



Noun-Verb Semantic Distance Analyses in Sentence Production of Alzheimer's Disease

Jee Eun Sung¹ Ph.D., Yoonseob Lim Ph.D.², Kimun Kim^{2,3}, and Sujin Choi¹

¹Department of Communication Disorders, Ewha Womans University, Seoul, South Korea

²The Center for Intelligent and Interactive Robotics, Korea Institute of Science and Technology, Seoul, Korea

³Department of HY-KIST Bio-convergence, Hanyang University, Seoul, Korea



The Purpose of the Study

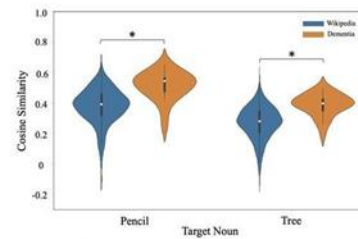
- The current study analyzed the sentence production behaviors of individuals with AD using the DementiaBank, which is a part of the TalkBank projects.
- The purpose of this study is to apply a machine-learning approach to analyze the noun-verb semantic distance of AD in a sentence construction task 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).

Results

1. Semantic Distance Between the Target Noun and Verb

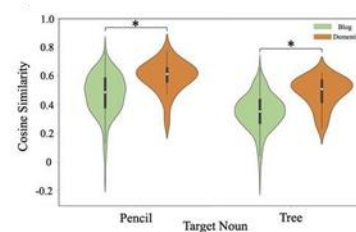
1) DementiaBank vs. Wikipedia Database

- 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 ($t_{595} = -5.050, p = 5.881e-7$).
- This indicates 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 ($t_{7498} = -7.888, p = 3.053e-8$).



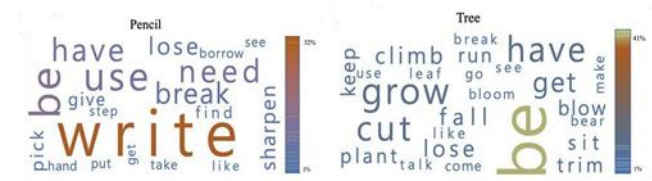
2) DementiaBank vs. Blog Database

- 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 ($t_{945} = -5.289, p = 1.526e-7$).



2. Verb Clustering and Regression Analysis

- 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. *Archives of Neurology*, 51, 585-594. doi:10.1001/archneur.1994.00540180063015.

Correspondence to: Jee Eun Sung, Ph.D. Email: jeesung@ewha.ac.kr

Acknowledgment: 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).

Methods

◆ Participants

- We extracted the data of 99 probable AD on the sentence construction task from the DementiaBank (Becker et al., 1994). Table 1 provides demographic information of participants.

<Table 1> Demographic information of participants

	Age (yr)	Education (yr)	MMSE	Gender (male : female)
Mean (SD)	71.86 (±8.80)	11.84 (±2.765)	19.13 (±4.193)	34 : 65
Range	53-88	6-20	8-27	

◆ Task

- Participants were asked to construct a sentence with given words (pencil, tree) which is similar to Altmann's (2004).

Target Noun	Examples
Pencil	I write with the pencil. / I need a pencil. / I use the pencil.
Tree	There is a tree. / The tree is green. / I cut the tree down.

◆ Analyses

1. Semantic Distance Analysis

- 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)).

2. Verb Clustering and Regression Analysis

- To identify whether the verb clusters are associated with demographic factors (age, education, and dementia severity), we conducted a stepwise logistic regression.



Noun-Verb Semantic Distance Analyses in Sentence Production of Alzheimer's Disease

Jee Eun Sung¹ Ph.D., Yoonseob Lim Ph.D.², Kimun Kim^{2,3}, and Sujin Choi¹

¹Department of Communication Disorders, Ewha Womans University, Seoul, South Korea

²The Center for Intelligent and Interactive Robotics, Korea Institute of Science and Technology, Seoul, Korea

³Department of HY-KIST Bio-convergence, Hanyang University, Seoul, Korea



The Purpose of the Study

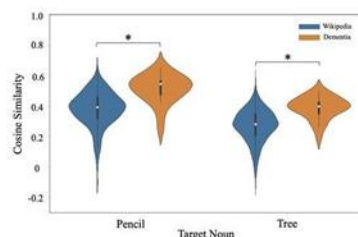
- The current study analyzed the sentence production behaviors of individuals with AD using the DementiaBank, which is a part of the TalkBank projects.
- The purpose of this study is to apply a machine-learning approach to analyze the noun-verb semantic distance of AD in a sentence construction task 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).

Results

1. Semantic Distance Between the Target Noun and Verb

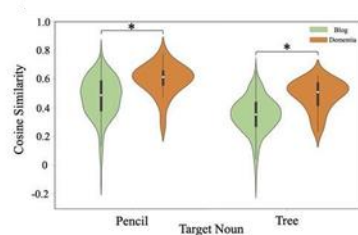
1) DementiaBank vs. Wikipedia Database

- 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 ($t_{595} = -5.050, p = 5.881e-7$).
- This indicates 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 ($t_{7498} = -7.888, p = 3.053e-8$).



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 ($t_{945} = -5.289, p = 1.526e-7$).



2. Verb Clustering and Regression Analysis

- 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. *Archives of Neurology*, 51, 585-594. doi:10.1001/archneur.1994.00540180063015.

Correspondence to: Jee Eun Sung, Ph.D. Email: jeesung@ewha.ac.kr

Acknowledgment: 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).

Methods

◆ Participants

- We extracted the data of 99 probable AD on the sentence construction task from the DementiaBank (Becker et al., 1994). Table 1 provides demographic information of participants.

<Table 1> Demographic information of participants

	Age (yr)	Education (yr)	MMSE	Gender (male : female)
Mean (SD)	71.86 (±8.80)	11.84 (±2.765)	19.13 (±4.193)	34 : 65
Range	53-88	6-20	8-27	

◆ Task

- Participants were asked to construct a sentence with given words (pencil, tree) which is similar to Altmann's (2004).

Target Noun	Examples
Pencil	I write with the pencil. / I need a pencil. / I use the pencil.
Tree	There is a tree. / The tree is green. / I cut the tree down.

◆ Analyses

1. Semantic Distance Analysis

- 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)).

2. Verb Clustering and Regression Analysis

- To identify whether the verb clusters are associated with demographic factors (age, education, and dementia severity), we conducted a stepwise logistic regression.