JSLHR

Research Article

Automated Proposition Density Analysis for Discourse in Aphasia

Davida Fromm,[a] Joel Greenhouse,[a] Kaiyue Hou,[a] G. Austin Russell,[a] Xizhen Cai,[a] Margaret Forbes,[a] Audrey Holland,[b] and Brian MacWhinney[a]

Purpose: This study evaluates how proposition density can differentiate between persons with aphasia (PWA) and individuals in a control group, as well as among subtypes of aphasia, on the basis of procedural discourse and personal narratives collected from large samples of participants.
Method: Participants were 195 PWA and 168 individuals in a control group from the AphasiaBank database. PWA represented 6 aphasia types on the basis of the Western Aphasia Battery–Revised (Kertesz, 2006). Narrative samples were stroke stories for PWA and illness or injury stories for individuals in the control group. Procedural samples were from the peanut-butter-and-jelly-sandwich task. Language samples were transcribed using Codes for the Human Analysis of Transcripts (MacWhinney, 2000) and analyzed using Computerized Language Analysis (MacWhinney, 2000), which automatically computes proposition density (PD) using rules developed for automatic PD measurement by the Computerized Propositional Idea Density Rater program (Brown, Snodgrass, & Covington, 2007; Covington, 2007).
Results: Participants in the control group scored significantly higher than PWA on both tasks. PD scores were significantly different among the aphasia types for both tasks. Pairwise comparisons for both discourse tasks revealed that PD scores for the Broca's group were significantly lower than those for all groups except Transcortical Motor. No significant quadratic or linear association between PD and severity was found.
Conclusion: Proposition density is differentially sensitive to aphasia type and most clearly differentiates individuals with Broca's aphasia from the other groups.

Measures of the density of ideas or propositions in discourse are thought to provide an indication of communicative adequacy despite possible disruption of language at the sentence level (Ulatowska, Freedman-Stern, Doyel, Macaluso-Haynes, & North 1983). Although language may be impaired in both complexity and quantity, essential narrative propositions may still be preserved. Ferguson, Spencer, Craig, and Colyvas (2014) trace the concept of the proposition back to Hughlings-Jackson's (1879) early writings on aphasia in 1879. Researchers have typically conducted propositional analyses of aphasic discourse using a variety of nonautomated approaches (Christiansen, 1995; Gleason et al., 1980; Ulatowska et al., 1983; Ulatowska, North, & Macaluso-Haynes, 1981). In general, results have shown reduced quantity and complexity of language, with relative preservation of discourse structure in the output of people with aphasia (PWA). These analyses, as well as related discourse analyses using measures such as content units (Yorkston & Beukelman, 1980) and correct information units (Nicholas & Brookshire, 1993), require long hours of coding following extensive training, as well as, in some cases, restricted content (using specific elicitation stimuli, for example) for which relevant and informative words can be predetermined (Ulatowska et al., 1983). For both clinical and research purposes, it is desirable that measures of propositional density be automated.

Automated analyses offer several advantages. First, they require less time than analyses by hand. Second, they are potentially more accurate and definitely more consistent in terms of the criteria they apply. Third, they can easily be replicated. Fourth, they facilitate the automated exploration of the effects of additional variables. For example, propositional analyses have been done using written samples and oral samples, different types of oral-language sampling (e.g., interviews, story recall, picture description, filmstrip narration, personal narratives), different sample lengths, and different measurement techniques (e.g., total number of essential propositions, proposition complexity index, proposition

[a]Carnegie Mellon University, Pittsburgh, PA
[b]University of Arizona, Tucson

Correspondence to Davida Fromm: fromm@andrew.cmu.edu

Editor: Rhea Paul
Associate Editor: Margaret Blake

Received November 20, 2015
Revision received January 29, 2016
Accepted March 7, 2016
DOI: 10.1044/2016_JSLHR-L-15-0401

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

density). Some studies of proposition density report the number of expressed propositions divided by the number of words (Brown, Snodgrass, & Covington, 2007; Bryant et al., 2013; Covington, 2007; Ferguson et al., 2013; Roark, Mitchell, Hosom, Hollingshead, & Kaye, 2011), whereas others report the number of ideas per 10 words (Chand, Baynes, Bonnici, & Farias, 2012; Cunha, Sousa, Mansur, & Aluisio, 2015; Farias et al., 2012; Kemper, Greiner, Marquis, Prenovost, & Mitzner, 2001) and in some cases use just the last 10 sentences of the samples (Engelman, Agree, Meoni, & Klag, 2010; Riley, Snowdon, Desrosiers, & Markesbery, 2005; Snowdon et al., 1996). Comparisons of results across studies require careful consideration of all these possible variables, which can make it challenging to synthesize, replicate, and build on the findings.

Covington and colleagues (Brown et al., 2007; Covington, 2007) developed Computerized Propositional Idea Density Rater (CPIDR), a software program for counting propositions using part-of-speech tags. The program was based on the seminal theoretical work of Kintsch (1974), as operationalized in a manual by Turner and Greene (1977), which provides detailed rules for counting propositions according to Kintsch's theory. Covington's research demonstrated that automated coding is more reliable than coding performed by human raters (Brown, Snodgrass, Kemper, Herman, & Covington, 2008). Bryant et al. (2013) and Ferguson et al. (2013) used CPIDR (version 3.2; Brown et al., 2007) to calculate proposition density (PD; the ratio of the number of propositions to the total number of words) in aphasic discourse. Systematic Analysis of Language Transcripts (SALT, version 8; Miller, 2003) was used to calculate type–token ratio, number of different words, mean length of utterance in words (MLU), and number of utterances to establish concurrent validity for the PD measure. The researchers analyzed language samples from interviews conducted by a speech-language pathologist who asked 50 PWA about their stroke, the impact of aphasia, and their goals for rehabilitation. The samples ranged from 103 to 6,484 words for the participants with aphasia and 1,780 to 6,533 words for those without. Results showed significantly decreased PD in the aphasia group compared with the 49 nonaphasic control participants, who were family members of the PWA group. Concurrent validity was demonstrated with significant differences between the two groups for number of different words, MLU, and type–token ratio. Severity of aphasia (as measured by the Aphasia Quotient of the Western Aphasia Battery–Revised [WAB-R]; Kertesz, 2006) was significantly and positively correlated with PD as well as number of different words and MLU. The authors recommended further research with larger numbers of participants, language samples collected from a wider range of communicative contexts, and consideration of differences in types of aphasia.

PD has been well studied in the area of aging and dementia, with several studies reporting results from the autobiographical writing samples in the Nun Study and demonstrating that PD is a sensitive predictor of Alzheimer's disease and language decline associated with aging (Butler & Snowdon, 1996; Kemper et al., 2001; Mortimer, 2012; Riley et al., 2005; Snowdon et al., 1996). This literature uses the term *idea density* to mean the exact same thing as proposition density, also sometimes called P-density. For this article, we will use the term *proposition density*. What is consistently reported, regardless of methodological differences across studies, is a significant relationship between PD in early adult life and cognition in late life, with low PD scores being associated with low cognitive test scores and/or increased risk of dementia (Engelman et al., 2010; Iacono et al., 2009; Kemper et al., 2001; Riley et al., 2005; Snowdon et al., 1996). Studies have also shown relationships between low PD scores in early adult life and greater severity of Alzheimer's disease pathology in the neocortex (Riley et al., 2005; Snowdon, Greiner, & Markesbery, 2000). Two other studies have confirmed that proposition density is not significantly different in groups with mild cognitive impairment and healthy older adults (Chand et al., 2012; Roark et al., 2011).

Working with a different population, Coelho, Grela, Corso, Gamble, and Feinn (2005) used propositional analysis to better understand the microlinguistic impairments associated with traumatic brain injury because it allowed for the examination of semantic complexity apart from sentence structure and grammaticality. On two story discourse tasks, participants with traumatic brain injury generated fewer propositions per T-unit (independent clause plus any associated subordinate clauses) than did a control group, despite showing no significant difference from the control group on syntactic complexity (subordinate clauses per T-unit) and cohesive adequacy (proportion of total cohesive ties that were complete) in storytelling and picture descriptions. The authors stress the importance of using multilevel analyses within the broader domains of micro- and macrolinguistic analysis.

Across these various populations, PD has been shown to be a sensitive and predictive measure of microlinguistic abilities in written and oral output. The extensive work begun by Kintsch and colleagues that has continued in the field for almost 40 years supports and explains the importance of propositions as the microstructure of the semantic system (Kintsch & van Dijk, 1978). Automating the measurement of these meaningful propositional constructs based on part-of-speech tagging allows for the added advantages of efficiency and replicability. In an examination of 87 studies that performed quantitative analyses of written or spoken discourse, Foster, Tonkyn, and Wigglesworth (2000) addressed the importance of defining units of analysis and the greater difficulty of doing so for oral versus written language samples. Their discussion identified the shortcomings of many units of analysis used frequently in the literature.

The current study was undertaken in an effort to further understand the PD measure in aphasia, where participants may have a variety of patterns of linguistic impairment in addition to different degrees of impairment. PD is intended to be a measure of connections among words, or “idea units” (Turner & Greene, 1977, p. 3), and thus captures a degree of complexity in language. It reflects a person’s ability to express relations among words rather than solely using them referentially. Using a larger sample size and multiple discourse tasks, this study was also intended to corroborate and extend the findings from Bryant et al. (2013) and Ferguson et al. (2013) about PD in aphasia using automated measures. The following questions were addressed:

- Do PWA differ on PD from individuals in a control group who do not have aphasia for both procedural discourse and personal narratives?
- How does PD differ across types of aphasia?

Method

Participants

The analysis included all PWA (n = 195) and all participants in the nonaphasic control group (n = 168) from the AphasiaBank database who were native speakers of English and who had responded to both the procedural-discourse prompt and the personal-narrative prompt (MacWhinney, Fromm, Forbes, & Holland, 2011). The WAB-R Aphasia Quotient (AQ) subtests were administered to the PWA group. Coexisting apraxia and dysarthria were not exclusionary criteria, but participants with dementia or with comorbidities associated with serious cognitive consequences were excluded. Participants in the control group were tested with the Mini-Mental State Exam (Folstein, Folstein, & McHugh, 1975) and the Geriatric Depression Scale (Brink et al., 1982) to rule out cognitive impairment and depression. The AphasiaBank project received the required approval from the institutional review board. All participants (or their representatives) signed consent forms for the testing and data collection and gave approval for the data to be available for research and teaching. Table 1 provides a summary of demographic characteristics of the sample.

On the basis of WAB-R AQ subtest scores, the PWA group included these six subgroups: Anomic (n = 77), Broca’s (n = 35), Conduction (n = 37), Wernicke’s (n = 13), Transcortical Motor (n = 8), and AQ > 93.8 (n = 25).

Table 1. Demographic characteristics of participants.

Characteristic	PWA (n = 195)	Control group (n = 168)
Mean (*SD*) age (years)	62.2 (11.8)	64.9 (17.1)
Mean (*SD*) education (years)	15.4 (2.7)	15.3 (2.4)
% men	57	48
% women	43	52
Mean (*SD*) time postonset (years)	5.4 (4.9)	
Mean (*SD*) WAB-R AQ score	76.0 (16.4)	

Note. PWA = people with aphasia; WAB-R AQ = Western Aphasia Battery–Revised Aphasia Quotient.

This latter group consisted of PWA who scored above the WAB-R AQ cutoff but still considered themselves (and were considered by their clinicians) to have aphasia.

Language-Sampling Procedure

All sessions were conducted by a licensed speech-language pathologist, with a small number of exceptions in which a trained and supervised graduate student in speech-language pathology ran the session. Sessions were video-recorded and typically took place in a room at a clinic or an aphasia center, with the investigator and the participant seated together at a table. From the larger AphasiaBank discourse protocol, two discourse genres were selected for this analysis: personal stroke narrative and procedural discourse. In place of the stroke narrative, the control group was asked to tell about an injury or illness. The stroke or illness narrative is comparable in content to the samples used by Bryant et al. (2013) and Ferguson et al. (2013). The procedural-discourse task requires participants to describe how to make a peanut-butter-and-jelly sandwich. This is a short, simple task that can be transcribed quickly by busy clinicians to provide potentially useful information concerning connected speech.

The prompts given by the investigator to PWA for the personal narrative of stroke and coping were the following:

1. “Do you remember when you had your stroke? Please tell me about it.”
2. “Tell me about your recovery. What kinds of things have you done to try to get better since your stroke?”

The investigator did minimal verbal prompting but was an attentive active listener, providing nonverbal encouragers and plenty of time for the participant to give as complete a response as possible. The second question was asked only after the participant had clearly finished answering the first question. For participants in the control group, the questions were the following:

1. “In this research project, I ask people who’ve had strokes to tell me what they remember about when they had their stroke. Since you haven’t had a stroke, I wonder if you could tell me what you remember about any illness or injury you’ve had.”
2. “Tell me about your recovery from that illness (or injury). What kinds of things did you do to get better?”

The procedural-discourse prompt for both groups was “Tell me how you would make a peanut-butter-and-jelly sandwich.”

In addition to the discourse protocol, four tests were administered to PWA: the AQ subtests from the WAB-R, the short form of the Boston Naming Test–Second Edition (Kaplan, Goodglass, & Weintraub, 2001), the Verb Naming Test from the Northwestern Assessment of Verbs and

Sentences–Revised (Cho-Reyes & Thompson, 2012), and the AphasiaBank Repetition test (available at the AphasiaBank website, http://aphasia.talkbank.org/), developed to assess word- and sentence-level repetition skills.

Language Transcription and Analysis

Transcriptions were completed by trained and experienced transcribers using Codes for the Human Analysis of Transcripts (CHAT) format, which allows for analyses using the Computerized Language Analysis (CLAN) program (MacWhinney, 2000). Following the guidelines of Berndt, Wayland, Rochon, Saffran, and Schwartz (2000), utterances were segmented on the basis of the following hierarchy of indices: syntax, intonation, pause, semantics. Two transcribers reviewed each transcription, and the two reached forced-choice agreement on any discrepancies. Word errors were coded using the error-coding system described on the AphasiaBank website (http://aphasia.talkbank.org/).

Utterances that were not relevant to the task (e.g., comments about the task, asking for a repeat or clarification of the instructions) were eliminated from the analyzed transcripts. This applied primarily to the procedural-discourse task, which was a more constrained task. Also not counted in the analyses were unintelligible content, word fragments, repetitions, content that was revised, and fillers (e.g., *uh*, *um*). These exclusions are consistent with the approaches used in analyzing language from PWA (Berndt et al., 2000; Nicholas & Brookshire, 1993). Paraphasic errors were transcribed with intended target-word replacements when it was obvious what they were. For example, if a person said "peanut buther," the transcription would be "peanut buther [: butter]." Analyses can be done with or without the target replacement. The advantage of using the replacement is that the English lexicon in the CLAN programs can identify the part of speech of the speaker's intended word for morphological analyses.

Using its own automatically generated part-of-speech analysis and the rules developed by Covington (2007), CLAN automatically computes the CPIDR PD index with an accuracy slightly exceeding that of Covington (2007). CLAN's PD results were compared with those from CPIDR 3 (using the speech-mode option) for 40 of the original 80 transcripts used in the study by Brown et al. (2008). The correlation between the PD scores from CPIDR 3 and CLAN was .998 for number of words and .992 for number of propositions. The correlation between CLAN's PD scores and those from human coding (also made available from the Brown et al., 2008, study) was .971, which is almost identical to the .973 correlation between CPIDR 3[1] and human coding. In CLAN, the EVAL command generates the density measure (Forbes, Fromm, Holland, & MacWhinney, 2014).

[1]A newer version of CPIDR, CPIDR 5.1 (Covington 2012), no longer relies on an external tagger and is slightly more accurate than CPIDR 3 in coding certain sentences.

Statistical Analysis

Hotelling's T^2 statistics, a multivariate analog of the two-sample t test, was used to address the first question of whether there were significant differences in the distribution of PD scores between the PWA group and the control group for both the procedural and narrative tasks. After determination of an overall statistically significant difference, univariate analyses were performed for differences for the procedural and narrative PD means, respectively, between the PWA and the control group.

A one-way analysis of variance with aphasia type as a factor was used to investigate how the distribution of PD differed across aphasia types. After significant differences in means for both the procedural and narrative tasks were determined, Tukey's honestly significant difference was used to identify which pairs of aphasia types were different from each other. Analysis of covariance was used to control for aphasia severity, as measured by the WAB-R AQ.

All tests and confidence intervals were assessed for deviations from normality and nonconstant variance. Analyses were run using the statistical package R, with $\alpha = .05$ as the cutoff for statistical significance.

Results

The study sample included 195 PWA and 168 participants in a control group. Mean PD scores and mean number of words for each task (narrative and procedural) are presented in Table 2 for each group.

The box plots in Figure 1 present the distributions of PD scores for each task by group. The distribution of PD scores for the control group relative to the PWA group is shifted for both discourse types. That is, for the procedural task, the median for the control group lines up with the 75th percentile of the PWA group, meaning that PD scores for 75% of PWA are below the PD scores for 50% of the control group. For the narrative task, the first quartile of the control group lines up with the third quartile of the PWA group, meaning that PD scores for 75% of the PWA group are lower than the PD scores for 25% of the control group.

A joint, bivariate test for differences in the distribution of PD scores for both tasks between the PWA and control groups was statistically significant on the basis of Hotelling's T^2 test ($p < .001$). The 95% confidence intervals for the difference in mean PD scores between the control group and the PWA groups for the narrative task were [.051, .072] and for the procedural task were [.025, .054]. In both cases, the PD means of the control group were greater than the PD means for the PWA group.

The box plots in Figure 2 display the distributions of PD scores for both discourse tasks by aphasia type. The control group generally has higher scores than all aphasia subgroups for both discourse tasks. The Broca's group appears to have more variability than any other group.

Multivariate analysis of variance revealed a statistically significant difference in the distribution of PD

Table 2. Proposition density (PD) and number of words by task and group.

	Narrative		Procedural	
Group	**PD, *M* (*SD*)**	**# words, *M* (*SD*)**	**PD, *M* (*SD*)**	**# words, *M* (*SD*)**
PWA (*n* = 195)	.453 (.069)	329.1 (356.0)	.416 (.090)	42.9 (33.4)
Controls (*n* = 168)	.514 (.031)	303.2 (216.6)	.456 (.045)	87.8 (54.9)

Note. PWA = people with aphasia.

scores among the groups for each discourse type ($p < .001$). Tukey's honestly significant difference was used to adjust for all pairwise multiple-comparisons tests. The statistically significant results are presented in Tables 3 and 4 in the form of square matrices in which the lower triangle gives the *p* values associated with each pairwise difference. The corresponding cells in the upper right triangle give the 95% confidence interval associated with each *p* value. The confidence intervals in the upper right triangle should be interpreted as follows: For the Anomic–Broca's comparison in the procedural task, the confidence interval is [.029, .120], which means there is 95% confidence that the difference between the PD means for the Anomic and Broca's groups is contained in this interval. For both discourse types, the average PD score for participants with Broca's aphasia was significantly lower compared with participants in the control group and all other aphasia subgroups except Transcortical Motor. Further, for the narrative task, the average PD score for the control group was not significantly different from those of the participants with Wernicke's aphasia or WAB-R AQ scores above 93.8. For the procedural task, the average PD score for the control group was not statistically different from those of any of the aphasia groups other than Broca's.

Next, analysis of covariance was used to compare PD scores across aphasia types, controlling for WAB-R AQ score—that is, aphasia severity—separately for each discourse task. There was no statistically significant quadratic or linear association between PD and AQ for either task and no changes in the significance of the pairwise comparisons just reported. However, the scatter plots of PD versus AQ for both discourse tasks presented in Figure 3 call attention to two interesting subgroups of the data as it relates to severity: the Broca's participants with low AQ and low PD scores and the Wernicke's participants with low AQ and high PD scores. On further investigation, these Broca's participants had very limited output due to coexisting apraxia of speech (according to their primary clinicians' assessments). If the output is primarily nouns, as is typical with Broca's aphasia, PD scores may approach 0, as in the case of the participant with a PD score of 0 on the procedural task and a WAB-R AQ score of 40.9. Her transcript (in CHAT format) looks like this:

*PAR: oh &=ges:open_cabinet.
*PAR: &g &um &=ges:bread &d &um jam.
*PAR: &=ges:bread bread [/] bread.
*PAR: &um &=ges:scoops jelly.
*PAR: &=ges:spreads jelly.
*PAR: &um &um jam [/] jam &=ges:spreads.
*PAR: &=ges:sandwich yeah.

Another participant with Broca's aphasia and an AQ of 36.2 had a PD score of .125 on the procedural task on the basis of the conjunction "and" being counted as the only proposition in his response. The "xxx" marking in his CHAT transcript given here means that the speech was unintelligible (not transcribable):

*PAR: one two three &=ges.
*PAR: xxx.
*PAR: penis [: peanut] butter (.) and jelly sandwich.
*PAR: &=ges:holding xxx.

In contrast, participants with Wernicke's aphasia with equally low AQ scores used more grammatical elements (i.e., function words), which resulted in higher PD scores. Many of these PD scores fell more than 1 *SD* above the mean for the control group (.416). Again, samples from the procedural task are brief enough to include here for illustrative purposes, and results did not differ for the longer narrative task. This transcript from a man with an AQ of 36.8 and a PD score of .556 on the procedural task provides a good example. Words counted as propositions are in bold. In this transcript, the phonetically transcribed words were all coded as neologisms.

Figure 1. Distribution of proposition density for each discourse task by group. PWA = persons with aphasia.

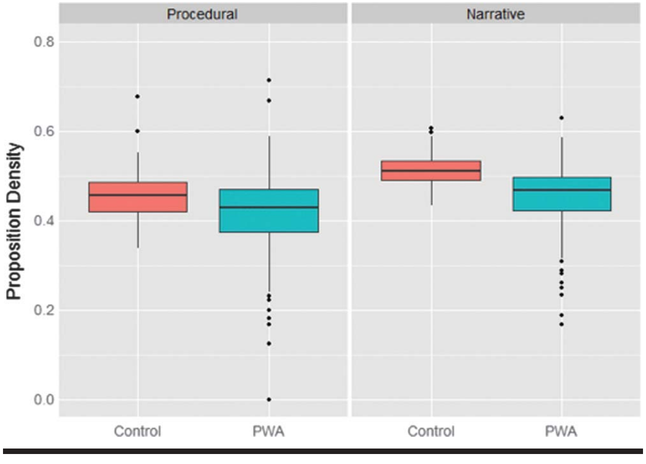
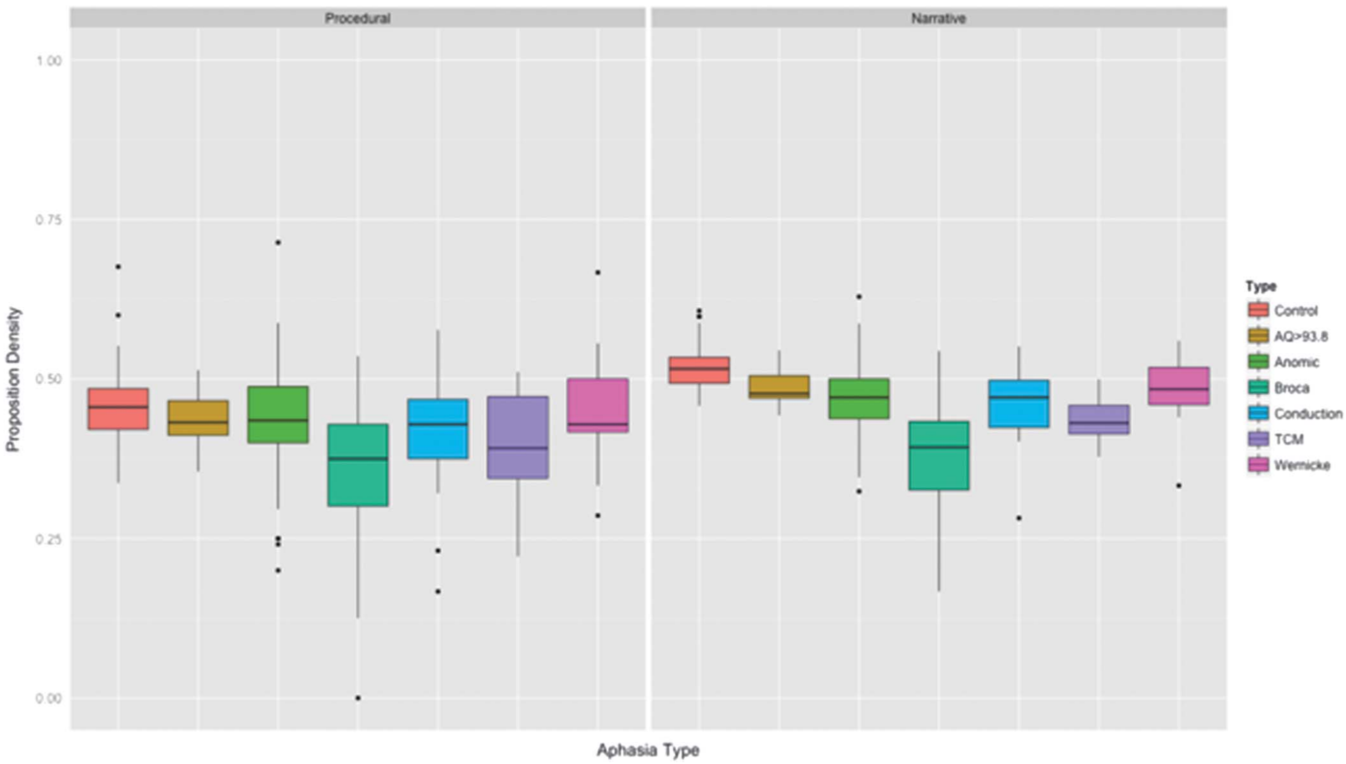


Figure 2. Distribution of proposition density by discourse task and aphasia type. TCM = transcortical motor.



*PAR: well **first** I **did** +…
*PAR: tæŋgɑr@u [: x@n] [//] tændʒɑr@u [: x@n].
*PAR: **but now that** everything **of** the **big** one &=imit:eating.
*PAR: I ofɪn@u [: **open**] it.
*PAR: pæ̃@u [: x@n] it &=pats:palm.
*PAR: &=pats:table **cut** the **bigger** one **of** it.
*PAR: I **usually pick** &=points:table **down of** it **a_little out_of a_little** xxx.
*PAR: **and take** it.

Another participant with Wernicke's aphasia and an even lower AQ of 28.2 had a short response but, interestingly, still had a very high PD score of .667:

*PAR: &b &d &=ges:spread.
*PAR: &=ges:spread **whatever** you (.) &k **hold and fold** it &=ges:fold.

Discussion and Conclusions

PWA and participants in the control group differed significantly on both discourse tasks for PD. PD appears differentially sensitive to aphasia type and severity. High PD scores can occur with lower WAB-R AQ scores, as in Wernicke's aphasia, or higher WAB-R AQ scores, as in anomic aphasia. The critical variable is the presence of words that count as propositions, such as verbs, participles,

Table 3. Significant differences in proposition density by aphasia type for narrative discourse task.

Aphasia type	Anomic	Broca's	Conduction	TCM	Wernicke's	AQ > 93.8	Control group
Anomic		(.057, .118)					(–.068, –.027)
Broca's	.000		(–.123, –.052)		(–.151, –.054)	(–.143, –.064)	(–.163, –.107)
Conduction		.000					(–.075, –.020)
TCM							(–.133, –.024)
Wernicke's		.000					
AQ > 93.8		.000					
Control group	.000	.000	.000	.000			

Note. TCM = transcortical motor; AQ = Western Aphasia Battery–Revised Aphasia Quotient.

Table 4. Significant differences in proposition density by aphasia type for procedural discourse task.

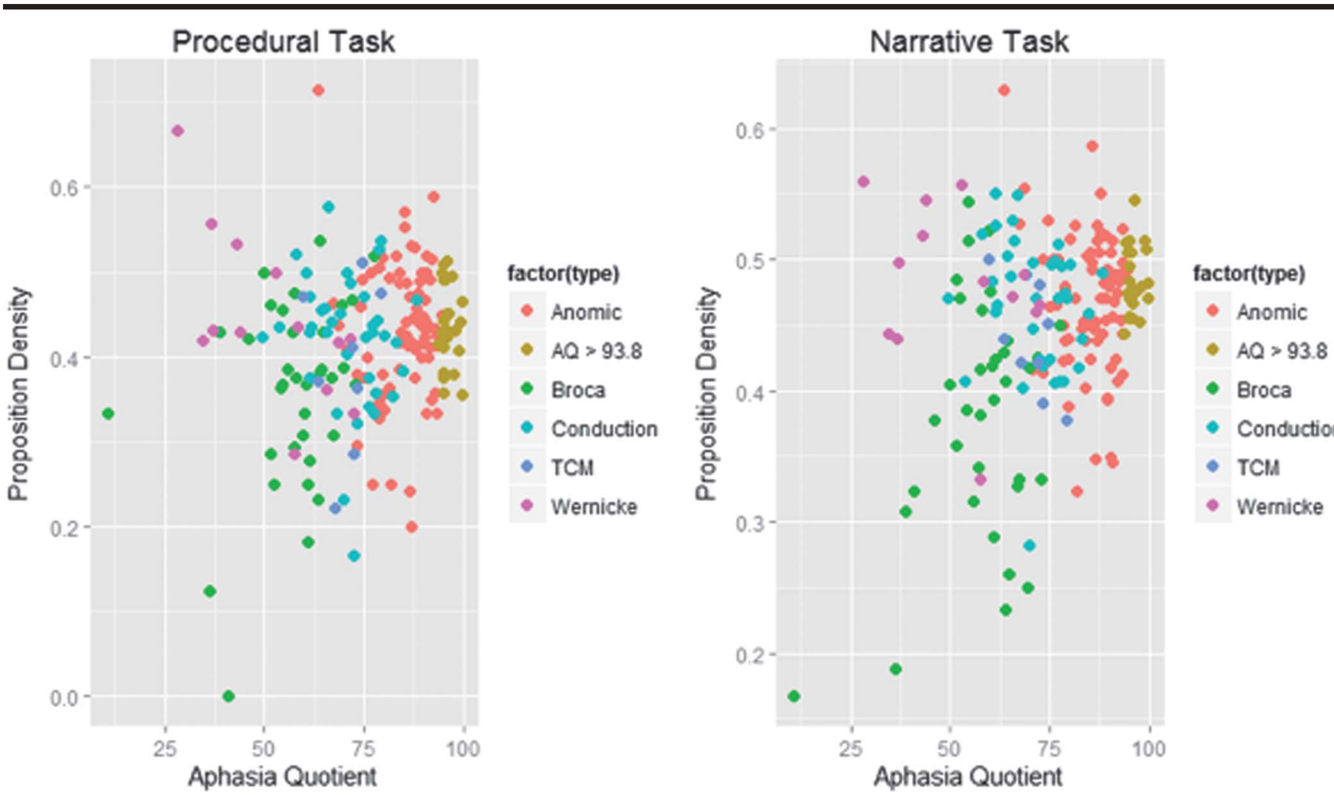
Aphasia type	Anomic	Broca's	Conduction	TCM	Wernicke's	AQ > 93.8	Control group
Anomic		(.029, .120)					
Broca's	.000		(−.116, −.011)		(−.160, −.015)	(−.137, −.020)	(−.140, −.057)
Conduction		.006					
TCM							
Wernicke's		.006					
AQ > 93.8		.001					
Control group		.000					

Note. TCM = transcortical motor; AQ = Western Aphasia Battery–Revised Aphasia Quotient.

prepositions, adjectives, and adverbs. Low PD scores occur mostly with Broca's aphasia, where WAB-R AQ scores can range from about 8 to 80 and many grammatical elements that contribute to PD may be reduced or absent, whereas nouns, which do not count as propositions, are more frequent. PD works best in distinguishing participants with Broca's aphasia from the other aphasia subgroups, with the exception of the Transcortical Motor group. Furthermore, when compared with the control group on the procedural-discourse task, the participants with Broca's aphasia were the only ones whose PD scores differed significantly. For the narrative task, participants with anomic, Broca's, conduction, and transcortical motor aphasia all had significantly lower PD scores. Analyses that controlled for severity did not reveal a significant association between PD and aphasia severity for procedural or narrative discourse.

These results are in some agreement and some disagreement with those reported in the poster by Ferguson et al. (2013) and the subsequent article by Bryant et al. (2013), which elaborated on the work presented in the poster. In comparing results, however, it is worth considering a few key differences. The language samples in the study by Bryant et al. were done as interviews and were much longer (approximately 2 hr in length, M = 2,831 words for PWA and 5,138 for the control group) than the samples in this study (approximately 3 min in length for PWA and 2 min for the control group for the stroke and illness narratives). Sample length has been shown to have

Figure 3. Scatter plots for Western Aphasia Battery–Revised Aphasia Quotient and proposition density for both discourse tasks. TCM = transcortical motor.



an effect on PD scores and variability in PD scores in studies of written texts (Ferguson et al., 2014; Spencer et al., 2015). To be specific, PD scores increase with larger text size, leveling out at about 60 words. PD variability also decreases with increased text size. The results from the present study of oral discourse conform to those findings from written discourse, with PD scores higher and variability lower for both groups in the task with more verbal output. Although the PWA group averaged fewer than 60 words on the procedural task, the control group averaged almost 90; both groups averaged over 300 words on the narrative task. For both tasks, no attempts were being made here to establish global norms or standards for PD scores in oral discourse. Instead, the focus was on comparing across groups performing the same tasks, one yielding a longer sample (stroke or illness narrative) and the other a shorter sample (procedural discourse). As demonstrated here, the number of propositions produced in a short, procedural task can still be compared across groups and provide results that do not differ from those for the longer, narrative-discourse task.

Another difference is that the comparison groups in the studies by Bryant et al. (2013) and Ferguson et al. (2013) were made up of family members of the PWA, who were also asked about the stroke, aphasia, and rehabilitation. In the current study, the comparison group was unrelated to the PWA group and produced narratives about an illness or injury instead of a stroke. The participants discussed a wide variety of problems, such as cancers, surgeries, heart problems, and broken bones. These illnesses and surgeries clearly were not as salient to many people or to these particular exchanges as strokes are to PWA talking with clinical aphasia researchers.

Despite the large difference in language-sample size and number of participants, this study supports the finding from Bryant et al. (2013) that PD was significantly different (lower) for PWA compared with the control group. Results from this study also indicated more variability in PD scores for PWA compared with the control group. Bryant et al. (2013) reported scores ranging from .009 to .528 propositions per word for the PWA and .505 to .573 for the control group. In the current study, the ranges for the stroke-narrative discourse task were .167 to .613 for the PWA group and .449 to .607 for the control group. The previous study reported a correlation of .475 between PD and WAB-R AQ score. Using a curve-estimation regression analysis, Bryant et al. reported a significant quadratic relationship. Results of the current study did not reveal any type of association between severity and PD.

It is interesting to note, however, that the scatter plots of PD and WAB-R AQ scores from both studies look very similar. Though Bryant et al. did not distinguish types of aphasia in their study, their scatter plot had the same low AQ scores associated with both very low and higher PD scores. Further inquiry revealed that the lowest AQ and PD scores in their study were also from participants with Broca's aphasia (L. Bryant, personal communication, May 2015). It is possible that removing these outlier data points would change the results of any relationship analyses. No aphasia-type confirmation was possible for the points in the upper left quadrant of the scatter plot by Bryant et al.; those points were found in the current study to represent primarily individuals with Wernicke's aphasia, who demonstrated a negative relationship between AQ and PD scores. With Wernicke's aphasia, higher PD scores occurred with lower AQ scores because output often included parts of speech that qualified as propositions (e.g., verbs, participles, adjectives, adverbs, prepositions.). Likewise, PWA with mild–moderate Broca's aphasia had relatively higher AQ scores but lower PD scores due to the predominant use of nouns (not counted as propositions) and the limited use of prepositions, adverbs, adjectives, and other parts of speech counted as propositions. These characterizations are consistent with the extensive literature on language characteristics of these aphasia types (Ardila, 2010; Armstrong, 2000; Brookshire, 1997; Saffran, Berndt, & Schwartz, 1989). The fact that PD did not distinguish the Broca's group from the Transcortical Motor group is likely because they are both nonfluent aphasia types that differ on the basis of WAB-R Repetition subtest scores only. Individuals with transcortical motor aphasia have been described as presenting with the cardinal features of Broca's aphasia (Turkstra, 2011). All of the other aphasia subgroups were fluent aphasias. Thus, although PD is measured per total words, the reduced output of nonfluent aphasia is also clearly limited in its propositional content. It is important to note, with specific reference to the individuals with Wernicke's aphasia, that a higher PD score does not necessarily mean that discourse is more successful or less impaired. As with MLU and other measures in aphasia, more is often, but not always, better.

In other adult language research, a number of studies have demonstrated relationships between PD in early adult life and cognition in late life (Engelman et al., 2010; Iacono et al., 2009; Kemper et al., 2001; Riley et al., 2005). Farias et al. (2012) recently reported a significant association between low PD measured in late life and steeper subsequent declines in cognitive function. Many authors describe proposition density as an index of "cognitive reserve" or a buffer against the effects of neuropathology (Chand et al., 2012; Engelman et al., 2010; Iacono et al., 2009; Snowdon et al., 1996). As mentioned earlier, PD has been shown to decline in cases of traumatic brain injury (Coelho et al., 2005).

In addition to the discourse findings about aphasia, this article presents a newly implemented version of the automated measure for PD in discourse. Unlike CPIDR, this new method runs on both Windows and Macintosh computers; both version can be downloaded freely from the web. With these types of tools, we recommend further research on the density measure to better understand its value in aphasia, with specific attention to its association with type and severity of aphasia. Moving forward, the goal is to explore other discourse measures that can be automated and then compare directly across measures, much like Fergadiotis, Wright, and West (2013) did for lexical diversity. Measures such as D-level analysis (Rosenberg & Abbeduto, 1987) for rating sentence complexity—which was revised by Covington, He, Brown, Naçi, and Brown

(2006) and more recently automated by Lu (2010)—would be interesting and relevant to such exploration using the grammatical relations program in CLAN. Access to automated analysis tools and the large AphasiaBank database should facilitate further investigation of PD and its relationship to other aphasia measures.

Acknowledgment

This work was funded by National Institute on Deafness and Other Communication Disorders Grant R01-DC008524 (2012–2017), awarded to Brian MacWhinney.

References

Ardila, A. (2010). A proposed reinterpretation and reclassification of aphasic syndromes. *Aphasiology, 24,* 363–394.

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

Berndt, R. S., Wayland, S., Rochon, E., Saffran, E., & Schwartz, M. (2000). *Quantitative production analysis: A training manual for the analysis of aphasic sentence production*. Hove, United Kingdom: Psychology Press.

Brink, T. L., Yesavage, J. A., Lum, O., Heersema, P. H., Adey, M., & Rose, T. L. (1982). Screening tests for geriatric depression. *Clinical Gerontologist, 1*(1), 37–43.

Brookshire, R. H. (1997). *Introduction to neurogenic communication disorders* (5th ed.). St. Louis, MO: Mosby-Yearbook.

Brown, C., Snodgrass, T., & Covington, M. (2007). Computerized Propositional Idea Density Rater 3 (CPIDR 3) [Computer software]. Athens, GA: The University of Georgia, CASPR Project, Artificial Intelligence Center.

Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods, 40,* 540–545.

Bryant, L., Spencer, E., Ferguson, A., Craig, H., Colyvas, K., & Worrall, L. (2013). Propositional Idea Density in aphasic discourse. *Aphasiology, 27,* 992–1009.

Butler, S. M., & Snowdon, D. A. (1996). Trends in mortality in older women: Findings from the Nun Study. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 51B,* S201–S208.

Chand, V., Baynes, K., Bonnici, L. M., & Farias, S. T. (2012). A rubric for extracting idea density from oral language samples. *Current Protocols in Neuroscience, 58,* 10.5.1–10.5.15.

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology, 26,* 1250–1277. doi:10.1080/02687038.2012.693584

Christiansen, J. A. (1995). Coherence violations and propositional usage in the narratives of fluent aphasics. *Brain and Language, 51,* 291–317.

Coelho, C. A., Grela, B., Corso, M., Gamble, A., & Feinn, R. (2005). Microlinguistic deficits in the narrative discourse of adults with traumatic brain injury. *Brain Injury, 19,* 1139–1145.

Covington, M. A. (2007). *CPIDR 3 user manual*. Institute for Artificial Intelligence, The University of Georgia.

Covington, M. A. (2012). *CPIDR® 5.1 user manual*. Athens, GA: Artificial Intelligence Center, The University of Georgia.

Covington, M. A., He, C., Brown, C., Naçi, L., & Brown, J. (2006). *How complex is that sentence? A proposed revision of the Rosenberg and Abbeduto D-Level scale* (CASPR Resarch Report No. 2006-01). Athens, GA: Artificial Intelligence Center, The University of Georgia.

Cunha, A. L. V. da, Sousa, L. B. de, Mansur, L. L., & Aluisio, S. M. (2015). Automatic proposition extraction from dependency trees: Helping early prediction of Alzheimer's disease from narratives. In C. Traina Jr., P. P. Rodrigues, B. Kane, P. M. de A. Marques, & A. J. M. Traina (Eds.), *28th IEEE International Symposium on Computer-Based Medical Systems* (pp. 127–130). Los Alamitos, CA: Conference Publishing Services.

Engelman, M., Agree, E. M., Meoni, L. A., & Klag, M. J. (2010). Propositional density and cognitive function in later life: Findings from the Precursors Study. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 65B,* P706–P711.

Farias, S. T., Chand, V., Bonnici, L., Baynes, K., Harvey, D., Mungas, D., . . . Reed, B. (2012). Idea density measured in late life predicts subsequent cognitive trajectories: Implications for the measurement of cognitive reserve. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 67,* P677–P686.

Fergadiotis, G., Wright, H. H., & West, T. M. (2013). Measuring lexical diversity in narrative discourse of people with aphasia. *American Journal of Speech-Language Pathology, 22,* S397–S408.

Ferguson, A., Spencer, E., Bryant, L., Craig, H., Colyvas, K., & Worrall, L. (2013, June). *Computerized analysis of Propositional Idea Density: Effects of presence and severity of aphasia*. Poster presented at the Clinical Aphasiology Conference, Tucson, AZ.

Ferguson, A., Spencer, E., Craig, H., & Colyvas, K. (2014). Propositional Idea Density in women's written language over the lifespan: Computerized analysis. *Cortex, 55,* 107–121.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198. doi:10.1016/0022-3956(75)90026-6

Forbes, M., Fromm, D., Holland, A., & MacWhinney, B. (2014, May–June). *EVAL: A tool for clinicians from AphasiaBank*. Paper presented at the Clinical Aphasiology Conference, St. Simons Island, GA.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21,* 354–375.

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

Hughlings-Jackson, J. (1879). On affections of speech from disease of the brain. *Brain, 2,* 203–222.

Iacono, D., Markesbery, W. R., Gross, M., Pletnikova, O., Rudow, G., Zandi, P., & Troncoso, J. C. (2009). The Nun Study: Clinically silent AD, neuronal hypertrophy, and linguistic skills in early life. *Neurology, 73,* 665–673.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test–Second Edition*. Austin, TX: Pro-Ed.

Kemper, S., Greiner, L. H., Marquis, J. G., Prenovost, K., & Mitzner, T. L. (2001). Language decline across the life span: Findings from the Nun Study. *Psychology and Aging, 16,* 227–239.

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

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363–394.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15,* 474–496.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25,* 1286–1307.

Miller, J. (2003). Systematic Analysis of Language Transcripts (Version V8.0) [Computer software]. Madison: University of Wisconsin–Madison, Waisman Center, Language Analysis Laboratory.

Mortimer, J. (2012). The Nun Study: Risk factors for pathology and clinical-pathologic correlations. *Current Alzheimer Research, 9,* 621–627.

Nicholas, L. E., & 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,* 338–350.

Riley, K. P., Snowdon, D. A., Desrosiers, M. F., & Markesbery, W. R. (2005). Early life linguistic ability, late life cognitive function, and neuropathology: Findings from the Nun Study. *Neurobiology of Aging, 26,* 341–347.

Roark, B., Mitchell, M., Hosom, J.-P., Hollingshead, K., & Kaye, J. (2011). Spoken language derived measures for detecting mild cognitive impairment. *IEEE Transactions on Audio, Speech, and Language Processing, 19,* 2081–2090.

Rosenberg, S., & Abbeduto, L. (1987). Indicators of linguistic competence in the peer group conversational behavior of mildly retarded adults. *Applied Psycholinguistics, 8,* 19–32.

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language, 37,* 440–479.

Snowdon, D. A., Greiner, L. H., & Markesbery, W. R. (2000). Linguistic ability in early life and the neuropathology of Alzheimer's disease and cerebrovascular disease: Findings from the Nun Study. *Annals of the New York Academy of Sciences, 903,* 34–38.

Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., & Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the Nun Study. *JAMA, 275,* 528–532.

Spencer, E., Ferguson, A., Craig, H., Colyvas, K., Hankey, G. J., & Flicker, L. (2015). Propositional idea density in older men's written language: Findings from the HIMS study using computerised analysis. *Clinical Linguistics & Phonetics, 29,* 85–101.

Turkstra, L. (2011). Transcortical motor aphasia. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (pp. 2539–2540). New York, NY: Springer New York.

Turner, A., & Greene, E. (1977, April). *The construction and use of a propositional text base* (Technical Report No. 63). Boulder, CO: University of Colorado, Institute for the Study of Intellectual Behavior.

Ulatowska, H. K., Freedman-Stern, R., Doyel, A. W., Macaluso-Haynes, S., & North, A. J. (1983). Production of narrative discourse in aphasia. *Brain and Language, 19,* 317–334.

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

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.