Propositional Idea Density in written descriptions of health: Potential clinical applications

INTRODUCTION

In order to assess the effect of word finding difficulties for the spontaneous discourse of people with aphasia, a number of different measures of informativeness have been developed for clinical application (Doyle, Goda, & Spencer, 1995; Nicholas & Brookshire, 1993; Oelschlaeger & Thorne, 1999; Wright, Silverman, & Newhoff, 2003). The main challenges for the assessment of discourse (written or spoken) relate to issues of validity and reliability (AUTHOR DELETED). There is a need for valid and authentic sampling which is personally relevant to individuals and additionally, able to be repeated for the same individual on successive occasions, and comparable to other individuals. The use of a consistent elicitation task that could be widely used for adult populations would be beneficial to both allow comparisons of the same individual over time and also across individuals.

Also, there is limited normative data to assist in the interpretation of linguistic measures from adult language samples. As part of interpreting the clinical significance of the findings from the assessment of written discourse, clinicians need to have access to evidence of the expected range and variation of performance in individuals without brain damage (Bromley, 1991) particularly given the research suggesting age-related decline in healthy older adults (e.g., Kemper & Sumner, 2001; Ritchie, Artero, & Touchon, 2001). See Table 1 for a comparison of the number of non-brain-damaged participants and text size from selected studies of informativeness.

Additionally, analyses of language data needs to have high intra- and inter-analyst reliability to increase confidence in the significance of any changes observed with recovery or in response to treatment. In order to be clinically feasible analyses must be low cost, quick, and have minimal training requirements. (AUTHOR DELETED) have explored the use of a freely available computerized program for analyzing Propositional Idea Density (PD) for participants in the Australian Longitudinal Study on Women's Health (ALSWH) - a joint project between the University of Newcastle, Australia, and the University of Queensland (Lee et al., 2005). In repeated surveys over 12 years, the 1973-78 cohort provided over 8,000 written comments, the 1946-51 cohort provided over 12,000 written comments, and the 1921-26 cohort provided over 17,000 written comments, with a total word count for the written corpus of over 3 million words. Written comments by proxies were reported for 141 comments for the 1921-26 cohort (from surveys 3, 4, 5) and these comments were excluded from the analysis. A total of 37,853 written responses of 10 or more words were analyzed. Overall, this research demonstrated that PD is stable across age ranges with a very small decline in late old age (see Figure 1).

The present paper presents the results of a computerized method to measure informativeness, CPIDR 3 for Propositional Idea Density, in longitudinal written data obtained from repeated sampling for five older participants from the ALSWH study. The method and results will be discussed to highlight clinical relevance and potential of the method, and with reference to the ongoing development of a large normative reference set.

METHOD

Participants

The data for this paper focuses on five participants in the ALSWH project born between 1921-26 who provided written survey responses on each of five survey occasions every three years. The five participants were randomly selected from amongst the women who responded to all five surveys with responses of at least 35 words in length over the 12 year period. All

five participants were aged between 70-71 years at the time of the first survey, and between 82-83 at the time of the fifth survey.

Data collection

In this study, as part of the survey, the women respondents were invited to respond to one open ended question at the end of the survey:

"If there is ANYTHING else you would like to tell us about changes in your health (especially in the LAST THREE YEARS) please write on the lines below."

The question remained the same on each survey and therefore provided repeated elicitation of written language within a natural context.

Data analysis

Computerized analysis was employed to determine Propositional Idea Density (PD) using the Computerized Propositional Idea Density Rater (CIPDR3 version 3.2.2785.24603). CPIDR3 applies a part-of-speech tagger to determine propositions based on the definitions of propositions developed by Turner and Green (1977) and has been used in previous research (e.g. Engelman et al., 2010). It has been shown to have 100% intra-rater and 97% inter-rater reliability when compared with human raters (Brown, Snodgrass, Kemper, Herman, & Covington, 2008).

RESULTS

Results to date from the analysis of the written discourse from the five participants demonstrated the stability of the measure across the survey periods from early old age to later old age. The mean PD scores for each survey period for the five participants ranged from 0.508 (5.08 per 10 words) to 0.556 (5.56 per 10 words). The average PD between the participants at survey period 5 was 0.562 (5.62 per 10 words) with a range 0.623 (6.23 per 10 words) to 0.527 (5.27 per 10 words). Overall, the range of PD across the survey periods was 0.615 (6.15 per 10 words) to 0.441 (4.41 per 10 words). See Table 2 and Figure 2 for individual participant's measures of PD.

DISCUSSION

The elicitation question for the written comments in this research is suggested to provide an authentic personal question of relevance to the clinical population, which could be used for test re-test purposes and so to offer a valid alternative to picture description or story retell tasks in both verbal and written modes. The use of CPIDR in analyzing PD is suggested to offer promise as a clinical tool, given that it offers a freely available, reliable, easy to use computerized analysis of linguistic features highly relevant to aphasia. The five case studies presented in this paper illustrate the longitudinal application of PD for individuals' language over time, with findings that reflected the large normative reference sample. The potential of this measure as a possible barometer of linguistic health and therefore cognitive health is considered to be of relevance to those working with normal ageing and language disordered populations. Further research is ongoing to continue the development of a reference set to include other population groups (e.g., men, acquired language disorders of varying etiologies and severity, as well as to continue the present research program in establishing concurrent validity of the measure.

REFERENCES

Bromley, D. B. (1991). Aspects of written language production over adult life. *Psychology & Aging, 6*(2), 296-308.

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.

Doyle, P. J., Goda, A. J., & Spencer, K. A. (1995). The communicative informativeness and efficiency of connected discourse by adults with aphasia under structured and conversational sampling conditions. *American Journal of Speech-Language Pathology, 4*, 130-134.

Kemper, S., & Sumner, A. (2001). The structure of verbal abilities in younger and older adults. *Psychology & Aging, 16*(2), 312-322.

Lee, C., Dobson, A. J., Brown, W. J., Bryson, L., Byles, J., Warner-Smith, P., & Young, A. F. (2005). Cohort profile: The Australian Longitudinal Study on Women's Health. *International Journal of Epidemiology, 34*(5), 987-991.

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.

Ritchie, K., Artero, S., & Touchon, J. (2001). Classification criteria for mild cognitive impairment: A population-based validation study. *Neurology, 56*(1), 37-42.

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.

Table 1. Comparison of number of participants and text size for selected studies of informativeness of discourse

Study	Year	PWA n	NBD n	Informative-ness Measure	Text type	PWA Text size - words	NBD Text size - words
AUTHOR DELETED	(sub)	50	49	Propositional Idea Density	Interview	Mean 2,831 Range 103 - 6,454	Mean 5,138 Range 1,780 - 6,533
AUTHOR DELETED	2012	NI	127	Propositional Idea Density	Description of health (written)	NI	Mean 118 Range 35 - 574
Wright et al.	2003	18	NI	Lexical Diversity	Picture description, conversation	Range 208 - 655	NI
Oelshlaeger & Thorne	1999	1	NI	Correct Information Unit	Natural conversation	Mean 1,014 Range 272 - 1,595	NI
Doyle, Goda & Spencer	1995	20	NI	Correct Information Unit	Picture description & elicitation (multiple)	Mean 748 Range 92 - 2,609	NI
					Conversation	Mean 540 Range 171 - 1,408	NI
Nicholas & Brookshire	1993	20	20	Correct Information Unit	Picture description & elicitation (multiple)	Means 73, 77, 78 Range 30 - 160	Means 101, 104, 113 Range 59 - 176

Key: PWA – Person with aphasia; NBD – non brain-damaged

Table 2. Means and ranges for linguistic measures for five ALSWH participants across five surveys

Measure		N1	N2	N3	N4	N5
Response Length (clauses)	Mean Range	30.6 24 - 36	16.8 10 - 23	18.8 5 - 26	20 13 - 25	20 11 - 29
Length of Utterance (words)	Mean Range	20 19 - 21	14 11 - 20	16.19 13 - 21	13.61 11 - 15	20.31 16 - 24
Number of Different Words (as % of total words)	Mean Range	65 62 - 70	75.8 66 - 83	68.2 60 - 85	69.6 62 - 79	70.20 64 - 80
Propositional Idea Density (PD per 10 words)	Mean Range	0.554 (5.44) 0.5-0.615	0.556 (5.56) 0.545-0.561	0.545 (5.45) 0.441-0.623	0.508 (5.08) 0.463-0.559	0.537 (5.37) 0.509-0.565

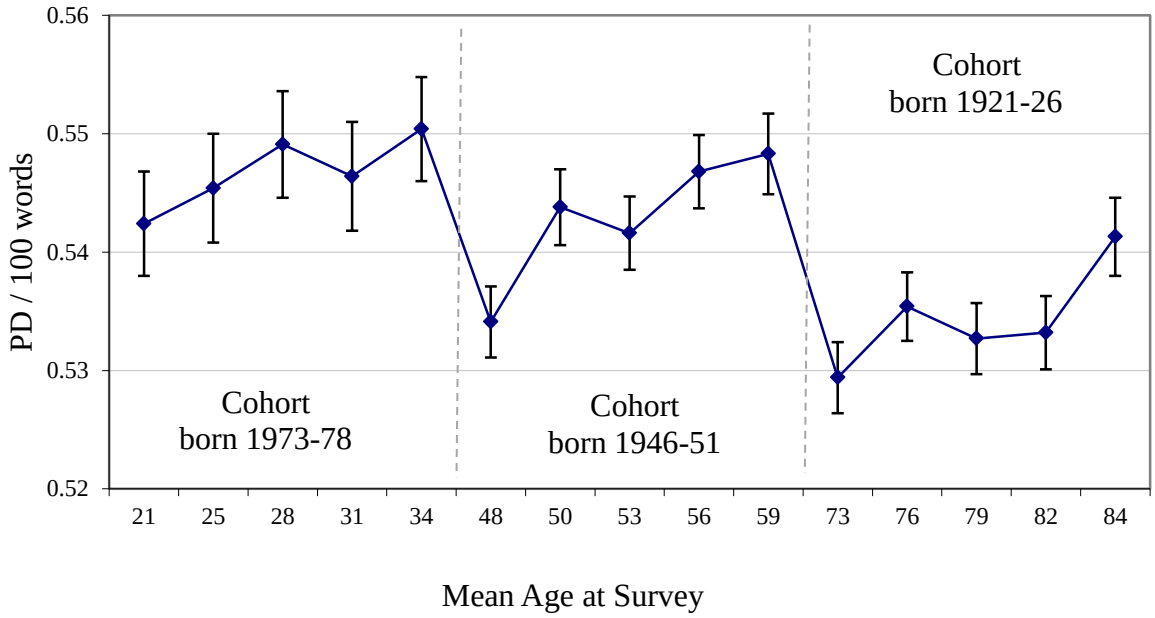

Figure 1. Measures of Propositional Idea Density (PD) for three age cohorts of ALSWH participants across 5 survey periods (n = 19,512) - (Fitted LMM model)

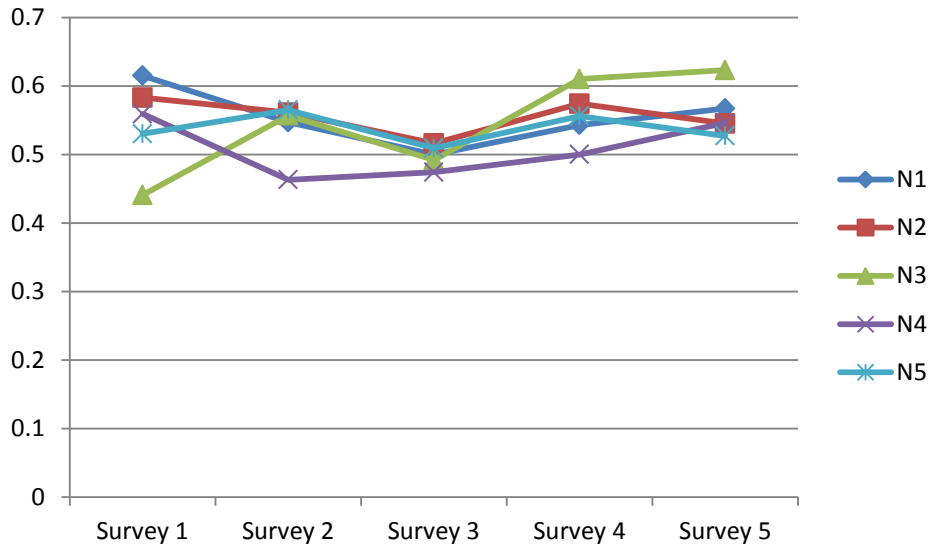

Figure 2. Propositional Idea Density for five participants from the ALSWH 1921-26 cohort