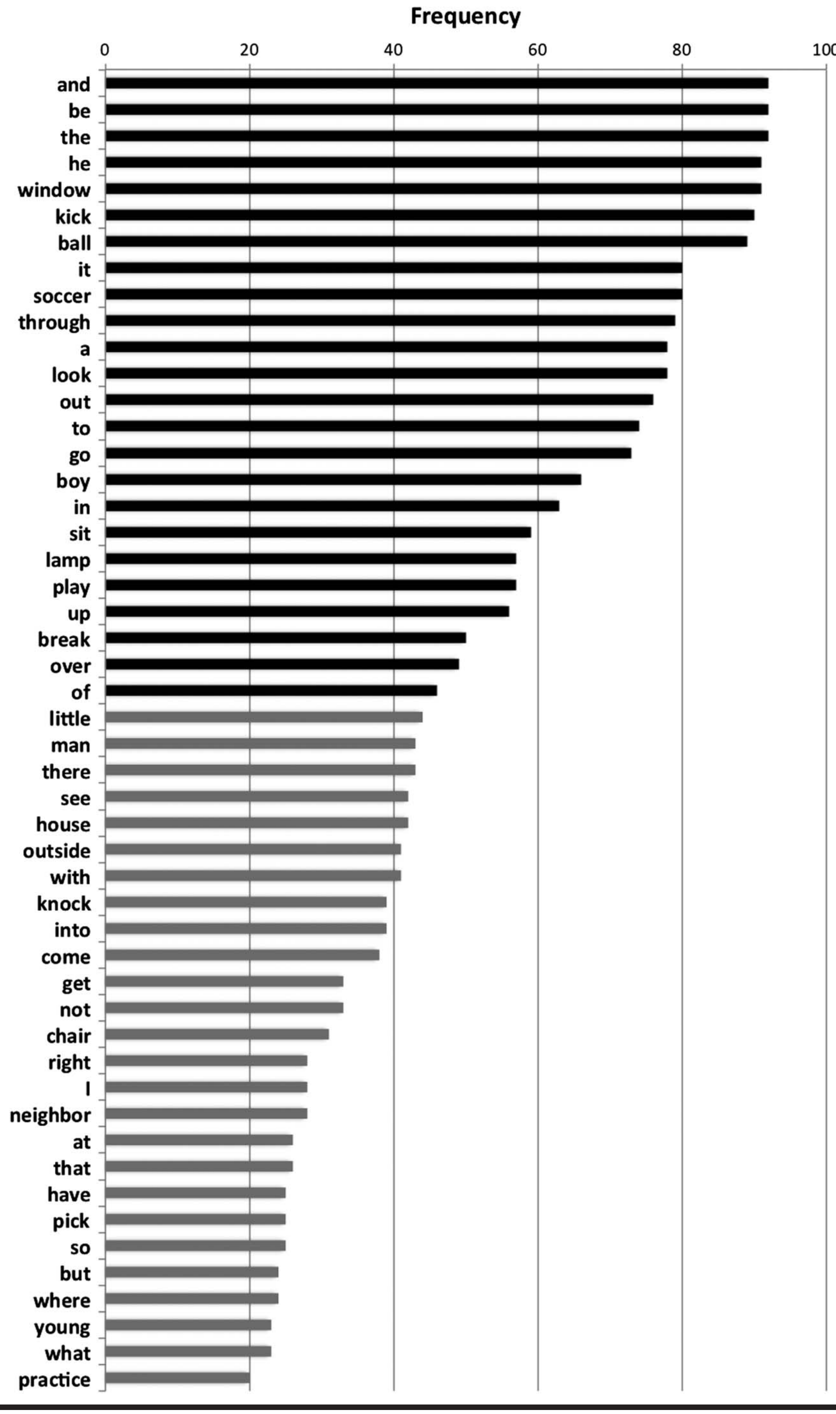
Table 2. Demographic information for all individuals with aphasia and by aphasia subtype.

Group	Age (yrs)	Time since stroke (yrs)	WAB Aphasia Quotient[c]	Gender		Education (yrs)	Race/ethnicity				
				Female	Male		White	African American	Hispanic/Latino	Other	Unknown
All[a] (*n* = 235)	60.7 (±12.8)	5.3 (±4.8)	70.9 (±20.3)	97	138	15.5 (±2.8)	204	16	9	5	1
Anomic (*n* = 79)	61 (±21.6)	4.5 (±4)	85.2 (±7.3)	32	47	15.6 (±2.7)	74	2	3	*na*	*na*
Broca's (*n* = 63)	57 (±13)	6.4 (±5.4)	49.7 (±15.9)	21	43	14.9 (±2.7)	49	8	4	2	*na*
Conduction[b] (*n* = 46)	62.7 (±12.1)	15.31 (±4.7)	69.9 (±9.8)	22	24	15.6 (±3.1)	40	3	*na*	3	*na*
NABW (*n* = 25)	61.4 (±14.1)	5.1 (±3.7)	96.4 (±1.7)	15	10	16.1 (±2.9)	22	1	2	*na*	*na*
Wernicke's[b] (*n* = 22)	65.7 (±10.9)	5.3 (±6.9)	52.6 (±14.2	8	14	16 (±2.5)	19	2	*na*	*na*	1

Note. NABW = not aphasic by Western Aphasia Battery–Revised (WAB); *na* = not applicable.

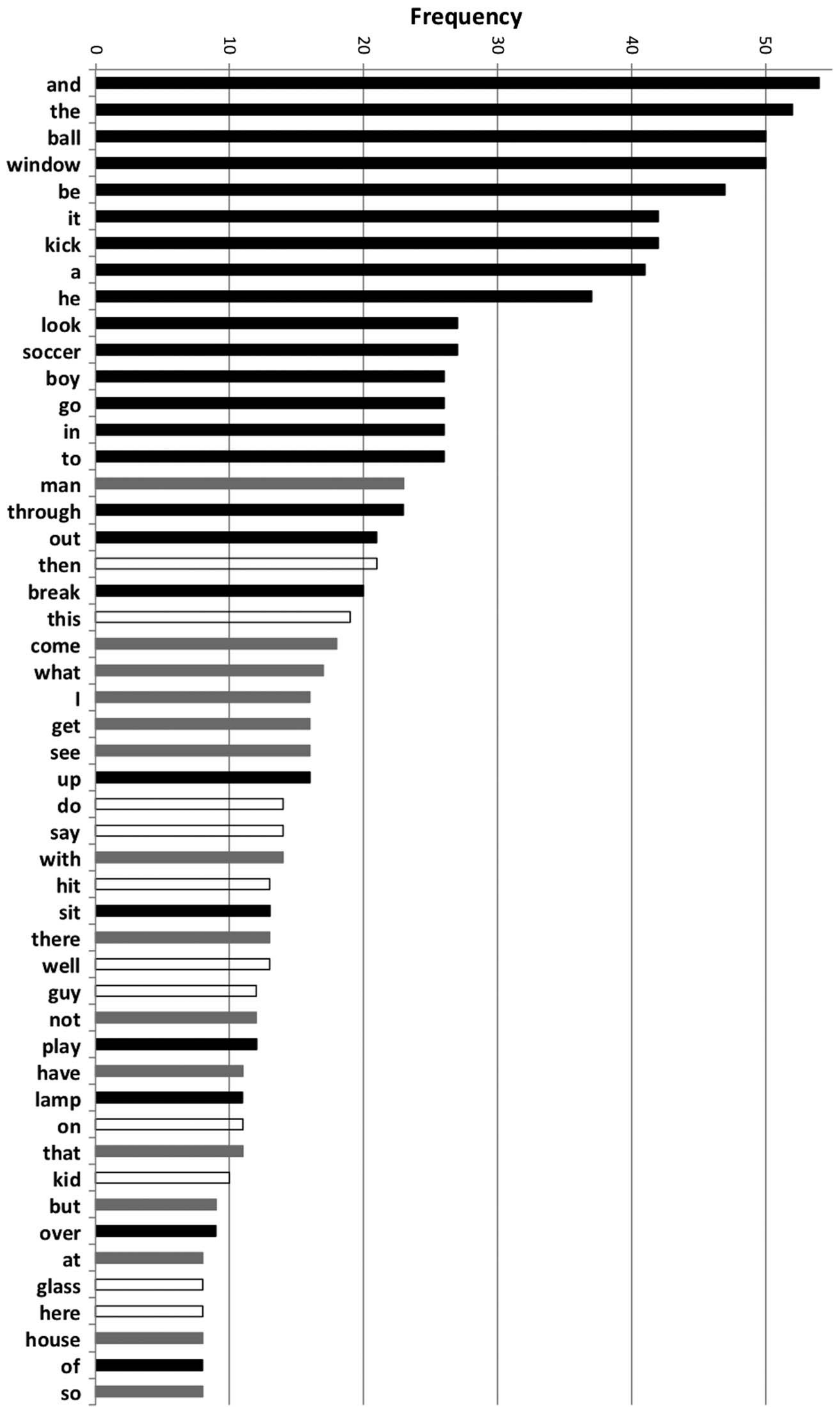
[a]Two individuals (one conduction, one Wernicke's) were missing data for age, education, and time since stroke. [b]One individual was missing data for age, education, and time since stroke. [c]Two individuals (one anomic, one Broca's) were missing data for WAB Aphasia Quotient.

Figure 1. Frequency counts of the top 50 lemmas produced by control speakers. Black bars indicate lemmas that reached the criterion threshold for the core lexicon. Gray bars indicate lemmas that were commonly produced by control speakers but did not reach threshold.



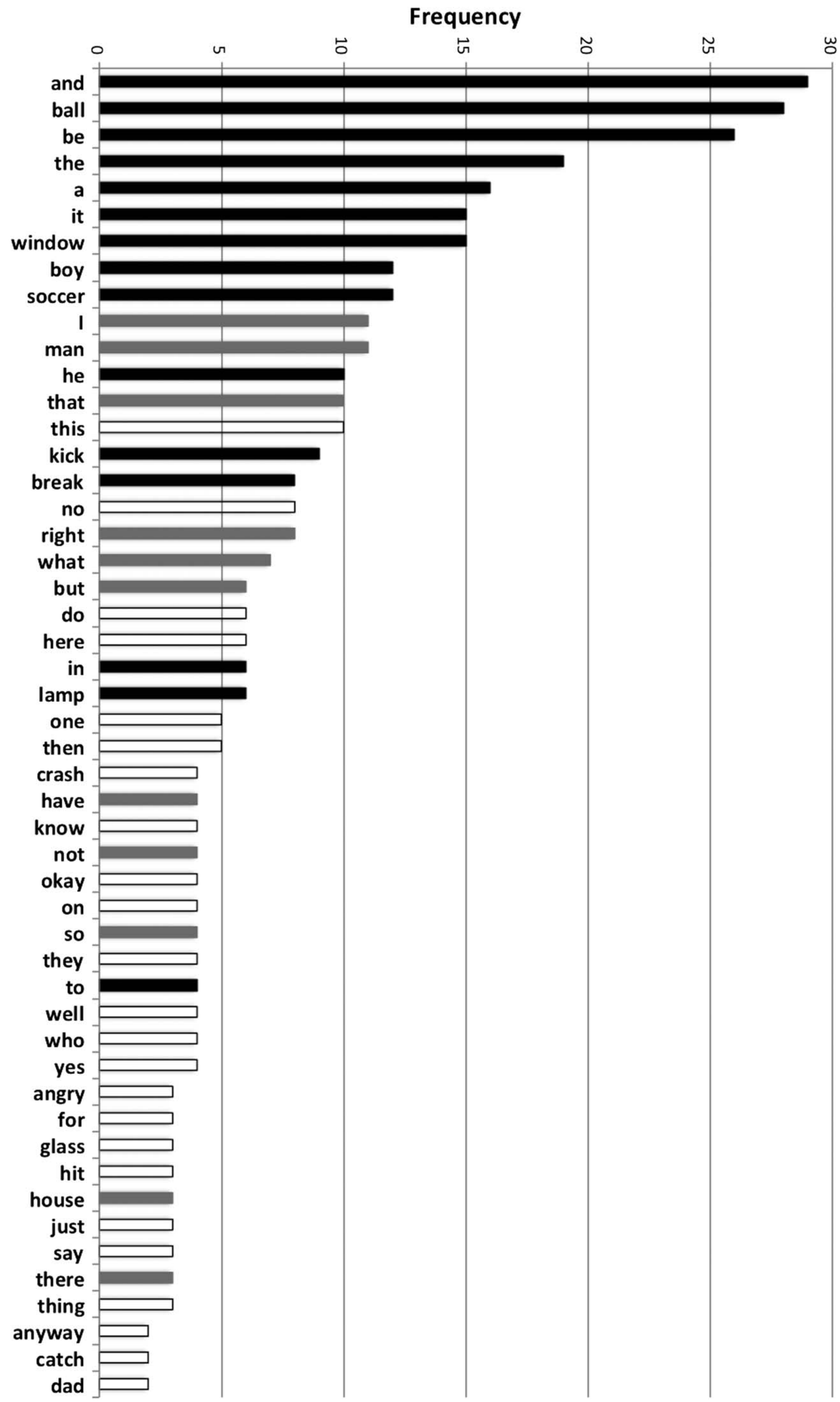
lemmas, whereas PWAs produced 13,323 words and 858 unique lemmas. The 50 lemmas most commonly produced by the speakers in the normative sample are shown in Figure 1 and included 24 structure (pronouns, articles, prepositions, auxiliaries, and conjunctions) and 26 content (nouns, verbs, adverbs, and adjectives) words. Of the 50 top lemmas, 24 (12 structure words, 12 content words) reached criterion for inclusion into the CoreLex (indicated by black columns in Figure 1), and 26 did not reach criterion (indicated by gray columns in Figure 1).

Figure 2. Frequency counts of the top 50 lemmas produced by individuals with anomic aphasia. Black bars indicate core-lexicon lemmas, gray bars indicate lemmas present in the top 50 produced by control speakers, and white bars indicate lemmas frequently produced by individuals with anomic aphasia and not by control speakers.



Omnibus median tests revealed a significant difference between controls and PWAs for CoreLex, $\chi^2(6, 413) = 240.252$, $p < .001$. Further testing revealed significant differences between all aphasia subtypes (including those categorized as NABW) and controls for CoreLex ($p < .001$ for all comparisons). Planned comparisons to examine differences between aphasia subtypes were performed following recalculation of the median (excluding controls)

Figure 3. Frequency counts of the top 50 lemmas produced by individuals with Broca's aphasia. Black bars indicate core-lexicon lemmas, gray bars indicate lemmas present in the top 50 produced by control speakers, and white bars indicate lemmas frequently produced by individuals with Broca's aphasia and not by control speakers.



to avoid median inflation. Significant differences were found for CoreLex scores between Broca's aphasia and all other subtypes studied: anomic, $\chi^2(2, 142) = 57.202$, $p < .001$; conduction, $\chi^2(2, 109) = 27.238$, $p < .001$; NABW, $\chi^2(2, 88) = 74.988$, $p < .001$; and Wernicke's, $\chi^2(2, 85) = 20.954$, $p < .001$. CoreLex scores also differed significantly when comparing persons judged NABW to aphasia subtypes of anomic, $\chi^2(2, 104) = 9.337$, $p = .002$; conduction, $\chi^2(2, 71) = 18.959$, $p < .001$; and Wernicke's, $\chi^2(2, 47) = 16.982$, $p < .001$.

Figure 4. Frequency counts of the top 50 lemmas produced by individuals with conduction aphasia. Black bars indicate core-lexicon lemmas, gray bars indicate lemmas present in the top 50 produced by control speakers, and white bars indicate lemmas frequently produced by individuals with conduction aphasia and not by control speakers.

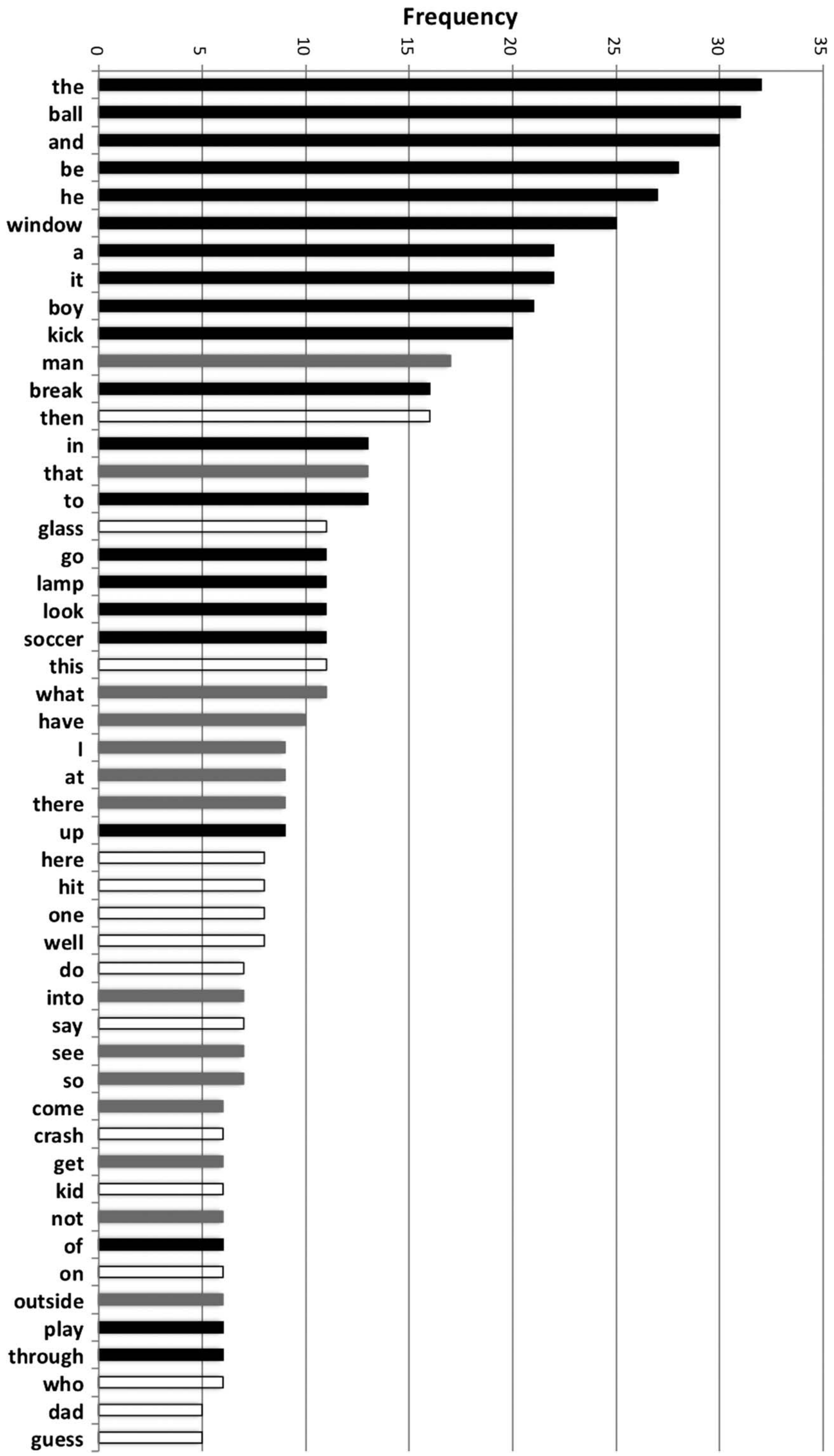
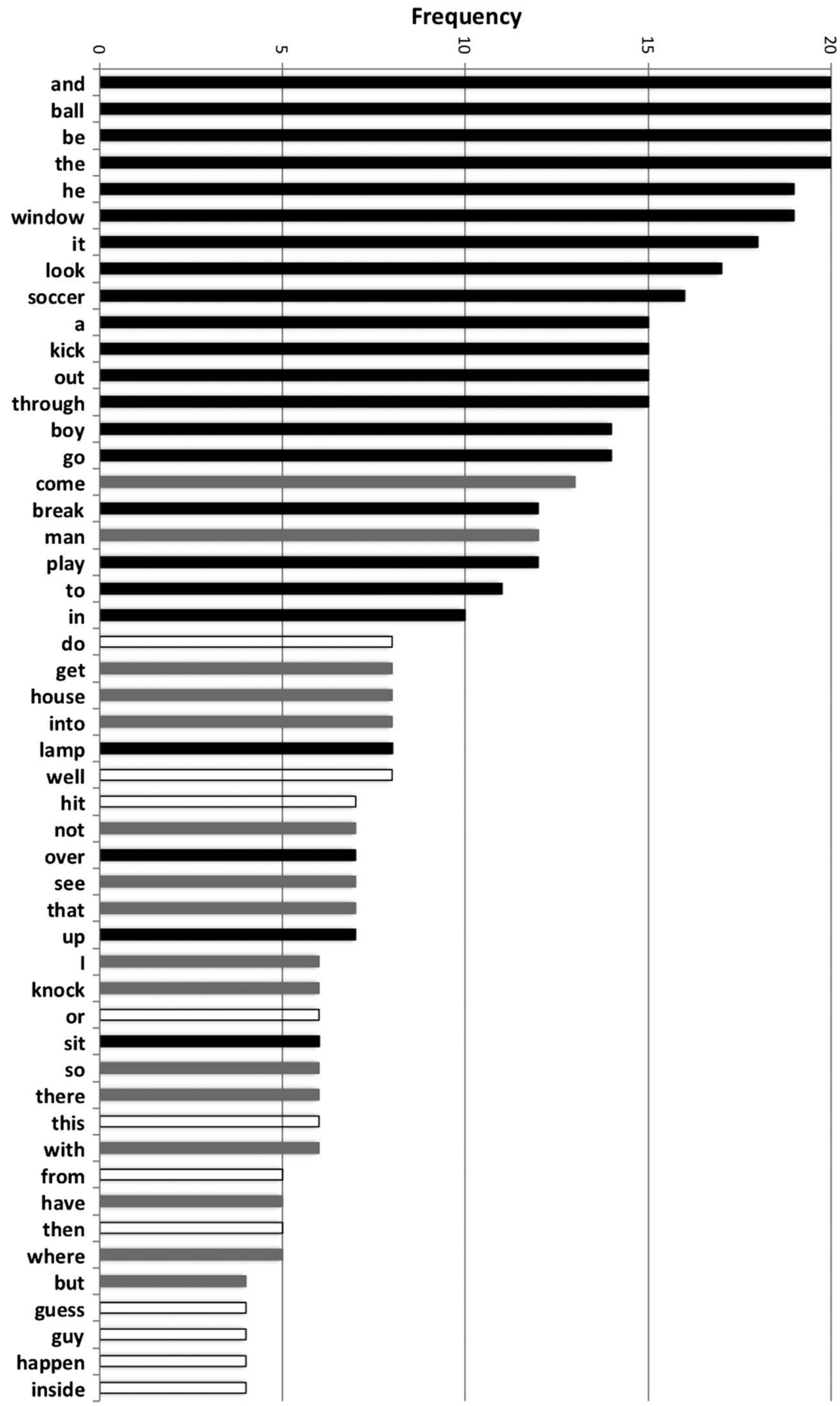
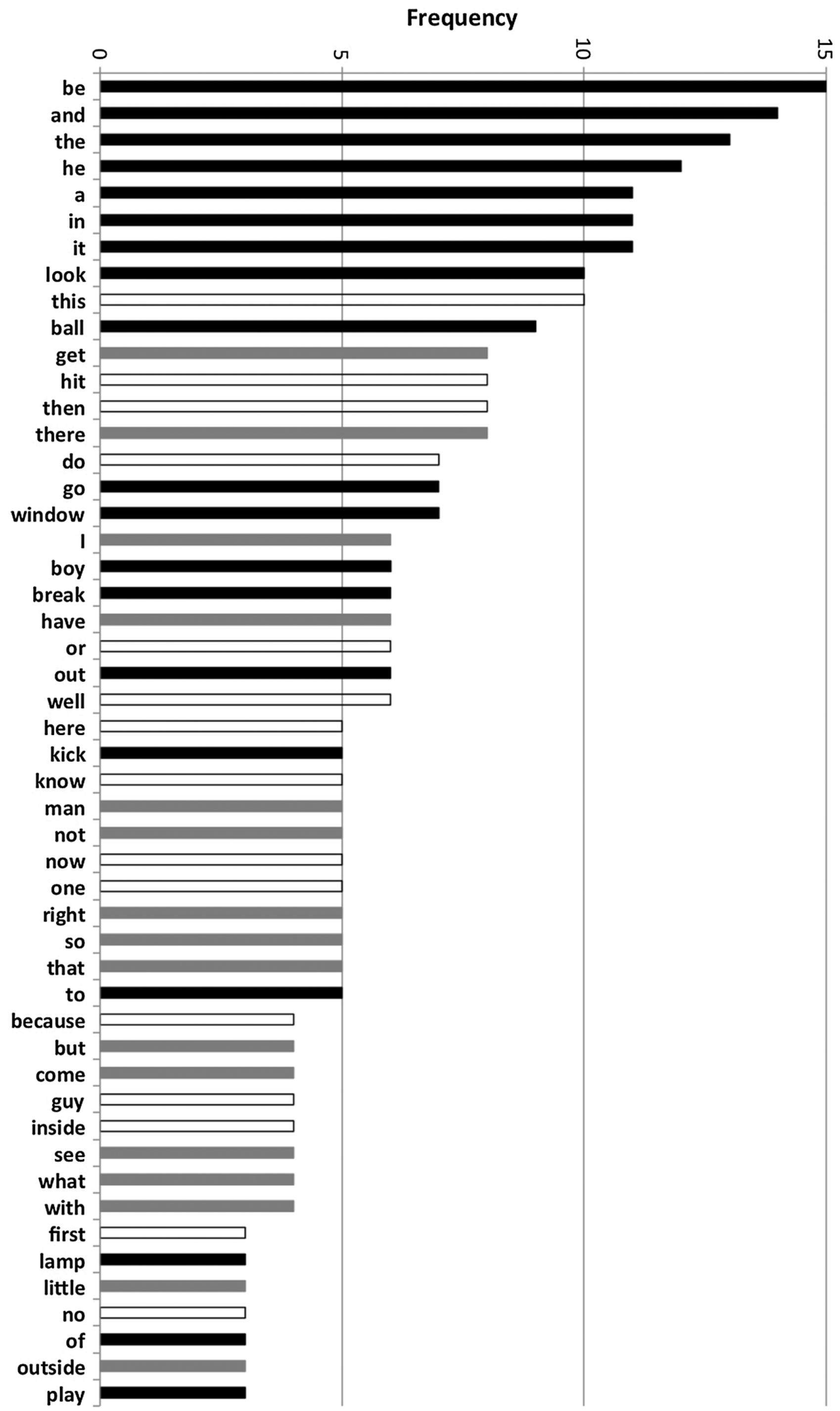


Figure 5. Frequency counts of the top 50 lemmas produced by individuals judged not aphasic by the Western Aphasia Battery–Revised. Black bars indicate core-lexicon lemmas, gray bars indicate lemmas present in the top 50 produced by control speakers, and white bars indicate lemmas frequently produced by individuals judged not aphasic by the Western Aphasia Battery–Revised and not by control speakers.



Differences were observed between all subtypes of aphasia and controls for the 50 most frequently used lemmas. Persons with anomic aphasia produced all 24 CoreLex lemmas, 15 of the top 50 lemmas, and 11 unique lemmas. Persons with Broca's aphasia produced 15 CoreLex lemmas, 11 of the top 50 lemmas, and 24 unique lemmas. Twenty-one CoreLex items, 14 top 50 lemmas, and 15 unique lemmas were produced by persons with conduction aphasia, and 23 CoreLex items, 16 top 50 lemmas, and 11 unique lemmas were present for individuals judged NABW. Last,

Figure 6. Frequency counts of the top 50 lemmas produced by individuals with Wernicke's aphasia. Black bars indicate core-lexicon lemmas, gray bars indicate lemmas present in the top 50 produced by control speakers, and white bars indicate lemmas frequently produced by individuals with Wernicke's aphasia and not by control speakers.



individuals with Wernicke's aphasia produced 19 CoreLex items, 15 top 50 lemmas, and 17 unique lemmas. Of the unique lemmas frequently produced by PWAs, five were shared across all subtypes: *then, do, well, hit,* and *this*.

MCs

Omnibus median tests revealed a significant difference between controls and PWAs for MC scores, $\chi^2(6, 413) = 1.509$, $p < .001$. Further testing revealed significant differences

Table 3. Percentages of each subgroup that attempted to produce a main concept.

Main concept	%						
	Anomic	Broca's	Conduction	NABW	Wernicke's	Normative sample (n = 92)	All control speakers (n = 166)
The boy was outside.	13.9	12.7	23.9	24	31.8	65.2	64.4
The boy was playing soccer.	92.4	66.7	93.5	100	68.2	100	99.4
The ball breaks the man's window.	92.4	60.3	93.5	100	72.7	93.4	97
The man is sitting in the house.	51.9	28.6	54.3	56	36.4	68.4	63.3
The man was startled.	21.5	6.3	15.2	24	22.7	45.6	46.4
The ball broke a lamp.	25.3	22.2	37	48	27.3	63	62
The man picked up the ball.	34.2	17.5	46	40	18.2	67.4	67.5
The man looked out of the window.	49.4	6.3	39.1	64	45.5	71.7	71

Note. NABW = Not aphasic by the Western Aphasia Battery–Revised. The main concepts in Table 3 are reprinted with permission of the publisher, Taylor & Francis Ltd. (http://www.tandfonline.com), from the following: Richardson, J. D., & Dalton, S. G. (2015). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology*, 1–29. doi:10.1080/02687038.2015.1057891

between persons with all aphasia subtypes and control speakers for MC measures ($p < .001$ for all comparisons except NABW for MC scores, where $p = .008$). Planned comparisons to examine differences between aphasia subtypes were performed following recalculation of the median (excluding controls) to avoid median inflation. MC scores showed a similar pattern of significant differences between Broca's aphasia and all other subtypes studied: anomic, $\chi^2(2, 142) = 41.806$, $p < .001$; conduction, $\chi^2(2, 109) = 28.709$, $p < .001$; NABW, $\chi^2(2, 88) = 45.7$, $p < .001$; and Wernicke's, $\chi^2(2, 85) = 10.171$, $p = .001$. The only other comparison that was significantly different was between NABW and Wernicke's, $\chi^2(2, 47) = 9.252$, $p = .002$.

As illustrated in Table 3, the group of persons with Broca's aphasia had the lowest percentages of attempted MCs compared to all other groups. Examination of the MC list concept by concept reveals that "The boy was outside" and "The man was startled" were produced with the least frequency by PWAs. The most frequently produced MCs across all groups were "The boy was playing soccer" and "The ball breaks the man's window." Other clear patterns of similarity or difference were difficult to discern; for example, persons with Broca's aphasia performed quite similarly to persons with anomic aphasia for some concepts but not for others. There are obvious differences between PWAs and control speakers, and it is important to remember that the percentages represent attempts, whether they were accurate and complete or not. Differences are likely more clear and pronounced when accuracy/inaccuracy and completeness/incompleteness are considered (Nicholas & Brookshire, 1995).

Relationship Between CoreLex and MCs

Using a two-tailed Spearman correlation, CoreLex and MC scores showed a significant positive relationship after collapsing across groups, $r_s(401) = .868$, $p < .001$. There was also a significant positive correlation between CoreLex and MC scores for individuals with aphasia, $r_s(235) = .783$, $p < .001$, and individuals with normal language, $r_s(166) = .630$, $p < .001$. In order to identify which groups were driving this finding, we conducted correlations for each subtype. For all subtypes, correlations were significant and positive—anomic: $r_s(79) = .710$, $p <$ 001; Broca's: $r_s(63) = .742$, $p < .001$; conduction: $r_s(46) = .463$, $p = .001$; NABW: $r_s(25) = .519$, $p = .008$; and Wernicke's: $r_s(22) = .707$, $p < .001$.

Discussion

We introduced a CoreLex checklist for the commonly used Broken Window picture-sequence discourse task (Menn et al., 1998). Speakers with aphasia had significantly lower CoreLex scores than control speakers. CoreLex and MC scores were also able to distinguish between individuals with fluent and nonfluent aphasia, though it is important to note that only persons with Broca's aphasia were included in the nonfluent group. In addition, a subgroup of individuals with brain injury who did not meet standardized-test cutoffs for aphasia (i.e., NABW) scored significantly differently on CoreLex and MC compared to control speakers and to persons with several aphasia subtypes. Last, CoreLex and MC scores were highly correlated for all groups and subtypes included here.

For the Broken Window narrative, CoreLex (but not MC) values differed significantly between individuals who were NABW versus persons with anomic and conduction aphasia, which may indicate that these subtypes are comparable in conveying the gist of the story but may differ in the typicality of the lexical items retrieved during narrative. In contrast, NABW individuals performed differently from individuals with Wernicke's aphasia on both CoreLex and MC measures. This may suggest that word-retrieval deficits in Wernicke's aphasia have a significant effect both on conveying the gist of the story and on the typicality of the lexical items. The CoreLex and MC measures we used did not allow us to measure local and global coherence, but it is possible that differences in CoreLex scores may be reflected in decrements in local and global coherence, as argued by Andreetta et al. (2012).

Consistent with the findings of the median tests, visual inspection of the 50 most frequently produced lemmas in each aphasia subtype showed distinct differences from the list representing control speakers. There appeared to be a continuum of typicality corresponding roughly to fluency and aphasia severity. It is interesting that individuals who were NABW performed similarly to persons with anomic aphasia, indicating that standardized tests may be missing mild language impairments that may nevertheless have a negative impact on functional communication. Although at first glance these groups (anomic, NABW) appeared to have productions similar to control speakers in regard to CoreLex and top 50 lemmas, differences became apparent upon examining the number of individuals in each group that produced a lemma. PWAs in this investigation used less typical language, and the language used was less consistent from one individual to the next compared to healthy control speakers. This is congruent with the notion that lexical-retrieval difficulties may contribute to discourse impairments. Lemmas produced frequently by individuals with aphasia but not by control speakers included more general versus specific nouns (e.g., *guy*) and "soft" verbs (e.g., *do, know*), similar to findings reported by MacWhinney et al. (2010) during analysis of the Cinderella narrative. However, these CoreLex results should be interpreted with caution—no attempts were made in this analysis to match for age, gender, or education across groups. It is possible that these person variables could lead to differences in lemma selection (e.g., Kavé, Samuel-Enoch, & Adiv, 2009; Singh, 2001) and could thus be a potential contributor to these group differences. A strength of this study is that we used a picture-description task that is highly constrained, and therefore lexical selection is also likely constrained across person variables (E. Armstrong, 2000; Capilouto et al., 2005; Newman, Groom, Handelman, & Pennebaker, 2008; Ulatowska et al., 2003). This more constrained discourse task reduces much of the variability in responses and historically has reduced group (age, gender) differences found on other discourse measures, such as lexical diversity (e.g., Cooper, 1990; Fergadiotis, Wright, & Capilouto, 2011; Newman et al., 2008).

It was not surprising that, given knowledge of Broca's aphasia profiles and CoreLex results, persons with Broca's aphasia attempted to produce fewer MCs compared to all other groups. PWAs in general seemed to focus their efforts on attempting the following MCs: "The boy was playing soccer," "The ball breaks the man's window," "The man is sitting in the house," and "The man looked out of the window." These concepts seem to correspond to the most salient and most clearly illustrated events depicted in the four-picture sequence.

Though not examined in this study, the different MC codes could be directly related to the types of lexical-retrieval difficulties experienced. Individuals who are able to access at least some of the typical vocabulary for a narrative would likely be able to produce more AC or AI statements, whereas those who are able to access nontypical vocabulary may produce more IC or II statements. Individuals with significant word-retrieval difficulties, or those who do not attempt productions that they believe will be in error, would be expected to have a high number of concepts coded as AB. The reduction of attempted concepts overall, coupled with concepts of varying inaccuracy and incompleteness, would result in a less informative narrative that would be difficult for a listener to follow or understand. The significant positive correlations between CoreLex and MC scores reported for all groups support this interpretation. These strong correlations indicate that CoreLex may be capturing information related to the structure of the narrative as well as its content, likely due to the inclusion of function words in our list. In particular, the strong correlations between CoreLex and MCs for individuals with anomic, Broca's, and Wernicke's aphasia suggest that CoreLex might be particularly appropriate for use in predicting concept-level discourse abilities in these groups. On the other hand, correlations between the two measures were weaker for individuals with conduction aphasia and those judged NABW, supporting the need for multilevel approaches to narrative assessment (E. Armstrong, 2000; Marini et al., 2011; Sherratt, 2007; Wright & Capilouto, 2012).

Limitations and Future Directions

CoreLex examines an individual's ability to produce the most typical or common words used to tell a story. Individuals thus do not receive credit when a correct alternative (synonym) is used. A CoreLex score does not necessarily provide precise information regarding the severity of discourse deficits; instead, it provides information about the typicality of the language an individual is able to use. It is possible that greater lexical diversity could negatively affect an individual's CoreLex score if the individual is not using the most typical language to produce the story. Individuals who use greater numbers of synonyms could also be experiencing more word-retrieval difficulties, which might be accompanied by pauses, fillers, and revisions, leading to a loss of coherence (e.g., Andreetta et al., 2012; Verhaegen & Poncelet, 2013). If the latter were the case, then the predictive strength of a measure such as CoreLex might actually be improved by counting only the most typical productions.

As a picture-sequence description task (four pictures), the Broken Window narrative is significantly shorter and more constrained than other discourse tasks and may have sufficient power to differentiate between individuals with and without aphasia, and perhaps between individuals with different subtypes of aphasia. Using CoreLex and/or MC checklists, this task may be able to serve as a stand-alone discourse assessment option to assist with clinical decision making in settings where clinicians have high caseloads and little time to collect and analyze discourse samples, and/or in settings where there may be extreme time constraints and individuals who fatigue easily (e.g., acute inpatient settings).

Specific participant characteristics for CoreLex and MC analysis should be investigated to determine whether subgroups of individuals with normal language perform differently from one another. Age-related differences in discourse production have been reported in individuals with normal language in relating the gist of a story (Capilouto et al., 2005; Richardson & Dalton, 2015; Wright, Capilouto, Wagovich, Cranfill, & Davis, 2005); given that CoreLex is correlated well with MC, one such measure of gist, it would be important to assess age-related differences in this measure as well. Gender, race/ethnicity, and education level may also be sources of variability in discourse production and should be investigated further. If significant differences exist on these lines, then corresponding stratified checklists and normative information should be developed to ensure that PWAs' performance is compared to the most appropriate control groups' performance. This should be followed by research to confirm and characterize the validity, reliability, sensitivity, and specificity of such measures. MC analysis has proven to have high inter- and intrarater reliability, and it is expected that CoreLex scoring will also, given that raters need only record the presence or absence of the lexical items. We used transcripts available from the AphasiaBank to establish and score CoreLex and MCs. An important next step is to determine the reliability of CoreLex and MC scoring without transcriptions, both online (as individuals produce the story) and using audio recordings, in order to ensure that this measure will be clinically practicable.

Before moving forward with development, it would be helpful to know whether these measures provide information of value beyond existing standardized, norm-referenced measures of overall language performance (e.g., WAB-R AQ). If the differences between aphasia subtypes in discourse scores in this investigation were driven solely by aphasia severity, we might expect that the same group differences found in discourse measures would also be found in AQ scores, and AQ scores would correlate very highly with discourse measures. Though not the aim of this investigation, we feel it is important to explore these questions and to share our findings with the research community to help guide future research. First, we conducted median tests to examine differences in AQ scores by subtype. Though significant differences in discourse measures were observed between individuals with Broca's and Wernicke's aphasia, for example, they were not observed for the AQ scores for these subtypes. Conversely, significant differences in discourse measures were not observed between individuals with anomic, conduction, and Wernicke's aphasia but were observed for the AQ scores for these subtypes. Second, we conducted a series of Spearman correlations (two-tailed, $p = .05$) to examine the relationship between AQ and each discourse measure. Collapsed across groups, AQ was highly correlated with CoreLex ($r_s = .70$) and MC ($r_s = .66$) scores. A closer examination by subtype revealed that AQ was significantly correlated with CoreLex and MC scores for individuals with anomic, Broca's and Wernicke's aphasia, but only with MC scores for individuals with conduction aphasia. (No significant correlations were found for individuals NABW, which is not surprising given their ceiling performance on the WAB-R.) Furthermore, the strength of the correlations varied by subtype, from $r_s = .33$ to $r_s = .77$. Although these relationships need to be directly and thoroughly investigated with more homogeneous groups, this exploratory post hoc analysis supports continued development of these discourse measures.

Difficulty with access and/or retrieval of the core lexicon associated with a story may place a greater strain on working memory as the amount of time between narrative conception and narrative production increases. Difficulty retrieving a word for production may cause interference in working-memory storage of later words that have been retrieved. For example, if a PWA attempts to say "I need to go to the store to get milk and eggs," difficulty producing "store" may interfere with the storage of "eggs" or "milk." This working-memory interference may reduce local and global coherence by increasing the number of repetitions, revisions, and fillers required to produce a narrative (Andreetta et al., 2012). Future studies could use CoreLex or similar checklists to examine the relationship between lexical retrieval and working-memory capacity, which is currently a topic of interest (Wright & Fergadiotis, 2012).

Conclusions

Using information about discourse abilities to plan and evaluate treatment outcomes is an important move in the profession to improve the ecological validity of our assessments. Traditional outcome measures, such as naming, do not appear to accurately reflect response to treatment (Mayer & Murray, 2003) or predict listener ratings of functional improvement (Ross & Wertz, 1999) compared to discourse measures. Use of the CoreLex and MC analysis, alone or in addition to other analyses, may help to better predict response to treatment and functional improvements than traditional measures (e.g., naming ability) or standardized tests.

Acknowledgments

We thank the Chapman Foundation for financial support and the AphasiaBank developers and contributors.

References

Andreetta, S., Cantagallo, A., & Marini, A. (2012). Narrative discourse in anomic aphasia. *Neuropsychologia, 50,* 1787–1793.

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

Armstrong, L., Brady, M., Mackenzie, C., & Norrie, J. (2007). Transcription-less analysis of aphasic discourse: A clinician's dream or a possibility? *Aphasiology, 21,* 355–374.

Bates, E., & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes, 12,* 507–584.

Boyle, M. (2014). Test–retest stability of word retrieval in aphasic discourse. *Journal of Speech, Language, and Hearing Research, 57,* 966–978.

Capilouto, G., Wright, H. H., & 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.

Caramazza, A., & Berndt, R. S. (1985). A multicomponent deficit view of agrammatic Broca's aphasia. In M.-L. Kean (Ed.), *Agrammatism* (pp. 27–63). New York, NY: Academic Press.

Carey, J. W. (1988). *Communication as culture: Essays on media and society*. New York, NY: Psychology Press.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review, 13,* 181–197.

Cooper, P. V. (1990). Discourse production and normal aging: Performance on oral picture description tasks. *Journals of Gerontology, 45,* P210–P214.

Dimbleby, R., & Burton, G. (1998). *More than words: An introduction to communication* (3rd ed.). New York, NY: Routledge.

Duncan, E. A. S., & Murray, J. (2012). The barriers and facilitators to routine outcome measurement by allied health professionals in practice: A systematic review. *BMC Health Services Research, 12*(96), 1–9.

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

Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across discourse types. *Aphasiology, 25,* 1261–1278.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1187–1199.

Fromm, D., Forbes, M., Holland, A., & MacWhinney, B. (2013, May–June). *PWAs and PBJs: Language for describing a simple procedure*. Poster presented at the Clinical Aphasiology Conference, Tucson, AZ.

Goodglass, H., & Menn, L. (1985). Is agrammatism a unitary phenomenon. In M.-L. Kean (Ed.), *Agrammatism* (pp. 1–26). New York, NY: Academic Press.

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

Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject–verb agreement. *Language and Cognitive Processes, 21,* 181–204.

Kavé, G., Samuel-Enoch, K., & Adiv, S. (2009). The association between age and the frequency of nouns selected for production. *Psychology and Aging, 24,* 17–27.

Kertesz, A. (2006). *Western Aphasia Battery–Revised*. New York, NY: Pearson.

Kong, A. P.-H. (2009). The use of main concept analysis to measure discourse production in Cantonese-speaking persons with aphasia: A preliminary report. *Journal of Communication Disorders, 42,* 442–464.

Kong, A. P.-H. (2011). Family members' report on speech-language pathology and community services for persons with aphasia in Hong Kong. *Disability and Rehabilitation, 33,* 2633–2645.

Larfeuil, C., & Le Dorze, G. (1997). An analysis of the word-finding difficulties and of the content of the discourse of recent and chronic aphasic speakers. *Aphasiology, 11,* 783–811.

MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated analysis of the Cinderella story. *Aphasiology, 24,* 856–868.

Marini, A., Andreetta, S., del Tin, S., & Carlomagno, S. (2011). A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology, 25,* 1372–1392.

Mayer, J., & Murray, L. (2003). Functional measures of naming in aphasia: Word retrieval in confrontation naming versus connected speech. *Aphasiology, 17,* 481–497.

Menn, L., Reilly, K. F., Hayashi, M., Kamio, A., Fujita, I., & Sasanuma, S. (1998). The interaction of preserved pragmatics and impaired syntax in Japanese and English aphasic speech. *Brain and Language, 61,* 183–225.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes, 45,* 211–236.

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

Nicholas, L. E., & Brookshire, R. H. (1993b). A system for scoring main concepts in the discourse of non-brain-damaged and aphasic speakers. *Clinical Aphasiology Conference, 21,* 87–99.

Nicholas, L. E., & 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,* 145–156.

Olness, G. S., Ulatowska, H. K., Wertz, R. T., Thompson, J. L., & Auther, L. L. (2002). Discourse elicitation with pictorial stimuli in African Americans and Caucasians with and without aphasia. *Aphasiology, 16,* 623–633.

Owens, R. E., Jr. (2008). *Language development: An introduction* (7th ed.). New York, NY: Pearson.

Richardson, J. D., & Dalton, S. G. (2015). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology,* 1–29. doi:10.1080/02687038.2015.1057891

Ross, K. B., & Wertz, R. T. (1999). Comparison of impairment and disability measures for assessing severity of, and improvement in, aphasia. *Aphasiology, 13,* 113–124.

Sherratt, S. (2007). Multi-level discourse analysis: A feasible approach. *Aphasiology, 21,* 375–393.

Simmons-Mackie, N., Threats, T. T., & Kagan, A. (2005). Outcome assessment in aphasia: A survey. *Journal of Communication Disorders, 38,* 1–27.

Singh, S. (2001). A pilot study on gender differences in conversational speech on lexical richness measures. *Literary & Linguistic Computing, 16,* 251–264.

Ulatowska, H. K., North, A. J., & Macaluso-Haynes, S. (1981). Production of narrative and procedural discourse in aphasia. *Brain and Language, 13,* 345–371.

Ulatowska, H. K., Olness, G. S., Wertz, R. T., Samson, A. M., Keebler, M. W., & Goins, K. E. (2003). Relationship between discourse and Western Aphasia Battery performance in African Americans with aphasia. *Aphasiology, 17,* 511–521.

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

Verna, A., Davidson, B., & 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.

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

Wright, H. H., Capilouto, G. J., Wagovich, S. A., Cranfill, T. B., & Davis, J. E. (2005). Development and reliability of a quantitative measure of adults' narratives. *Aphasiology, 19,* 263–273.

Wright, H. H., & Fergadiotis, G. (2012). Conceptualising and measuring working memory and its relationship to aphasia. *Aphasiology, 26,* 258–278.

Yorkston, K. M., & Beukelman, D. R. (1980). An analysis of connected speech samples of aphasic and normal speakers. *Journal of Speech and Hearing Disorders, 45,* 27–36.

Appendix

Sample Transcripts of a Control Speaker and a Person With Aphasia Retrieved From the AphasiaBank Database

Broken Window Transcript of a 75-Year-Old Male With Broca's Aphasia
@G: Window
*PAR: the ball &=points:picture_1 . [+ gram]
*PAR: &=points:picture_2 well &uh &uh &=touches:forehead &um &uh &um +…
*PAR: god_damn_it . [+ exc]
*PAR: &uh (…) &=points:picture_2 oh .
*PAR: &=points:picture_3 (…) &=points:picture_1 &uh (..) the ball &uh &=points:picture_3 here . [+ gram]
*PAR: and &uh &=taps:picture_4 +…
*PAR: yeah . [+ exc]
*PAR: yes . [+ exc]
*PAR: yes [/] yes &=laughs . [+ exc]

Broken Window Transcript of a 75-Year-Old Male With Normal Language
@G: Window
*PAR: a young boy is practicing playing soccer .
%mor: det|a adj|young n|boy aux|be&3S part|practice-PROG part|play-PROG n|soccer
*PAR: kicking the ball up and keeping it i(n) the air .
%mor: part|kick-PROG det|the n|ball adv:loc|up coord|and n:gerund|keep-PROG pro|it prep|in det|the n|air .
*PAR: he miskicks .
%mor: pro:sub|he mis#v|kick-3S .
*PAR: and [/] &scr and it fall [//] &br goes and breaks the window of his house .
%mor: coord|and pro|it v|go-3S coord|and v|break-3S det|the n|window prep|of pro:poss:det|his n|house .
*PAR: of the living+room actually .
%mor: prep|of det|the n|+n|living+n|room adv|actual&dadj-LY .
*PAR: and bounces into the living room knocking a lamp over where his father is sitting .
%mor: coord|and v|bounce-3S prep|into det|the part|live-PROG n|room part|knock-PROG det|a n|lamp prep|over adv:wh| where pro:poss:det|his n|father cop|be&3S part|sit-PRESP .
*PAR: the father picks up the &b soccer ball .
%mor: det|the n|father v|pick-3S prep|up det|the n|soccer n|ball .
*PAR: looks out the window .
%mor: n|look-PL prep|out det|the n|window .
*PAR: and calls for the little boy &t to come and explain .
%mor: coord|and n|call-PL prep|for det|the adj|little n|boy inf|to v|come coord|and ex#n|plain .
@End