CORTEX XXX (2013) 1–15

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Special issue: Research report

Propositional Idea Density in women's written language over the lifespan: Computerized analysis

Alison Ferguson [a,b,*], *Elizabeth Spencer* [a], *Hugh Craig* [a,c,d] *and Kim Colyvas* [e]

[a] *Faculty of Education and Arts, University of Newcastle, Australia*
[b] *CCRE in Aphasia Rehabilitation, Australia*
[c] *Centre for Linguistic and Literary Computing, University of Newcastle, Australia*
[d] *Humanities Research Institute, University of Newcastle, Australia*
[e] *Faculty of Science and Information Technology, University of Newcastle, Australia*

ARTICLE INFO

Article history:
Received 30 December 2012
Reviewed 25 March 2013
Revised 6 May 2013
Accepted 30 May 2013
Published online xxx

Keywords:
Language decline
Aging
Propositional Density
Idea Density
Corpus linguistics

ABSTRACT

The informativeness of written language, as measured by Propositional Idea Density (PD), has been shown to be a sensitive predictive index of language decline with age and dementia in previous research. The present study investigated the influence of age and education on the written language of three large cohorts of women from the general community, born between 1973 and 1978, 1946–51 and 1921–26. Written texts were obtained from the Australian Longitudinal Study on Women's Health in which participants were invited to respond to an open-ended question about their health. The informativeness of written comments of 10 words or more (90% of the total number of comments) was analyzed using the Computerized Propositional Idea Density Rater 3 (CPIDR-3). Over 2.5 million words used in 37,705 written responses from 19,512 respondents were analyzed. Based on a linear mixed model approach to statistical analysis with adjustment for several factors including number of comments per respondent and number of words per comment, a small but statistically significant effect of age was identified for the older cohort with mean age 78 years. The mean PD per word for this cohort was lower than the younger and mid-aged cohorts with mean age 27 and 53 years respectively, with mean reduction in PD 95% confidence interval (CI) of .006 (.003, .008) and .009 (.008, .011) respectively. This suggests that PD for this population of women was relatively more stable over the adult lifespan than has been reported previously even in late old age. There was no statistically significant effect of education level. Computerized analyses were found to greatly facilitate the study of informativeness of this large corpus of written language. Directions for further research are discussed in relation to the need for extended investigation of the variability of the measure for potential application to the identification of acquired language pathologies.

* *Corresponding author*. Speech Pathology Discipline, Faculty of Education and Arts, University of Newcastle, Callaghan, NSW 2308, Australia.
E-mail address: Alison.Ferguson@newcastle.edu.au (A. Ferguson).
0010-9452/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cortex.2013.05.012

1. Introduction

As the proportion of older adults in the community increases, there has been an escalation of research investigating the changes associated with healthy aging and age-related pathology (Abrams and Farrell, 2011; de Bot and van der Hoeven, 2011). Oral and written language of older adults has been proposed as a sensitive early barometer of compromised cognitive health as well as a potential indicator of cognitive reserve (Christensen et al., 2007). Researchers with a primary interest in language as reflective and predictive of cognition have tended to work with extended spontaneous or elicited written language samples. The present research has been informed by this approach and seeks to make a contribution to the emerging evidence regarding language change across adulthood through the investigation of written language produced by a large cohort of over 19,000 women aged between 18 and 90 years of age.

The work of a number of key researchers in the humanities has informed the development and application of linguistic measures that provide indices of language change over the lifespan of well-known writers (Lancashire, 2010), including Iris Murdoch and Agatha Christie (e.g., Garrard et al., 2005; Lancashire and Hirst, 2009; Le, 2010; Le et al., 2011) and political figures (Garrard, 2009). For example, Williams et al. (2003) studied 57 letters of King James VI/I written over a twenty year period from when he was 38–58 years (shortly before his death). They observed that the decline in linguistic measures of complexity observed from the age of 53 years may have been related to historically known health and cognitive problems that arose at that time, and they considered these findings to be suggestive of a diagnosis of vascular dementia.

A common concern in both literary and clinically focused studies of aging is the recognition of the need to contextualize identification of potential pathology within a wider understanding of bio-psycho-social constructs of aging (Duszak and Okulska, 2011). In particular, such notions as healthy aging and typical ('normal') language performance are slippery terms, often defined by the absence of overt pathology. For example, Ritchie et al. (2001) discussed the difficulties in distinguishing mild cognitive impairment (MCI) which is seen as a subtle early indicator of future decline associated with dementia, and what they term as 'age-associated cognitive decline', which is seen as essentially benign rather than suggestive of future dementia. As studies of language and aging are most directed to the detection of subtle and subclinical indicators of potential cognitive decline, there is a constant need for recognition of the variability and context-bound nature of language use (Labov and Auger, 1993). The two methodological strategies to manage the inherent ambiguities in this area are the adoption of longitudinal research design and/or the incorporation of large numbers of participants, whose characteristics are either well defined and/or as representative as possible of the wider population. One of the pioneering research programs in this area has been the Nun Study (Snowdon, 2001; Snowdon et al., 1996; Tyas et al., 2007) and the research related to this study will be discussed in more detail in this Introduction. The present research adopts both a longitudinal and large cohort research strategy in order to contribute to this area.

The other common feature of literary and clinically focused studies is their emphasis on written language. There are a few studies that have investigated the differences in oral and written language associated with aging through direct comparison of these modes of discourse. Croisile et al. (1996) compared oral and written descriptions of a picture commonly used for aphasia testing ('Cookie Theft' picture – Goodglass and Kaplan, 1983) by 22 men and women with probable Alzheimer's disease (AD) and 24 matched controls. Participants with AD produced fewer words across all word types, fewer of the predetermined 'content units', and showed reduced grammatical complexity (in terms of fewer complete clauses). Their study found no statistically significant difference between the task modes for either group in the inclusion of predetermined 'content units'. However, the increased presence of semantic intrusions and grammatical errors in the written task prompted the researchers to conclude that the written mode appeared to be better suited to identify the linguistic difficulties present in AD. Mitzner and Kemper (2003) compared oral and written autobiographical accounts from 118 women in the Nun Study in 1995–6 when they were aged between 78 and 91 years. They found a stronger relationship between cognition [as measured by the Mini-Mental State Examination (MMSE) – Folstein et al., 1975] and the linguistic measures for written discourse. They suggested written discourse was potentially more sensitive in detecting cognitive status. The present research is similarly focused on written discourse.

It has been suggested that language decline with healthy aging and with identified cognitive pathologies such as dementia relates strongly to working memory function (Kemper, 2012; Kemper et al., 1989; Wingfield and Grossman, 2006; Wingfield et al., 2000). Kemper and Sumner (2001) compared a range of linguistic measures from the analysis of short oral personally relevant descriptions such as description of an influential person, or an interesting experience, with tests of cognitive function for 100 younger healthy adults (aged between 18 and 28 years) and 100 older healthy adults (aged between 63 and 88 years). They found the expected differences between older and young participants for the measures, but of more interest were their findings (for both older and younger groups) that the measures of working memory (digits forward and backward) covaried with Type Token Ratio (TTR – as an index of lexical diversity) and D-Level (a rating scale of clausal complexity), while Mean Length of Utterance in words (MLU) and PD covaried with measures of verbal fluency and reading rate. They interpreted these findings for D-Level as being related to the need for working memory to manage complex clausal embedding. They interpreted the findings for PD as reflecting processing efficiency and discussed the logical relevance of this suggestion, given Kintsch's research suggesting propositions as the basic unit for memory for textual comprehension and given the relationship between number of propositions and reading rate (Kintsch and Keenan, 1973).

PD is one amongst a number of measures applied to the analysis of information content. The informativeness of written language, as measured by PD, has been shown to be a sensitive and predictive index of language decline with age and dementia in a range of well-known longitudinal studies

Please cite this article in press as: Ferguson A, et al., Propositional Idea Density in women's written language over the lifespan: Computerized analysis, Cortex (2013), http://dx.doi.org/10.1016/j.cortex.2013.05.012