Spring 2019

Can Semantic-Phonemic Discrepancy in Verbal Fluency Help to Detect Alzheimer's Dementia?

Haoxuan Chen

CAN SEMANTIC-PHONEMIC DISCREPANCY IN VERBAL FLUENCY HELP TO DETECT ALZHEIMER'S DEMENTIA?

by

Haoxuan Chen

A thesis submitted in partial fulfillment of the requirements
for graduation with Honors in the Speech Pathology and Audiology

Jean. K. Gordon
Thesis Mentor

Spring 2019

All requirements for graduation with Honors in the
Speech Pathology and Audiology have been completed.

Yu-Hsiang Wu
Speech Pathology and Audiology Honors Advisor

College of Liberal Art and Sciences
The University of Iowa
Iowa City, Iowa

CAN SEMANTIC-PHONEMIC DISCREPANCY IN VERBAL FLUENCY HELP TO DETECT ALZHEIMER'S DEMENTIA?

By

Haoxuan Chen

A thesis submitted in partial fulfillment of the requirements for graduation with Honors in the Department of Communication Sciences and Disorders

Jean. K. Gordon

Thesis Mentor

Spring 2019

Abstract

Word retrieval difficulty is one of the early signs of Alzheimer's disease, although such difficulties can also occur in typically aging. Therefore, it is necessary to find a task that differentiates the early stages of Alzheimer's dementia from typically aging. Verbal fluency is a widely used measure to assess subjects' cognitive processes following neurological damage, and often includes two subtests: semantic fluency, in which participants are asked to produce words which meet a semantic criterion, such as food or animals; and letter fluency, which requires participants to produce words starting with a certain letter, such as F or S. People with Alzheimer's disease have more difficulty with semantic than letter fluency, although this pattern has also been shown in typically aging. In the current research, we investigate whether the semantic-letter discrepancy can differentiate Alzheimer's dementia from typically aging.

Introduction

Alzheimer's disease is a degenerative disease of the aging brain associated with dementia. Memory will be affected first; later on, language deficits occur. It is the most common dementia and has no cure. Ultimately, individuals with Alzheimer's disease lose abilities to perform daily activities of living (Bear, Conners, & Paradiso, 2015). Eventually, it can lead to death. It's difficult to detect Alzheimer's disease early, because people tend to show some memory degradation when they get older, and it is difficult to discriminate whether it is dementia or an aging process. In addition to the memory deficit, an AD individual's language ability is also significantly more impaired comparing to typically aging (Ting, Hameed, Earnest, & Tan, 2013). Since language impairment is another sign of Alzheimer's disease, language measures are often used to detect early Alzheimer's disease. In this study, we investigate how individuals with Alzheimer's disease show a degradation of word retrieval.

Verbal fluency is used for this research to assess participants' ability to retrieve words. It is a widely used measure to assess subjects' cognitive processes following any neurological damage, because it requires cognitive functions of the frontal lobe (e.g., working memory, executive functioning) and language abilities of the temporal lobe (e.g., word retrieval, vocabulary storage, language comprehension) (Henry, Crawford, & Phillips, 2004). This test often includes two kinds of subtests: one is semantic fluency, which asks participants to produce words which have associative meanings or meet a semantic criterion, such as food or animals; another is letter fluency, which requires participants to produce words starting with a certain letter, such as F or S (Henry et al., 2004). In addition to the number of items a subject can produce, there are another two important components of the verbal fluency task performance:

clustering (i.e., generating words within a same subcategories) and switching (i.e., shifting from one subcategory to another) (Troyer, Moscovitch, Winocur, Leach, & Freedman, 1998).

Verbal fluency in studying Alzheimer's disease

In individuals with brain damage, performance on the verbal fluency tasks is associated with participants' affected brain areas. It is hypothesized that the semantic and phonemic fluency tasks can be affected differently based on the location of the lesion. Henry and Crawford (2004) reported in their meta-analysis that, although phonemic and semantic fluency both require processing in frontal lobe, semantic fluency appears to rely more on the function of the temporal lobe. The temporal lobe is involved in processing memories and language comprehension, and semantic fluency requires word retrieval and access to word meaning. Therefore, semantic fluency is hypothesized to be more sensitive to the temporal pathology, while phonemic fluency is proposed to be less impaired by lesions in the temporal lobe. In a meta-analysis, Henry and his colleagues (2004) reported that in individuals with Alzheimer's disease, semantic fluency is more impaired than phonemic fluency (Henry et al., 2004). Therefore, impaired semantic fluency has been viewed as an early sign of semantic degradation (Chen, Ratcliff, Belle, Cauley, DeKosky, & Ganguli, 2001). Similarly, Adlam and his colleagues (2006) suggested that semantic fluency is better than letter fluency to differentiate participants with Mild Cognitive Impairment from typically aging controls (Adlam, Bozeat, Arnold, Watson, & Hodges, 2006), since MCI is considered to be a "prodromal state" of Alzheimer's disease (Grundman et al., 2004).

Factors that influence task performance in typically aging adults

It is important to find out how the performance of typically aging individuals can be affected by different variables, before investigating how individuals with AD would perform. Sex is one factor which is mentioned in some studies, but the effect is not yet clear. Capitani and his colleagues (1998) reported that female subjects tended to show better performance in phonemic fluency than male subjects (Capitani, Laicona, & Basso,1998). On the contrary, Tombaugh and his colleagues (1999) reported that there is no sex difference between adults in FAS and animal naming tasks (Tombaugh, Kozak, & Rees, 1999). Laws (2004) found that, in certain categories, such as tools and vehicles, men showed significantly better performance in the semantic fluency task, while women outperformed with categories of fruits. However, there was no significant difference between men and women for animal tasks, which may be because animals is one of the most familiar categories in people's knowledge (Laws, 2004).

Higher levels of education typically lead to better fluency in typically aging participants (Mathuranath, George, Cherian, Alexander& Sarma, 2003). In particular, the concept of 'cognitive reserve' suggests that education may protect subjects' performance from age-associated decline (Stern, 2002). Ting and colleagues (2013) compared a group of AD participants who had education levels of 0-6 years to a group who had education levels of over 6 years, and found that breakdown of semantic fluency appears less in higher educated subjects than lower educated subjects in specific category fluency tasks. They mentioned that food is a common category of semantic knowledge that appears often in everyone's daily life, whereas animals are not. The finding showed that education level may offer more opportunities to know those items which are less common, such as animals (Ting et al., 2013).

Those studies compared subjects with and without dementia from the same age groups to help eliminate the possibility that some results may be due to age, because age itself has an effect on verbal fluency. Studies have found that, compared to younger participants, older participants produce fewer items during the verbal fluency tasks (Gordon, Young, & Garcia, 2017; Kavé, 2015; Troyer et al., 1998). Marsolais and colleagues (2014) did fMRI experiments. They reported that, although age didn't have a significant effect on participants' behavioral results, fMRI indicated reduced functional connectivity of semantic networks, especially when the required semantic categories were difficult (e.g., sports, clothing) (Marsolais, Perlbarg, Benali & Joanette, 2014). Moreover, a meta-analysis written by Rodriguez-Aranda and Martinussen (2006) reviewed 26 studies about the performance of typical adults in different age ranges. They indicated that the number of items which a participant can produce in the phonemic fluency task declines slowly before the late 60s; but after late 80s, the number of items declines rapidly.

Pattern of age effects in two different tasks

A specific pattern of differences in the two kinds of the fluency tasks has also been demonstrated: the age effect is significantly greater in the semantic fluency task than in the letter one. In semantic fluency, typically aging subjects tend to generate fewer words than younger subjects, while this difference is less distinct on phonemic fluency (Gordon et al., 2017). Kavé (2015) suggested that, when retrieving words in the animal category, participants need to process the meaning of the word, whereas letter fluency only requires them to retrieve lexical form. Meinzer and colleagues (2009) proposed a similar idea: that the response in semantic fluency tends to be more constrained. That is, the nature of the semantic category limits responses to nouns. Contrary to that, letter categories only have the constraint of the letter, so participants don't need to consider the word class during the tests (Meinzer, et al., 2009).

This pattern of different performances in the semantic and letter tasks may also be because of various cognitive contributors to the two tests. Gordon et al. (2017) found that lexical retrieval speed heavily affects the results of semantic fluency, whereas letter fluency depends largely on vocabulary knowledge, which can protect performance from age effects. They also mentioned that visualization strategies help during the semantic fluency task, such as visualization of animal environments; however, this skill might decline with age (Gordon, et al., 2017).

Why study semantic-letter discrepancy

Suhr and Jones (1998) compared semantic and letter fluency test performances of subjects with Alzheimer's or Parkinson's to older typical adults. They found that, compared to the typically aging control group, AD and PD subjects' performances were significantly worse on both tasks. Gomez and White (2006) did a study which included 76 older typical adults and 77 individuals with very mild Alzheimer's dementia. They reported that the typically aging group performed better than the very mild AD group across semantic and letter fluency. The control group generated more words with more switches and clusters and larger clusters for the most part. The only exceptions they found were the number of clusters and cluster size when producing words starting with "S" (Gomez & White, 2006). Troyer et al. (1998) also examined clustering and switching in people with Alzheimer's disease and typical controls. They found that, when doing the semantic fluency tasks, older participants with AD tended to generate fewer items in total than the control group. They also switched less frequently than typical controls. When doing the phonemic fluency tasks, overall, the size of clusters was smaller for the AD subject group compared to the older typical subject group (Troyer et al., 1998).

Discrepancy is the parameter to use when researchers try to compare two values. In the verbal fluency task, some researchers have focused on the discrepancy between semantic and letter fluency, which equals the score of the semantic fluency task minus the score of the letter fluency task, because they were trying to see if the onset of AD can be detected by comparing the discrepancy in AD to typically aging. A meta-analysis analyzed the degree of impairment in both the semantic and letter fluency tasks between AD and typical control groups (Laws, Duncan, & Gale, 2010). Surprisingly, they reported that across 50 studies, there was no significant difference in discrepancy scores between typically aging groups and AD participants, which indicated that the pattern of the semantic-letter discrepancy might just be an exaggerated normal tendency. In addition, none of the following moderator variables significantly predicted discrepancy scores: severity of dementia, participants' ages, education levels or proportion of female participants.

In order to detect the onset of Alzheimer's disease, a discriminant analysis is needed to determine sensitivity and specificity. Storandt and Hill (1989) found three psychometric tests which showed good discrimination of very mild AD from typically aging: Logical Memory (Wechsler, 1974), Digit Symbol Task (Wechsler, 1955), and Boston Naming Test (Kaplan, Goodglass & Weintraub, 1983). Together, these tasks showed 68% sensitivity and 74% specificity in terms of distinguishing very mild AD from typically aging. In a stepwise procedure, Gomez and White (2006) added letter fluency (P), letter fluency (S), and animal fluency into Storandt and Hill's discriminative analysis. Their results showed that animal fluency had the greatest strength in predicting group (i.e., very mild AD or typically aging), followed by Logical Memory. None of the other variables significantly improved prediction. They got a final sensitivity score of 78% and a specificity score of 74% through semantic fluency and Logical Memory. Notably, letter fluency was not a significant predictor in either analysis.

Suhr and Jones (1998) calculated the optimal cutoff score for each task that discriminated between individuals with AD and heathy age-matched control participants. For the letter fluency tasks, they asked participants to produce any word started with C, F and L. Each task is produced separately in one minute. The raw score is the total correct responses across the three tasks, and the optimal cutoff was 33 for LF, which had a maximum of 94% sensitivity and 76% specificity. The semantic fluency tasks required subjects to name any animals, fruits/vegetables, and tools/kitchen utensils. Each category was produced in one minute, and the optimal cut-off (total of 40 across the three tasks) had 87% sensitivity and 88% specificity. From these numbers, it is not clear which task had a better discriminative power to detect AD from typically aging, because the semantic fluency task contributes to a higher specificity score, whereas the letter fluency task contributes to a higher sensitivity score.

From previous research, we know that performance on the verbal fluency tasks can vary because of the subjects' age, education, gender; and typically, age affects the semantic more than letter fluency task. Compared to typical controls, AD subjects tend to produce fewer items in both tasks. Previous studies have also shown that AD subjects' semantic fluency is more impaired than letter fluency compared to typically aging subjects (Henry et al., 2004); however, Laws' study reported that there is no difference between these two groups' discrepancy scores (Laws et al., 2010). Thus, the effect of AD on semantic-letter discrepancy still unclear.

In the current research, we investigate how the comparison of semantic-letter discrepancy between AD and typically aging distinguishes between typically aging and dementia. We examined both a typically aging group and individuals with Alzheimer's dementia, because the typical controls help factor out age-related performance. We analyzed the performance of both groups on the verbal fluency test, and compared their performances in the semantic and letter tasks. Eliminating the age effect would help determine the factors that only relate to Alzheimer’s disease.

Method

Participants

In this study, we selected two groups of participants to compare their discrepancy scores. The data of the dementia group, which were downloaded from DementiaBank (Becker et al., 1994), included audio files and transcriptions of the verbal fluency tasks. The 134 individuals with AD who did both animal and F tasks were selected to compare to the control group. The age range of AD subjects was from 49 to 88 with a mean age of 71 years old. The data of the typically aging group was selected from a previous study (Gordon et al., 2017). In order to match the age range of the dementia group, we selected 66 typical individuals ranging in age from 49 to 89 with a mean of 70.21 years old. Although the typically aging subjects did F, A and S for the letter fluency task, we only used their responses from the F task to compare with the group of AD participants. All the participants in this study were native English speakers, including both the AD group and the typical control group. Demographic information about both groups is listed in Table 1.

Tasks and Procedure

Two verbal fluency tasks were examined in this study. One was semantic fluency using the category of animals. Instructions like "Name all the animals you know" were given to the subjects, and the task was 60 seconds long. The other one was letter fluency using the letter "F". Subjects were asked to say all the words they know that begin with letter "F." The task was also 60 seconds long. Since the typical group did both tasks over 90 seconds, we took out all the items typical participants produced after 60s.

Table 1. Demographic Characteristics of the Two Groups

		AD group (n=134)	Control group (n=66)
Age (yrs)	Average (range)	71.11 (49-88)	70.21 (49-89)
Education (yrs)	Average	20.01 (8-30)	12.91 (9-16)
Sex (%)	Male (%)	53 (39.8%)	29 (43.9%)
	Female (%)	80 (60.2%)	37 (56.1%)
Diagnosis (#)	Possible AD	15	N/A
	Probable AD	94	
	Vascular	4	
	MCI	17	
	Memory	2	
	Other	1	

Coding

In order to analyze the total score of each subject's performance, we counted the number of correct answers. We categorized incorrect answers in the semantic fluency task as several types: 1) out-of-category errors (OC), responses that are not in the requested category (e.g., *car* for animal task); 2) proper names (PROP), which are names of people or places (e.g. *South Africa*); 3) form-based errors (FOR), recognizable responses that have an error in form (e.g. *dramadery* for *dromedary*); 4) perseverations (PER), repetitions of an earlier response, including singular and plural repetitions (e.g. *rabbit, rabbits*); 5) redundant errors (RED), which can be either synonyms of a previous response (e.g. *cat, kitty*) or category labels for which subordinate labels are also provided (e.g. *bird*, *falcon*); 6) other errors (OTH), which includes fragments, nonwords, and unintelligible responses.

For letter fluency, incorrect answers were categorized under the same types of errors as we listed above for the semantic fluency. However, there are some types of errors which are unique to letter fluency. One category is variation (VAR), which means the morphological variations of a previous response (e.g. *frat, fraternity, fraternization; forty, forty-one, forty-two*). The other category is noun phrases (NP). These are responses made up of one noun as a head and a modifier which are embedded on the noun (e.g. *farm animals, farm buildings; apple butter, apple sauce*).

We also coded clusters. A cluster is defined as a group of words consisting of successively generated words belonging to the same subcategory (e.g., marine animals such as *dolphin*, *shark* and *octopus*). Clustering is considered a reflection of how participants apply strategies to maximize responding in the verbal fluency tasks. Some responses which are not counted towards the total word count may still count in clustering.

Analyses

We conducted *t*-tests on participants' demographic information in both group to show that both groups' age and education levels are matched in this study. A chi-square test was calculated to determine whether the two groups were matched in terms of sex distributions.

We conducted analyses of variance (ANOVAs) comparing the typically aging and the dementia group's performances across the semantic and letter fluency tasks. The outcome measures included total correct responses, which reflects the overall level of subjects' performance. Errors were analyzed to reflect where the cognitive processing is problematic. We counted the number of errors each subject made and calculated the proportion of errors produced as a percentage of total responses. Clusters and singletons were also analyzed: we counted the number of clusters produced by each subject to indicate how well subjects can use the strategy of producing words by subcategory in this test. In addition, we calculated the number and the proportion of singletons produced to indicate whenever the subjects are not using this strategy to retrieve words. We conducted 6 2x2 ANOVAs, one for each outcome measure. The two factors were group (typical vs AD) and task (semantic vs letter fluency). In addition, we conducted 4 one-way ANOVAs to determine if the four discrepancy scores are significantly different between groups. Total discrepancy is calculated as the total semantic responses minus the total letter responses. Percentage of total discrepancy is calculated as the total semantic responses minus the total letter responses then divided by the total semantic responses. Total cluster discrepancy is calculated as the total semantic clusters minus the total letter clusters. Percentage of total cluster discrepancy is calculated as the total semantic clusters minus the total letter clusters and divided by the total semantic clusters.

We also conducted a descriptive analysis of the error types produced by each group to compare the differences between the two groups' types of errors. All the error types are listed above in the coding section.

Results

Total responses

On average, typically aging subjects produced 22 items during the semantic fluency task and 14 items during the letter fluency task. The average number of total items a subject with dementia produced during the semantic fluency task was 10, and their average number of total responses during the letter fluency task was 7. In terms of task, both groups produced significantly more items during the semantic fluency than letter fluency task ($p < 0.01$); and in terms of group, subjects with dementia tended to produce fewer responses than typically aging subjects ($p < 0.01$). There was also a significant interaction between group and task ($p < 0.01$) which arises because the task differences were bigger for the typically aging group than for the AD group.

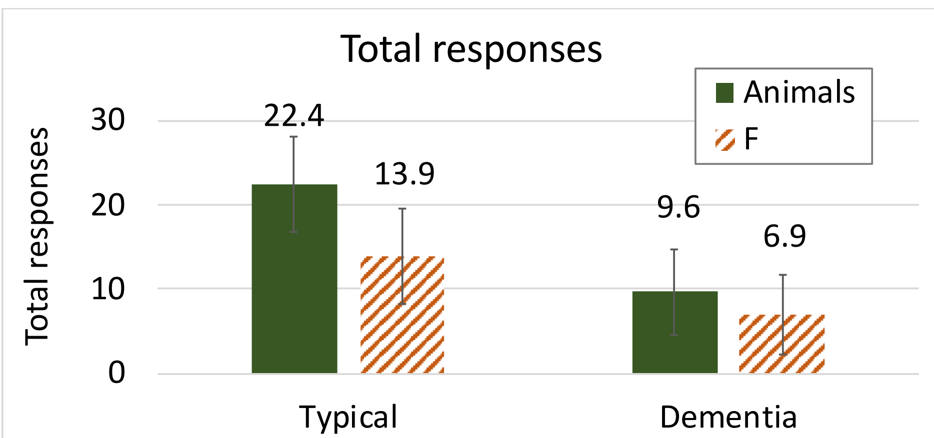

Figure 1. Two groups' total responses in the semantic and letter fluency tasks

Clusters

On average, typical subjects produced 10.7 clusters during the semantic fluency task, and 5.1 clusters during the letter fluency task. Dementia subjects produced 5 clusters on average during the semantic fluency task and 2.4 clusters during the letter fluency task. Both groups produced more clusters in semantic than letter fluency ($p < 0.01$), but overall the typically aging group produced more clusters than the dementia group across both tasks ($p < 0.01$). There is also an interaction effect between task and group shown by the ANOVA results ($p < 0.01$). The difference of total clusters between tasks was bigger in the typically aging group than the dementia group.

Similarly, the dementia group produced significantly more singletons during each task than the typically aging group ($p < 0.01$), and both groups produced significantly more singletons in the letter than semantic fluency task ($p < 0.01$), but there was no interaction between group and task ($p = 0.97$). We also analyzed the proportion of singletons each subject produced among all their responses. The ANOVA results showed a significant group effect ($p < 0.01$) and a significant task effect ($p < 0.01$) but no significant interaction ($p = 0.27$).

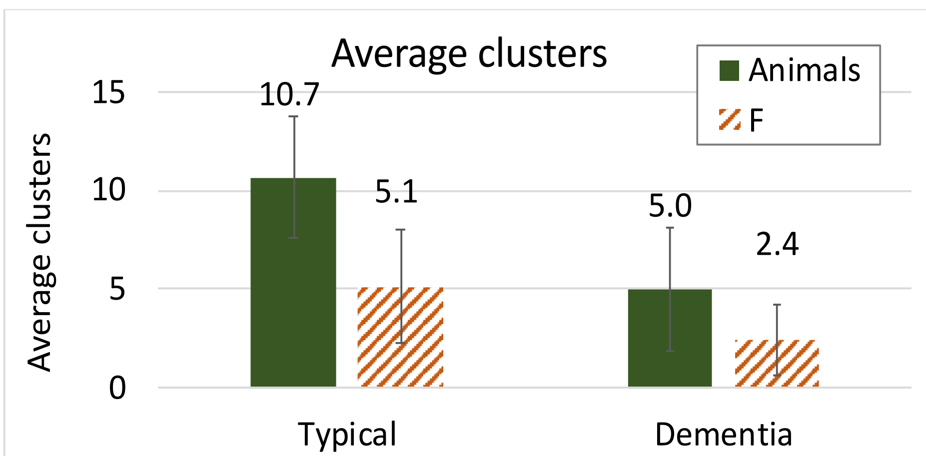

Figure 2. Two groups' average clusters in the semantic and letter fluency task

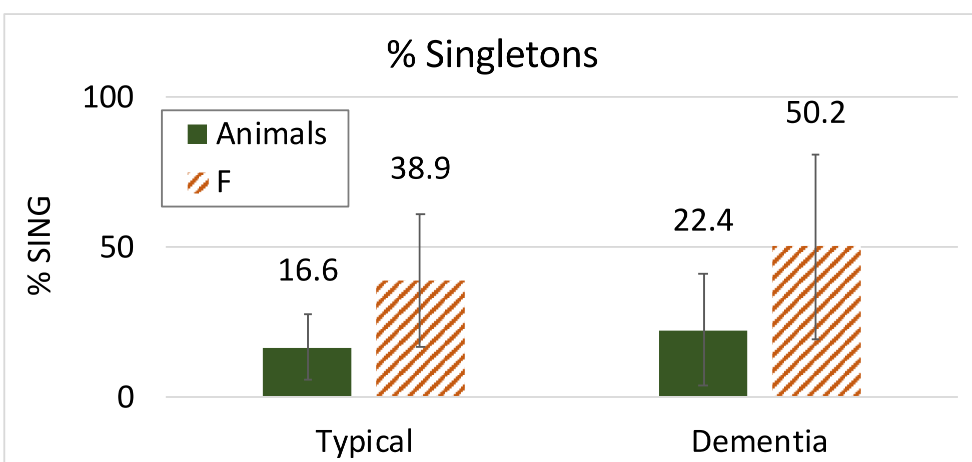

Figure 3. Proportion of singleton responses over all responses in two groups' semantic and letter fluency task

Errors

The dementia group produced more errors during both tasks than typically aging subjects ($p < 0.01$); however, the number of errors across two tasks was not significantly different ($p = 0.56$). There was also no significant interaction between group and task ($p = 0.14$). The results in terms of proportion of errors was the same. A larger proportion of errors was produced by the dementia group than the typically aging group for both tasks ($p < 0.01$); however, there was no difference between the proportion of errors which two groups produced across the two tasks ($p = 0.19$), and there was no significant interaction between group and task ($p = 0.81$).

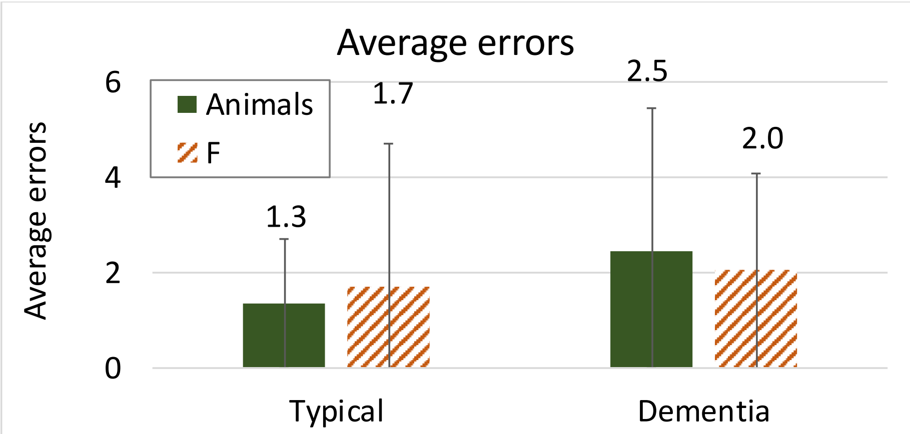

Figure 4. Two groups' average error responses in the semantic and letter fluency task

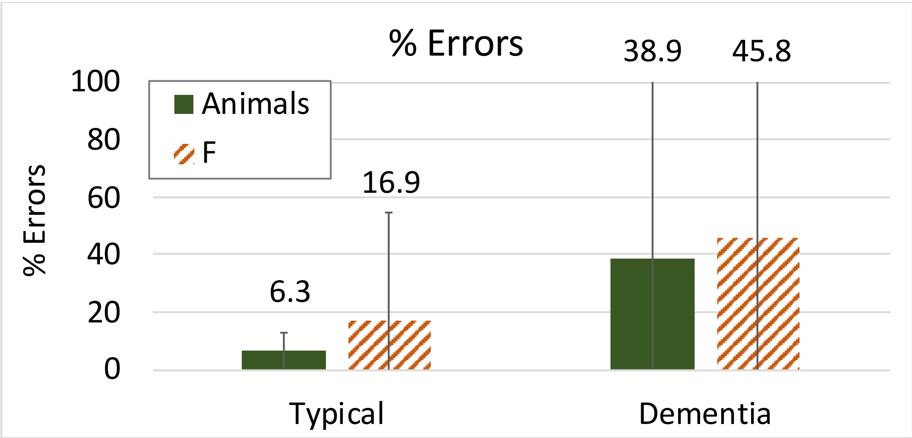

Figure 5. Proportion of error responses over all responses in two groups' semantic and letter fluency task

Error types

The most common error type across both tasks was perseveration, but the dementia group produced even more repetition (94.8% on semantic fluency; 67.9% on letter fluency) than the typically aging group (62.5% on semantic fluency; 36.3% on letter fluency). On the semantic fluency task, the dementia group also produced relatively more out-of-category errors (4.0% vs 0.0% for the typically aging group), whereas the typically aging group produced relatively more redundant errors (36.4% vs 1.2% for the dementia group).

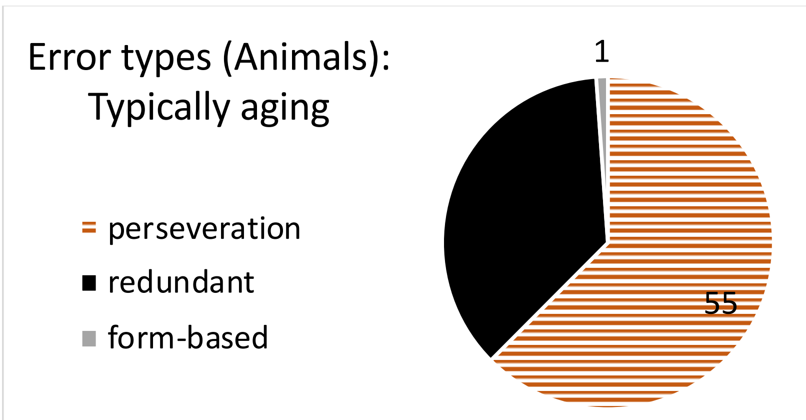

Figure 6. Error types of the typically aging group in the semantic fluency task

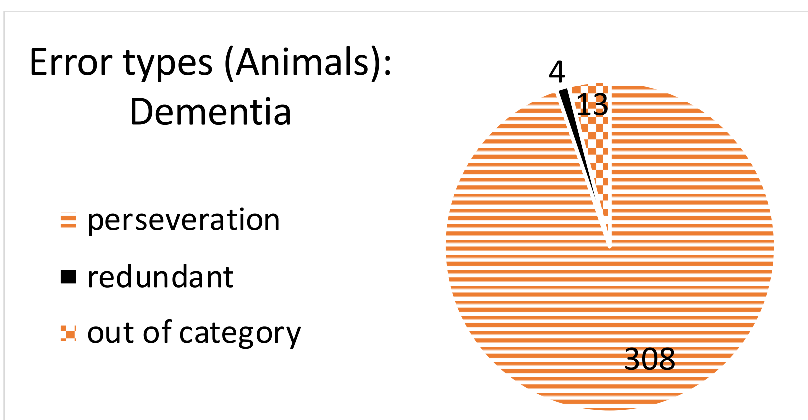

Figure 7. Error types of the dementia group in the semantic fluency task

On the letter fluency task, the dementia group produced relatively more out-of-category errors (16.2% vs 5.3% for the typical group), and the typically aging group produced relatively more variation (38.9% vs 4.4% for the dementia group) and proper name errors (15.9% vs 9.6% for the dementia group).

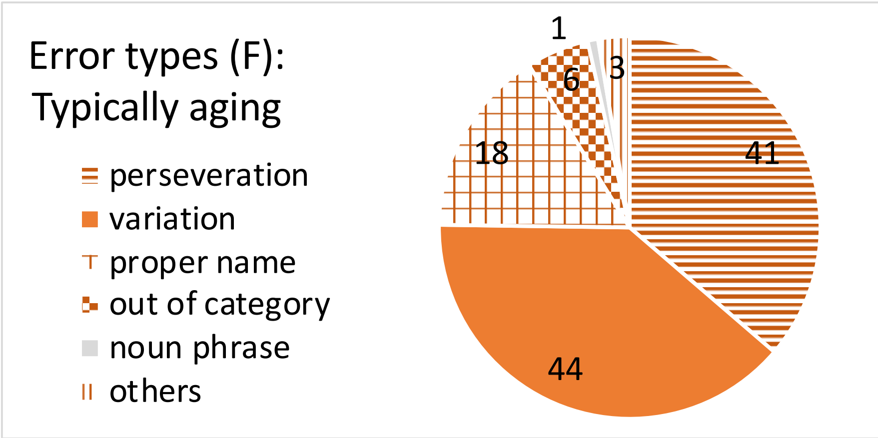

Figure 8. Error types of the typically aging group in the letter fluency task

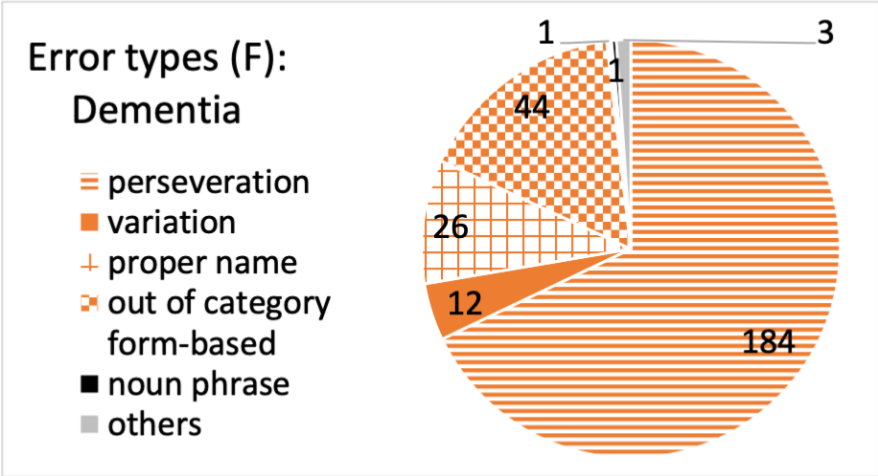

Figure 9. Error types of the dementia group in the letter fluency task

Discrepancy

Both the dementia and the typically aging group produce significantly more total correct responses and clusters in semantic than letter fluency. Figures 1 and 2 visually show that, using the raw number of total responses, the discrepancy is significantly different ($F = 47.33$, $p < 0.01$). Similarly, the total cluster discrepancy calculated by the raw numbers is significantly different between two groups ($F = 33.96$, $p < 0.01$). The smaller semantic-letter discrepancy of the dementia group is also shown by the error results that there is no significant difference between the dementia group's scores in the semantic and letter fluency task, but the typically aging group produces significantly more errors in the letter fluency task. However, if we calculate the percentage of total discrepancy ((total semantic responses - total letter responses) / total semantic responses), the total discrepancy has a borderline significance ($F = 3.72$, $p = 0.06$), and the percentage of total cluster discrepancy is not significant ($F = 2.38$, $p = 0.13$).

Discussion

In all, as we expected, participants with dementia produced fewer responses, fewer clusters, but made more errors than the typically aging group; besides, our findings also matched with the previous studies showing that both groups produced more responses and more clusters in the semantic fluency task than letter fluency task (Gomez & White, 2006; Suhr & Jones, 1998). Our study also analyzed semantic-letter discrepancy. We found that the difference value of total correct responses and total clusters between the typically aging group and dementia group was significantly bigger in the semantic fluency task. This means that, although the dementia group produced fewer items in both tasks compared to the typically aging group, they produced even fewer items in the semantic fluency task. This supports the result of previous studies that semantic fluency is impaired more than letter fluency (Henry et al., 2004). However, when we calculated the discrepancy scores using the percentage of total responses and clusters to eliminate the effect of more responses produced by the typically aging group, we found that neither of the two discrepancy scores is significantly different between groups. This supports the conclusion of the previous meta-analysis (Laws et al., 2010). On the other hand, since the percentage of total discrepancy has border-line significance, it could become significant if we had more samples in our study.

In addition to the statistical analyses, the descriptive analysis of error types is also important in our study. For both groups, the most common type of error involved repeated responses, and the dementia group produced even more repetition errors than the typically aging group. These two findings are same as the conclusion in the previous study by Suhr and Jones (1998). However, this study showed that there was no significant difference between groups in terms of the intrusion errors (which is named as out-of-category error in our study), whereas our study showed that participants with dementia produced many more out-of-category errors than the typically aging group in both tasks. They also tended to repeat verbatim and forget what the requirements were, which led to errors like out-of-category and form-based errors. These observations suggest that participants with dementia produce more severe errors which can be more easily avoided by a typically aging subject. On the contrary, typically aging participants tended to produce variations, which we considered more like strategies they used in order to produce more responses in total.

Our study has a sufficient sample size, and we conducted analyses that took into account the raw difference in total responses between people with dementia and typically aging adults. However, one limitation in our study is that the data is not first-hand. Since we did not administer the tests, we could not control the instructions which were given to the subjects. Moreover, the transcripts and the audio files could not help us fully understand what was going when subjects made errors. For example, sometimes subjects repeat responses because they forget they said it before, but sometimes they may be practicing or talking to themselves. Another limitation is that the subjects with dementia had different types of dementia, but the data did not tell us their lesion sites. Future research can address this by comparing certain types of dementia with their brain scans, so that their lesion sites and task performance can be linked when discussing the results.

In summary, our study found that people with dementia performed worse than the typically aging group in both tasks, but both groups performed better in semantic fluency than letter fluency. The critical analysis of calculating responses using percentage discrepancy showed that both tasks were affected similarly by dementia. For both groups, the most common type of error involved repeated responses. However, typical participants tended to use strategies such as producing morphological variations, while participants with dementia produced more severe errors like perseverations and out-of-category errors. In conclusion, the dementia group shows the smaller semantic-letter discrepancy because of the raw differences of total responses between groups. Therefore, in the future study, this overall difference should be considered to determine whether the smaller discrepancy can be used as a sign of Alzheimer's dementia.

Acknowledgements

This research was supported by a grant from the John Templeton Foundation and the University of Chicago. Additional thanks goes to the developers of and contributors to DementiaBank, especially Davida Fromm for her help with data extraction.

References

Adlam, A. L., Bozeat, S., Arnold, R., Watson, P., & Hodges, J. R. (2006). Semantic knowledge in mild cognitive impairment and mild Alzheimer's disease. *Cortex, 42*(5), 675–684.

Bear, M.F., Conners, B.W., & Paradiso, M.A. (2015). *Neuroscience: Exploring the Brain.* (4th ed.) Lippincott, Williams, & Wilkes, Baltimore.

Becker et al. (1994). The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis. Archives of Neurology, 51(6), 585-594.

Capitani, E., Laicona, M., & Basso, A. (1998). Phonetically cued word-fluency, gender differences and aging: A reappraisal. *Cortex, 34*,779–783

Chen, P., Ratcliff, G., Belle, S. H., Cauley, J. A., DeKosky, S. T., & Ganguli, M. (2001). Patterns of cognitive decline in presymptomatic Alzheimer's disease: A prospective community study. *Archives of General Psychiatr*y, 58, 853– 858.

Gomez, R. G., & White, D. A. (2006). Using verbal fluency to detect very mild dementia of the Alzheimer type. *Archives of Clinical Neuropsychology, 21*(8), 771–775.

Gordon, J. K., Young, M., & Garcia, C. (2018). Why do older adults have difficulty with semantic fluency? Aging, *Neuropsychology & Cognition, 25*(6), 803–828.

Grundman M., Petersen R.C., Ferris S.H., Thomas R.G., Aisen P.S., Bennett D.A., Foster N.L., Jack C.R., Galasko D.R., Doody R., Kaye J., Sano M., Mohs R., Gauthier S., Kim H.T., Jin S., Schultz A.N., Schafer K., Mulnard R., van Dyck C.H., Mintzer J., Zamrini E.Y., Cahn-Weiner D., & Thal L.J. (2004). Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Archives of Neurology, 61*(1), 59-66.

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. *Neuropsychologia, 42*(9), 1212–1222.

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology, 18*(2), 284–295.

Kavé, G. (2015). Lifespan development of phonemic and semantic fluency: Universal increase, differential decrease. Journal of Clinical and Experimental *Neuropsychology, 37*(7), 751–763.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston naming test.* Philadelphia, PA: Lea & Febiger.

Laws, K. R. (2004). Gender differences in lexical size across categories. *Personality and Individual Differences, 36,* 23–32.

Laws, K. R., Duncan, A., & Gale, T. M. (2010). 'Normal' semantic-phonemic fluency discrepancy in Alzheimer's disease? a meta-analytic study. *Cortex, 46*, 595–601.

Marsolais Y., Perlbarg V., Benali H., Joanette Y.(2014). Age-related changes in functional network connectivity associated with high levels of verbal fluency performance. *Cortex, 58*, 123–138.

Mathuranath, P.S., George, A., Cherian, P.J., Alexander, A., Sarma, S.G., & Sarma, P.S., (2003). Effects of age, education and gender on verbal fluency. *Journal of Clinical and Experimental Neuropsychology, 25,* 1057–1064.

Meinzer, M., Flaisch, T., Wilser, L., Eulitz, C., Rockstroh, B., Conway, T., Rothi, L.J.G., & Crosson, B. (2009). Neural signatures of semantic and phonemic fluency in young and old adults. *Journal of Cognitive Neuroscience, 21*(10), 2007–2018.

Rodriguez-Aranda, C. & Martinussen, M. (2006). Age-related differences in performance of phonemic verbal fluency measured by Controlled Oral Word Association Task (COWAT): A meta-analytic study. *Developmental Neuropsychology. 30,* 697–717.

Stern Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society. 8,* 448–460.

Storandt, M., & Hill, R. D. (1989). Very mild senile dementia of the Alzheimer's type: II. Psychometric test. *Archives of Neurology, 46,* 383–386.

Suhr, J. A., & Jones, R. D. (1998). Letter and semantic fluency in Alzheimer's, Huntington's, and Parkinson's dementias. *Archives of Clinical Neuropsychology, 13,* 447–454.

Ting, S. K. S., Hameed, S., Earnest, A., & Tan, E. K. (2012). Dissociative semantic breakdown in Alzheimer's disease: Evidence from multiple category fluency test. *Clinical Neurology and Neurosurgery, 115*(7), 1049-1051.

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology, 14*(2), 167–177.

Troyer, A. K., Moscovitch, M., Winocur, G., Leach, L., & Freedman, M. (1998). Clustering and switching on verbal fluency tests in Alzheimer's and Parkinson's disease. *Journal of International Neuropsychological Society, 4,* 137–143.

Wechsler, D. (1955). *Manual of the Wechsler Adult Intelligence Scale*. New York, NY: The Psychological Corporation.

Wechsler, D. (1974). *Wechsler Memory Scale manual*. San Antonio, TX: The Psychological Corporation.