

A Computerized Measure of Idea Density in Two Genres of Written Language Samples



Katie Moser, B.A. & Suzanne Hungerford, Ph.D., CCC-SLP | Plattsburgh State University of New York

Purpose and Research Questions

The purpose of this study is to investigate idea density (ID) in written language samples of normal adults, using the Computerized Propositional Idea Density Rater (CPIDR) 5.1 (Brown et al., 2012).

The following questions are explored through this study:

- 1) Does idea density differ depending on the type of language sample elicited (expository vs. narrative)?
- 2) Is idea density, when calculated from the narrative or expository language genre, correlated with of any common measure(s) of academic success (e.g., GPA, SAT, ACT scores) or vocabulary knowledge?

Background

- The analysis of language samples in the field of speech-language pathology is routinely used to detect language pathology in a wide range of populations. Idea (or proposition) density (ID) is a language sample measure rarely used in speech-language pathology.
- Propositional analysis that yields ID has been developed out of psycholinguistic research and now can be calculated using a computer program known as CPIDR 5.1 (Brown et al., 2012). A proposition is defined as a main verb and all its arguments as well as additional descriptive elements within the text (Brown et al., 2008). Each proposition can be found as true or false apart from the others (Brown et al., 2008). ID is a measure of propositions expressed relative to the number of words used (Kemper & Sumner, 2001), and is therefore seen as a measure of economy of expression (Kemper & Sumner, 2001) and language complexity (Spencer, Craig, Ferguson, & Colyvas, 2012).
- ID has been shown to have predictive validity regarding cognitive decline in aging (Snowden et al., 1996; Engelman, Agree, Meoni, & Klag, 2010; Farias et. al 2012), and to have validity as a measure of linguistic ability across the lifespan (Kemper, Greiner, Marquis, Prenevost, & Mitsner, 2001). Research has also shown that high ID early in life is correlated with intact cognitive ability in late life, even in the presence of Alzheimer's-related brain lesions (Iacono et al., 2009).
- Little is known about ID in younger populations and how it might relate to other measures of language, academic, and cognitive ability. It is also unclear as to how ID is influenced by the type of language sample elicited (e.g., expository vs. narrative). Further study of idea density may yield a new tool for language analysis that may provide information as to a person's ability to effectively convey ideas and information during communication, and may be used as an indicator for decline in this ability.

Methods

- This study examined written language samples collected from 43 individuals age 17-26. Participants spoke English as their primary language, had completed 13-17 years of education, and were currently enrolled in undergraduate college courses. Each subject produced two written language samples, one expository (language sample A) and one personal narrative (language sample B) (see Figure 1 for writing prompts). Each participant was given one writing prompt at a time (in counterbalanced order) and allowed up to 15 minutes per writing prompt to type their responses.

Methods cont.

- ID and writing sample length was calculated for each written language sample using CPIDR 5.1 (Brown et al., 2012).
- Measures related to academic success were obtained with participant consent. These included: overall GPA, transfer GPA, SAT (critical reading, writing, and math), and ACT scores. Subjects were also administered the Peabody Picture Vocabulary Test, and standard scores were calculated.

Figure 1: Writing prompts used to elicit written language samples.

Expository Writing Prompt	Narrative Writing Prompt
<ul style="list-style-type: none"> "A lot of medical research today focuses on developing medicinal cures to ageing. Presumably, with the right breakthrough, humankind would live forever. How do you feel about this? Write a short essay describing the advantages and disadvantages of extremely long life." 	<ul style="list-style-type: none"> "When we reflect upon our childhood, we often come back to a few key events that had a major impact on us. Write a short essay describing one of those defining events from your childhood, and the impact that it had on you."

Results

- A paired samples t-test showed that mean idea density for personal narrative (B) ($M=.55400$, $SD=.025073$) was significantly greater than mean idea density for expository text (A) ($M=.53849$, $SD=.026459$), $t(42)=-3.29$, $p<.002$. The eta squared statistic (.347) indicated a large effect size. Thus, idea density is influenced by genre of writing sample (see Tables 1 and 2 and Figure 1).

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 Idea Density A	.53849	43	.026459	.004035
Idea Density B	.55400	43	.025073	.003824

	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Pair 1 Idea Density A - Idea Density B	-.015512	.030913	.004714	-.025025	-.005998	-3.290	42	.002

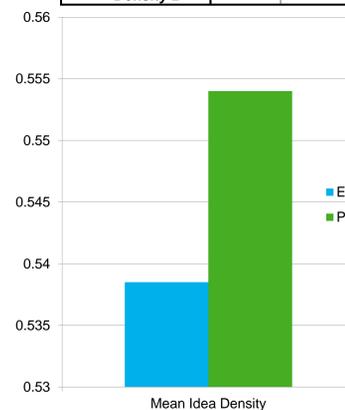


Figure 1

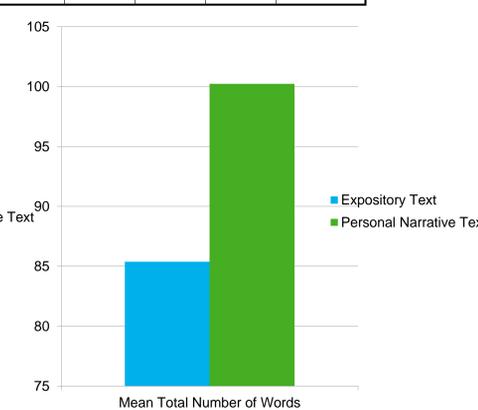


Figure 2

Results cont.

- Another paired samples t-test showed mean total number of words for personal narrative (B) ($M=309.53$, $SD=100.221$) was significantly greater than mean total number of words for expository text (A) ($M=247.26$, $SD=85.382$), $t(42)=-4.22$, $p<.0005$. The eta squared statistic (.298) indicated a large effect size (see Figure 2 and Table 3).

Paired Samples Test

	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Pair 1 Total # words - A - Total # words - B	-62.279	96.724	14.750	-92.046	-32.512	-4.222	42	.000

- Finally, correlations showed that there was very little relationship among idea density and education and vocabulary-related variables (PPVT scores, SAT scores, ACT scores, and grade point average). Only SAT Critical Reading scores were significantly correlated with idea density of personal narrative text.

Summary and Conclusions

- Previous research has shown that language genre affects language performance (e.g. Scott & Windsor, 2000) and that the more structured expository discourse imposes a greater demand on cognitive processing capacity and overall linguistic performance (Berman & Nir-sagiv, 2007). In this study, mean ID for expository texts was significantly lower when compared to that of narrative texts. This suggests that genre also has an effect on an individual's overall efficiency in expressing their ideas in writing as measured by ID. If, as suggested by Kemper & Sumner (2001), ID is a measure of economy of expression, it can be concluded that individuals are less efficient in their expression of ideas when those ideas required them to draw from their knowledge-base as compared to personal experience. Not only were participants able to put more ideas into their personal narrative writing, but they were able to produce a significantly greater number of words for the this text as compared to the expository.
- Only Critical Reading SAT scores were correlated with ID (only narrative ID). Other education and vocabulary variables were not significantly correlated with ID (although not all students had all of these scores). Kemper and Sumner (2001) stated that ID is related to reading rate and that both reflect "how much information can be processed at one time" (p. 319). Both reading ability and ID may reflect processing efficiency. Kemper et al. (2001) also found that ID had little relationship to grades in English or math or to level of educational attainment.
- Further investigation of ID density is needed to determine how ID interacts with different kinds of language samples, and to see if ID has utility in identification of language disorders.

The authors have no relevant financial or nonfinancial relationships to disclose.

Plattsburgh
STATE UNIVERSITY OF NEW YORK

Works Cited

- Berman, R. A. & Katzenberger, I. (2004) Form and function in introducing narrative and expository texts: A developmental perspective. *Discourse Processes*, 38(1), 57-94. doi:10.1207/s15326950dp3801_3
- Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods*, 40(2), 540-5. doi: 10.3758/BRM.40.2.540
- Brown, C., Snodgrass, T., Covington, M., Han, J., Boisclair, C., Drucker, E., ... & Nadig, K. (2012) Computerized Propositional Idea Density Rater (Version 5.1) [computer software]. CASPR Project: Artificial Intelligence Centre, University of Georgia.
- Engelman, M., Agree, E. M., Meoni, L. A., & Klag, M. J. (2010). Propositional density and cognitive function in later life: Findings from the precursors study. *Journal of Gerontology: Psychological Sciences*, 65B(6), 706–711. doi:10.1093/geronb/gbq064
- Farias, S. T., Chand, V., Bonnici, L., Baynes, K., Harvey, D., Mungas, D., ... & Reed, B. (2012). Idea density measured late in life predicts subsequent cognitive trajectories: Implications for the measurement of cognitive reserve. *The Journals of Gerontology, Series B: Psychological and Social Sciences*, 67(6), 667-686. doi:1093/geronb/gbr162
- Iacono, D. , Markesbery, W. R., Gross, M., Pletnikova, O., Rudow, G., Zandi , P., & Troncoso, J. C. (2009). The nun study: Clinically silent AD, neuronal hypertrophy, and linguistic skills in early life. *Neurology* , 73(9), 665 – 673. doi:10.1212/WNL.0b013e3181b01077
- Kemper, S., Greiner, L. H., Marquis, J. G., Prenovost, K., & Mitzner, T. L. (2001). Language decline across the life span: Findings from the nun study. *Psychology & Aging*, 16(2), 227-239. doi:10.1037/0882-7974.16.2.227
- Kemper, S., & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology & Aging*, 16(2), 312-322. doi:10.1037/0882-7974.16.2.312
- Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research*, 43(2), 324-339. doi:10.1044/jslhr.4302.324
- Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D.R., & Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the Nun Study. *Journal of the American Medical Association*, 275(7), 528-532. doi:10.1001/jama.1996.03530310034029
- Spencer, E., Craig, H., Ferguson, A., & Colyvas, K. (2012). Language and ageing – exploring propositional density in written language – stability over time. *Clinical Linguistics & Phonetics*, 26(9). 743-754. doi:10.3109/02699206.2012.673046