## INTRODUCTION

This paper presents research that aimed to extend the available analyses of informativeness of aphasic discourse. A ‘proposition’ can be defined as a linguistic relation and its associated arguments (Kintsch & Keenan, 1973; Turner & Greene, 1977), and has been used as an index of informativeness in research on language and aging. The proportion of propositions in a text (Propositional Idea Density – PD) has been found to be a sensitive index of age-associated cognitive impairment and dementia (Riley, Snowdon, Desrosiers, & Markesbery, 2005). The research on PD has primarily used manual analysis methods, noting high training needs for raters to ensure adequate inter-coder and intra-coder reliability, as has also been found in analyses of informativeness in the field of aphasia (Nicholas & Brookshire, 1993; Oelschlaeger & Thorne, 1999; Yorkston & Beukelman, 1980). The development of a computer program, Computerized Propositional Idea Density Rater known as CPIDR (Brown, Snodgrass, & Covington, 2007; Brown, Snodgrass, Kemper, Herman, & Covington, 2008) has made the process of calculating PD accessible to untrained individuals. The benefits of a computer-based program are further seen in reliability, with 100% consistency when re-counting a single sample, and inter-rater reliability of 97% when compared to manual calculations which is more reliable than most human coders (Brown, et al., 2008). The present research made use of this computerised analysis of PD to investigate the effects of aphasia on informativeness. It was hypothesised that information content, as measured by PD, would be significantly reduced in the oral discourse of people with aphasia when compared to non-aphasic controls, and that PD would decrease with increasing aphasia severity as determined by Western Aphasia Battery - Aphasia Quotient (Kertesz, 2006).

# METHOD

### Description of data and linguistic analyses

De-identified transcriptions of separately conducted interviews with 50 individuals with aphasia and 49 family members (used as non-brain damaged matched controls for the purposes of the present study) from the (REMOVED FOR DE-IDENTIFICATION) were analyzed for the purposes of the present research. Transcripts were stripped of all interviewer data, leaving only conversational contributions made by the participants, with the average text size of 2,831 words (PWA) and 5,138 words (Controls). (Refer to Table 1). These formatted transcripts were analyzed using CPIDR version 3.2 (Brown, et al., 2007; Brown, et al., 2008) for the analysis of PD. Systematic Analysis of Language Transcripts (SALT Version 8, Miller, 2003) was used for the analysis of Type Token Ratio (TTR), Number of Different Words (NDW), Mean Length of Utterance (MLU) and Number of Utterances (NU) in order to compare findings for PD against those obtained through established measures.

### Statistical analyses

All statistical analyses were conducted using SPSS (version 20) for Windows (IBM, 2011) Between groups analysis was performed to determine if PD and other discourse measures differed significantly between participants with aphasia and non-aphasic control participants. A series of paired-samples t-tests were conducted, and as the data did not conform to the assumed normal distribution, results were confirmed with non-parametric Wilcoxon signed rank tests (Story, 2004). Spearman correlations were used to analyze the relationship between aphasia severity and PD (Story, 2004). Graphic analyses and curve estimation regression analyses were applied to significant results to define the nature of the relationship between aphasia severity and the discourse measures.

## RESULTS

The presence of aphasia had a significant impact on the information content of language, as measured by PD ($p < .001$). The concurrent validity of these findings was supported by established discourse measures of NDW and MLU. However, TTR was found to be significantly higher for the aphasic participants (see Discussion).There was no significant difference in overall NU. Refer to Table 2.

Univariate correlations were performed to analyze the relationships between discourse measures and age, gender, and location (for all participants); as well as level of education, and time post onset for participants with aphasia. Due to the number of tests performed, a Bonferroni adjustment was applied to reduce the risk of type 1 error. Accordingly, results with a $p$ value < .001 were considered significant. None of the potentially confounding variables had a significant effect on PD or other discourse measures.

Spearman correlation was conducted to determine if a relationship existed between aphasia severity and PD. This non-parametric correlation analysis was used as the data was not normally distributed. As shown by the results in Table 3, PD was shown to have a significant positive correlation with aphasia severity ($p < .001$) in that as severity of aphasia increased, PD decreased. This effect was supported when Spearman correlations were used to compare aphasia severity to the other discourse measures. Significant relationships were established with the measures of NDW and MLU. Each of these were indicated to be significant at the $p$<.001 level. NDW and MLU were positively correlated with aphasia severity, with the measures decreasing as aphasia increased in severity.

With significant relationships determined, graphic analyses were applied to define the nature of the relationships between aphasia severity and significant discourse measures. Scatter graphs were generated and a non-parametric smoother was used to visually represent the correlation. For PD, the graphic representation indicated that the relationship with aphasia severity was not linear. For more information, a curve estimation regression analysis indicated a significant quadratic relationship ($p$<.05). When applied to the graphic analysis, the quadratic curve accounted for the greatest amount of variation in the data (R-square =.453). Graphic analyses of NDW and MLU indicated linear relationships with aphasia severity, as measured by the WAB-AQ. These were confirmed using curve estimation regression analyses, with both indicating significance to $p$<.001. The linear model for NDW accounted for 49% of variation in the data (R-square = .490), while that for MLU explained 25.6% of variation (R-square = .256).

## DISCUSSION

The findings of this research supported the hypothesis that the presence of aphasia would result in a decrease in PD. The severity of aphasia was shown to significantly decrease informativeness as measured by PD. However, the nature of the correlation, best represented by a quadratic model, indicated that mild aphasia had relatively little impact on the PD of conversational discourse. The validity of the effect of aphasia on PD was confirmed when compared to other automated discourse measures. Where aphasia was present, a reduction was seen in the lexical diversity, as measured by NDW which was consistent with previous research (Wright, Silverman, & Newhoff, 2003). However, the other measure of vocabulary, TTR, provided an unexpected result, showing an increase where aphasia was present. Given the well-recognised confounds for TTR in relation to sample size (Wright, et al., 2003), the present findings can be queried as being associated with the very large sample sizes of aphasic language data in the present research in comparison with previous research.

Further research is needed to compare these results to language samples collected from a wider range of communicative contexts (partners, settings, genres) and could usefully consider differences in type of aphasia. This paper will discuss the potential and limitations associated with the clinical application of this measure of informativeness in aphasia.

**Acknowledgements**: (REMOVED FOR DE-IDENTIFICATION OF PAPER)

## REFERENCES


Brown, C., Snodgrass, T., & Covington, M. (2007). *Computerized Propositional Idea Density Rater 3 (CPIDR 3).* CASPR Project, Artificial Intelligence Centre, University of Georgia.

Brown, C., Snodgrass, T., Kemper, S., Herman, R., & Covington, M. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods, 40*(2), 540-545.

IBM. (2011). IBM SPSS Statistics for Windows (Version 20.0). Armonk, NY: IBM Corp.

Kertesz, A. (2006). *Western Aphasia Battery - Revised.* San Antonio, TX: Harcourt Assessment.

Kintsch, W., & Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology, 5*(3), 257-274.

Miller, J. F. (2003). Systematic Analysis of Language Transcripts - SALT (Version V8.0). Madison, WI: Language Analysis Laboratory, University of Wisconsin-Madison.

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 & Hearing Research, 36*(2), 338-350.

Oelschlaeger, M. L., & Thorne, J. C. (1999). Application of the correct information unit analysis to the naturally occurring conversation of a person with aphasia. *Journal of Speech, Language, and Hearing Research, 42*, 636-648.

Riley, K. P., Snowdon, D., Desrosiers, M. F., & Markesbery, W. R. (2005). Early life linguistic ability, late life cognitive function, and neuropathology: findings form the Nun Study. *Neurobiology of Aging, 26*(3), 341-347.

Story, I. (2004). A brief summary of univariate group comparison procedures. In G. Minichello, K. Sullivan, K. Greenwood & R. Axford (Eds.), *Handbook of research methods for nursing and health science* (2nd ed., pp. 543-573). Frenchs Forrest: Pearson Education Australia.

Turner, A., & Greene, E. (1977). *The construction and use of a propositional text base.* Boulder, CO: University of Colorado Psychology Department (accessed 22 Nov 2012, from The Institute of Cognitive Science, University of Colorado Boulder, http://www.colorado.edu/ics/sites/default/files/attached-files/77-63.pdf ).

Wright, H. H., Silverman, S. W., & Newhoff, M. (2003). Measures of lexical diversity in aphasia. *Aphasiology, 17*(5), 445-452.

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


## TABLES

Table 1

*Information about participants and texts*

| *Partici-pants* | *Gender* | | *Age (years)* | | | *Level of Education** | | | *Text size in words* |
|---|---|---|---|---|---|---|---|---|---|
| | Male | Female | 32-59 | 60-69 | 70 + | 7-10 years | 11-12 years | Tertiary | Mean (Range) |
| With aphasia (n=50) | 30 | 20 | 14 | 20 | 16 | 15 | 12 | 7 | 2,831 (103 - 6,484) |
| Controls (n=49) | 13 | 37 | 19 | 18 | 12 | na | na | na | 5,138 (1,780 – 6,533) |

Note: Education level information was not available for 16 participants with aphasia.

Table 2

Comparison of results for discourse measures

| | ***With aphasia (n=50)*** | ***Controls (n=49)*** | ***Level of Significance*** |
|---|---|---|---|
| **Propositional Idea Density (PD)** | | | |
| mean (SD) | 0.455 (0.12) | 0.547 (0.015) | ** |
| minimum | 0.009 | 0.505 | |
| maximum | 0.582 | 0.573 | |
| **Number of Different** | | | |
| **Words (NDW)** | 448.44 (306.745) | 933.67 (277.223) | ** |
| mean (SD) | 10 | 423 | |
| minimum | 1225 | 1520 | |
| maximum | | | |
| **Type Token Ratio (TTR)** | | | |
| mean (SD) | 0.178 (0.08) | 0.148 (0.036) | * |
| minimum | 0.02 | 0.09 | |
| maximum | 0.41 | 0.25 | |
| **Mean Length of Utterance in Words (MLU)** | | | |
| mean (SD) | 8.036 (7.211) | 22.777 (10.887) | ** |
| minimum | 1.19 | 7.88 | |
| maximum | 39.19 | 52.03 | |
| **Number of Utterances (NU)** | | | |
| mean (SD) | 381.6 (153.486) | 342.04 (185.292) | |
| minimum | 57 | 86 | |
| maximum | 1161 | 950 | |
| * Significant at p<.05; ** Significant at p<.001, Wilcoxon signed-rank test | | | |

Table 3
Correlations between aphasia severity and language measures

| **Measures** | **Correlation with aphasia severity r** | **Level of significance** |
|---|---|---|
| Propositional Idea Density (PD) | .475 | ** |
| Number of Different Words (NDW) | .696 | ** |
| Type Token Ratio (TTR) | .030 | |
| Mean Length of Utterance in Words (MLU) | .659 | ** |
| Number of Utterances (NU) | .045 | |

** indicates significance at $p<.001$, Spearman correlation