

Original Research Report

A Large-Scale Semantic Analysis of Verbal Fluency Across the Aging Spectrum: Data From the Canadian Longitudinal Study on Aging

Vanessa Taler, PhD,^{1,2,*} Brendan T. Johns, PhD,³ and Michael N. Jones, PhD⁴

¹School of Psychology, University of Ottawa, Canada. ²Bruyère Research Institute, Ottawa, Canada. ³Department of Communicative Disorders and Sciences, University at Buffalo, New York. ⁴Department of Psychological and Brain Sciences, Indiana University, Bloomington.

*Address correspondence to: Vanessa Taler, PhD, School of Psychology, University of Ottawa, 136 Jean Jacques Lussier Vanier Hall, Ottawa, Ontario, Canada K1N 6N5. E-mail: vtaler@uottawa.ca

Received: August 10, 2018; Editorial Decision Date: December 25, 2018

Decision Editor: Angela Gutches, PhD

Abstract

Objectives: The present study aimed to characterize changes in verbal fluency performance across the lifespan using data from the Canadian Longitudinal Study on Aging (CLSA).

Methods: We examined verbal fluency performance in a large sample of adults aged 45–85 ($n = 12,686$). Data are from the Tracking cohort of the CLSA. Participants completed a computer-assisted telephone interview that included an animal fluency task, in which they were asked to name as many animals as they could in 1 min. We employed a computational modeling approach to examine the factors driving performance on this task.

Results: We found that the sequence of items produced was best predicted by their semantic neighborhood, and that pairwise similarity accounted for most of the variance in participant analyses. Moreover, the total number of items produced declined slightly with age, and older participants produced items of higher frequency and denser semantic neighborhood than younger adults.

Discussion: These findings indicate subtle changes in the way people perform this task as they age. The use of computational models allowed for a large increase in the amount of variance accounted for in this data set over standard assessment types, providing important theoretical insights into the aging process.

Keywords: Aging, CLSA, Computational modeling, Verbal fluency

Verbal fluency is one of the tasks most commonly used in clinical practice to assess semantic and executive function. It has recently attracted theoretical interest as a central task to explore the mechanisms of memory search (e.g., [Abbot, Austerweil, & Griffiths, 2015](#); [Hills, Jones, & Todd, 2012](#)). In this task, the participant is asked to name as many items meeting a given criterion as they can in 1 min. Typically the criterion will be orthographic (words starting with a given letter, e.g., F) or semantic

(words falling into a given semantic category, e.g., *animals* or *vegetables*).

Traditionally, fluency performance is assessed simply in terms of total number of items produced. However, a more sophisticated approach was proposed by [Troyer, Moscovitch, and Winocur \(1997\)](#), who observed that fluency output is often organized in terms of semantic categories, and that verbal fluency performance usually entails the production of clusters of semantically related items (e.g.,

farm animals, pets; see the original article for a full list of items by category). Thus, in addition to the total items produced, verbal fluency performance can also be examined in terms of average cluster size and number of switches between clusters. Using this approach, Troyer, Moscovitch, Winocur, Leach, and Freedman (1998) were able to identify qualitative alterations in fluency output in different populations (e.g., Alzheimer's disease, Parkinson's disease).

Although the clustering/switching approach represented a major step forward in our understanding of verbal fluency performance, a number of drawbacks to this approach have been observed. Coding is time-consuming and reliability is an issue. Moreover, the scoring system is binary—items are scored as either belonging to a given cluster or not. Recently, a number of novel computational approaches have been developed to analyze and better understand semantic fluency output (e.g., Hills, Jones, & Todd, 2012). These approaches overcome some of the limitations of the clustering/switching approach. The uniqueness of these new computational approaches is that they are able to quantify the pathway that individuals take through memory. For example, these models have been used to analyze differences in fluency strategy between different populations, such as monolinguals versus bilinguals (Taler, Johns, Young, Sheppard, & Jones, 2013) and cognitively healthy older adults versus people who went on to develop mild cognitive impairment (Johns et al., 2018).

The novelty of the approaches used by Hills and colleagues (2012), and subsequently by Taler and colleagues (2013) and Johns and colleagues (2018), is that the models utilize semantic representations of words learned from the natural language environment. This class of model is referred to as a distributional model of semantics (for a review, see Jones, Wilits, & Dennis, 2015). Distributional models were originally developed as theoretical tools, and were designed to better understand the mechanisms with which semantic representations were constructed from lexical experience (e.g., Landauer & Dumais, 1997). However, they are increasingly being used to analyze various aspects of language and memory, such as episodic memory (Johns, Jones, & Mewhort, 2012, 2014; Mewhort, Shabahang, & Franklin, 2018), lexical organization (Hollis, 2017; Hsaoi & Nation, 2018; Johns et al., 2012; Johns, Dye, & Jones, 2016; Johns, Sheppard, Jones & Taler, 2016; Jones, Johns, & Recchia, 2012), and individual differences in language usage (Johns & Jamieson, 2018), among many others. The effort to combine theories of knowledge representation with theories of cognitive and memory processing have begun to push the boundaries of psychological theory—allowing researchers to match the behavior of models to people's behavior in laboratory studies at both the item and participant level.

Although prior work has demonstrated the promise of distributional models in understanding verbal fluency performance, it remains unclear how strongly the natural language environment, as encoded in the representations

of distributional models, accounts for fluency performance. A recent large-scale study, the Canadian Longitudinal Study of Aging (CLSA; Raina et al., 2009), has collected data on tens of thousands of participants, including a semantic (animal) fluency task (Tuokko, Griffith, Simard, & Taler, 2017). These data allow us to examine, at both the item and participant level, how variability in verbal fluency production reflects the semantic associations learned by distributional models. In the present study, we examine verbal fluency performance at a large scale, and determine the most important information sources underlying the behavioral data. We do this through the lens of a distributional model of semantics.

Although there are currently a number of competing process models of fluency (e.g., Abbot et al., 2015; Hills et al., 2012), in this study, we do not attempt to differentiate between process mechanisms. Rather, the goal is to determine the degree to which verbal fluency performance is connected at the population level to objective corpus-derived lexical variables (i.e., variables that are derived from the natural language environment). This knowledge will allow us to determine how much of the variance in this task can be explained by representation and how much must be accounted for by a prospective process model, a current goal within big data approaches to cognition (e.g., Johns, Mewhort, & Jones, 2017; Johns, Jones, & Mewhort, in press; Jones, Hill, & Todd, 2015; Jones, 2017).

In this study, the use of distributional modeling has both an applied and theoretical purpose. From an applied perspective, the study will demonstrate the utility of this modeling approach, which allows for lexical behavior to be easily quantified, enabling additional and unique information to be derived from a set of language-based data. Additionally, these analyses can be done at scale (i.e., at the population level) because they do not require human raters, meaning that they are appropriate to assess large-scale trends in human behavior.

However, the most important contribution will be theoretical. The basis of distributional modeling is that a person's accumulated episodic experiences of words results in the formation of a sophisticated representation of word meanings (Jones, 2018; Landauer & Dumais, 1997; McDonald & Shillcock, 2001). This class of theory has much in common with a recent proposal on cognitive aging by Ramscar, Hendrix, Shaoul, Milin, and Baayen (2014), who propose that much age-related cognitive decline (e.g., Deary et al., 2009) simply reflects the larger amount of information that older adults have accumulated in comparison to younger adults. By comparing older adults to younger adults, age effects are confounded with the amount of previously acquired information that must be processed, which is larger for older than younger adults. Ramscar, Sun, Hendrix, and Baayen (2017) recently empirically validated this approach.

The basis of the information accumulation perspective on aging is that learning does not end once adulthood

begins, but extends throughout the lifespan (Lindenberger, 2014). This is reflected in recent computational and empirical work on semantic memory (e.g., Dubossarsky, De Deyne, & Hills, 2017) and lexical organization (e.g., Johns, Sheppard, Jones, & Taler, 2016). Distributional models inherently embody this perspective, proposing that each episodic experience that a person has with language is used to update the representations contained in semantic memory (Jones, 2018; Landauer & Dumais, 1997; McDonald & Shillcock, 2001). This perspective is consistent with findings that knowledge tends to increase across the lifespan (e.g., Salthouse, 2003).

Combining the predictions of distributional modeling with the information accumulation perspective on aging, we predict that the changes in semantic memory that occur during aging should be observable with distributional models, due to increased language experience in older adults. The CLSA contains data from thousands of participants across a wide range of ages, meaning that CLSA verbal fluency data is ideal to conduct a large-scale analysis of the effect of accumulated experience on verbal fluency performance across the lifespan.

Method

We report here on the verbal fluency performance of a large group of participants from the CLSA, an ongoing long-term study consisting of a national stratified random sample of more than 50,000 Canadians aged 45–85 (Raina et al., 2009). Ethical review of the CLSA protocol was conducted by the Ethical, Legal, and Social Issues Committee, falling under the jurisdiction of the Canadian Institutes of Health Research (CIHR), and research ethics board approval was then acquired from each research site. Baseline testing has been completed for all participants. The present study uses baseline data from the Tracking cohort ($n > 20,000$) and was approved by the research ethics boards at Bruyère Research Institute and University of Ottawa. The current data request is part of the Psychological Working Group's proposal for access to CLSA data for preliminary analysis.

Participants

Detailed information on CLSA design and methodology are provided in Raina and colleagues (2009). Briefly, the CLSA is a 20-year prospective cohort study that includes >50,000 Canadian residents aged 45–85 years at baseline. Twenty-thousand of these participants provide data through telephone interviews (Tracking Cohort), while the remaining 30,000 participate in data collection through in-home interviews and assessment at a data collection site (Comprehensive Cohort). The present analysis includes participants in the Tracking Cohort who completed the animal fluency task in English. For the purposes of the present study, we excluded participants who reported a diagnosis of dementia or Alzheimer's disease, a memory

disorder, Parkinson's disease, or brain cancer, and those who reported having previously suffered a stroke or traumatic brain injury. This procedure left a total of 12,686 participants in the analysis.

Procedure

Detailed discussion of the psychological health component of the CLSA is available in Taler et al. (2016). As part of the cognitive battery, all participants completed an animal fluency task, in which they were asked to name all the animals that they could think of in 1 min. Descriptive analyses of overall performance in this task are available in Tuokko, Griffith, Simard, and Taler (2017). In our selected sample, the average age of the participants was 62.57 ($SD = 10.54$). Figure 1 shows the spread of the participant group; as can be seen, there is a wide sampling across the aging spectrum within the sample.

Model Description and Corpus

The distributional semantic model that will be used here is BEAGLE (Jones & Mewhort, 2007; Recchia, Sahlgren, Kanerva, & Jones, 2015). Broadly, the model works by encoding each word's meaning into a set of corresponding vectors. To train BEAGLE, each of the i unique words in a text corpus is represented by a unique n -dimensional environment vector e_i , with each element assigned a random deviate from a normal distribution with mean zero and variance $1/n$ (in the simulations that follow, dimensionality was set to $n = 1,024$). Environment vectors are stable over a simulation and are meant to serve as unique identifiers for the words in the corpus.

Next, the model "reads" the corpus one sentence at a time to build a semantic memory vector, m_i , for each word. The memory vector for each word is composed of two

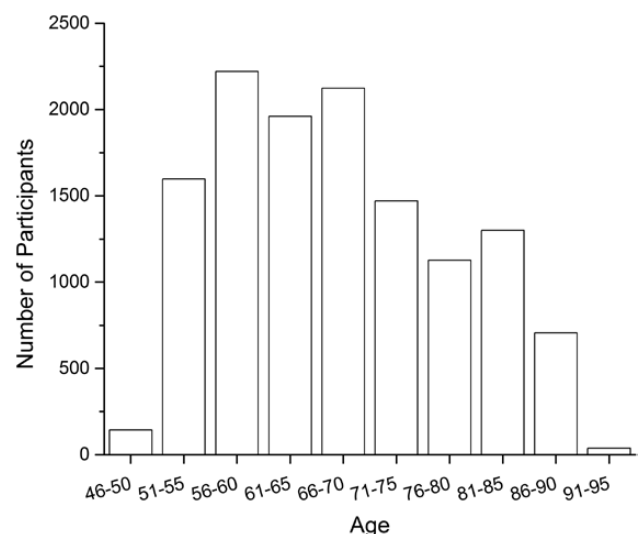


Figure 1. Age of participants in our study.

kinds of information: *context information* and *order information*. Context information is computed by summing the environmental vectors for all other words in the same sentence into the representation for that word. The summing of environmental vectors in this manner causes the memory vectors for all words in the same sentence to grow more similar to one another.

Order information encodes how a word is used within a sentence and is computed by encoding all of the n -grams (up to a certain size) to which a word belongs within a sentence. [The standard is to use n -grams up to size 7 (Jones & Mewhort, 2007); this standard will also be used here.] The computation of order-information relies on noncommutative circular convolution (Plate, 1995) to bind environmental vectors into unique n -gram vectors, which are then summed into the target word's order representation. The order representation encodes how a word is used in combination with others (see Jones & Mewhort, 2007 for a complete explication).

In sum, a word's context vector represents pure co-occurrence information, while order information is encoding simplified syntactic relations. The composite representation used here is the sum of a word's context and order representation; hence, it is a vector pattern that represents the word's history of co-occurrence with, and position relative to, other words. Despite its simplicity, BEAGLE explains a broad range of language phenomena (Johns et al., 2018; Jones & Mewhort, 2007; Recchia et al., 2015).

To train the model, a two-billion word corpus composed of fiction and nonfiction books was used. This corpus has previously been shown to give an excellent account of lexical behaviors (Johns, Jones, & Mewhort, 2016; Johns & Jamieson, in press). The corpus was preprocessed such that multiword animal names were concatenated into a single lemma in the corpus (e.g., "polar bear" was recoded as "polarbear"). Word frequency information was computed from the same corpus.

The main metric taken from BEAGLE will be the similarity between two words, calculated with a vector cosine (normalized dot product). Similarity will be used in different ways depending on the analysis being conducted.

Results

Given the large scale of data available from the CLSA, we were able to perform both a participant-level analysis and an item-level analysis. The goal of the participant-level analysis is to determine how well word-level lexical variables predict the changes in verbal fluency that occur with age. The goal of the item-level analysis is to determine how well word-level lexical properties predict word production probabilities.

Participant-Level Analysis

This analysis aims to determine the sensitivity of different variables to participant age, and to determine how the use of these variables in memory search changes with age. Two

semantic variables were included in this analysis: (a) the average pairwise similarity of the items a participant produced, and (b) the number of Troyer categories produced (Troyer et al., 1998; adopted from Hills et al., 2012). The comparison of these two variables will provide evidence as to whether distributional similarity provides any unique information over class techniques. Given that the use of word frequency is known to be an important variable in verbal fluency (e.g., Hills et al., 2012; Johns et al., 2018), the average log word frequency of the items produced was also included. Orthographic similarity (defined as average edit distance of items produced) was included in case there were sequential effects of orthographic overlap in verbal fluency production. Edit distance is the minimum number of letter insertions, deletions, and substitutions that would be required to transform one word into another, and has been shown to be a better metric of orthographic distance than traditional metrics (Yarkoni, Balota, & Yap, 2008). Finally, the number of items produced by a participant was also included.

Results are presented in Table 1. Overall, there is a negative correlation between age and number of items produced, indicating that older people produce fewer items in verbal fluency tasks. Additionally, there is a negative correlation between age and number of categories produced, consistent with the findings of Troyer and colleagues (1998). There is a positive correlation between age and pairwise semantic similarity as well as between age and average word frequency, indicating that with increasing age, participants generate items that are closer together in semantic space and of higher environmental frequency. We have found a similar pattern longitudinally in a group of older adults who went on to develop MCI (Johns et al., 2018), as well as in bilinguals who were required to switch between languages in a verbal fluency task (Taler et al., 2013). Finally, there was also a small negative correlation between average pairwise edit distance and age, signaling that there might also be a small change in the use of orthographic information across the age spectrum.

However, the table also shows that many of these variables are intercorrelated. To identify the variables that account for the most variance in category fluency, a stepwise regression analysis was performed where we calculated the amount of unique variance accounted for by each variable. The analysis we conducted is standard and provides a measure of the predictive gain (i.e., measured as percent ΔR^2 improvement) for one predictor over competing predictors (Adelman, Brown, & Quesada, 2006; Johns, Sheppard, Jones, & Taler, 2016). The overall fit of the regression was $R^2 = .14$ ($p < .001$). The amount of unique variance each variable accounts for is shown in Figure 2. Average pairwise semantic similarity accounts for the greatest amount of unique variance in predicting the age of a participant, followed closely by number of items the participant produced. The number of Troyer categories also accounted for a significant amount of variance, followed by average word frequency and average edit distance. That is, the semantic

Table 1. Correlations Between Fluency Variables and Age

	1	2	3	4	5
1. Age	—	—	—	—	—
2. No. of produced	-.312	—	—	—	—
3. No. of categories	-.268	.595	—	—	—
4. Word frequency	.193	-.514	-.189	—	—
5. Orthographic similarity	-.15	.278	.11	-.597	—
6. Semantic similarity	.285	.466	-.263	.739	-.468

Note. $N = 12,686$; all correlations significant at the $p < .001$ level.

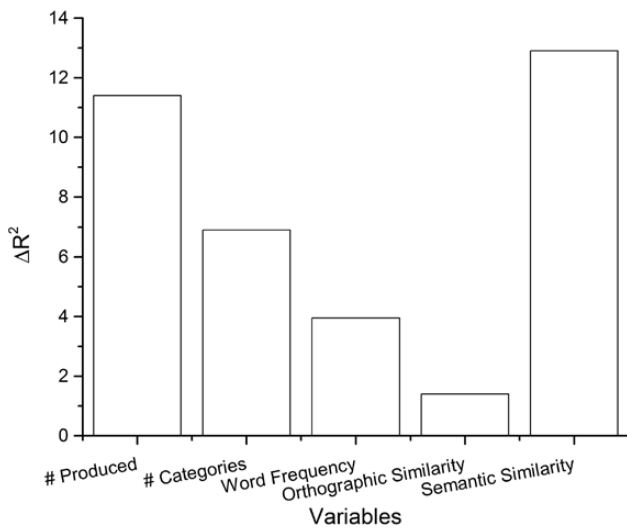


Figure 2. Results of the participant-level regression determining the most important lexical variables to account for age from verbal fluency performance.

model provides more unique power in predicting a person’s age from verbal fluency performance than the standard approaches used to score fluency data.

The finding that semantic similarity and word frequency are significant predictors of age is consistent with the information accumulation perspective of Ramscar and colleagues (2014, 2017) and the distributional perspective on aging. Together, these theories suggest that older participants’ lexical behavior should be more reflective of environmental information (as encoded in average word frequency and semantic similarity from a distributional model) than that of younger participants, given that older adults have lived longer and have accumulated more experience. To test this prediction, we split the data into two groups: an older group (aged over 60 years; $n = 6,764$) and a younger group (aged 60 years or younger; $n = 5,922$). For the older group, a correlation of $r(6,764) = .21, p < .001$ was found between average pairwise semantic similarity and participant age, and a correlation of $r(6,764) = .18, p < .001$ was found between average word frequency and participant age. For the younger group, correlations were weaker: a correlation of $r(5,922) = .11, p < .001$ was found between average pairwise semantic similarity and

participant age, and a correlation of $r(5,922) = .06, p < .001$ was found between average word frequency and participant age. Using a Fisher r -to- z test, it was confirmed that both of these correlations were significantly greater for the older subjects at the $p < .001$ level. This finding supports Ramscar et al.’s proposal that older participants have accumulated a greater amount lexical experience, resulting in their behaviors being relatively more tied to environment statistics than those of younger adults.

Next, an item-level analysis was conducted to gain a better understanding of the underlying variables that drive the production of words in a verbal fluency task, and how these change as a function of experience.

Item-Level Analysis

To eliminate incorrect responses, only words that were produced by two or more participants were included in this analysis. This procedure resulted in 1,306 animals being included in the analyses. The number of times each animal was produced was recorded. To visualize this frequency distribution, the frequency values were plotted on a Zipf (1935) scale (see also Ferrer i Cancho & Sole, 2003; Piantadosi, 2014; Zipf, 1949), where the x -axis corresponds to the rank of a word within the distribution and the y -axis corresponds to the frequency of that word (Figure 3). This figure shows that animal productions show a similar pattern to natural language word frequency, where there is a small number of high frequency productions (e.g., *dog, cat*) and a long tail of low frequency productions (e.g., *boa, koi*). The goal of this analysis is to determine which lexical information sources best map onto this pattern of word production. As is typical in examining word frequency (e.g., Adelman & Brown, 2008), production frequencies were transformed with a logarithm before analysis.

Three different lexical variables were included: (a) log word frequency, (b) semantic neighborhood size, and (c) orthographic neighborhood size. The main goal of the present study is to determine how well distributional models can distinguish older and younger adults’ differential lexical experience; these three variables were selected because they can be directly derived from corpora, and all were shown to predict unique variance in the participant-level analysis. Other variables, such as age of acquisition (AoA;

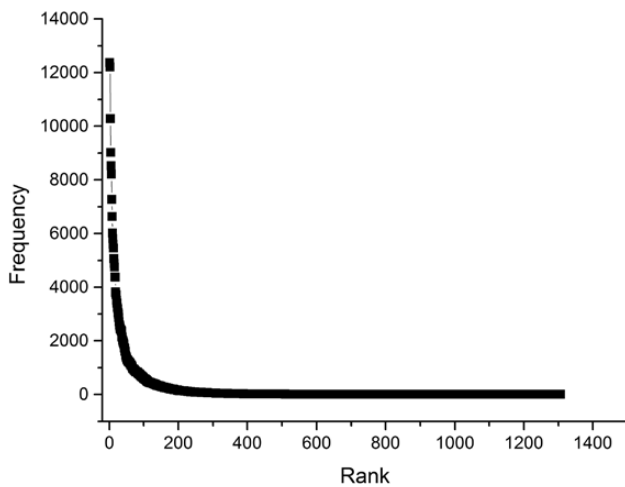


Figure 3. A Zipf scale of the frequency of productions from the CLSA data. This figure shows that production frequency in a verbal fluency task closely resembles word frequency distributions from natural language, with a small set of very high frequency words followed by a long tail of low frequency words.

Kuperman, Stadthagen-Gonzalez, & Gonzalez, 2012) were not included because they contained less than half of the words included in this analysis. For the AoA norms provided by Kuperman and colleagues (2012), only 593 of the 1,306 words used in this analysis were contained in the published norms, meaning that including this variable would provide an incomplete view of verbal fluency performance. We chose to use the standard orthographic edit distance, rather than phonological edit distance, in these analyses. There is a large degree of overlap between phonological and orthographic neighborhood metrics; for example, in the English Lexicon Project (Balota et al., 2007) the correlation between phonological and orthographic neighborhood was $r(40,480) = .78, p < .001$. Moreover, given that the current analysis concerns semantic verbal fluency, orthography, and phonology are likely to make very small contributions to performance (as seen in the participant-level analysis).

The correlation between log word frequency and log production was strong $r(1,305) = .54, p < .001$, demonstrating that animal names that are more common in natural language have a greater likelihood of being produced in a verbal fluency task, as expected.

Semantic neighborhood size was calculated by determining the number of associates a single word has above a set similarity value. This was computed by taking the pairwise similarity between one animal's representation and all other animals in the set, and determining the number of animals that exceed a certain parameter. To determine the best neighborhood parameter value, all similarity values from 0 to 1, in steps of 0.05, were tested, and the correlation between the log production frequency value and resulting semantic neighborhood size was taken. The result of this simulation is shown in the top panel of Figure 4. This figure shows that the best semantic parameter is a similarity

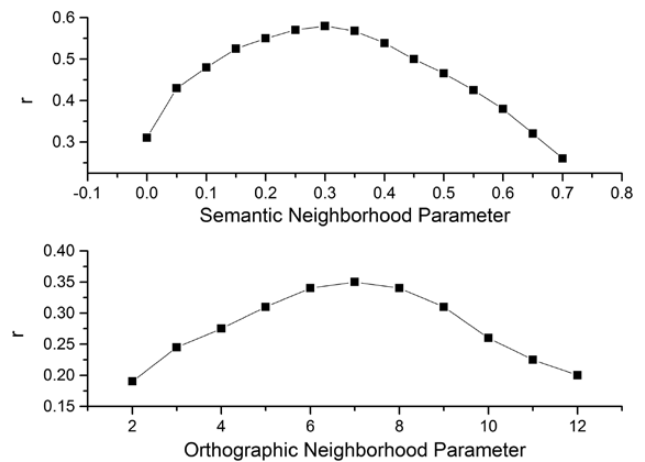


Figure 4. Results of the simulation for semantic and orthographic neighborhood size.

criterion of 0.3. At this parameter level, there is a correlation of $r(1,305) = .58, p < .001$, demonstrating both that the BEAGLE model provides a good account of semantic structure of the animal category, and that people are more likely to produce animals that are semantically related to a greater number of other animals. It is not possible to interpret the semantic neighborhood parameter value, because the similarity distributions that different distributional models construct are based on underlying assumptions about word representations (see Johns & Jones, 2010 for a systematic analysis). That is, different distributional models could have different semantic neighborhood values, but very similar semantic neighborhoods.

Orthographic neighborhood size was calculated in the same way as semantic neighborhood size, but with edit distance (Yarkoni et al., 2008) as the similarity metric. To count orthographic neighborhood size, the edit distance between each animal word was computed; the number of words that fall below a certain edit distance (controlled with a parameter) signals a word's orthographic neighborhood size. The optimal orthographic neighborhood size was determined by manipulating the orthographic neighborhood parameter from 2 to 12 in steps of 1, and the resulting simulation is contained in the bottom panel of Figure 4. This figure shows that the optimal edit distance was 7, with a maximal fit of $r(1,305) = .35, p < .001$.

However, there is likely a great deal of shared variance across these three variables. Thus, to measure the unique predictive power of these variables, we again used stepwise regression to quantify the unique variance accounted for by the three variables. The overall fit of the regression was $R^2(1,305) = .34, p < .001$. Figure 5 displays the results of this analysis. This figure shows that semantic neighborhood accounts for the lion's share of the variance in category production frequency, with word frequency accounting for some unique variance, while orthographic neighborhood does not account for any unique variance. This finding suggests that semantic category production is primarily driven

by the semantic relatedness of category exemplars, with baseline environmental frequency also playing a role.

The finding that semantic neighborhood size is the most important variable in predicting verbal fluency output is consistent with our findings in the participant-level analysis, where pairwise semantic similarity was found to be the most important predictor of the age of a participant. Together, these findings serve as an important step towards understanding the effects of lexical experience on verbal fluency during aging.

If the accumulation of distributional information underlies changes in semantic behavior across the lifespan, then the lexical behavior of adults of different ages should be better predicted by corpora originating from sources with differing lexical experience. For example, Johns and colleagues (in press) recently showed that the lexical decision reaction times of young adult participants are more in line with the lexical statistics contained in young adult novels, while older adults' lexical decision times were more strongly predicted by more advanced fiction novels.

One method of organizing lexical sources is to use time of publication (Johns & Jamieson, 2018). Given that language changes across time, it is likely that older participants have different lexical experience simply due to living during a different time. Specifically, older adults' behavior should be better predicted by corpora based on earlier writings, compared with younger adults. To examine the effect of time-specific lexical experience on verbal fluency performance, we conducted an additional item-level analysis. The first step was to organize some of the materials used to train the BEAGLE model described above into different time periods.

Given the size of the language database under question, it was infeasible to label each individual book by the year

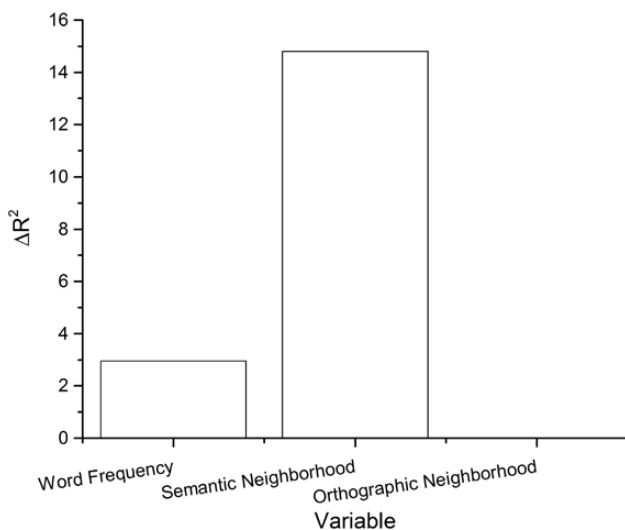


Figure 5. Results of the regression determining the most important lexical variable to account for production frequency in verbal fluency. Semantic neighborhood accounts for the most unique variance, followed by environmental word frequency.

of publication. However, it was possible to label the author of each book by their date of birth. Specifically, 2,088 authors had their date of birth recorded in the book corpus used here; birth year ranged from 1801 to 1998. The total corpus size of these authors was 1.5 billion words. Figure 6 presents the number of words by author year of birth. This figure shows that the majority of authors in this sample were born between 1925 and 1975. To enable a split for an *old* and *new* corpus, the materials were split at an author date of birth of 1942, resulting in *old* and *new* corpora of approximately equal size: the *new* corpus consisted of 792 million words, while the *old* corpus had 762 million words.

BEAGLE was trained on each corpus separately and new semantic neighborhoods were computed from each corpus using a similarity criterion of 0.3. To compare these values to the CLSA data, the verbal fluency frequency data were split into four quantiles from youngest to oldest. The splits that enabled the most equal splitting of the materials was: 44–53 ($n = 3,606$), 55–62 ($n = 3,153$), 63–71 ($n = 2,983$), and 71–88 ($n = 2,944$). Separate verbal fluency frequency distributions were then assembled from the data from these groups from the CLSA data.

To examine how well the semantic neighborhoods from the two corpora accounted for the data, a stepwise regression was conducted, calculating the amount of unique variance accounted for by the semantic neighborhood from the *new* corpus over the *old* corpus, and vice versa. This is an equivalent to the analysis presented in Figure 5 but for the two semantic neighborhood values. There was little variation in the overall fits of the regression of the first three age groups, with an $R^2(1,305) = .29, p < .001$ for the 44–53 group, an $R^2(1,305) = .30, p < .001$ for the 55–62 group, and an $R^2(1,305) = .30, p < .001$ for the 63–71 group. However, the fit for the oldest group of 72–88 was larger than the other groups, with an $R^2(1,305) = .34, p < .001$, suggesting that the semantic representations derived from the models provide a better fit to this group, corroborating the above finding that older adults' lexical behavior is more

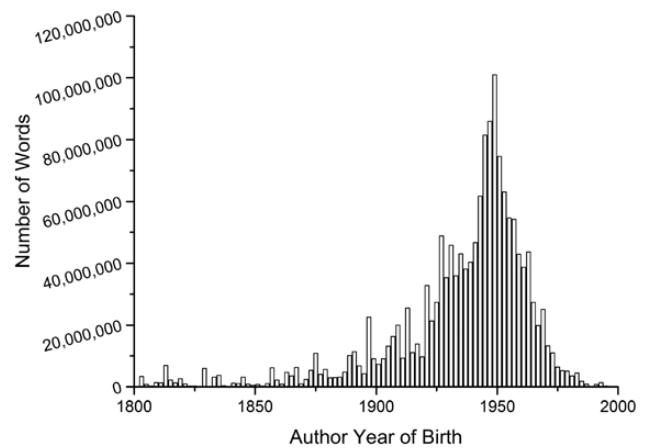


Figure 6. A histogram displaying the number of words contained in the book collection by author date of birth.

connected to lexical statistics attained from the natural language environment.

Figure 7 shows that there is a clear pattern in terms of the semantic neighborhood value that accounted for the bulk of these fits. This figure shows that as the age of the group increased there was also a constant increase in the variance accounted for by the *old* corpus, and a constant decrease in the variance accounted for by the *new* corpus. For the youngest group (44–53), the semantic neighborhood variable derived from the *new* corpus accounted for more variance (R^2 change = .017, $F = 33.15$, $p < .001$) than the semantic neighborhood values derived from the *old* corpus (R^2 change = .012, $F = 20.3$, $p < .001$). Beginning at the 55–62 group, this pattern switched, whereby the values derived from *old* corpus accounted for more variance (R^2 change = .018, $F = 29.9$, $p < .001$) than the values derived from the *new* corpus (R^2 change = .015, $F = 25.9$, $p < .001$). This pattern was strengthened in the 63–71 group, where the variance accounted for by the *old* corpus further increased (R^2 change = .019, $F = 30.5$, $p < .001$) compared with the *new* corpus (R^2 change = .013, $F = 24.3$, $p < .001$). This trend is the largest for the 72–88 group, where the semantic neighborhood from *old* corpus accounts for the bulk of the variance (R^2 change = .025, $F = 50.4$, $p < .001$), compared with the *new* corpus (R^2 change = .01, $F = 17.7$, $p < .001$). This finding demonstrates that the better a distributional model estimates participants' likely experience, the better the fit attained by that model, consistent with the distributional perspective on aging.

Discussion

Verbal fluency is a common task used for both clinical (Taler & Phillips, 2008) and theoretical (Hills et al., 2012) purposes. The Canadian Longitudinal Study of Aging (Raina et al., 2009) collected animal fluency from tens of thousands of participants, providing the opportunity to analyze

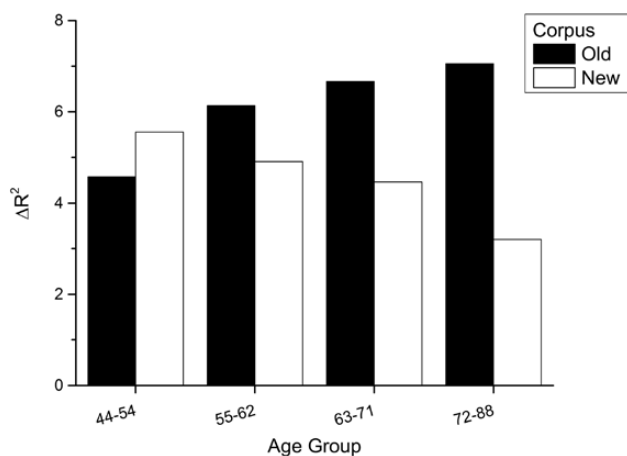


Figure 7. Amount of unique variance that the semantic neighborhood values from the old and new corpus accounts for across four age groups.

this type of data at a large scale. Using data at this scale enables an examination of this task at both the item and the participant level, allowing us to understand how variability in verbal fluency production reflects different lexical variables. Specifically, a common distributional model of semantics (the BEAGLE model; Jones & Mewhort, 2007; Recchia et al., 2015) was used to determine how well people's behavior on this task reflected the usage of words in the natural language environment.

At the participant level, several variables were assembled to examine how memory search changes across the age spectrum. The best predictor of aging was pairwise similarity (taken from BEAGLE), followed by total number of items produced, number of Troyer categories, and average item frequency. These results suggest that differences are observed in verbal fluency output across different age groups using environmentally-derived lexical variables.

At the item level, the best predictor of the frequency of an item's production was the word's semantic neighborhood, followed by its environmental frequency, suggesting that the words most likely to be produced are those that are commonly experienced and have a high similarity to many other animals in semantic space. Additionally, semantic neighborhood values derived from corpora organized by year of author birth differentially accounted for the verbal fluency distributions from different age groups, providing direct evidence that the different experience that participants have accumulated across their life is reflected in their verbal fluency performance.

At first glance, the finding that older participants tend to produce fewer items appears to provide evidence for cognitive decline across aging. However, this is not necessarily the case. The information accumulation perspective of Ramscar and colleagues (2014) suggests that the age-related slowing on many psychometric tests is not due to any systematic decline in cognitive processing, but instead reflects the accumulation of linguistic knowledge across time. Ramscar, Sun, Hendrix, and Baayen (2017) further demonstrated that this buildup results not only in stronger associations between items that commonly co-occur (e.g., *dog-wolf*), but also in a greater level of negative associations between words that do not commonly co-occur together (e.g., *dog-salmon*). The buildup of these negative associations would make switching to unrelated words in memory search much more unlikely as an individual ages: older adults would exhibit less switching between categories, and produce fewer items overall.

Thus, the changes observed in cognitively healthy adults in verbal fluency may not be due to any underlying cognitive decline, but rather to the effects of experience on the memory and language systems. Distributional models allow us to evaluate the impact of experience on language processing by examining the differences that are observed across the aging spectrum at both the participant and the item level. We have demonstrated significant effects of experience on verbal fluency performance, suggesting that

changes in fluency performance with age may reflect not only cognitive decline, but also the accumulation of lexical experience. Language is a dynamic system impacted by both experiential and cognitive factors, and isolating the contribution of one or the other system is challenging. However, the techniques used in this article shed light on potential avenues to examine the impact of experience on aging.

More generally, the results of both the item-level and participant level analysis point to the utility of both large-scale behavioral data sets and corpus-based cognitive models. Population-level data sets (such as the CLSA) provide a relatively complete picture of the variance contained in a given behavioral task, and corpus-based models (such as BEAGLE) allow for a determination of how representations derived from the natural language fit into this picture at both the item and the participant level. By combining the two approaches, a better understanding of human behavior can be attained than would be possible using either approach in isolation.

Funding

Funding for the Canadian Longitudinal Study on Aging (CLSA) is provided by the Government of Canada through the Canadian Institutes of Health Research (CIHR) under grant reference: LSA 9447 and the Canada Foundation for Innovation.

Acknowledgments

This research was made possible using the data/biospecimens collected by the Canadian Longitudinal Study on Aging (CLSA). This research has been conducted using the CLSA Baseline Tracking version 3.2, Baseline Comprehensive version 3, under Application Number 150813. The CLSA is led by Drs Parminder Raina, Christina Wolfson, and Susan Kirkland. The opinions expressed in this article are the authors' own and do not reflect the views of the Canadian Longitudinal Study on Aging.

Conflict of Interest

None reported.

References

- Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. *Psychological Review*, 122, 558–569. doi:10.1037/a0038693
- Adelman, J. S., & Brown, G. D. (2008). Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review*, 115, 214. doi:10.1037/0033-295X.115.1.214
- Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, 17, 814–823. doi:10.1111/j.1467-9280.2006.01787.x
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B.,...Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445–459. doi:10.3758/BF03193014
- Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E.,...Starr, J. M. (2009). Age-associated cognitive decline. *British Medical Bulletin*, 92, 135–152. doi:10.1093/bmb/ldp033
- Dubossarsky, H., De Deyne, S., & Hills, T. T. (2017). Quantifying the structure of free association networks across the life span. *Developmental Psychology*, 53, 1560–1570. doi:10.1037/dev0000347
- Ferrer i Cancho, R., & Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 788–791. doi:10.1073/pnas.0335980100
- Hills, T., Jones, M., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, 119, 431–440. doi:10.1037/a0027373
- Hollis, G. (2017). Estimating the average need of semantic knowledge from distributional semantic models. *Memory & Cognition*, 45, 1350–1370. doi:10.3758/s13421-017-0732-1
- Johns, B. T., Dye, M. W., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review*, 4, 1214–1220. doi:10.3758/s13423-015-0980-7
- Johns, B. T., & Jamieson, R. K. (2018). A large-scale analysis of variance in written language. *Cognitive Science*, 42, 1360–1374. doi:10.1111/cogs.12583
- Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, 17, 662–672. doi:10.3758/PBR.17.5.662
- Johns, B. T., Jones, M. N., & Mewhort, D. J. (2012). A synchronization account of false recognition. *Cognitive Psychology*, 65, 486–518. doi:10.1016/j.cogpsych.2012.07.002
- Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2014). A continuous source reinstatement model of true and illusory recollection. In: *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin TX: Cognitive Science Society.
- Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (in press). Using experimental optimization to build lexical representations. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-018-1501-2
- Johns, B. T., Mewhort, D. J. K., & Jones, M. N. (2017). Small worlds and big data: Examining the simplification assumption in cognitive modeling. In M. N. Jones (Ed.), *Big data in cognitive science: From methods to insights*. New York: Taylor & Francis.
- Johns, B. T., Sheppard, C. L., Jones, M. N., & Taler, V. (2016). The role of semantic diversity in word recognition across aging and bilingualism. *Frontiers in Psychology*, 7, 703. doi:10.3389/fpsyg.2016.00703
- Johns, B. T., Taler, V., Pisoni, D. B., Farlow, M. R., Hake, A. M., Kareken, D. A.,...Jones, M. N. (2018). Cognitive modeling as an interface between brain and behavior: Measuring the semantic decline in mild cognitive impairment. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, 72, 117–126. doi:10.1037/cep0000132
- Jones, M. N. (2017). Developing cognitive theory by mining large-scale naturalistic data. In M. N. Jones (Ed.), *Big data in cognitive science: From methods to insights*. New York: Taylor & Francis.

- Jones, M. N. (2018). When does abstraction occur in semantic memory: Insights from distributional models. *Language, Cognition and Neuroscience*. doi:10.1080/23273798.2018.1431679
- Jones, M. N., Hills, T. T., & Todd, P. M. (2015). Hidden processes in structural representations: A reply to Abbott, Austerweil, and Griffiths (2015). *Psychological Review*, 122, 570–574. doi:10.1037/a0039248
- Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, 66, 115–124. doi:10.1037/a0026727
- Jones, M. N., & Mewhort, D. J. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114, 1–37. doi:10.1037/0033-295X.114.1.1
- Jones, M. N., Willits, J. A., & Dennis, S. (2015). Models of semantic memory. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.) *Oxford handbook of mathematical and computational psychology* (pp. 232–254). New York: OUP.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44, 978–990. doi:10.3758/s13428-012-0210-4
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211. doi:10.1037/0033-295X.104.2.211
- Lindenberger, U. (2014). Human cognitive aging: corriger la fortune? *Science (New York, N.Y.)*, 346, 572–578. doi:10.1126/science.1254403
- McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: the neglected role of distributional information in lexical processing. *Language and Speech*, 44, 295–323. doi:10.1177/00238309010440030101
- Mewhort, D. J. K., Shabahang, K. D., & Franklin, D. R. J. (2018). Release from PI: An analysis and a model. *Psychonomic Bulletin & Review*, 25, 932–950. doi:10.3758/s13423-017-1327-3
- Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: a critical review and future directions. *Psychonomic Bulletin & Review*, 21, 1112–1130. doi:10.3758/s13423-014-0585-6
- Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks*, 6, 623–641. doi:10.1109/72.377968
- Raina, P. S., Wolfson, C., Kirkland, S. A., Griffith, L. E., Oremus, M., Patterson, C., ... Brazil, K. (2009). The Canadian longitudinal study on aging (CLSA). *Canadian Journal on Aging = La Revue Canadienne du Vieillessement*, 28, 221–229. doi:10.1017/S0714980809990055
- Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, 6, 5–42. doi:10.1111/tops.12078
- Ramscar, M., Sun, C. C., Hendrix, P., & Baayen, H. (2017). The mismeasurement of mind: Life-span changes in paired-associate-learning scores reflect the "cost" of learning, not cognitive decline. *Psychological Science*, 28, 1171–1179. doi:10.1177/0956797617706393
- Recchia, G., Sahlgren, M., Kanerva, P., & Jones, M. N. (2015). Encoding sequential information in semantic space models: Comparing holographic reduced representation and random permutation. *Computational Intelligence and Neuroscience*, 2015, 986574. doi:10.1155/2015/986574
- Salthouse, T. A. (2003). Interrelations of aging, knowledge, and cognitive performance. In Staudinger, U. M., & Lindenberger, U. (Eds.), *Understanding human development* (pp. 265–287). Boston, MA: Springer.
- Taler, V., Johns, B., Young, K., Sheppard, C., & Jones, M.N. (2013). A computational analysis of semantic structure in bilingual fluency. *Journal of Memory and Language*, 69, 607–618. doi:10.1080/13803390701550128
- Taler, V., & Phillips, N.A. (2008). Language performance in Alzheimer's disease and mild cognitive impairment: A comparative review. *Journal of Clinical and Experimental Neuropsychology*, 30, 501–556. doi:10.1080/13803390701550128
- Taler, V., Sheppard, C., Raina, P., Kirkland, S. (2016). Canadian Longitudinal Study on Aging: A Platform for Psychogeriatric Research. In: Pachana N. (Ed.), *Encyclopedia of Geropsychology*. Springer, Singapore. doi:10.1007/978-981-287-080-3_56-1
- Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, 11, 138–146. doi:10.1037//0894-4105.11.1.138
- 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 the International Neuropsychological Society*, 4, 137–143.
- Troyero, H., Griffith, L. E., Simard, M., & Taler, V. (2017). Cognitive measures in the Canadian Longitudinal Study on Aging. *The Clinical Neuropsychologist*, 31, 233–250. doi:10.1080/13854046.2016.1254279
- Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: a new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15, 971–979. doi:10.3758/PBR.15.5.971
- Zipf, G. K. (1935). *The psycho-biology of language*. Oxford, UK: Houghton, Mifflin.
- Zipf, G. K. (1949). *Human behaviour and the principle of least-effort*. Cambridge, MA: Addison-Wesley.