Journal of Memory and Language 69 (2013) 607–618

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

A computational analysis of semantic structure in bilingual verbal fluency performance

Vanessa Taler [a,b,*], Brendan T. Johns [c], Katherine Young [a,b], Christine Sheppard [a,b], Michael N. Jones [c]

[a] University of Ottawa, Canada
[b] Bruyère Research Institute, Canada
[c] Indiana University, Bloomington, United States

ARTICLE INFO

Article history:
Received 13 August 2012
revision received 19 August 2013
Available online 26 September 2013

Keywords:
Verbal fluency
Bilingualism
Semantic memory
Computational modeling

ABSTRACT

Groups of English monolingual and English–French bilingual participants completed letter and category fluency tasks, either only in English (monolinguals) or in English, French, free-switch and forced-switch conditions (bilinguals). Response patterns were modeled using a semantic space approach that estimates the weight of frequency and semantic similarity information in determining output patterns. Overall, semantic similarity had a stronger influence on output patterns than did frequency. In the forced English–French switching condition, the weight of similarity information was reduced and the weight of frequency information was increased, suggesting that the increased executive demands related to language switching result in alterations in the semantic structure of fluency output. Moreover, the frequency and similarity model parameters were negatively correlated in all tasks, indicating that they may be in competition during verbal fluency tasks.

Introduction

Bilingualism is extremely prevalent, with more than 50% of the world's population being bilingual or multilingual (Grosjean, 2008). Recent research has demonstrated differences in cognitive function between bilinguals and monolinguals, with bilinguals exhibiting lower performance than monolinguals in language-related tasks, but better performance on tasks of executive control (for a review, see Bialystok, 2009; Bialystok, Craik, & Luk, 2008). Verbal fluency tasks are among the most commonly used measures to assess language function in neuropsychological testing. These tasks require the participant to generate as many words as possible according to a specific criterion within a given time period (typically 1 min). Most commonly, the criterion is either a phonemic cue (e.g., words starting with the letter F) or a category (e.g., animals). Verbal fluency requires both language and executive function; the subject must organize verbal retrieval and recall, initiate responses, monitor prior responses, and inhibit inappropriate responses (Henry, Crawford, & Phillips, 2004). However, the demands on semantic knowledge vary depending on the criterion. Category fluency intrinsically requires rapid access to semantic knowledge. Letter fluency, in contrast, can be performed without access to semantic knowledge (Rohrer, Salmon, Wixted, & Paulsen, 1999), although output is typically influenced by semantic organization to some extent in healthy adults (Schwartz, Baldo, Graves, & Brugger, 2003).

One influential approach to analyzing fluency output examines measures of clustering and switching (Troyer, Moscovitch, & Winocur, 1997; Troyer, Moscovitch, Winocur, Leach, & Freedman, 1998). Clustering refers to the grouping together of items from a given subcategory; for example, a participant may produce a subgroup of farm

* Corresponding author. Address: School of Psychology, University of Ottawa, 136 Jean Jacques Lussier, Vanier Hall, Ottawa, Ontario K1 N 6N5, Canada.

E-mail address: vtaler@uottawa.ca (V. Taler).

0749-596X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jml.2013.08.004

animals, and then switch to African animals, and so on throughout the test period. By examining the number of clusters and the number of switches between clusters that a participant makes, a measure of the semantic coherency of the output can be measured. This approach has been successfully used to examine deficits in verbal fluency in various disorders, including Alzheimer's and Parkinson's disease (Troyer, Moscovitch, Winocur, Leach, et al., 1998), Huntington's disease (Ho et al., 2002), amyotrophic lateral sclerosis (Lepow et al., 2010), and traumatic brain injury (Zakzanis, McDonald, & Troyer, 2011). Based on data from patients with frontal and temporal lobe lesions, it has been claimed that clustering in normal fluency output reflects temporal lobe function, strongly related to semantic representations, while switching is related to frontal lobe function (Troyer, Moscovitch, Winocur, Alexander, & Stuss, 1998).

However, the traditional clustering procedure suffers from a number of drawbacks. First, a clustering analysis relies on subjective judgments of category membership, which brings into question issues of reliability and validity. Second, it does not allow a fine-grained analysis of fluency output, because a binary decision is made on each item (i.e., whether the item is a member of given category or not). Most category structures are more continuous than this, and so this need for a binary decision on each word may not accurately reflect actual category structure. Third, the procedure cannot differentiate between a deficit in switching and general decline in processing speed (Mayr, 2002), which recent research suggests is one of the major contributors to fluency performance (McDowd et al., 2011).

Research into verbal fluency across languages indicates that language of administration can affect performance. For example, higher fluency output is observed in Vietnamese relative to Spanish, possibly because animal names are typically monosyllabic in Vietnamese and multisyllabic in Spanish (Kempler, Teng, Dick, Taussig, & Davis, 1998). Moreover, bilinguals—both younger and older—typically exhibit lower category fluency performance relative to monolinguals, even when the task is administered in the participant's native language (Bialystok et al., 2008; Gollan, Montoya, & Werner, 2002; Portocarrero, Burright, & Donovick, 2007; Rosselli et al., 2002). In balanced French–English bilinguals, similar performance across languages has been reported (Roberts & Le Dorze, 1997), and in Spanish–English bilinguals and English monolinguals, similar animal subcategories and semantic associations were produced in all groups and languages (Rosselli et al., 2002). In letter fluency, some studies report lower performance in bilinguals than monolinguals (Bialystok et al., 2008; Gollan et al., 2002), while others report similar performance in the two language groups (Portocarrero et al., 2007; Rosselli et al., 2000). Vocabulary size also affects fluency performance in bilinguals: those with relatively higher vocabularies outperform those with lower vocabularies on letter (but not category) fluency (Luo, Luk, & Bialystok, 2010).

A few studies have examined bilingual fluency output in terms of semantic variables. A greater number of subcategory exemplars—that is, increased semantic clustering—was observed in the native language in French–English and Spanish–English bilinguals (Roberts & Le Dorze, 1997; Rosselli et al., 2002). Salvatierra, Rosselli, Acevedo, and Duara (2007) also found more semantic clusters (defined as two or more consecutive items from the same subcategory) in Spanish than in English, both in normal older adults and in patients with Alzheimer's disease (AD). However, de Picciotto and Friedland (2001) found no differences across language conditions in Afrikaans–English bilingual older adults and AD patients. It should be noted, however, that these studies did not provide detailed information about the semantic clusters produced by participants, nor how the output was coded.

Gollan et al. (2002) tested Spanish–English bilingual college students using both an English-only and a free-choice condition (i.e., the participant could say words in both languages, using whichever words came to mind). Surprisingly, performance did not improve in the both-languages condition relative to the English-only condition, although participants did make use of both languages. The authors interpret these results in terms of both interference between languages in bilinguals, and weaker links between semantic and phonemic representations in bilinguals relative to monolinguals. These weaker links are postulated to result from reduced use of word forms in both languages, relative to monolingual speakers of that language. This reduced use also leads to differences in frequency effects in bilinguals and monolinguals. For example, bilinguals differ from monolinguals in a semantic association task when the strongest associate is of low frequency (e.g., in response to "bride" they produce "dress", while monolinguals produce "groom"), but perform similarly to monolinguals when the associate is of high frequency (Antón-Méndez & Gollan, 2010). In picture naming, bilinguals show the greatest disadvantage in low-frequency items, although this effect is weaker in older adults in the non-dominant language (Gollan, Montoya, Cera, & Sandoval, 2008). This "weaker-links hypothesis" (Gollan et al., 2008) predicts that use of frequency information should differ between bilinguals and monolinguals in verbal fluency tasks.

In order to disentangle the different mechanisms potentially contributing to the effects of bilingualism on verbal fluency, Sandoval, Gollan, Ferreira, and Salmon (2010) conducted a time-course analysis of fluency output in monolinguals and bilinguals. They found that bilinguals' lexical retrieval during the fluency task was delayed relative to monolinguals; bilinguals produced more cognates and lower-frequency items than monolinguals; and cross-language intrusion occurred when bilinguals were tested in the non-dominant language, but not when testing was in the dominant language. These results were taken to indicate that between-language interference plays a major role in the verbal fluency disadvantage observed in bilinguals.

In the present paper, we use a novel approach to explore the effects of frequency and semantic relatedness on fluency output in bilinguals, employing similarity structure learned from a distributional model of lexical semantics. Distributional models do not use the traditional approach to coding semantic similarity based on hand-coded relationships between words (e.g., Collins & Quillian, 1972; Fellbaum, 1998). Rather, the models infer a word's meaning by observing its statistical co-occurrences with other words in linguistic contexts, learned from a large corpus of text.

A variety of different algorithms have been proposed to infer distributional semantics (for a review, see Bullinaria & Levy, 2007; McRae & Jones, in press; Riordan & Jones, 2011), but all are based on the simple idea that the set of linguistic contexts in which a word does and does not appear provides a set of mutual constraints with which to infer its similarity to other words (Harris, 1970; Landauer & Dumais, 1997). For example, robin and egg may become related because they tend to co-occur frequently within contexts. Robin and sparrow, in contrast, become related because they frequently appear in similar contexts (with the same words), even if they rarely co-occur directly. Ostrich may be less related to robin due to a lower overlap of their contexts compared to sparrow, and stapler is likely to have very little contextual overlap with robin. Distributional models have successfully explained a range of semantic behaviors, including synonymy judgments (Landauer & Dumais, 1997), free association norms (Griffiths, Steyvers, & Tenenbaum, 2007), typicality effects (Jones & Mewhort, 2007), and semantic priming (Jones, Kintsch, & Mewhort, 2006).

The present study used a distributional model to characterize English–French bilinguals' category fluency output under different conditions. Fluency data were collected in each of the bilingual's languages (French and English); a mixed-language condition, in which participants provided responses in whichever language they preferred; and a forced-switch condition, in which participants alternated between French and English (e.g., "cat", "chien" [dog], "frog", etc.). These conditions differ in terms of their cognitive demands: single-language output entails inhibition of the language not in use, while free-switch does not, and the forced-switch condition places greater demands on executive resources than other conditions, while maintaining similar semantic requirements. We used a computational approach to model semantic structure in each task, allowing us to quantify the relative importance of frequency and semantic information in the different conditions.

Unlike other modeling approaches to category fluency, such as optimal foraging models (Hills, Jones, & Todd, 2012) or random walk models (Abbott, Austerweil, & Griffiths, 2012), the present method was designed to examine the differential information used in memory search in bilinguals. A similar approach has been taken by Johns et al. (2013) to examine changes in the use of semantic memory seen in the development of Alzheimer's disease. There has been a wide range of research within the bilingual word retrieval literature (for a review, see Degani & Tokowicz, 2010), that can provide insights into the expected outcomes of the modeling work.

The most common distinction in theories of bilingual word retrieval is between selective-access accounts (e.g. Gerard & Scarborough, 1989), and nonselective-access accounts (e.g. de Groot & Delmaar, 2000). Selective accounts assume that individuals can inhibit the non-target language. In terms of category fluency, such a model predicts that the search spaces used across languages are separate and should not impede each other. In contrast, nonselective accounts assume that all linguistic information sources across languages are activated simultaneously, and compete for selection. For category fluency, this type of access model would predict that there should be interference between languages, particularly when one language must be selected over another.

Given that nonselective-access accounts are currently the dominant class of theories in bilingual word retrieval (Degani & Tokowicz, 2010), we expect to find an increase in the amount of interference between languages when language information must be actively selected. In the four category fluency conditions used here, only the forced-switch condition requires active selection of language information: interference would be manifested in this condition because it requires individuals to iteratively inhibit one language and sample from the other. This requirement would lead to a change in searching strategies, because the equal activation of both languages from semantic context would be harmful when switching between languages, due to the need to inhibit the non-selected language. A selective model would not make this prediction, because it would assume that people are capable of sampling from the different languages independently, meaning that interference between languages should be minimal.

However, this prediction does not specify the form of the change that would be expected when there is an interaction between the two languages. One possibility comes from reordered-access models (Duffy, Morris, & Rayner, 1988). These models propose that word retrieval (typically measured in terms of ambiguity resolution within a sentence processing task) specifies that words are activated both by the preceding context—including semantics—and by frequency information. Reordered-access models have also been proposed to account for bilingual ambiguity resolution (Arêas Da Luz Fontes & Schwartz, 2010; Degani, Prior, & Tokowicz, 2011; Degani & Tokowicz, 2010). In the realm of category fluency, these models predict that the semantic context and frequency of category words should drive the searching process. However, as discussed, when one is required to switch between languages, activation from semantic context becomes harmful in forced switching due to the nonselective nature of the activation process, which leads to interference from the non-target language. The presence of this interference would alter the processes that people use to retrieve words from the lexicon. Specifically, because semantic context causes interference, the use of semantic information should be reduced, which in turn will increase the importance of other linguistic variables that are used in retrieval (e.g. frequency). Thus, a reordered nonselective-access based model predicts a decrease in the use of semantics and an increase in the use of frequency information when bilinguals must actively switch between languages. A selective access model would predict that the strategies should not change a great deal, because it is possible to sample from each lexicon independently.

The free-switch condition will also provide diagnostic information to differentiate between selective and nonselective accounts. Specifically, if semantic context is used to activate both languages, a nonselective account would predict that there should be no (or minimal) difference in the processing strategy that is used in this condition vs. the single language tasks, because both languages can be equally selected based on context. In contrast, a selective account would assume that the two languages would need to be sampled independently, indicating that the use of semantic context should be reduced, while use of frequency information should increase.

By examining parameter changes that dictate the use of semantics and frequency across the four fluency tasks described above, it is possible to test these predictions with the computational model described below.

Methods

Participants

Two groups of participants took part in the present study: monolingual ($n = 32$) and bilingual ($n = 38$) young adults. Monolingual participants spoke only English, and bilinguals spoke French and English but no other languages. All participants had good self-reported health and no neurological or psychiatric history. Bilingual participants acquired a high degree of proficiency in both English and French before age 13. They provided a self-reported rating, on a 5-point Likert scale, of their English and French proficiency in the areas of auditory comprehension, reading, speaking and writing (1 = no ability and 5 = native-like ability). Mean self-reported proficiencies for all modalities in both languages are provided in Table 1; self-rated proficiency in English was equal to or higher than French proficiency in 35 of 38 bilingual participants. Participants were recruited in the Ottawa–Gatineau region through advertising and word of mouth. Demographic and neuropsychological data for each group are provided in Table 2.

Table 2
Participants' demographic, neuropsychological, and language proficiency characteristics (reported as mean ± standard deviation).

	Monolinguals	Bilinguals	*p*-Value
N	32 (men = 15)	38 (men = 15)	
Age	21.63 ± 1.56	21.53 ± 2.31	.83
Education	15.56 ± 1.08	15.50 ± 1.57	.84
Animacy judgment (CV)			
English	0.21 ± 0.09[a]	0.25 ± 0.14[c]	.09
French	N/A	0.29 ± 0.12[c]	
MoCA	28.50 ± 1.52	28.00 ± 1.32	0.15
Digit Span (/30)	17.94 ± 4.25	19.52 ± 4.43	.05[*]
WCST (categories)	4.19 ± 1.12	4.64 ± 0.68[4]	.05[*]
LM (/50)			
Immediate	27.94 ± 4.73[a]	28.94 ± 7.49[d]	.51
Delay	25.83 ± 4.38[b]	25.47 ± 7.99[d]	.82
Stroop 1 (words)	108.52 ± 17.33[a]	107.11 ± 13.92[d]	.72
Stroop 2 (colours)	78.35 ± 15.03[a]	76.49 ± 11.15[e]	.57
Stroop 3 (interference)	52.16 ± 12.73[a]	53.41 ± 11.59[e]	.68
CVLT (*t*-score)	52.25 ± 8.98	51.08 ± 8.30[d]	.58
BNT	53.25 ± 3.30	49.43 ± 6.92	.004[**]
FAS	40.28 ± 13.53	37.37 ± 10.33	.32
Category fluency	24.28 ± 5.82	24.03 ± 6.93[e]	.87

[a] $n = 31$.
[b] $n = 30$.
[c] $n = 34$.
[d] $n = 36$.
[e] $n = 37$.
[*] $p < 0.05$.
[**] $p < 0.01$.

Measure of bilingualism

Participants' proficiency in English and French was further assessed using the animacy judgment task developed by Segalowitz and Frenkiel-Fishman (2005). In this task, participants are asked to decide whether stimuli are living (animate) or non-living (inanimate). Stimuli comprise 32 animate and 32 inanimate nouns in each language, with no translation equivalents in each language set. Monolingual participants completed only the English block, while bilingual participants completed first the English block and then the French block. The task was run using E-Prime software (Version 2.0) on a Dell laptop computer running Windows XP. We then calculated the coefficient of variability (CV) for each language administration by dividing the mean standard deviation of response time for correct trials by the mean response time for correct trials in each language. This measure is taken to reflect cognitive efficiency, due to reduced variability when processing is relatively more automatic, even when average response latencies are the same (for a discussion, see Segalowitz & Frenkiel-Fishman, 2005).

Neuropsychological battery

All participants completed a neuropsychological battery that included the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005); the forward and backward digit span subtests of the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; Wechsler, 1997); the Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948); a version of the Stroop color-word interference test in which the

Table 1
Mean ranking (±standard deviation) for proficiency by modality for both English and French for bilingual participants. Ranking followed a 5 point Likert scale (1 = no ability; 5 = native-like ability). Monolingual participants were not asked to rank their language proficiency, as they were assumed to be a "1" in French and a "5" in English, across all modalities. ($n = 37$; data missing for one bilingual participant).

	English language proficiency	French language proficiency	*p*-Value
Auditory comprehension	4.92 ± 0.28	4.77 ± 0.48	.07
Reading	4.92 ± 0.28	4.62 ± 0.59	.003
Speaking	4.92 ± 0.28	4.46 ± 0.65	<.001
Writing	4.76 ± 0.49	4.13 ± 0.92	<.001

number of items produced in 45 s was recorded in each condition; the Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 1983); and letter (FAS) and category (animal) verbal fluency tasks. Scores by participant group are provided in Table 2. The two groups did not differ on any demographic variables. Bilinguals outperformed monolinguals on the WCST and Digit Span tests, and monolinguals outperformed bilinguals on the BNT.

Procedure

Monolingual participants completed a single 1-min category fluency trial, in which they were asked to name as many animals as they could, without repeating items, in 1 min. Bilingual participants completed four separate 1-min category fluency trials. In each trial, they were asked to name as many animals as they could in 1 min without repeating items. In all cases, output was recorded using Audacity and subsequently transcribed. The trials were: (1) English-only; (2) French-only; (3) free-switch, in which participants could produce items in whichever language they preferred; and (4) forced-switch, in which participants had to switch between English and French on each item, without producing translation equivalents. All participants also completed three letter fluency trials (F, A, S). In the letter fluency trials, participants were asked to produce as many words as they could in 1 min that began with the specified letter, without repeating items, providing numbers, proper nouns or words with the same root (e.g., sing, singing). Monolinguals completed letter fluency trials in English, and bilinguals completed trials in three conditions: English, French, and free-switch. Thus, monolinguals completed four trials in total (animal, F, A, and S) and bilinguals completed 13 (animals—English, animals—French, animals—free-switch, animals—forced-switch; and F, A, and S in English, French, and free-switch).

For both category and letter fluency, all English, French, and free-switch trials were completed in a different randomized order for each participant, and the forced-switch animal category fluency trials were always completed after the free-switch animal category fluency trials (separated by letter and vegetable-musical instrument switching trials). Participants were remunerated for their participation. The study received ethical approval from the Research Ethics Board at the Bruyère Research Institute, Ottawa, Ontario, Canada, and participants signed an informed consent form prior to the study.

Computational analyses

Participants' output was transcribed and French items were translated into English to allow comparison across conditions. Our computational model of category fluency is based on early work by Romney, Brewer, and Batchelder (1993) predicting semantic clustering in free recall. Romney et al. presented participants with multiple words from various natural categories, and examined the classic effect of semantic clustering in free recall (Bousfield, 1953). A separate group of participants rated the semantic similarity of all pairs of words used on the list, and Romney et al. conducted multidimensional scaling on the similarity matrix to estimate the semantic space their participants traversed when recalling items. They then evaluated various cognitive process models in their ability to generate the sequences of items produced by participants in free recall in this estimated similarity space. Irrespective of presentation organization, items tend to be recalled in sequences related to their semantic similarity.

Given the large number of animals produced by our participants, pairwise similarity ratings are infeasible (Romney et al., 1993, were only able to collect ratings on 41 items across all categories). In free recall, the experimenter selects the set of words present on the list; hence the response set is predetermined. In contrast, category fluency allows the participant to generate the responses without any experimenter control over the items that are produced. In addition, using similarity ratings raises the validity issue of predicting behavior from behavior, and does not afford a general method by which we could generalize our process model to new participants who produce items that we have not normed. Rather, to model our semantic space we use a distributional model trained on a large corpus of text. The space we use in the present study has been validated on other tasks that utilize semantic similarity (e.g., Hills et al., 2012; Johns et al., 2013; Jones & Mewhort, 2007), and provides a representation for almost any animal name that could be produced by a participant. We then evaluate various process models to explain how participants are mentally moving through this space, in a similar spirit to Romney et al. (1993). These process models allow us to explore various information sources that may be used by participants to generate their responses in the task, and examine how these sources shift across groups and fluency instructions.

Representation

To represent the semantic space of animals, we constructed a context co-occurrence vector for each possible animal word. Using a corpus of 250,000 documents from Wikipedia[1] (Willits, D'Mello, Duran, & Olney, 2007), a target word's vector has 250,000 elements—the element is coded as 1 if the word occurs in the document (irrespective of frequency), and 0 otherwise. This type of representation is commonly referred to as a context vector (Dennis & Humphreys, 2001), and has been consistently shown to produce excellent fits to human semantic similarity ratings, similar to pointwise mutual information (Bullinaria & Levy, 2007; Recchia & Jones, 2009). This representation type has been used to examine a number of semantic behaviors, including recognition memory (Johns, Jones, & Mewhort, 2012) and perceptual inference (Johns & Jones, 2012), lending credence to this approach. The set of target words consisted of all produced items in the behavioral data.

The similarity between two words is then computed as the cosine of the angle between their vectors. Functionally, this behaves very much like a correlation coefficient. The cosine between a word and itself is one, and the expected

[1] The corpus was modified such that multiword animal names were concatenated into a single lemma in the corpus (e.g., "polar bear" was recoded as "polarbear").

cosine between two randomly selected words is zero (e.g., *sparrow-stapler*). Due to the use of only positive values in this representation type, the cosine will range from 0 (no contextual overlap) to 1 (complete contextual overlap). As in Romney et al. (1993), our semantic space for the category of animals is then the (symmetric) similarity matrix of all animal pairs.

In addition to extracting semantic information about a word, the frequency of animal names across the corpus was also computed. Frequency reflects the ease of lexical access, and accounts for unique variance beyond semantic organization in other lexical tasks (Forster & Chambers, 1973; Murray & Forster, 2004). The natural logarithm of word frequency will be used here, as it has been demonstrated to provide a close correspondence to human lexical access latency (Adelman & Brown, 2008). It should be noted that frequency and semantic similarity measures are on different scales, which may make the resulting analysis of these values difficult to compare. To partially address this concern, a normalization procedure was conducted, whereby each value was divided by the maximum corresponding value from the word set, standardizing both scales between 0 and 1.

A process model for verbal fluency

Following the earlier work of Romney et al. (1993; see also, Hills et al., 2012) we model the path taken by participants through semantic space using Luce's (1959, 1977) choice axiom, a ubiquitous decision rule in cognitive psychology and economics. The axiom defines how humans select an item from possible alternatives (in our case, the word produced from all the words that could have been produced). Given a set of stimulus similarities, Luce's axiom states that the probability of responding to stimulus S_i with response R_j is defined as:

$$P(R_j|S_i) = \frac{\beta_j^x S(i,j)^y}{\sum_{k=1}^{n} \beta_k^x S(i,k)^y}, \tag{1}$$

where β_j is the response bias for item j, and $S(i,j)$ is the similarity between item i and j. The parameters x and y control the relative contributions of base rate (frequency) and similarity in producing the response (both are positive real values). In our simulations, the "stimulus" is the previous word produced, and the "response" is the word that is about to be produced (from the set of alternatives that could be produced). The set of alternatives considered in the denominator at time t in a participant's trial is the full set of items produced across all participants, omitting items produced by the participant prior to time t.

For example, if the participant produced the transition *dog–cat*, the model would predict this transition as highly likely and determined by both semantic similarity and frequency. A transition of *dog–hyena* would be based more on semantic similarity than frequency, and a transition of *dog–robin* would be based more on frequency than semantic similarity. These would be reflected by the parameters that control attention to similarity and frequency, respectively. A transition of *dog–iguana* would be best predicted by a model that reduces both x and y close to zero (which is essentially a model that selects items at random, but ignores similarity and frequency). Hence, if the current word is referred to as the "origin" and the subsequent word as the "destination," the Luce model may be illustrated for our purposes as:

$$P(\text{destination}|\text{origin}) = \frac{\mathit{freq}(\text{destination})^x \times \mathit{similarity}(\text{origin}, \text{destination})^y}{\sum_{k \varepsilon \text{Animals}} \mathit{freq}(k)^x \times \mathit{similarity}(\text{origin}, k)^y}. \tag{2}$$

For each participant's sequence of items produced in each task, we determine the most likely set of parameters that would have generated the observed data if the model were correct. We compare a variety of nested models in their goodness-of-fit to the fluency data, each representing a potential cognitive process that may have generated the data. The simplest model, used as our baseline measure of chance, simply has x and y fixed to zero—this model essentially predicts that word sequences will be selected at random, ignoring both frequency and semantic similarity. The performance of the other higher parameter cognitive models will be compared to this baseline model to assess their ability to explain the data beyond chance.

The second and third models allow one of the two parameters to be estimated from the data. In the "frequency only" model, the x parameter (controlling attention to frequency) is free to be estimated from the data, and the y parameter (controlling attention to semantic similarity) is fixed to zero, forcing the model to produce items based on frequency but ignoring semantic similarity. In the "similarity only" model, the y parameter is free to vary and the x parameter is fixed. Each of these models is expected to outperform the random model because they each have an additional free parameter over the baseline model controlling attention to additional information—in the limit, each one-parameter model should be able to reproduce any pattern of data generated by the baseline model.

The final two models each allow both of the control parameters to be estimated from the data. These two-parameter models propose that both frequency and semantic similarity are necessary to produce the fluency data, but the relative use of the information sources may differ across tasks and language groups. The first of these models is essentially the full Luce rule presented above: a fluency sequence is the product of frequency and similarity, each raised to the power of a cognitive control parameter. This type of control process is at the heart of most successful models of higher order memory retrieval (e.g., Anderson, 1993; Raaijmakers & Shiffrin, 1981). The second version of a two-parameter model uses an exponential sensitivity function rather than a power function, (e.g., $exp^{y \cdot Sim(i,j)}$ rather than $Sim(i,j)^y$ in Eq. (1)). Whether internal psychological representations of stimulus intensities (frequency or similarity) are transformed via a power law or an exponential function has been the subject of recent controversy (see Heathcote, Brown, & Mewhort, 2000). Hence, both two-parameter models will be explored here. The use of exponential scaling is based on Shepard's law of scaling (Shepard, 1987). In an exponential model, the control parameter will control how much scaling is taking place – the higher the value, the greater the impact of high similarity words on the search operation. The importance of exponential similarity has also been seen in other areas in the study of language, such as lexical access (Jones, Johns, & Recchia, 2012). Both models allow for differential weighting of the two information sources when generating sequences of items.

Parameter estimation and model comparison

Parameters were fit for each participant under each of the above models using maximum likelihood estimation (Myung, 2003). Specifically, a grid-search algorithm was used to find the optimal set of parameters to maximize the log-likelihood that the model generated the data. All parameter values between 0 and 20, in steps of 0.25 were tested. The value of 20 was selected because no subject exceeded this value, for either the semantic similarity or the frequency parameter. This method allows the best-fitting parameters of a single model for a particular individual to be determined.

Models were compared using the Akaike information criterion (AIC; Akaike, 1974), a standard and reliable method to compare models' ability to quantitatively fit human data (Shiffrin, Lee, Kim, & Wagenmakers, 2008). The AIC compares the quantitative fit of a model to human data (based on log-likelihood), intrinsically penalizing models as a function of the number of free parameters. Models with the lowest AIC value are preferred, and we use this value to select among the different proposed models. For more details on model comparison techniques, we refer the reader to Busemeyer and Diederich (2010) or Lewandowsky and Farrell (2011).

Parameters were fit to each individual and, in the case of the bilingual participants, across all four conditions separately. This analysis allows us to determine the change in parameter values across the different fluency tasks for bilinguals. If there is no shift in the searching process due to the different task demands, no changes in the parameters are expected. However, if there is a change in the type of information being used in the search process then this change should be apparent in a corresponding change in the parameters of the model. There is a small difference in the number of monolingual and bilingual participants (with each population having a fairly large number of participants), but this imbalance is unimportant because parameters are calculated at the individual participant level.

Results

Mean total correct responses for letter and category fluency by condition and group are shown in Fig. 1. Performance across conditions within bilinguals was analyzed using a repeated-measures ANOVA. When applicable, comparisons between monolingual and bilingual performance were analyzed by substituting the appropriate between-subjects error term and controlling for familywise error rate. In animal fluency, bilinguals exhibited a significant effect of condition on number of items produced, $F(3,69) = 15.46$, $p < .001$. LSD posthoc tests indicated significant differences between forced-switch and all other conditions ($p < .01$ in all cases) and between English and French ($p < .01$); marginal differences were observed between English and free-switch ($p = .085$) and between French and free-switch ($p = .063$). It is noteworthy that most participants (25 of 38) in fact used both languages in the free-switch condition; overall, 66% of items were produced in English and 34% in French. An independent-samples *t*-test revealed no difference between monolinguals and bilinguals in the English condition ($p = 0.64$).

In letter fluency, bilinguals exhibited a significant effect of condition on number of items produced, $F(2,74) = 13.12$, $p < .001$. LSD post hoc tests indicated significant differences between French administration and all other conditions ($p < .01$ in all cases). An independent-samples *t*-test revealed no difference between monolinguals and bilinguals in the English condition ($p = 0.32$).

In order to rule out the possibility of order of administration effects, a series of one-way ANOVAs was conducted with total correct items in each condition as the dependent variables, and sequential order of the trial as the independent variable. No effect of order of administration was observed in any of the conditions ($p > 0.1$ in all cases).

Model fits

The AIC fits of the four models described previously, along with a baseline random model (where each word has an equal probability of occurring), for both the monolingual subjects and the four bilingual conditions are displayed in Fig. 2. As the figure shows, all four models provide a large improvement over the random model.[2] The semantics-only model gave a consistently better fit than the frequency-only model, indicating that the semantic connection between adjacent words is a more important source of information than frequency in a semantic fluency task. Moreover, this figure also shows that the multiplicative model gave a consistently better fit than either of the individual models on their own, across all conditions.

Additionally, the multiplicative exponential model provided a modest gain in AIC values in four of the five conditions (approximately 5.0 on average), indicating that this transformation provides a better fit to the data, likely because this model provides a more realistic internal scaling of similarity and frequency values. However, it is worth noting that in the forced-switch condition the two models fit the data equally well. This finding may indicate a processing difference in this task relative to the other three conditions, in addition to the parameter differences that are described below. As discussed previously, the superior fit for the exponential model is not surprising, given the previous research on the importance of scaling and exponential transformations across multiple cognitive tasks (Heathcote et al., 2000; Jones et al., 2012; Shepard, 1987). This transformation increases the strength between more similar words, and reduces the strength among dissimilar words, in a nonlinear fashion. The finding that this transformation provides a superior fit suggests that the semantic search spaces that are used in fluency tasks are

[2] A smaller AIC indicates a better quantitative fit.

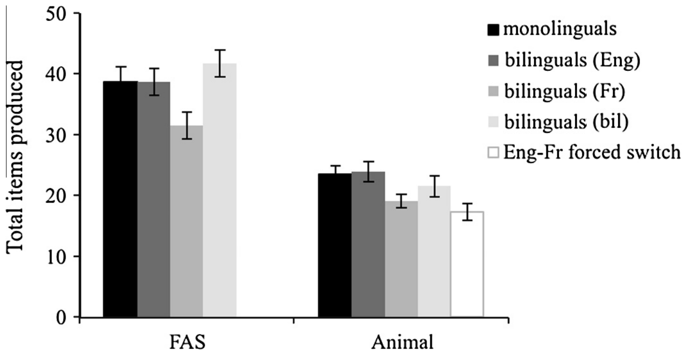

Fig. 1. Mean total items produced by condition. Eng = English administration; Fr = French administration; bil = bilingual (free switch) administration; Eng–Fr = English–French.

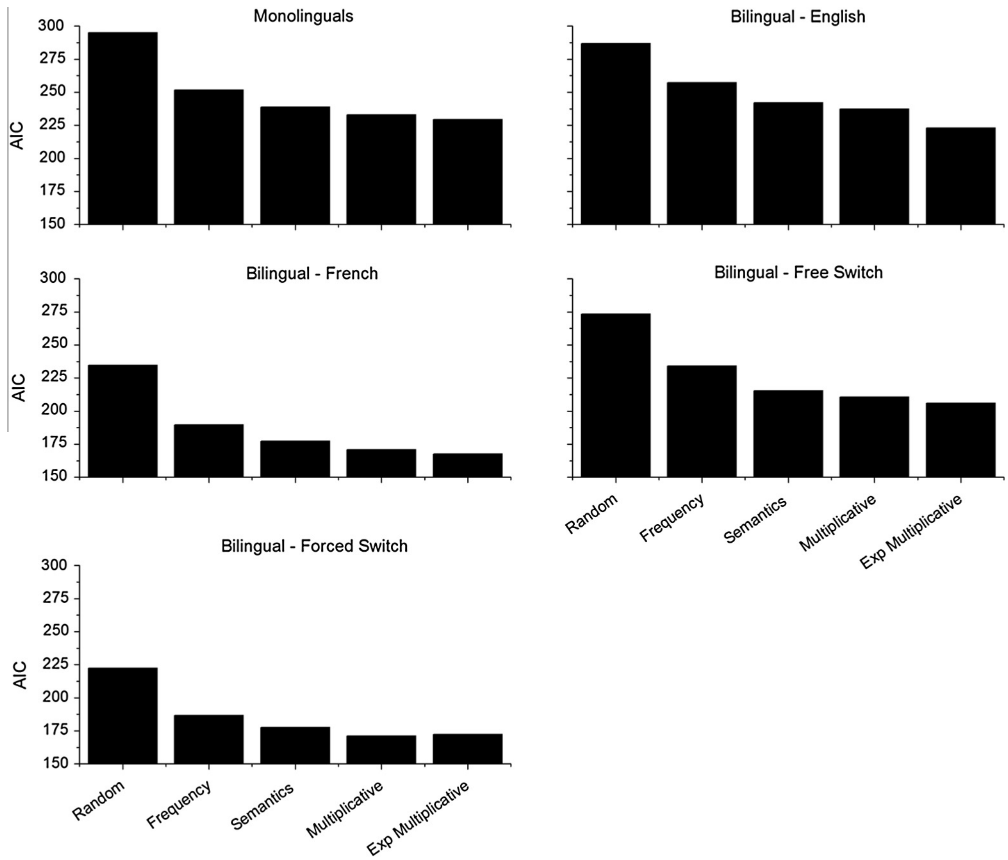

Fig. 2. Fit of the four different models across the five tasks. This includes a random model for comparison purposes, and this demonstrates that all models perform better than chance. Additionally, it was found that the exponential multiplicative model is the best performing model across 4 of the 5 tasks.

biased towards selecting high similarity items. Because the exponential multiplicative model provides the best fit to these data, it is used in all subsequent analyses.

Individual parameter change by condition

To examine the effect of condition (English, French, free-switch, forced-switch) on fluency output, the parameter values for each participant were computed across each condition for the category fluency data; mean parameter values are displayed in Fig. 3. The main result of this analysis is that the importance of semantic similarity is lower and the importance of frequency is higher in the forced-switch task relative to other tasks.

A 4 (fluency conditions) × 2 (parameters) repeated measures ANOVA was used to compare changes in parameter values across the different fluency tasks, with the dif-

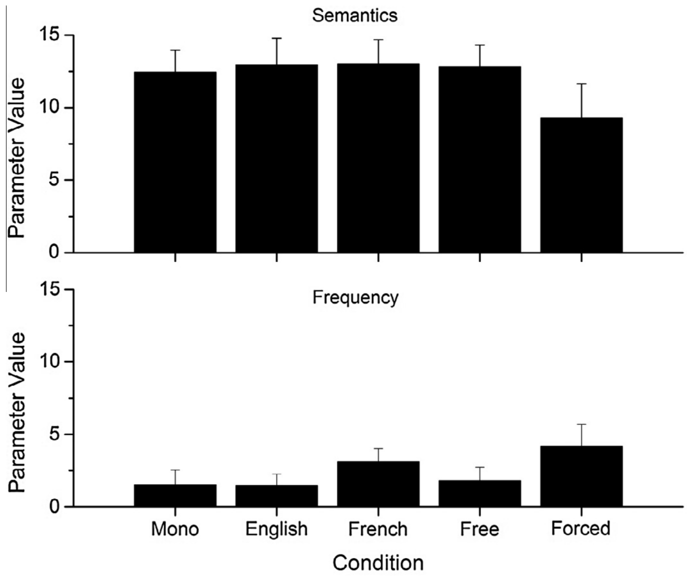

Fig. 3. The averaged individual parameter values across the different tasks. This figure shows that the importance of the different information sources switches for the forced switch task, with lower importance of semantic information and higher importance of frequency information.

ferent fluency conditions being the independent variable and the parameter values being the dependent variables. There was a main effect of fluency condition, $F(3,99) = 2.99$, $p < .05$, and also of parameter type, $F(1,33) = 297.41$, $p < .001$. The large difference for parameter type is due to the semantic parameter being greater than the frequency parameter for almost every subject, across all conditions. Additionally, there was a significant interaction effect, $F(3,99) = 7.492$, $p < .001$, suggesting that there is a change in parameter values across the different tasks.

To examine the interaction between parameters and fluency tasks, an LSD post hoc test was done. For the semantic parameter, the value in the forced-switch condition was lower than the value found across the other three conditions ($p < .01$). No other differences were found for this parameter. The frequency parameter for the forced-switch task was significantly greater than both the English-only and free-switch tasks at the .001 level, and was significantly different from the French-only task at the .05 level. Additionally, the parameter value for the French-only task was slightly significantly greater than the values found for the English-only and Free-switch task ($p < .05$).

A between-subjects analysis was conducted to compare the parameter values for monolinguals and bilinguals, in order to determine if there were any differences in searching processes between these two participant groups. This was accomplished with a 2 (participant type) × 2 (parameter value) multivariate ANOVA, with participant type being the independent variable and the parameter values being the dependent variable. No main effect was found for participant type ($F(1,68) = 0.002$, *n.s.*), nor for parameter value ($F(1,68) = 0.002$, *n.s.*), suggesting that there is no difference in the searching process used by monolinguals and bilinguals in the English condition.

These results suggest that when the task requires an individual to actively select among languages (i.e. in the forced-switch condition), the blending of information sources used in the searching operation is modified. As suggested by nonselective reordered access models of word retrieval (Arêas Da Luz Fontes & Schwartz, 2010; Degani & Tokowicz, 2010), this pattern of parameter change likely occurs due to semantic context activating both languages equally. This activation means that semantic context causes interference in selecting a specific target language. Thus, semantic information is inhibited in the forced-switch task, while base-rate frequency becomes more important in the sampling process, due to both languages being activated by the semantic context.

In order to examine how the two parameters interact across the conditions, a correlational analysis was con-

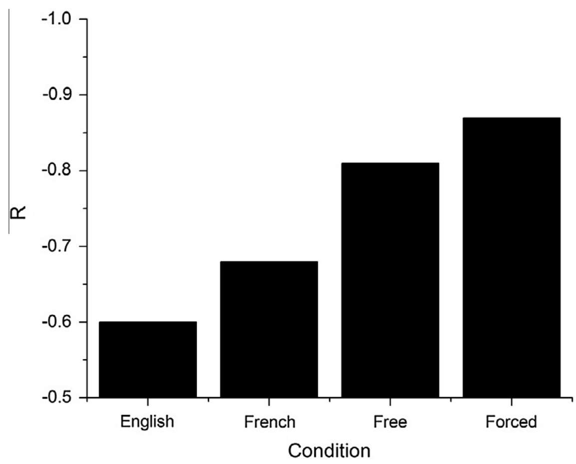

Fig. 4. Negative correlations between the similarity and frequency parameters, in the different bilingual conditions.

ducted. Fig. 4 presents the Pearson correlation of the two parameter values across the different conditions. This analysis allows us to measure the inter-dependency of the two information sources, by examining how they are interacting across the different fluency conditions. The first notable result of this analysis is that the parameters are negatively correlated in all conditions. That is, as one information source increases in prominence, the other tends to decrease. The second observation is that the magnitude of the correlation is altered across conditions: the weakest correlation is in the English-only task ($r = -0.38$, $p < .05$) and the strongest is in the forced-switch task ($r = -0.8$, $p < .001$). The correlation coefficients were compared using a Fisher r-to-z transformation (Cohen & Cohen, 1975), and a corresponding z-test indicated that difference between these two correlations was significant ($p < .01$). This finding suggests that there is competition between these information sources, and that competition is stronger in the more difficult tasks. Thus, when one source becomes less important in the searching process (e.g. semantics in the forced-switch condition), the other source compensates by becoming relatively more important.

Discussion

Overall, similar semantic and letter fluency performance was observed in monolingual and bilingual participants, both when responses were required in English and when participants could switch freely between English and French. These findings are in contrast to previous research reporting lower category fluency performance in bilingual than monolingual speakers (Bialystok et al., 2008; Gollan et al., 2002; Portocarrero et al., 2007; Rosselli et al., 2002), possibly reflecting the high proficiency of the bilingual speakers in our sample. Previous research has used Spanish–English bilinguals in the United States (Gollan et al., 2002; Portocarrero et al., 2007; Rosselli et al., 2000) or mixed samples with different native languages and ages of English acquisition (Bialystok et al., 2008). Differences in demographic and language characteristics between our participants and those studied in previous research may thus account for these conflicting findings.

The literature is more mixed with respect to the effects of bilingualism on letter fluency performance, with some studies finding lower performance in bilingual than monolingual participants (Bialystok et al., 2008; Gollan et al., 2002), while others, as in the present study, find no difference between language groups. (Portocarrero et al., 2007; Rosselli et al., 2000). In French, however, bilingual participants' performance was significantly worse than that of monolinguals for both semantic and letter fluency, and forced-switch (English–French) category fluency performance was lower than all other administrations, with the exception of French-only administration. This finding likely reflects the very high English proficiency of our participant sample.

The best fit to the category fluency data was provided by an exponential multiplicative model that used frequency and semantic similarity information to model performance. Even though this transformation provided only a moderate gain over a multiplicative power model, it suggests that higher similarity exerts an exaggerated influence on output over lower similarity words, indicating that participants are biased towards searching within a limited semantic space around a word (consistent with Shepard's, 1987, exponential scaling law of similarity). Overall, the best-fitting parameters suggest that participants relied more heavily on semantic similarity than frequency information in their output; that is, as expected, participants produced items that are semantically related to the previous item, and were more likely to produce a lower-frequency related item than a higher-frequency unrelated item (although output is also influenced by frequency). As predicted, there is a slight increase in the importance of frequency information in the French-only administration relative to English-only, although this effect did not reach significance; that is, the structure of output in French was similar to English and free-switch conditions, although the total number of items produced was lower.

Notably, a clear difference was observed in the forced-switch task relative to the other conditions: participants produced fewer items, reduced their reliance on semantic information, and increased their reliance on frequency information (although the overall pattern of a greater reliance on semantic than frequency information persisted). This finding indicates that the increased executive demands elicited by switching between languages led to alterations in the semantic structure of the output. The most plausible explanation of this effect is given by a non-selective reordered access model, which is described in greater detail below.

A second important finding relates to the links between frequency and similarity information. In all conditions, a significant negative correlation was observed between the two information sources; thus, a decrease in reliance on semantic information in high-demand conditions (French-only or forced-switch) leads to a concomitant increase in reliance on frequency information. This finding suggests that the two information sources are in competition—alterations in fluency output in different populations (e.g., unbalanced bilinguals, patients with cognitive impairment) may occur through differing relative importance of these two parameters. That is, semantic context and frequency information both play a role in fluency, but they also seem to be used in a competitive fashion, such that when executive demands are highest, frequency increases in importance while semantic similarity decreases. Different populations may use these two information sources differently, and cognitive modeling techniques such as the ones described here may be used to reveal important differences in memory search strategies. In our lab, we are currently collecting norms for English and French verbal fluency performance in different language groups with the goal of further exploring this finding.

It should be noted that the research reported here focuses on a subgroup of highly proficient English–French bilinguals, and the findings may not generalize to other groups of bilinguals or language combinations. This issue should be explored in future research. Moreover, frequency and semantic similarity information for computational modeling of French and switching data was based on English translation equivalents for the French terms. While this information may vary in French, recent analyses in computational linguistics suggest that type and token frequency (Bybee & Hopper, 2001) and semantic similarity among items (Ploux & Hyungsuk, 2003) are both surprisingly consistent between English and French. Similarly, studies that have examined free association to targets between languages (*dog*-___ vs. *chien*-___) have found remarkable consistency in items produced, irrespective of the language in which the task was performed (Rosenzweig, 1961).

To revisit models of word retrieval, the current research supports a nonselective, reordered-access based model (Degani & Tokowicz, 2010). This model proposes that during retrieval, words from both languages are activated across both languages, based upon meaning frequency. As discussed previously, this model predicts that there should be an interaction between languages, but only when one language needs to be actively selected, as in the forced-switch fluency task. More importantly, it was also predicted that, because both languages are activated by semantic context, this information source should become less important in the searching task, while base-rate frequency should become more important (suggesting a more random sampling process). This prediction is similar to the pattern produced by our participants, and suggests that the retrieval operations that occur during lexical access and sentence processing may apply to other language retrieval tasks, such as category fluency.

Given the success of general psycholinguistic theories in explaining the use of different information sources in category fluency described here, and the recent interest in computational models of this task (e.g. Johns et al., 2013; Abbot et al., 2012; Hills et al., 2012), it may be possible to integrate these two approaches to form more complete models of language search. Word retrieval models propose the type of information that is important in memory search, while computational models formally specify how the search process is conducted. By integrating the two approaches a better understanding of the language system can be attained across a number of different tasks.

The present findings elucidate the roles of different information sources in fluency performance across conditions that vary in terms of executive demands, while holding semantic demands constant. Word frequency and semantic similarity information appear to play competing roles in semantic fluency production, with semantic information playing a less important role when executive demands are highest. Language switching fluency tasks may thus be a useful experimental paradigm to examine the interaction between languages in bilinguals, above and beyond information provided by single-language administration, especially when combined with cognitive modeling techniques.

Acknowledgments

The present research was supported by a Young Investigator Award to V.T. from the Alzheimer Society of Canada and by NSF BCS-1056744 to M.J. We thank Jenna Schultz, Chloé Corbeil and Kathryne Holmes for assistance in recruitment, data collection, and coding.

References

Abbott, J., Austerweil, J., & Griffiths, T. (2012). Human memory search as a random walk in a semantic network. *Advances in Neural Information Processing Systems, 25*, 3050–3058.

Adelman, J. S., & Brown, G. D. A. (2008). Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review, 115*, 214–227.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Antón-Méndez, I., & Gollan, T. H. (2010). Not just semantics: Strong frequency and weak cognate effects on semantic association in bilinguals. *Memory & Cognition, 38*, 723–739.

Arêas Da Luz Fontes, A. B., & Schwartz, A. I. (2010). On a different plane: Cross-language effects on the conceptual representations of within-language homonyms. *Language and Cognitive Processes, 25*, 508–532.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*, 3–11.

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 859–873.

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology, 49*, 229–240.

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods, 39*, 510–526.

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Los Angeles, CA: Sage.

Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure: Typological Studies in Language* (Vol. 45). Amsterdam: John Benjamin.

Collins, A. M., & Quillian, M. R. (1972). How to make a language user. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York, NY: Academic Press.

de Groot, A. M. B., & Delmaar, P. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory. *Quarterly Journal of Experimental Psychology, 53A*, 397–428.

de Picciotto, J., & Friedland, D. (2001). Verbal fluency in elderly bilingual speakers: Normative data and preliminary application to Alzheimer's disease. *Folia Phoniatrica et Logopaedica, 53*, 145–152.

Degani, T., Prior, A., & Tokowicz, N. (2011). Bidirectional transfer: The effect of sharing a translation. *Journal of Cognitive Psychology, 23*, 18–28.

Degani, T., & Tokowicz, N. (2010). Semantic ambiguity within and across languages: An integrative review. *Quarterly Journal of Experimental Psychology, 63*, 1266–1303.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*, 452–478.

Duffy, S., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27*, 429–446.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge, Massachusetts: MIT Press.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627–635.

Gerard, L. D., & Scarborough, D. L. (1989). Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 305–315.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58*, 787–814.

Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish–English bilinguals. *Neuropsychology, 16*, 562–576.

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology* (38), 404–411.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114*, 211–244.

Grosjean, F. (2008). *Studying bilinguals*. Oxford: Oxford University Press.

Harris, Z. (1970). Distributional structure. In *Papers in structural and transformational linguistics* (pp. 775–794). Dordrecht, Holland: D. Reidel Publishing Company.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7*, 185–207.

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*, 1212–1222.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review, 119*, 431–440.

Ho, A. K., Sahakian, B. J., Robbins, T. W., Barker, R. A., Rosser, A. E., & Hodges, J. R. (2002). Verbal fluency in Huntington's disease: A longitudinal analysis of phonemic and semantic clustering and switching. *Neuropsychologia, 40*, 1277–1284.

Johns, B. T., Taler, V., Pisoni, D. B., Farlow, M. R., Hake, A. M., Kareken, D. A., et al. (2013). Using cognitive models to investigate the temporal dynamics of semantic memory impairments in the development of Alzheimer's disease. In *Proceedings of the 12th international conference on cognitive modeling* (ICCM).

Johns, B. T., & Jones, M. N. (2012). Perceptual Inference through global lexical similarity. *Topics in Cognitive Science, 4*, 103–120.

Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2012). A synchronization account of false recognition. *Cognitive Psychology, 65*, 486–518.

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology, 66*, 121–132.

Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language, 55*, 534–552.

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114*, 1–37.

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

Kempler, D., Teng, E. L., Dick, M., Taussig, I. M., & Davis, D. S. (1998). The effects of age, education, and ethnicity on verbal fluency. *Journal of the International Neuropsychological Society, 4*, 531–538.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211–240.

Lepow, L., Van Sweringen, J., Strutt, A. M., Jawaid, A., MacAdam, C., Harati, Y., et al. (2010). Frontal and temporal lobe involvement on verbal fluency measures in amyotrophic lateral sclerosis. *Journal of Clinical and Experimental Neuropsychology, 32*, 913–922.

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. Sage.

Luce, D. R. (1959). *Individual choice behavior*. New York: John Wiley & Sons.

Luce, D. R. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology, 15*, 215–233.

Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition, 114*, 29–41.

Mayr, U. (2002). On the dissociation between clustering and switching in verbal fluency: Comment on Troyer, Moscovitch, Winocur, Alexander and Stuss. *Neuropsychologia, 40*, 526–566.

McDowd, J., Hoffman, L., Rozek, E., Lyons, K. E., Pahwa, R., Burns, J., et al. (2011). Understanding verbal fluency in healthy aging, Alzheimer's disease, and Parkinson's disease. *Neuropsychology, 25*, 210–225.

McRae, K., & Jones, M. N. (in press). Semantic memory. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology*.

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review, 111*, 721–756.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology, 47*, 90–100.

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatric Society, 53*, 695–699.

Ploux, S., & Hyungsuk, J. (2003). A model for matching semantic maps between languages (French/English, English/French). *Computational Linguistics, 29*, 1–24.

Portocarrero, J. S., Burright, R. G., & Donovick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology, 22*, 415–422.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.

Recchia, G. L., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information to latent semantic analysis. *Behavior Research Methods, 41*, 657–663.

Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science, 3*, 303–345.

Roberts, P. M., & Le Dorze, G. (1997). Semantic organization, strategy use, and productivity in bilingual semantic verbal fluency. *Brain and Language, 59*, 412–449.

Rohrer, D., Salmon, D. P., Wixted, J. T., & Paulsen, J. S. (1999). The disparate effects of Alzheimer's disease and Huntington's disease on semantic memory. *Neuropsychology, 13*, 381–388.

Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science, 4*, 28–34.

Rosenzweig, M. R. (1961). Comparisons among word-association responses in English, French, German, and Italian. *American Journal of Psychology, 74*, 347–360.

Rosselli, M., Ardila, A., Araujo, K., Weekes, V. A., Caracciolo, V., Padilla, M., et al. (2000). Verbal fluency and repetition skills in healthy older Spanish–English bilinguals. *Applied Neuropsychology, 7*, 17–24.

Rosselli, M., Ardila, A., Salvatierra, J., Marquez, M., Matos, L., & Weekes, V. A. (2002). A cross-linguistic comparison of verbal fluency tests. *International Journal of Neuroscience, 112*, 759–776.

Salvatierra, J., Rosselli, M., Acevedo, A., & Duara, R. (2007). Verbal fluency in bilingual Spanish/English Alzheimer's disease patients. *American Journal of Alzheimer's Disease and Other Dementias, 22*, 190–201.

Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition, 13*, 231–252.

Schwartz, S., Baldo, J., Graves, R. E., & Brugger, P. (2003). Pervasive influence of semantics in letter and category fluency: A multidimensional approach. *Brain and Language, 87*, 400–411.

Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory & Cognition, 33*, 644–653.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*, 1317–1323.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E. J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science, 32*, 1248–1284.

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.

Troyer, A. K., Moscovitch, M., Winocur, G., Alexander, M. P., & Stuss, D. (1998). Clustering and switching on verbal fluency: The effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia, 36*, 499–504.

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.

Wechsler, D. (1997). *Wechsler adult intelligence scale* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Willits, J. A., D'Mello, S. K., Duran, N. D., & Olney, A. (2007). Distributional statistics and thematic role relationships. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual meeting of the cognitive science society* (pp. 707–712). Austin, TX: Cognitive Science Society.

Zakzanis, K. K., McDonald, K., & Troyer, A. K. (2011). Component analysis of verbal fluency in patients with mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 33*, 785–792.