NIH Public Access
Author Manuscript
*Aphasiology*. Author manuscript; available in PMC 2012 November 01.

Published in final edited form as:
*Aphasiology*. 2011 ; 25(11): 1414–1430. doi:10.1080/02687038.2011.603898.

# Lexical diversity for adults with and without aphasia across discourse elicitation tasks

**Gerasimos Fergadiotis** and **Heather Harris Wright**
Department of Speech and Hearing Science, Arizona State University, Tempe, AZ, USA

## Abstract

**Background—**Differences in lexical diversity (LD) across different discourse elicitation tasks have been found in neurologically intact adults (NIA) (Fergadiotis, Wright, & Capilouto, 2010) but have not been investigated systematically in people with aphasia (PWA). Measuring lexical diversity in PWA may serve as a useful clinical tool for evaluating the impact of word retrieval difficulties at the discourse level.

**Aims—**The study aims were (a) to explore the differences between the oral language samples of PWA and NIA in terms of LD as measured by dedicated computer software (*voc-D*), (b) to determine whether PWA are sensitive to discourse elicitation task in terms of LD, and (c) to identify whether differences between PWA and NIA vary in magnitude as a function of discourse task.

**Method & Procedures—**Oral language samples from 25 PWA and 27 NIA were analysed. Participants completed three commonly used discourse elicitation tasks (single pictures, sequential pictures, story telling) and voc-*D* was used to obtain estimates of their LD.

**Outcomes & Results—**A mixed 2 × 3 ANOVA revealed a significant group task interaction that was followed by an investigation of simple main effects and tetrad×comparisons. Different patterns of LD were uncovered for each group. For the NIA group results were consistent with previous findings in the literature according to which LD varies as a function of elicitation technique. However, for PWA sequential pictures and story telling elicited comparable estimates of LD.

**Conclusions—**Results indicated that LD is one of the microlinguistic indices that are influenced by elicitation task and the presence of aphasia. These findings have important implications for modelling lexical diversity and selecting and interpreting results from different discourse elicitation tasks.

### Keywords

Discourse; Productive vocabulary; Computational linguistics

---

© 2011 Psychology Press, an imprint of the Taylor & Francis Group, an Informa business

Address correspondence to: Heather Harris Wright, Arizona State University, Department of Speech and Hearing Science, P.O. Box 870102, Tempe, AZ 85287-0102, USA. heather.wright.1@asu.edu.

**Publisher's Disclaimer:** This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

$watermark-text
$watermark-text
$watermark-text

$watermark-text

$watermark-text

$watermark-text

An important aspect of aphasia assessment is the analysis of discourse production. Discourse is successfully produced when the listener is able to reconstruct the message sent and interpret its intended meaning. People with aphasia (PWA) present with word retrieval deficits that can significantly impact how successfully their discourse production is conveyed. For example, their verbal output may include non-referential terms, fillers, paraphasias, and/or neologisms. Discourse analysis offers an opportunity to observe and study, through a wide variety of analyses, these complex cognitive/linguistic behaviours in the most naturally occurring and commonly used form of communication. As such, language sample analysis has been used as a clinical tool for differential diagnosis (e.g., Fleming & Harris, 2008; Murray, 2009), a key indicator for determining the efficacy of treatment approaches for individuals with aphasia (e.g., Cameron, Wambaugh, Wright, & Nessler, 2006; del Toro et al., 2008; Rider, Wright, Marshall, & Page, 2008); as well as an indicator of social validity (e.g., Ballard & Thompson, 1999).

Microlinguistic processes that give rise to specific discourse features, such as lexical-semantics, can be evaluated with different content analyses. These techniques may include assessments of informativeness, efficiency, and lexical diversity (LD) of a speaker's production. Reduced informativeness and reduced efficiency have been well documented in the aphasia literature with different discourse elicitation techniques. In addition, these properties have been explored in PWA speaking different languages, ranging in severity of language impairment, and ranging in time post onset of aphasia (Honda, Mitachi, & Wafamori, 1999; Larfeuil & Dorze, 1997; McNeil, Doyle, Fossett, Park, & Goda, 2001; McNeil et al., 2007). LD has also been quantified in PWA but less frequently and often estimated incorrectly.

## LEXICAL DIVERSITY IN DISCOURSE PRODUCTION BY PWA

For the purposes of this study, LD will be defined within Chapelle's (1994) model of vocabulary. Chapelle's model consists of four dimensions: (a) vocabulary size, (b) word knowledge (e.g., phonology, syntactic properties, etc), (c) density of the semantic network, and (d) processes that are involved in lexical access and retrieval. Within this framework, language performance is assumed to depend on both the implicit knowledge one possesses (e.g., variety of vocabulary) and the mechanisms that allow her/him to process it (e.g., access and retrieval). What becomes evident, then, is that *knowledge of* vocabulary and the *capacity to* demonstrate that knowledge cannot be equated (Chomsky, 1980). Based on these premises LD can be defined as the range of vocabulary deployed in a text by a speaker that reflects his/her capacity to access and retrieve target words from a relatively intact knowledge base (i.e., lexicon) for the construction of higher linguistic units. It is noteworthy that this view of LD is consistent with “performance” definitions that de-emphasise loss of language knowledge as the primary deficit in aphasia which is instead recast as an access deficit (e.g., McNeil & Pratt, 2001).

The efficiency of lexical access during discourse production depends on the synergy of several highly dynamical systems (Bock, 1982; Dell, 1986; Levelt, 1989) that may be impaired in PWA. Word retrieval depends on lexical characteristics (e.g., word frequency) and the interaction at different linguistic levels (e.g., phonologic, semantic, syntactic) of competing items in the lexicon. For example, Gordon and Dell (2003) have suggested that at the lemma selection stage of sentence generation (Levelt, Roelofs, & Meyer, 1999) at which words are selected as semantic/syntactic entities (Kempen & Hoenkamp, 1987), the selection of competing lexical targets depends on two separate networks. First, as the sentence unfolds, semantic nodes set off a target along with multiple semantic competitors through spreading activation. Subsequently, lexical items receive input from a network that applies syntactic constraints. During the same stage, phonologically related competitors may

$watermark-text

$watermark-text

$watermark-text

interfere further with the selection of target words (Schwartz, Dell, Martin, Gahl, & Sobel, 2006 in Dell, Oppenheim, & Kilttredge, 2008). In addition, the selection of the appropriate vocabulary may be influenced by supralinguistic factors such as the topic of the discourse, the genre a speaker chooses to respond, and the interlocutors (Haliday & Hassan, 1989). The relative strength of the signal that comes from different sources eventually determines the selection of linguistic representations.

Disruptions in one or more of these mechanisms can have direct effects on word retrieval during discourse production. Omissions can occur after a failure of the system to make words reach their activation threshold. Or paraphasias may be observed when competitors prevail over the target words. This interference from competition may arise from different sources. So, in “The cat eats the cheese”, the selection of “cat” might be hindered from semantic or phonological competitors such as “dog” or “rat”. Or interference may arise from other words in the sentence or the broader context in which the phrase occurs. Also, according to Gordon and Dell's *division of labour hypothesis* (2003), a lesion in the network responsible for boosting or inhibiting syntactic representations (or in the semantic network) can influence the ease with which certain lexical items are retrieved. Although it may often be difficult to pinpoint the exact mechanism of lexical processing breakdown, the nature and extent of impairment can potentially influence the range of vocabulary a speaker exhibits.

## MEASURING LEXICAL DIVERSITY

The most obvious way to capture LD in a sample would be to count the number of different words (NDW or *types*). However, only when the number of total words (*tokens*) is kept constant, are comparisons of NDW across different samples meaningful. Otherwise the NDW would reflect both LD as well as the contribution of length; that is, individuals who produced more verbal output would be credited with higher LD. To overcome this obstacle one could consider the ratio of the types divided by the tokens (TTR) to control for length. However, even though TTR is an improvement compared to the NDW, it is still inherently flawed because it also varies as a function of sample length. As the sample increases the probability of introducing new words decreases; and as a result the growth of the numerator in the TTR decelerates. However, the denominator (i.e., the number of tokens) always increases steadily with every additional word and as a result, overall, the TTR decreases monotonically. Therefore comparisons across speakers or even across different samples produced by the same speaker will be confounded by sample length. In an example that highlights the central weakness of TTR, McNeil et al. (2007) found that their participants with aphasia had significantly lower TTR on the Story Retelling Procedure (SRP; Doyle et al., 1998) compared to other discourse samples (e.g., procedural descriptions). McNeil et al. acknowledged that sample length variations might have contributed to this counterintuitive finding. Indeed, given that the number of words elicited using the SRP was significantly greater than any other elicitation procedure, lower TTR were expected.

To solve similar problems, researchers have used various algebraic transformations of TTR —e.g., Split TTR (Engber, 1995); Root TTR (Guiraud, 1960), Corrected TTR (Carroll, 1964), Log TTR (Herdan, 1960)—some of which have been applied to language samples produced by adults with aphasia (e.g., Prins, Snow, & Wagenaar, 1978; Wachal & Spreen, 1973). In these studies PWA produced significantly less lexically diverse samples compared to their NI counterparts. However, some of these tools have also been found to covary with sample length, thus yielding mathematically and conceptually spurious results. Others require very large samples to produce stable estimates that prohibit their use with PWA (Jarvis, 2002; Malvern & Richards, 1997; Tweedie & Baayen, 1998; Vermeer, 2000).

$watermark-text

$watermark-text

$watermark-text

To overcome these difficulties, some researchers have proposed standardising sample size through truncation. However, even though "standard" lengths have been proposed (Brookshire & Nicholas, 1994; Prins & Bastiaanse, 2004) often it is not feasible to obtain a predetermined number of tokens across studies. One reason is that sample length may depend on the elicitation task. When individuals describe a single picture or a procedure such as planting a flower in the garden, NIA and PWA often produce samples that are less than 200 tokens (Nicholas & Brookshire, 1993; Wright & Capilouto, 2011). It is common, then, for researchers to ignore consensus and restrict the number of tokens to be equal to the shortest sample in the study. As an example, Gordon (2008) used samples that consisted of 200 content word tokens. An additional issue that arises with truncation is that discarding any amount of a language sample reduces the samples' integrity and therefore reduces the validity of the analysis (McCarthy & Jarvis, 2010). Rather, it is possible that restricting the sample might obscure the findings due to clustering of content words (Prins & Bastiaanse, 2004).

Recently a new measure, *D*, was developed that combines an algebraic transformation model and curve fitting to estimate LD. *D* appears to be relatively robust to length variation, a feature that allows for comparisons of discourse samples within and between participants as well as across studies (McKee, Malvern, & Richards, 2000) without requiring truncation. The validity of *D* has been explored in several studies and has been found to be satisfactory (Malvern & Richards, 1997; McCarthy & Jarvis, 2010; McKee et al., 2000). *D* can be estimated using the voc-*D* program in Computerised Language Analysis (CLAN; MacWhinney, 2000) and the process is reasonably automated and straightforward once the samples have been transcribed.

## EFFECTS OF LANGUAGE SAMPLING TECHNIQUE

A variety of elicitation techniques have been used to obtain language samples. For example, researchers have asked participants to describe single or sequential pictorial stimuli (Christiansen, 1995; Nicholas & Brookshire 1993; Olness, 2006; Wright & Capilouto, 2009). Further, narratives have been elicited through re-telling of familiar stories (i.e., story tellings) and/or sharing past experiences through personal narratives (Ash et al., 2006; Coelho, Grela, Corso, Gamble, & Feinn, 2005; Hough & Barrow, 2003; Ulatowska, North, & Macaluso-Haynes, 1981). It is generally accepted that different elicitation techniques may impose different cognitive and linguistic demands (Bliss & McCabe, 2006; Brady, Armstrong, & Mackenzie, 2005; Nicholas & Brookshire, 1993; Ulatowska, Allard, & Chapman, 1990). For example, tasks may facilitate lexical access if they are associated with a high propensity for eliciting concrete, high-imageability, and high-frequency words or if they provide rich contextual information (Balota & Chumbley, 1984; Grosjean, 1980; Kroll & Merves, 1986; Tyler & Wessels, 1983). As a result, the performance on some microlinguistic indices such as LD may vary depending on the discourse type or the elicitation technique.

In a recent study Fergadiotis et al. (2010) investigated the effect of various elicitation techniques, and the discourse type most often elicited by them, on LD in cognitively healthy adults. Four discourse tasks were included; procedures, eventcasts, story telling, and recounts. Procedures are activity-focused, step-by-step descriptions of how to achieve a goal (e.g., how to plant a flower; Longacre, 1996). The other three tasks were designed to elicit three different types of narrative discourse. Eventcasts are narratives that explain a scene of activities (e.g., Cookie Theft Picture), stories are fictionaliSed, highly structured forms (e.g., Cinderella), and recounts are verbal reiterations of an event (e.g., what one did last weekend) (Heath, 1986). Study participants included 86 cognitively healthy adults comprising two age groups: young (20–29 years old) and old (70–89 years old). Using *D* as

$watermark-text

$watermark-text

$watermark-text

the method for estimating LD, Fergadiotis et al. found a LD hierarchy that was similar across age groups for the four discourse types. Procedural discourse yielded the lowest LD followed by event-casts and story telling, with recounts yielding the highest *D* values. Fergadiotis et al. concluded that, in cognitively healthy adults, LD varies as a function of elicitation technique and discourse type.

Wright, Silverman, and Newhoff (2003) measured TTR, NDW, and *D* in picture descriptions produced by adults with fluent and nonfluent aphasia. Sample length, measured by number of words, varied across the two aphasia groups. When sample length was not controlled, the participants with fluent aphasia yielded significantly higher *D* and NDW values compared to the participants with nonfluent aphasia indicating greater LD in their discourse samples. Once samples were truncated to 100 and 200 word samples, groups differed significantly for all measures (i.e., TTR, NDW, *D*). According to Wright et al., the findings demonstrate that *D* is a sensitive measure of subtle differences in LD in adults with aphasia that are not detected by TTR. Further, *D* is not as sensitive to input sample size as TTR is. However, the study included only one discourse elicitation task and a small number of study participants ($N$ = 9 per group).

The current study aims to expand on Fergadiotis et al.'s and Wright et al.'s work by: (a) establishing whether there are differences between the oral language samples of PWA and NIA in terms of LD as measured by a novel approach, (b) determining whether PWA are sensitive to discourse elicitation task in terms of LD, and (c) identifying whether differences between PWA and NIA vary in magnitude as a function of discourse task. Based on the well-documented word retrieval difficulties experienced by PWA that can severely affect their discourse production it is anticipated that language samples of PWA will demonstrate lower LD across all discourse tasks compared to NIA. Further, based on the effect of discourse task on LD in NIA, it is hypothesised that PWA will also demonstrate differences across tasks. Finally, we predict that when language samples are elicited using story telling, PWA will demonstrate poorer LD in the samples compared to the NIA group due to the increased demands that are associated with this elicitation technique.

# METHOD

## Participants

Language samples from 27 PWA with mild to moderate level of aphasia severity were retrieved from AphasiaBank, an online shared database that collects and analyses digital recordings of discourse from PWA across a series of tasks. All participants had acquired aphasia secondary to a single left hemisphere stroke. In addition, PWA met criteria for the study that included (a) chronic aphasia (minimum = 11 months post onset of stroke and maximum = 164 months), (b) no reported history of psychiatric or neurodegenerative disorders, (c) aided or unaided normal hearing acuity, (d) corrected or uncorrected normal visual acuity, (e) English as their primary language, and (f) classification of anomic or conduction aphasia as determined by performance on the *Western Aphasia Battery-Revised* (*WAB-R*; Kertesz, 2007). Further, several sub-tests from the Reading Comprehension Battery for Aphasia, Second Edition (*RCBA–2*; LaPointe & Horner, 1998) were administered as part of the AphasiaBank protocol but these were not analysed for the current study.

Only individuals with either anomic or conduction aphasia from AphasiaBank were included for several reasons. First, we were interested in obtaining a relatively large, yet homogeneous sample that would allow us to study LD in relative isolation from other impaired production processes. For example, individuals with nonfluent aphasias often exhibit deficits of articulatory-phonetic nature that render the coding of paraphasias difficult

$watermark-text
$watermark-text
$watermark-text



(e.g., apraxia of speech; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; McNeil, 1997). Second, estimating LD using voc-*D* requires a minimum of 50 tokens and therefore the inclusion of individuals with nonfluent aphasia who sometimes demonstrate very limited verbal output could become problematic.

All language samples for the NIA group came from an ongoing study of normal discourse under the direction of one of the authors (HHW). These participants were selected to closely match the PWA in terms of gender, age, and years of education. In addition they met the following inclusion criteria for participation in the study: (a) no history of stroke, head injury, or neurogenic disorder, per self-report, (b) aided or unaided hearing acuity within normal limits; (c) normal or corrected visual acuity; (d) monolingual speakers of English; (e) normal cognitive functioning as indicated by performance on the *Mini Mental State Examination* (*MMSE*; Folstein, Folstein, & McHugh, 1975); and (f) no signs of depression as indicated by a passing score (0–4) on the *Geriatric Depression Scale* (Brink et al., 1982). All participants lived independently and were recruited from the Lexington, KY and Phoenix, AZ communities. Table 1 summarises participants' characteristics.

## Discourse elicitation

**Stimuli and instructions—**Discourse samples were collected in a single session and three commonly used tasks for eliciting discourse for clinical and research purposes were employed: sequential pictures, single pictures, and telling of the story of *Cinderella* (Grimes, 2005). The three tasks were designed to elicit two different types of narrative discourse: eventcasts using the former two and story telling using the latter. The sequential picture stimuli included a four-frame strip from Menn et al. (1998) and a six-frame cartoon strip. The first is referred to as "*Broken Window*" and depicts a boy kicking a ball; the ball goes through a closed window and knocks down a lamp before a man picks up the ball and looks out of the window. The second, referred to as "*Umbrella*" depicts the story of a student who refuses to take the umbrella his mother offers him as he leaves for school; on his way to school it starts raining and he returns home to take the umbrella.

The single picture stimuli each showed a situation with a central theme and animate and inanimate pictured elements interacting. Each picture implied temporal sequencing of events prior and after the pictured scene. One of the single pictures is a photograph by Annie Wells of an emergency rescue of a girl from flood waters (Rubin & Newton, 2001). The second single picture stimulus is Nicholas and Brookshire's "*Cat Rescue*" (Nicholas & Brookshire, 1993). In this picture a little girl's cat is in a tree and her dad has tried to rescue the cat but got stuck in the tree; the fire department is arriving to rescue the cat and the man from the tree. The single and sequential picture stimuli can be found at: www.talkbank.org/AphasiaBank/protocol/pictures.html

For the eventcasts participants were first presented with the two sequential pictures and were asked to produce a story that was based on temporal sequencing. The tester used the following script: "Take a little time to look at these pictures. They tell a story. Take a look at all of them, and then I'll ask you to tell me the story with a beginning, a middle, and an end. You can look at the pictures as you tell the story." If the participants did not respond within 10 seconds, the tester prompted them to "*Take a look at this picture* (pointing to first picture) *and tell me what you think is happening*." If needed, the tester pointed to each picture sequentially, giving the prompt: "*And what happens here?*" Subsequently participants were presented with the two single pictures and were asked to produce a story with temporal sequencing using the following script: "Here is a picture. Look at everything that's happening and then tell me a story about what you see. Tell me the story with a beginning, a middle, and an end." Feedback was provided to avoid eliciting a simple description of objects, characters, and/or their physical characteristics. Participants also viewed and told

$watermark-text

$watermark-text

$watermark-text

the story depicted in a wordless picture book of *Cinderella*. Participants were presented with the stimulus book and were allowed as much time as they desired to view it and remind themselves of the story. Then, the book was removed and they were asked to “Tell the story in their own words”.

**Transcription and language sample preparation—**Samples were digitally recorded and then orthographically transcribed in the CHAT format that is compatible with a set of programs called CLAN (MacWhinney, 2000). Samples were then segmented into c-units. A c-unit is a communication unit and includes an independent clause with its modifiers (Loban, 1976); it is commonly used to segment oral discourse samples (Hughes, McGillivray, & Schmidek, 1997). The following is an example of a c-unit:

**Pre-c-unit segmented sample:** Oh little Sally's cat Chico got up in the tree again and daddy who is afraid of heights did his best to try to rescue him.

#### C-unit segmented

1. Oh little Sally's cat Chico got up in the tree again.
2. And daddy who is afraid of heights did his best to try to rescue him.

Approximately 10% of the samples from each group were randomly selected and transcribed again for reliability purposes. For both groups, intra- and inter-rater word-by-word transcription reliability was above 90%.

### Lexical diversity

Following segmentation, each word was tagged morphosyntactically. Because function words have little or ambiguous meaning and convey predominantly grammatical relationships, they were removed from the samples, leaving only content words (i.e., nouns, verbs, adjectives and -*ly* adverbs). Further, to avoid conflating LD with grammaticality, only unique lexical representations were counted as separate types. This was approximated by performing a lemma-based analysis of LD. To determine which affixes to retain we used Kiparsky's (1982) levels of morphological derivation. Level three inflections that do not alter the meaning of a stem were disregarded. For example, eat, eats, and ate were all considered three tokens of the same type.

Also, for both groups, repetitions, repairs, and filler words were excluded from further analysis.

Samples from PWA were further coded and all paraphasias were excluded from analysis (a) to ensure that estimates of LD reflected the efficiency of word access and retrieval as discussed within Dell and Chapelle's model (specifically, Chapelle's fourth dimension) and (b) because they distort the common code between inter-locutors that is prerequisite for efficient and successful communication (Jakobson, 1971). Finally, *D* was calculated using the voc-*D* program in CLAN. The code used for the analysis can be found at www.asu.edu/clas/shs/aald/current_research.html. See MacWhinney, Fromm, Holland, Forbes, and Wright (2010) for a more detailed example of automated analyses performed using CLAN.

*D* has been described in detail elsewhere, and therefore only a brief review of the approach is offered here (cf. MacWhinney, 2000; McCarthy & Jarvis, 2007; McKee et al., 2000). For each language sample, calculating *D* involves a series of random text samplings to plot an empirical TTR versus number-of-tokens curve (see Figure 1). First, 35 tokens are randomly drawn from the text without replacement and the TTR is computed. This process is repeated 100 times and the average TTR for 35 tokens is estimated and plotted. The same routine is

$watermark-text

$watermark-text

$watermark-text

then repeated for subsamples from 36 to 50 tokens. The average TTR for each subsample of increasing token size is subsequently plotted to form the empirical curve. Then the estimation of *D* proceeds by producing a theoretical curve that maximises the fit to the empirical TTR curve using the least squares approach. Because *D* is the product of a sampling, stochastic process, its value varies each time the program is run. For that reason the whole process is repeated three times and the final *D* value is the average of the three runs.

## RESULTS

### Preliminary analysis

Prior to performing the statistical analyses for addressing the study aims, data were examined through various PASW Statistics 18.0.1 (SPSS Inc.) programs for accuracy of data entry, missing values, univariate outliers with extreme *z* scores (larger than 3.3), and fit between variables' distributions and the assumptions of univariate analysis (i.e., gross violations of normality and homoscedasticity). The data were examined separately for the two groups. No outliers were identified for either group. For the PWA, two cases had fewer than 50 tokens after function words and paraphasias were excluded from their language samples, and as a result *D* was not estimated. These cases were not included in the analysis, leaving 25 cases in the PWA group. For the NIA group no missing data were observed. The shape of the distributions of the dependent variable *D* was assessed using histograms and was found satisfactory (skewness and kurtosis ranged from –1 to +1). The assumption regarding homogeneity of variances, as assessed using Levene's test, was also met. Regarding the language samples, the average numbers of tokens across different conditions are shown in Table 2.

### Lexical diversity across group and discourse task

A 2 × 3 mixed ANOVA was conducted that included group (NIA, PWA) and task (single pictures, sequential pictures, and story telling). The former was the between-participants factor and the latter the within-participants factor. Significant results were found for the discourse task main effect, $\Lambda = .43$, $F(2, 49) = 33.52$, $p < .01$, *partial* $\eta^2 = .57$, and the interaction, $\Lambda = .65$, $F(2, 49) = 13.42$, $p < .01$, *partial* $\eta^2 = .35$. The group main effect was also significant, $F(1, 50) = 78.23$, $p < .01$, *partial* $\eta^2 = .61$.

To better understand the significant interaction, group and discourse task simple effects were examined by conducting a series of independent sample *t*-tests and paired-sample *t*-tests, respectively. To control for Type I error, familywise alpha was set to .0125 and .025 for each age and discourse task simple main effect, respectively (Maxwell & Delaney, 2004). Subsequently, familywise error rate across the *t*-tests was controlled using the Holm's sequential Bonferroni approach.

According to the results, NIA demonstrated higher LD across all three elicitation techniques of discourse compared to PWA. For both groups, single pictures were associated with the lowest LD. However, for the NIA group, story telling yielded higher LD than the sequential pictures; whereas PWA exhibited similar LD for the story and sequential pictures (see Figure 2). All comparisons were significant with the exception of the comparison of LD between story telling and sequential pictures for PWA, $t(24) = .64$, $p = .53$ (see Tables 3 and 4).

To further investigate LD patterns across groups and task elicitation techniques, the significant interaction was further followed up with tetrad comparisons. The mean difference in LD for the Cinderella story was significantly larger between the two groups compared to the mean difference in LD for the single pictures as well as the sequential pictures, $t(40) = 4.42$, $p < .001$ and $t(40) = 3.63$, p = .001. The mean differences for the single pictures and the sequential pictures were not significantly different, $t(40) = .69$, p = .49.

## DISCUSSION

The purposes of this study were to explore differences in LD between language samples produced by PWA and NIA, determine whether PWA are sensitive to discourse elicitation technique in terms of LD, and investigate the magnitude of the differences in LD between PWA and NIA. Three elicitation techniques that are often used in clinical and research practice were compared (single pictures, sequential pictures, and story telling) across PWA and NIA. Using dedicated software, which allows for comparisons of LD that minimises the impact of length variation, significant differences in LD were found for the discourse tasks used for both PWA and NIA. To the best of our knowledge this was the first time that, after quantifying LD with a sophisticated computational tool, it was demonstrated that LD appears to be one of the microlinguistic indices that are influenced by the elicitation technique in PWA. In what follows is a discussion of the results, their methodological and clinical implications, and future directions.

For the NIA, results replicate and extend previous findings in the literature (Fergadiotis et al., 2010; Wright & Capilouto, 2009). As in the aforementioned studies, in our case NIA demonstrated highest levels of LD for story telling, followed by sequential pictures; and, single pictures elicited the least lexically diverse samples. However, one limitation of the previous studies is that neither one had conducted a lemma-based analysis. We addressed this issue by utilising an analysis that excluded functions words, and considered variations of the same stem as the same type. By accounting only for content words in this way, *D* is believed to better reflect LD.

One of the main goals of the study was to establish the difference between NIA and PWA in terms of LD during discourse production. Discourse production entails the activation and interaction of multiple interconnected cognitive and linguistic subsystems that are often impaired in PWA. In our study PWA demonstrated significantly lower LD than NIA across discourse tasks. Given that the two groups were matched for age, gender, and education, it is unlikely that this finding was driven by differences in the first three dimensions of Chapelle's (1994) model. If one accepts that aphasia is caused primarily by an access deficit that affects language processing rather than loss of language per se, then our results most likely reflect the limited resources of PWA for accessing and retrieving lexical items. Specifically, limited LD may be a manifestation of (a) the erroneous responses that take the form of lexical and sublexical paraphasias (e.g., literal and semantic paraphasias), and (b) the inability to map from the conceptual representation to the phonological form that could lead to omissions and often lends itself to avoidance strategies and circumlocutions. The latter point highlights a characteristic behaviour that is more prominent to language production at the discourse level as opposed to confrontational naming. That is, when PWA anticipate or use problematic vocabulary, they may resort to more familiar words that can result in altering or simplifying their message.

Regarding the impact of discourse task on LD, PWA showed both similarities as well as differences compared to NIA. First, PWA demonstrated significantly higher LD for sequential pictures than single pictures, similarly to NIA. Capilouto, Wright, and Wagovich (2005) have suggested that sequential stimuli provide the participants with additional temporal and causal information about the depicted story. As a result, participants' verbal output in response to sequential pictures can be more lexically diverse because it attempts to convey representations that entail a larger number and/or more complex relationships



$watermark-text $watermark-text $watermark-text

between characters and events. Expressing a larger number and/or more complicated interactions would require the introduction of specific vocabulary; in turn, increasing the likelihood of sampling from a wider variety of lexical items and thus producing new vocabulary. In the absence of such additional information though (i.e., in the case of the single pictures) participants may have a higher probability to simply “list” events without considering the underlying relationships, and therefore be more prone to repetition. A relevant implication that pertains both to the clinical and research fields is that sequential pictures can be used to obtain a more diverse language sample in terms of lexical items that translates into more opportunities to observe behaviours of interest.

Further, PWA demonstrated an analogous decrease in LD for single and sequential stimuli compared to their matched controls. For these types of tasks, results of the three-tier analysis are suggestive of an undifferentiated limitation in PWAs’ ability to demonstrate a range of vocabulary that would be comparable to that exhibited by NIA. In other words, the impact of aphasia in terms of LD appears to be the same regardless of whether one chooses to use single or sequential pictures to elicit the discourse samples.

However, as opposed to single and sequential pictures, when narratives were elicited using the story-telling format, PWA demonstrated a significantly larger difference (see Figure 2) compared to NIA. Whereas NIA showed evidence of higher LD for story telling compared to sequential pictures, PWA failed to demonstrate a similar pattern. Instead, PWA exhibited similar LD for story telling and sequential pictures. This finding suggests that PWA may be more susceptible to the demands associated with specific elicitation techniques (in this case story telling). Several possible and potentially inclusive explanations could be proposed to account for these results. First, the story-retelling task is performed without pictorial support during discourse generation. So, PWA have to produce a narrative without the scaffolding provided by the illustrations that may serve as a cognitive map or schema. The absence of immediate visual support may force participants to allocate more resources on memory, planning, and organisational processes. Such an account rests on the assumption that the autonomous low-level cognitive process of lexical access shares resources with higher-order cognitive processes. Some recent findings appear to be in favour of this hypothesis that nevertheless merits further investigation (Rabovsky, Álvarez, Hohlfeld, & Sommer, 2008). At the same time it possible that the illustrations are priming the semantic/conceptual content of the depicted words, thus boosting their activation and allowing them to be more easily retrieved. This account, which is consistent with spreading activation models, would suggest that tasks that lack the contextual support in the form of pictures might be harder for PWA. In addition to the theoretical significance, this finding is clinically important and should be explored further in future investigations because it suggests that story telling has a greater discriminatory potential that could be tapped for diagnostic purposes.

## CONCLUSIONS

Taken as a whole, these results provide additional support for the notion that the nature of the task contributes to the cognitive and linguistic demands that are imposed on the speakers (Armstrong, 2000; Coelho, 2002; Cooper, 1990). In other words, individuals’ LD performance can vary both as a function of the individual's latent trait, as well as from the discourse type that the participant chooses to produce in response to a discourse elicitation technique. A critical methodological implication is that if LD is to be explored in discourse samples produced by adults with or without neurogenic communication impairments, it might be useful to include a “type of discourse” factor in the analysis rather than collapsing or aggregating the data. Including such a factor may present an advantage because it would facilitate a more accurate modelling of the data.

$watermark-text
$watermark-text
$watermark-text

$watermark-text

$watermark-text

$watermark-text

One limitation of this study is that LD was quantified using a single index (i.e., *D*). By including multiple indicators (such as the measure of textual lexical diversity; McCarthy, 2005), it would be possible to use statistical methods for a richer and more accurate investigation of LD. For example, using a latent variable modelling approach, one would be able to (1) operationally define LD as an unobserved trait that characterises individuals, (2) consider their discourse as the observed manifestation that enables us to make inferences about the individuals' level of LD, and (3) obtain error-free estimates of individuals' LD. This interactionalist approach would allow researchers to put forward questions about individuals' discourse abilities by examining their language outputs. Further, if the distinction between the trait and the manifest variables is made clear, then researchers can conduct a wide variety of analyses to understand both the nature of cognitive-communicative deficits and take into account the "peculiarities" of the methods that are used to elicit language samples.

Considering that the word retrieval difficulties experienced by PWA can severely limit their participation in communicative activities and negatively affect their discourse production, indices that assess their performance at the discourse level can prove to be very informative. LD, which is quantified in this study using a method that is robust to length variation, can complement other discourse measures in an attempt to assess the impact of word finding deficits at the discourse level. The way *D* was employed in this study does not provide direct insight into the complex interactions of the linguistic systems that are involved in accessing and retrieving lexical items. However, it does allow us to construct a more detailed account of the richness of their verbal output. Future investigations using LD and measures such as story-telling propositions and information units will further enhance our understanding of how the capacity to employ a large lexicon is associated with more efficient communication; and in turn, they may lead to the development of more valid and reliable treatment outcome measures.

## Acknowledgments

This research was partially supported by the National Institute on Aging Grant R01AG029476. We are especially grateful to the study participants. We also thank the volunteers in the Aging and Adult Language Lab at ASU for assistance with transcription and language analyses, and Leonid Spector, Davida Fromm, and Margaret Forbes from Carnegie Mellon University for assisting with CLAN coding and facilitating exchange of data.

## REFERENCES

Armstrong E. Aphasic discourse analysis: The story so far. Aphasiology. 2000; 14(9):875–892. doi: 10.1080/02687030050127685.

Ash S, Moore P, Antani S, McCawley G, Work M, Grossman M. Trying to tell a tale: Discourse impairments in progressive aphasia and frontotemporal dementia. Neurology. 2006; 66(9):1405–1413. doi:10.1212/01.wnl.0000210435.72614.38. [PubMed: 16682675]

Ballard K, Thompson K. Treatment and generalization of complex sentence production in agrammatism. Journal of Speech, Language, and Hearing Research. 1999; 42(3):690–707.

Balota DA, Chumbley JI. Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. Journal of Experimental Psychology: Human Perception and Performance. 1984; 10(3):340–357. doi: 10.1037/0096-1523.10.3.340. [PubMed: 6242411]

Bliss LS, McCabe A. Comparison of discourse genres: Clinical implications. Contemporary Issues in Communication Science and Disorders. 2006; 33(2):126–137.

Bock JK. Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. Psychological Review. 1982; 89(1):1–47.

Brady M, Armstrong L, Mackenzie C. Further evidence on topic use following right hemisphere brain damage: Procedural and descriptive discourse. Aphasiology. 2005; 19(8):731–747. doi: 10.1080/02687030500141430.

$watermark-text $watermark-text $watermark-text

Brink TL, Yesavage JA, Lum O, Heersema P, Adey MB, Rose TL. Screening tests for geriatric depression. Clinical Gerontologist. 1982; 1:37–44.

Brookshire RH, Nicholas LE. Speech sample size and test–retest stability of connected speech measures for adults with aphasia. Journal of Speech & Hearing Research. 1994; 37(2):399–407. [PubMed: 8028321]

Cameron RM, Wambaugh JL, Wright SM, Nessler CL. Effects of a combined semantic/phonologic cueing treatment on word retrieval in discourse. Aphasiology. 2006; 20(2–4):269–285. doi: 10.1080/02687030500473387.

Capilouto G, Wright HH, Wagovich SA. CIU and main event analyses of the structured discourse of older and younger adults. Journal of Communication Disorders. 2005; 38(6):431–444. doi: 10.1016/j.jcomdis.2005.03.005. [PubMed: 16199238]

Carroll, JB. Language and thought. Prentice-Hall; Englewood Cliffs, NJ: 1964.

Chapelle CA. Are C-tests valid measures for L2 vocabulary research? Second Language Research. 1994; 10(2):157–187.

Chomsky, N. Rules and representations. Columbia University Press; New York, NY: 1980.

Christiansen JA. Coherence violations and propositional usage in the narratives of fluent aphasics. Brain and Language. 1995; 51:291–317. [PubMed: 8564473]

Coelho CA. Story narratives of adults with closed head injury and non-brain-injured adults: Influence of socioeconomic status, elicitation task, and executive functioning. Journal of Speech, Language, and Hearing Research. 2002; 45(6):1232–1248. doi:10.1044/1092-4388(2002/099.

Coelho CA, Grela B, Corso M, Gamble A, Feinn R. Microlinguistic deficits in the narrative discourse of adults with traumatic brain injury. Brain Injury. 2005; 19(13):1139–1145. doi: 10.1080/02699050500110678. [PubMed: 16286327]

Cooper PV. Discourse production and normal aging: Performance on oral picture description tasks. Journals of Gerontology. 1990; 45(5):210–214.

del Toro CM, Altmann LJP, Raymer AM, Leon S, Blonder LX, Rothi LJG. Changes in aphasic discourse after contrasting treatments for anomia. Aphasiology. 2008; 22(7–8):881–892. doi: 10.1080/02687030701844204.

Dell GGS. A spreading-activation theory of retrieval in sentence production. Psychological Review. 1986; 93(3):283–321. [PubMed: 3749399]

Dell GS, Oppenheim GM, Kittredge KA. Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. Language and cognitive processes. 2008; 23(4): 583–608. [PubMed: 20622975]

Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA. Lexical access in aphasic and non-aphasic speakers. Psychological Review. 1997; 104(4):801–838. [PubMed: 9337631]

Doyle PJ, McNeil MR, Spencer KA, Goda AJ, Cottrell K, Lustig AP. The effects of concurrent picture presentations on retelling of orally presented stories by adults with aphasia. Aphasiology. 1998; 12:561–574.

Engber CA. The relationship of lexical proficiency to the quality of ESL compositions. Journal of Second Language Writing. 1995; 4(2):139–155.

Fergadiotis, G.; Wright, HH.; Capilouto, G. Productive vocabulary across discourse types; Clinical Aphasiology Conference; Isle of Palms, SC, USA. May. 2010

Fleming VB, Harris JL. Complex discourse production in mild cognitive impairment: Detecting subtle changes. Aphasiology. 2008; 22(7–8):729–740. doi:10.1080/02687030701803762.

Folstein MF, Folstein SE, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research. 1975; 12(3):189–198. doi:10.1016/0022-3956(75)90026-6. [PubMed: 1202204]

Gordon JK. Measuring the lexical semantics of picture description in aphasia. Aphasiology. 2008; 22(7–8):839–852. doi:10.1080/02687030701820063. [PubMed: 22399832]

Gordon JK, Dell GS. Learning to divide the labor: An account of deficits in light and heavy verb production. Cognitive Science. 2003; 27:1–40.

Grimes, N. Walt Disney's Cinderella. Random House; New York, NY: 2005.

$watermark-text
$watermark-text
$watermark-text

Grosjean F. Spoken word recognition processes and the gating paradigm. Perception and Psychophysics. 1980; 28:267–283. [PubMed: 7465310]

Guiraud, P. Problèmes et méthodes de la statistique linguistique. D. Reidel; Dordrecht, The Netherlands: 1960.

Halliday, MAK.; Hasan, R. Language, context, and text: aspects of language in a social-semiotic perspective. Oxford University Press; Oxford, UK: 1989.

Heath SB. Taking a cross-cultural look at narratives. Topics in Language Disorders. 1986; 7:84–95.

Herdan, G. TypeToken mathematics. Mouton; The Hague, The Netherlands: 1960.

Honda R, Mitachi M, Watamori TS. Production of discourse in high-functioning individuals with aphasia—with reference to performance on the Japanese CADL. Aphasiology. 1999; 13(6):475–493. doi:10.1080/026870399402037.

Hough MS, Barrow I. Descriptive discourse abilities of traumatic brain-injured adults. Aphasiology. 2003; 17(2):183–191. doi:10.1080/729255221.

Hughes, D.; McGillivray, L.; Schmidek, M. Guide to narrative language: Procedures for assessment. Thinking Publications; Eau Claire, WI: 1997.

Jakobson, R. Two aspects of language and two types of aphasic disturbances.. In: Jakobson, R.; Halle, M., editors. Fundamentals of language. Mouton; The Hague, The Netherlands: 1971.

Jarvis S. Short texts, best-fitting curves and new measures of lexical diversity. Language Testing. 2002; 19(1):57–84.

Kempen G, Hoenkamp E. An incremental procedural grammar for sentence formulation. Cognitive Science. 1987; 11:201–258.

Kertesz, A. Western aphasia battery – revised. Grune & Stratton; New York, NY: 2007.

Kiparsky, P. From cyclic phonology to lexical phonology.. In: Hulst, HVD.; Smith, N., editors. The structure of phonological representations. Foris Publishing; Dordrecht, The Netherlands: 1982. p. 131-175.

Kroll JF, Merves JS. Lexical access for concrete and abstract words. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1986; 12(1):92–107.

Lapointe, LL.; Horner, J. Reading comprehension battery for aphasia. Pro-Ed; Austin, TX: 1998.

Larfeuil C, Le Dorze G. An analysis of the word-finding difficulties and of the content of the discourse of recent and chronic aphasic speakers. Aphasiology. 1997; 11(8):783–811. doi: 10.1080/02687039708250456.

Levelt, WJM. Speaking: From intention to articulation. MIT Press; Cambridge, MA: 1989.

Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. Behavioural and Brain Science. 1999; 22(1):991–975.

Loban, WD. Language development: Kindergarten through grade twelve. NCTE committee on research (Report No. 18). National Council of Teachers of English; Urbana, IL: 1976.

Longacre, RE. The grammar of discourse. 2nd ed.. Plenum Press; New York, NY: 1996.

MacWhinney, B. The CHILDES project: Tools for analysing talk, Volume 1: Transcription format and programs. 3rd ed.. Lawrence Erlbaum Associates; Mahwah, NJ: 2000.

MacWhinney B, Fromm D, Holland A, Forbes M, Wright H. Automated analysis of the Cinderella story. Aphasiology. 2010; 24(6–8):856–868. doi:10.1080/02687030903452632.

Malvern, DD.; Richards, BJ. A new measure of lexical diversity.. In: Ryan, A.; Wray, A., editors. Evolving models of language. Multilingual Matters; Clevedon, UK: 1997. p. 58-71.

Maxwell, SE.; Delaney, HD. Designing experiments and analysing data: A model comparison perspective. 2nd ed.. Lawrence Erlbaum Associates; Mahwah, NJ: 2004.

McCarthy, PM. An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual lexical diversity (Doctoral Dissertation). Retrieved from ProQuest Nursing & Allied Health Source Database; 2005.

McCarthy PM, Jarvis S. Vocd: A theoretical and empirical evaluation. Language Testing. 2007; 24(4): 459–488. doi:10.1177/0265532207080767.

McCarthy PM, Jarvis S. MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. Behaviour Research Methods. 2010; 42(2):381–392.

$watermark-text
$watermark-text
$watermark-text

McKee G, Malvern D, Richards B. Measuring vocabulary diversity using dedicated software. Literary and Linguistic Computing. 2000; 15(3):323–337.

McNeil, MR. Clinical management of sensorimotor speech disorders. Thieme Medical Publishers; New York, NY: 1997.

McNeil MR, Doyle PJ, Fossett TRD, Park GH, Goda AJ. Reliability and concurrent validity of the information unit scoring metric for the story retelling procedure. Aphasiology. 2001; 15(10–11): 991–1006. doi:10.1080/02687040143000348.

McNeil MR, Pratt SR. Defining aphasia: Some theoretical and clinical implications of operating from a formal definition. Aphasiology. 2001; 15(10):901–911. doi:10.1080/02687040143000276.

McNeil MR, Sung JE, Yang D, Pratt SR, Fossett TRD, Doyle PJ, et al. Comparing connected language elicitation procedures in persons with aphasia: Concurrent validation of the Story Retell Procedure. Aphasiology. 2007; 21(6–8):775–790. doi:10.1080/02687030701189980.

Menn L, Reilly KF, Hayashi M, Kamio A, Fujita I, Sasanuma S. The interaction of preserved pragmatics and impaired syntax in Japanese and English aphasic speech. Brain and Language. 1998; 61:183–225. [PubMed: 9468771]

Murray, L. Narrative differences among elderly with clinical depression versus early Alzheimer's. Poster session presented at the Convention of the American Speech-Language-Hearing Association; New Orleans, LA, USA: Nov. 2009

Nicholas LE, Brookshire RH. A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. Journal of Speech & Hearing Research. 1993; 36(2):338–350. [PubMed: 8487525]

Olness GS. Genre, verb, and coherence in picture-elicited discourse of adults with aphasia. Aphasiology. 2006; 20(2–4):175–187. doi:10.1080/02687030500472710.

Prins R, Bastiaanse R. Analysing the spontaneous speech of aphasic speakers. Aphasiology. 2004; 18(12):1075–1091. doi:10.1080/02687030444000534.

Prins RS, Snow CE, Wagenaar E. Recovery from aphasia: Spontaneous speech versus language comprehension. Brain and Language. 1978; 6(2):192–211. doi:10.1016/0093-934X(78)90058-5. [PubMed: 728786]

Rabovskya M, Álvarez CJ, Hohlfeld A, Sommer W. Is lexical access autonomous? Evidence from combining overlapping tasks with recording event-related brain potentials. Brain Research. 2008; 1222:156–165. [PubMed: 18585684]

Rider JD, Wright HH, Marshall RC, Page JL. Using semantic feature analysis to improve contextual discourse in adults with aphasia. American Journal of Speech-Language Pathology. 2008; 17(2): 161–172. doi:10.1044/1058-0360(2008/016). [PubMed: 18448603]

Rubin, C.; Newton, E. Capture the moment: The Pulitzer Prize photographs. W. W. Norton; New York, NY: 2001.

Schwartz MF, Dell GS, Martin N, Gahl S, Sobel P. A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. Journal of Memory and Language. 2006; 54:228–264.

Tweedie FJ, Baayen RH. How variable may a constant be? Measures of lexical richness in perspective. Computers and the Humanities. 1998; 32(5):323–352.

Tyler LK, Wessels J. Quantifying contextual contributions to word recognition processes. Perception and Psychophysics. 1983; 34:409–420. [PubMed: 6657445]

Ulatowska, HK.; Allard, L.; Bond Chapman, S. Narrative and procedural discourse in aphasia.. In: Joanette, Y.; Brownell, HH., editors. Discourse ability and brain damage: Theoretical and empirical perspectives. Springer-Verlag; New York, NY: 1990. p. 180-198.

Ulatowska HK, North AJ, Macaluso-Haynes S. Production of narrative and procedural discourse in aphasia. Brain and Language. 1981; 13(2):345–371. doi:10.1016/0093-934X(81)90100-0. [PubMed: 7260577]

Vermeer A. Coming to grips with lexical richness in spontaneous speech data. Language Testing. 2000; 17(1):65–83. doi:10.1191/026553200676636328.

Wachal RS, Spreen O. Some measures of lexical diversity in aphasic and normal language performance. Language and Speech. 1973; 16(2):169–181. [PubMed: 4762596]

Wright HH, Capilouto GJ. Manipulating task instructions to change narrative discourse performance. Aphasiology. 2009; 23(10):1295–1308. doi:10.1080/02687030902826844.

Wright HH, Capilouto GJ. Measuring age-related changes in discourse production. 2011 Unpublished raw data.

Wright HH, Silverman SW, Newhoff M. Measures of lexical diversity in aphasia. Aphasiology. 2003; 17(5):443–452. doi:10.1080/02687030344000166.

$watermark-text

$watermark-text

$watermark-text

$watermark-text
$watermark-text
$watermark-text

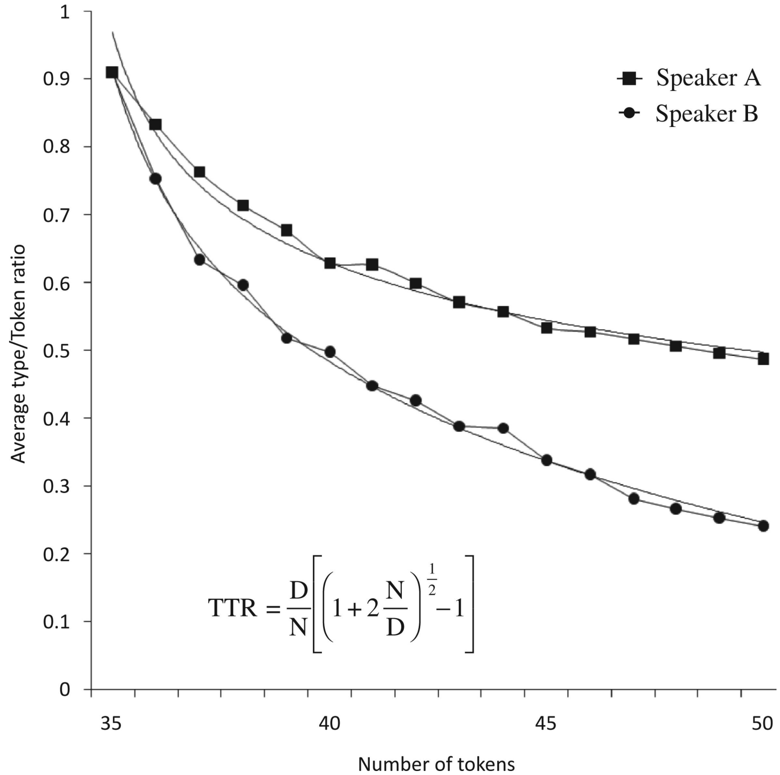


**Figure 1.**
Estimating *D*. Average type–token ratio (TTR) decreases as a function of sample length (N) for two speakers. Speaker A uses more diverse vocabulary (i.e., uses the same words less often) and his/her average TTR decreases at a slow pace. Speaker B uses the same words repeatedly and her/his TTR decreases faster as a function of the sample size. voc-*D* captures a speaker's lexical diversity by modelling how fast the average TTR decreases. The slope of the fitted nonlinear curve corresponds to different *D* values. The steeper the slope of the fitted line, the lower the *D* value. Selection of the best-fitting curve is determined using a least squares approach.

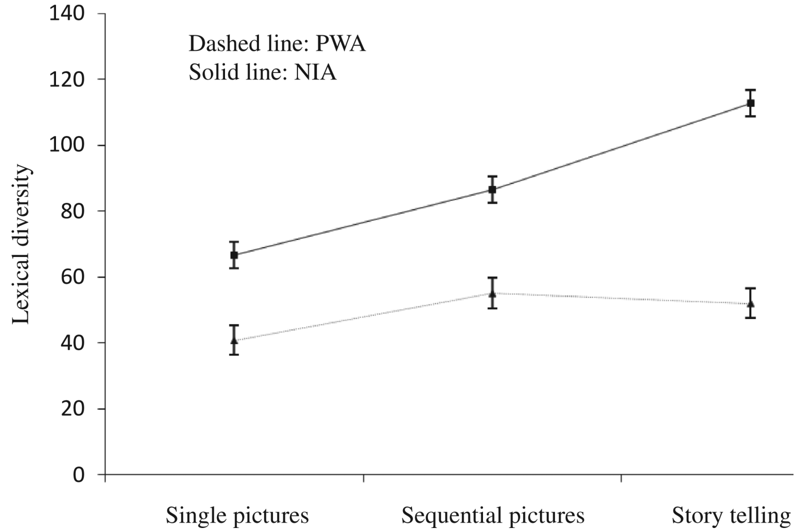


**Figure 2.**
Mean lexical diversity for each elicitation task within each group.

$watermark-text
$watermark-text
$watermark-text

$watermark-text

$watermark-text

$watermark-text

## TABLE 1

Participant characteristics

| | PWA group (N = 27) | NIA group (N = 27) |
|---|---|---|
| Age (yrs) | *M* = 66.81 (*SD = 12.83*) | *M* = 67.45 (*SD* = *13.20*) |
| Gender | 19 males, 8 females | 19 males, 8 females |
| Education | *M* = 15.72 (*SD* = *3.72*) | *M* = 15.02 (*SD* = *3.56*) |
| WAB-R[1] Aphasia type | Anomic = 19 | |
| Type | Conduction = 8 | |
| WAB-R AQ[2] | *M* = 76.82 (*SD* = *9.83*) | |
| BNT[3] | *M* = 7.45 (*SD* = *4.25*) | |

[1]Western Aphasia Battery – Revised

[2]Western Aphasia Battery – Revised Aphasia Quotient

[3]Boston Naming Test.

$watermark-text

$watermark-text

$watermark-text

**TABLE 2**

Mean number of content words for each elicitation task across groups

| | PWA group | NIA group |
|---|---|---|
| Single pictures | 76.33 (*SD* = 10.85) | 119.78 (*SD* = 9.64) |
| Sequential pictures | 68.89 (*SD* = 6.04) | 119.48 (*SD* = 8.29) |
| Story telling | 124.15 (*SD* = 15.59) | 289.40 (*SD* = 29.40) |

$watermark-text $watermark-text $watermark-text

## TABLE 3

Mean lexical diversity scores and standard errors (SE) for PWA and NIA and group differences for each discourse elicitation task

| Discourse task | PWA group | | NIA group | | | |
|---|---|---|---|---|---|---|
| | M | SE | M | SE | t(50) | p |
| Single pictures | 40.81 | 3.20 | 66.65 | 4.03 | –4.96 | .000* |
| Sequential pictures | 55.10 | 4.33 | 86.48 | 4.17 | –5.22 | .000* |
| Story telling | 52.07 | 4.94 | 112.53 | 4.75 | –8.82 | .000* |

To control for Type I Error, *alpha* was set to .0125 for each simple effect.

* $p < .0125$.

$watermark-text

$watermark-text

$watermark-text



## TABLE 4

Mean lexical diversity differences and standard errors (SE) and comparisons among discourse elicitation task for each group

| Comparison | Mean Δ | SE | t(24)[a] | p |
|---|---|---|---|---|
| PWA | | | | |
| Single vs Sequential pictures | –14.29 | 4.22 | –3.38 | .002 * |
| Single pictures vs Story telling | –11.26 | 3.91 | –2.08 | .008 * |
| Sequential pictures vs Story telling | 3.03 | 4.76 | .64 | .53 |
| NIA | | | | |
| Single vs Sequential pictures | –19.83 | 5.69 | –3.49 | .002 * |
| Single pictures vs Story telling | –15.87 | 5.75 | –7.99 | .000 * |
| Sequential pictures vs Story telling | –26.05 | 5.38 | –4.84 | .000 * |

To control for Type I Error, *alpha* was set to .008 for each simple effect.

[a]Paired sample *t* was conducted with 26 *df* for the NIA group.

[*] $p < .008$.