EasyChair Preprint

№ 6393

Noun-verb semantic distance analyses in sentence production of Alzheimer's disease

Jee Eun Sung, Yoonseob Lim, Kimun Kim and Sujin Choi

August 26, 2021

Noun-verb semantic distance analyses in sentence production of Alzheimer's disease

Jee Eun Sung[1*], Yoonseob Lim[2,3], Kimun Kim[2], and Sujin Choi[1]

[1]Department of Communication Disorders, Ewha Womans University, Seoul, Korea
[2]The Center for Intelligent and Interactive Robotics, Korea Institute of Science and Technology, Seoul, Korea
[3]Department of HY-KIST Bio-convergence, Hanyang University, Seoul, Korea

***corresponding author, Jee Eun Sung, jeesung@ewha.ac.kr**

Introduction

The current study analyzed the sentence production behaviors of individuals with AD using the DementiaBank, which is a part of the TalkBank projects. The purpose of this study is to apply a machine-learning approach to analyze the noun-verb semantic distance of AD in a sentence construction task using the DementiaBank. We further analyzed verb clusters with those nouns and investigated whether the verb clusters are associated with demographic factors(age, education, and dementia severity).

Methods

We extracted the data of 99 probable AD on the sentence construction task from the DementiaBank(Becker et al., 1994). Participants were asked to construct a sentence with given words(pencil, tree) which is similar to Altmann's(2004). To investigate the differences in semantic distances of different text corpora, we performed an independent samples t-test on the semantic distances for the two groups(DementiaBank vs. Wikipedia (or Blog)). In addition, to identify whether the verb clusters are associated with demographic factors(age, education, and dementia severity), we conducted a stepwise logistic regression.

Results

1. Semantic Distance between the target noun and verb

1.1 DementiaBank vs. Wikipedia database

For the analysis of the 'pencil', the noun-verb semantic distance, as indexed by the cosine similarity, was statistically higher in the DementiaBank than Wikipedia(t_{596} = -5.050, p = 5.881e-7), indicating that the semantic distance between the noun and verb is closer in the DementiaBank than the Wikipedia. For the 'tree', the cosine similarity was statistically higher in the DementiaBank than Wikipedia corpora($t_{24.98}$ = -7.888, p = 3.053e-8).

1.2 DementiaBank vs. Blog database

For the analysis of the 'pencil', the cosine similarity was statistically higher than Blog(t_{199} = -3.702, p = 2.764e-4); therefore, the DementiaBank had a higher proportion of verbs with semantic distances. For the 'tree', the cosine similarity is statistically higher in DementiaBank than the Blog(t_{946} = -5.289, p = 1.526e-7).

2. Verb Clustering and Regression Analyses

We found that education has a significantly positive(B = .661, Wald = 6.871, p = .009) effect on the choice of Write(baseline) or Be Verb for the ‘pencil’, but a marginally positive(B = .325, Wald = 3.553, p = .059) effect on the choice of verb ‘write’(baseline) or ‘use’ for the noun ‘pencil’. For MMSE scores, we found a significantly negative(B = -.294, Wald = 6.674, p = .01) effect on the choice of verb, either ‘write’(baseline) or ‘be’, for the ‘pencil’ and a significantly positive (B = .274, Wald = 3.921, p = .048) effect on the choice of either ‘be’(baseline) or ‘grow’ for the noun ‘tree’.

Conclusions

The current study found that the semantic distance between nouns and verbs is shorter in AD populations compared to the existing big databases. Furthermore, the semantic weight of verbs that AD participants used in a sentence construction task was significantly related to the severity of dementia, indicating that people with AD tend to use more light verbs as their disease progresses. We applied machine-learning techniques to the open-access big database under the framework of examining linguistically finite deficits in AD patients’ sentence production.

References

Altmann, L. J. (2004). Constrained sentence production in probable Alzheimer disease. Applied Psycholinguistics, 25(2), 145.

Becker, J. T., Boller, F., Lopez, O. L., Saxton, J., and McGonigle, K. L. (1994). The natural history of Alzheimer’s disease. Description of study cohort and accuracy of diagnosis. *Arch. Neurol*. 51, 585–594. doi:10.1001/archneur.1994.00540180063015.

Acknowledgments

This work was supported by the Technology Innovation Program-Industrialized Technology Innovation Project (10077553, Development of Social Robot Intelligence for Social Human-Robot Interaction of Service Robots) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea), a National Research Council of Science and Technology (NIST) grant from the Korean government (MSIP) (No. CRC-15-04-KIST).

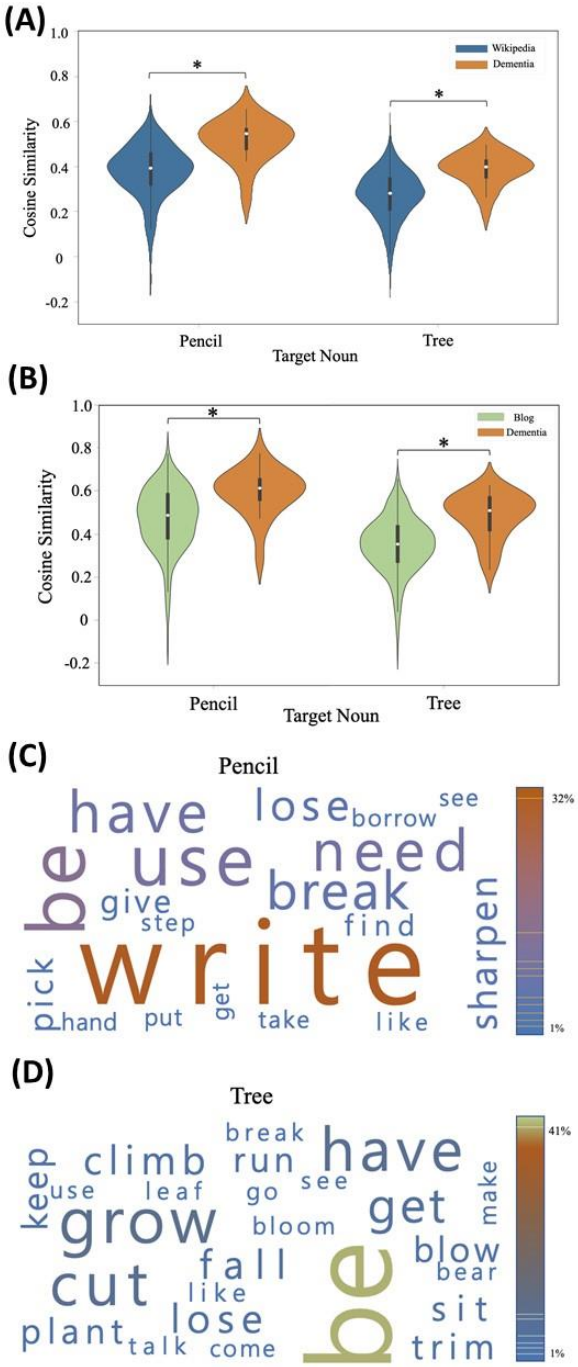

Figure 1. **Semantic distances of words in Wikipedia/Blog and verb clusters of DementiaBank** (A). Distribution of cosine similarity between the verbs and the target noun, pencil, or tree in Wikipedia. (B). Distribution of cosine similarity between the verbs and the target nouns pencil or tree in Blog. (C). Verbs associated with the noun pencil are shown in a word cloud format. (D). Verbs associated with the noun tree are shown in a word cloud format. The size or color of the text corresponds to the production frequency in DementiaBank.

Table 1. Summary of semantic distance for DementiaBank and Wikipedia/Blog

	Target Noun	Corpora	Mean	Standard Deviation	Number of Samples	Confidence Interval
Wikipedia	Pencil	Wikipedia	0.379	0.117	576	0.019
		DementiaBank	0.513	0.093	20	0.089
	Tree	Wikipedia	0.272	0.106	2642	0.008
		DementiaBank	0.388	0.0713	25	0.060
Blog	Pencil	Blog	0.472	0.142	181	0.042
		DementiaBank	0.594	0.105	20	0.101
	Tree	Blog	0.350	0.128	921	0.017
		DementiaBank	0.487	0.101	25	0.085