Multilingual Distributional Semantic Models:
Toward a Computational Model of the Bilingual Mental Lexicon

Akira Utsumi (utsumi@uec.ac.jp)

Department of Informatics, The University of Electro-Communications

1-5-1, Chofugaoka, Chofushi, Tokyo 182-8585, Japan

Abstract

In this paper, we propose a novel framework of a multilingual distributional semantic model to provide a psychologically plausible computational model of the bilingual mental lexicon. In the proposed framework, a monolingual semantic space for each target language is first generated from the corresponding monolingual corpus. These monolingual semantic spaces are then converted into ones with common dimensions, which are in turn integrated into a single multilingual semantic space. The language of dimensions, which we refer to as a pivot language, determines the type of bilinguals simulated by the model. We also tested the psychological plausibility of the proposed multilingual distributional semantic model by comparing the cosine similarity computed by the model with the cross-language word similarity ratings of L1 Japanese/L2 English sequential bilinguals. The result was that the bilingual semantic space with Japanese as a pivot language, which is predicted to be a model for L1 Japanese/L2 English sequential bilinguals, achieved better performance in simulating the similarity rating data. This suggests the plausibility of the proposed multilingual model.

Keywords: Multilingual distributional semantic model; Bilingual mental lexicon; Cross-language semantic similarity

Introduction

Distributional semantic models (henceforth, DSMs), or semantic space models, are models for semantic representations of words and for the way where semantic representations are constructed (Turney & Pantel, 2010). The semantic content is represented by a high-dimensional vector, and these vectors are constructed from a corpus by observing distributional statistics of word occurrence. Despite their simplicity, DSMs have provided a useful framework for cognitive modeling, especially for human semantic knowledge (e.g., Jones, Kintsch, & Mewhort, 2006; Landauer & Dumais, 1997).

However, these studies have explored the mental lexicon of a monolingual speaker and all DSMs used in these studies are monolingual. Given a recent growing interest in bilingualism in cognition (e.g., Bialystok, Craik, & Luk, 2012), it is quite reasonable to consider a multilingual extension of DSM toward a cognitive model of the bilingual (or multilingual) mental lexicon. This is what this study aims to accomplish.

In the field of natural language processing or computational linguistics, some studies (e.g., Bader & Chew, 2008; Wei, Yang, & Lin, 2008; Widdows, 2004) have proposed a multilingual extension of latent semantic analysis (LSA) for multilingual document clustering and cross-language word similarity computation. What these methods have in common is the use of a parallel corpus. A parallel corpus is a collection of bilingual (or multilingual) texts comprising sentences (or documents) in one language and their translations in other languages. By regarding aligned texts (i.e., a pair of an original text and its translations) as a single document,

the method for monolingual DSMs can be directly applied for multilingual DSMs. Against this advantage, however, the parallel-corpus-based approach to multilingual DSMs has some drawbacks. One serious drawback is that the use of parallel corpora is not psychologically plausible. It is extremely rare for bilinguals to be exposed to the same message in both languages simultaneously. Bilingual children often use different languages according to with whom to communicate (i.e., parents or friends) and where to communicate (i.e., at home or outside). It follows that bilingual lexical development and lexical knowledge is very unlikely to be explained by the distributional statistics obtained from a parallel corpus. The other drawback of the parallel-corpus-based approach is a practical one; parallel corpora are generally less easily available and of smaller size than monolingual corpora. Therefore, the multilingual semantic spaces generated from a parallel corpus cannot be expected to achieve a satisfactory performance.

In this paper, therefore, we propose a novel method for constructing multilingual DSMs toward a cognitive model of the bilingual mental lexicon. To overcome the drawbacks of the parallel-corpus-based approach, our method does not use any parallel corpora; it generates a monolingual semantic space for each target language using a monolingual corpus, and then integrates the multiple monolingual spaces into a single multilingual semantic space. The integration is carried out by aligning context words in different languages by direct correspondences between words (i.e., lexical links) or via a conceptual representation (i.e., conceptual links). This distinction is motivated by the psychological model of the bilingual mental lexicon (Kroll & Stewart, 1994). We then test the psychological plausibility of the proposed method for multilingual DSMs using the cross-language word similarity ratings of Japanese-English bilinguals (Allen & Conklin, 2014). Finally, we discuss the potential ability of the proposed method to simulate a variety of findings on bilingual lexicon and to provide a tool for bilingual research.
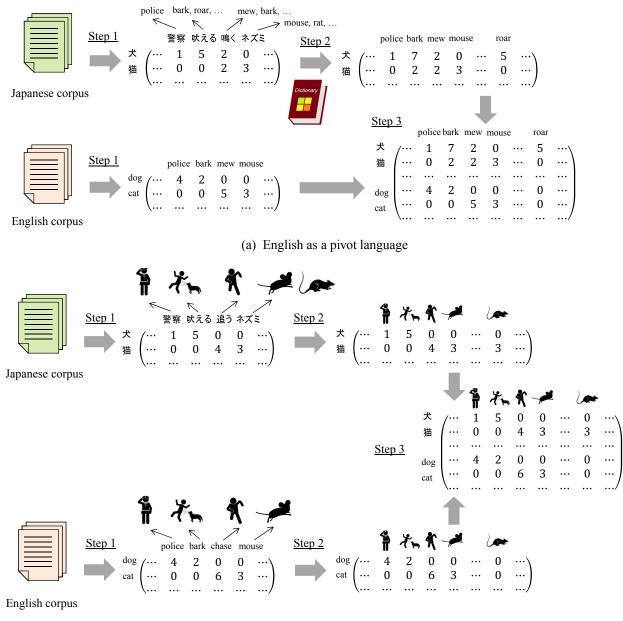
Computational Model

In this section, we first review a general method for constructing a monolingual semantic space. We then propose a novel method for constructing a multilingual semantic space, by which monolingual semantic spaces for target languages are integrated into a single multilingual semantic space.

Monolingual DSM

The method for constructing semantic spaces generally comprises the following three steps:

1. Initial matrix construction: n content words in a given corpus are represented as m-dimensional initial vectors whose

270

(a) English as a pivot language

(b) Concepts as a pivot language

Figure 1: A rough sketch of the multilingual distributional semantic model proposed in this paper: The case of English-Japanese bilingual semantic space.

elements are frequencies in a linguistic context. As a result, an n by m matrix $A = (a_{ij})$ is constructed using n word vectors as rows.

2. Weighting: The elements of the matrix A are weighted.

3. Smoothing: The dimension of the row vectors of A is reduced from the initial dimension m to r.

As a result, an r-dimensional semantic space including n words is generated.

For initial matrix construction in Step 1, two popular methods are used for computing the elements a_{ij} of A. In a "documents-as-contexts" method, an element a_{ij} is determined as the frequency of a word w_i in a document d_j (i.e., the number of times a word w_i occurs in a document d_j). On the other hand, in a "words-as-contexts" method, a_{ij} is calculated as the cooccurrence frequency of a target word w_i and a context word w_j within a certain context (i.e., the number of times two words w_i and w_j cooccur in a context). As a context for counting cooccurrence, we use a "window" spanning two words on either side of the target word. Note that the existing methods for multilingual LSA often employ a documents-as-contexts matrix by regarding aligned texts in a parallel corpus as a single document. On the other hand, our method does not

use a parallel corpus, and thus we apply a words-as-contexts method to initial matrix construction, as we will explain in the next subsection.

For Step 2, various weighting methods have been proposed. Two popular methods are entropy-based tf-idf weighting (Landauer & Dumais, 1997) and PPMI (positive pointwise mutual information) weighting (Bullinaria & Levy, 2007; Recchia & Jones, 2009). In this paper, we use a PPMI weighting method because it is suitable for words-as-contexts matrices and generally achieves good performance (Bullinaria & Levy, 2007; Recchia & Jones, 2009). PPMI is based on the pointwise mutual information (PMI) and replaces negative PMI values with zero. The last step (Step 3) of smoothing is optional and usually conducted using singular value decomposition (SVD). In this paper, we do not smooth the matrix because PPMI semantic spaces generally achieves good performance even though smoothing is not applied (Recchia & Jones, 2009).

Multilingual DSM

The basic idea underlying our multilingual DSM method is that word cooccurrence (i.e., words-as-contexts) matrices generated for each target language using a monolingual corpus are converted into cooccurrence matrices with the same set of context words; as a result, word vectors in different languages are placed in the single semantic space with common dimensions. The language of context words can be selected from target languages or other "language" representing concepts. We refer to this language as a pivot language.

Figure 1 illustrates our idea of the multilingual DSM method in the case of a Japanese-English bilingual DSM. The first step is to generate a word cooccurrence matrix for each target language using a monolingual corpus. In Figure 1 (a), two cooccurrence matrices, one for Japanese and the other for English, are constructed separately. For example, the Japanese cooccurrence matrix expresses that the target word 犬 cooccurs with the context word 警察 once and with the context word 吠える five times.

In the second step, a monolingual cooccurrence matrix for a target language is converted into a matrix expressing a pseudo-cooccurrence between target words in the target language and context words in the pivot language. As shown in Figure 1 (a), when English is a pivot language, the Japanese cooccurrence matrix must be converted into the matrix with English context words, while the English cooccurrence matrix does not need to be converted. Conversely, when Japanese is a pivot language, an English cooccurrence matrix is converted but a Japanese matrix is not. Furthermore, when a set of concepts is used as a pivot language as shown in Figure 1 (b), both cooccurrence matrices are converted into pseudo-cooccurrence matrices with concepts as contexts.

The matrix conversion in Step 2 can be done by translating context words into the pivot language using a dictionary or other lexical database, and by counting the "pseudo" cooccurrence frequency between a word in the target language and a context word in the pivot language. For example, as shown in Figure 1 (a), the context word 警察 has one translation equivalent *police* in English, and the cooccurrence frequency between the target word 犬 and 警察 is 1. It follows

that the cooccurrence frequency of 犬 and *police* is counted as 1. If a context word has more than one equivalent in the pivot language, the pseudo-cooccurrence frequencies for the other equivalents are also counted in the same way. For example, because the context word 吠える has at least two equivalents *bark* and *roar*, the cooccurrence frequency between 犬 and *roar* as well as between 犬 and *bark* is counted as 2. Note that the context word *bark* is also the equivalent of the context word 鳴く and thus the pseudo-cooccurrence frequency between 犬 and *bark* is 7 (= 5 from 吠える + 2 from 鳴く). Finally, in the last step (i.e., Step 3), the converted matrix for Japanese and the original cooccurrence matrix for English in Figure 1 (a) (or the converted English matrix in Figure 1 (b)) are concatenated into a single matrix expressing an English-Japanese bilingual semantic space.

In this framework of multilingual DSM, the pivot language determines the type of bilinguals for which the constructed semantic space is suitable. Because all the words in all target languages are represented through a pivot language, the pivot language can be regarded as the dominant language of bilinguals (or multilinguals). Hence, the multilingual DSM generated by this method can be regarded as a model of the mental lexicon of sequential bilinguals with a pivot language as L1. For example, the multilingual DSM with English as a pivot language shown in Figure 1 (a) is expected to be a cognitive model for L1 English/L2 Japanese bilinguals. When concepts are used as a pivot language as in the case of Figure 1 (b), the resulting DSM is assumed to be a model for simultaneous bilinguals, who are exposed to bilingual input from birth. This assumption is reasonable because simultaneous bilinguals do not have a dominant language and lexical development in two languages proceeds indifferently through concepts.

In the explanation given above, we use a Japanese-English bilingual DSM as an example, but our method is not specific to bilingualism. Formally, given k target languages $L_1, L_2, \cdots, L_k$, the method for constructing a multilingual semantic space on the basis of our idea can be described in the following steps:

1. The cooccurrence matrices $A_{11}, A_{22}, \cdots A_{kk}$ for target languages $L_1, L_2, \cdots, L_k$ are constructed from the corresponding monolingual corpora by the method for constructing monolingual DSMs.

2. Using conversion matrices D_{ip} ($1 \leq i \leq k$) from a target language L_i into a pivot language L_p, the cooccurrence matrices A_{ii} generated above are converted into the matrices A_{ip} with the same dimensions.

$$A_{ip} = A_{ii} \times D_{ip} \qquad (1)$$

Note that, if L_i is a pivot language L_p, then $A_{ip} = A_{ii}$.

3. All the converted matrices $A_{1p}, A_{2p}, \cdots A_{kp}$ are concatenated into a single matrix A.

$$A = \begin{pmatrix} A_{1p} \\ \vdots \\ A_{kp} \end{pmatrix} \qquad (2)$$

272

The resulting matrix A represents a multilingual semantic space.

At Step 2, we use a conversion (or term alignment) matrix D_{ip} to translate context words into a pivot language. The (s,t) entry of the matrix D_{ip} is 1 (or other nonzero value) if a word w_t in the pivot language L_p is a translation of a word w_s in the language L_i and otherwise 0.

Weighting (i.e., Step 2 of the monolingual DSM presented in the last section) can be applied either after Step 2 of the above algorithm or after Step 3 of the algorithm. Weighting after Step 2 implies that the converted matrices A_{ip} are weighted before they are concatenated into A, while weighting after Step 3 implies that the concatenated matrix A is weighted. Note that some weighting methods such as entropy-based one give the same matrix A regardless of whether weighting is applied before or after matrix concatenation, because in those methods word vectors (i.e., row vectors of A_{ip}) are weighted independently of each other. In the case of PPMI weighting, however, different matrices A are generated according to the timing of weighting.

Evaluation Experiment

Test Data

As test data for evaluating multilingual DSMs, we used the cross-linguistic similarity norms for Japanese-English translations provided by Allen and Conklin (2014). This data comprises semantic similarity and phonological similarity ratings of 193 Japanese-English word pairs and other relevant measures. Among these ratings, we used the semantic similarity rating on a 5-point scale ranging from 1 to 5, and compared it with the cosine similarity computed by multilingual DSMs to evaluate their modeling performance.

The 193 word pairs are divided into 98 cognates and 95 noncognates. Cognates are words in different languages that share both form and meaning. For example, the Japanese word カメラ /kamera/ and the English word "camera" are cognates. The cognates used in Allen and Conklin's (2014) study are all loanwords in Japanese, words borrowed from English and written in a separate script, katakana. Noncognates (e.g., 希望 and "hope") have the same meaning but do not share form. Cognates have been central to psycholinguistic research on bilingual language processing because they provide an effective way in examining an essential question of whether bilinguals selectively activate a single language or simultaneously both languages (Dijkstra, 2007).

Allen and Conklin's (2014) semantic similarity norm was collected from the native speakers of Japanese who also speak English as a second language, namely L1 Japanese/L2 English speakers. Hence, this semantic similarity data can be regarded as reflecting the mental lexicon of sequential bilinguals whose L1 is Japanese and L2 is English.

Materials for Multilingual DSM

As we explained before, the multilingual DSM proposed in this paper requires two kinds of language resources, namely a monolingual corpus for each target language and a dictionary (or lexical database) for converting between a pivot language and target languages. In this experiment, we used as a monolingual corpus Japanese newspaper articles (i.e., six years' worth of Mainichi newspaper articles) with 41.2M word tokens, and the written and non-fiction parts of the British National Corpus with 54.7M word tokens. In order to determine the vocabulary of the semantic space, we performed the widely used preprocessing steps, namely stopword removal and lemmatization. Concerning a dictionary, English WordNet 3.0 and Japanese WordNet 1.1 were used. WordNet is not a dictionary, but it can serve as a dictionary by connecting words in different languages via synsets. WordNet synsets are sets of cognitive synonyms, each expressing a distinct concept. Synsets provide an additional merit in using WordNet in that synsets can be used as a pivot language representing concepts (or more precisely a pivot concept). Japanese and English words to be included for bilingual semantic spaces were selected so that they can be translated into each other via WordNet. In other words, each of these Japanese words shares at least one synset with at least one of these English words. As a result, 22,416 Japanese words and 18,463 English words were selected as the vocabulary of the bilingual semantic space. These Japanese and English words are related via 23,421 synsets. Therefore, the size of the monolingual cooccurrence matrix in Step 1 of the proposed algorithm was 22,416 × 22,416 for Japanese and 18,463 × 18,463 for English.

Method

First of all, using the corpora and WordNet mentioned above, we constructed six bilingual semantic spaces from all combinations of three pivot languages (Japanese, English, and synset) and two timing of weighting (before or after matrix concatenation).

Using these six semantic spaces, we computed the cosine similarity of each pair of words in the test data. Note that 12 out of 193 word pairs of Allen and Conklin's (2014) data did not exist in the bilingual semantic spaces, and thus the remaining 181 word pairs (including 89 cognates and 92 noncognates) were used for similarity computation.

The performance of each bilingual semantic space was measured by Spearman's correlation coefficient between the computed cosine values and the semantic similarity ratings in the test data.

Prediction

If a bilingual semantic space is a plausible model of the L1 Japanese/L2 English bilingual's mental lexicon, the correlation coefficient is expected to take a high positive value. Furthermore, it is predicted that the semantic space with Japanese as a pivot language shows a higher correlation than that with an English pivot.

Result

Table 1 shows the correlation coefficients between the cosine similarity computed by the bilingual semantic spaces and the semantic similarity ratings of the test data. First of all, the correlation coefficients for all pairs were moderately high and statistically significant. This indicates that the proposed multilingual DSM framework provides a plausible model of

Table 1: Correlation coefficients between the cosine similarity computed by the bilingual semantic spaces and the semantic similarity ratings by Allen and Conklin (2014).

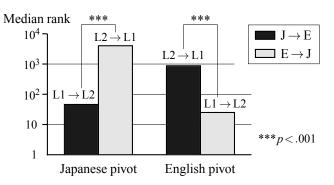| Pivot language	All pairs ($n = 181$)	Cognates ($n = 89$)	Noncognates ($n = 92$)
Weighting BEFORE concatenation			
Japanese	.294***	.284**	.316**
English	.247***	.221*	.282**
Synset	.291***	.290**	.304**
Weighting AFTER concatenation			
Japanese	.342***	.329**	.368***
English	.328***	.311**	.363***
Synset	.377***	.395***	.371***

*$p < .05$. **$p < .01$. ***$p < .001$

Figure 2: Median ranks of the target words in the ordering of cosine similarity to the prime words for the 181 word pairs used in the evaluation experiment. $J \to E$ ($E \to J$) denotes that Japanese (English) words in the pairs are used as primes and their paired English (Japanese) words are targets. Similarly, $L1 \to L2$ ($L2 \to L1$) denotes L1 (L2) primes and L2 (L1) targets, assuming that the pivot language of the multilingual DSM plays a role of L1. All the semantic spaces used here are weighted before matrix concatenation.

the bilingual mental lexicon. In addition, the semantic space with Japanese as a pivot language achieved higher correlations than those of the semantic space with an English pivot, regardless of whether they were calculated for all pairs or cognates/noncognates. This result is consistent with the prediction mentioned earlier, and thus suggests that the pivot language of the multilingual DSM can correctly model the dominant language of sequential bilinguals. The correlation coefficients for the DSM with synsets as a pivot language, which is expected to model a mental lexicon of simultaneous bilinguals, did not differ from (in the case of weighting before concatenation) or were slightly higher than (in the case of weighting after concatenation) those of the DSM with a Japanese pivot. We do not have a reasonable explanation of this result at the moment, but this result may reflect the fact that, as bilinguals become more proficient in L2, their L2 lexical knowledge is learned via a conceptual representation (Kroll & Stewart, 1994).

Comparison of the results between cognate and noncognate pairs shows that our proposed multilingual DSMs were more advantageous to noncognates. One possible reason would be due to word frequency effect; Japanese cognates are generally less frequent than noncognates (Allen & Conklin, 2014), and thus the cooccurrence statistics for cognates is less sufficient for plausible vector representations.

For the stage at which weighting is applied to a cooccurrence matrix, weighting after concatenation achieved better performance than weighting before concatenation. This result is not surprising because PPMI weighting requires to estimate the probability of context words across all target words, but weighting before concatenation computes the probability of context words separately for each language. However, it is an open question whether weighting before or after concatenation is plausible as a model of the bilingual mental lexicon. Although the proposed algorithm for constructing multilingual DSMs is not a psychological process model, weighting after concatenation may lend support to the view of a single integrated bilingual lexicon, rather than the view of two separate lexicons (for a review of two views, see French &

Jacquet, 2004).

Discussion

In this paper, we have proposed a novel method for constructing multilingual DSMs to provide a psychologically plausible computational model of the bilingual (or multilingual) mental lexicon. Its plausibility is tested and justified by comparing the cosine similarity computed by the multilingual DSMs with the semantic similarity data collected from Japanese-English bilinguals. In particular, the proposed method can provide a model that can discriminate between sequential bilinguals with different L1. Indeed, the evaluation experiment demonstrated that it can generate a semantic space appropriate for L1 Japanese/L2 English sequential bilinguals. However, the experiment presented in this paper is not so comprehensive and rather preliminary. Further justification of the modeling performance of the multilingual DSM must await further research, but in this section we discuss the potential ability of the multilingual DSM to explain other psycholinguistic findings on bilingual lexical processing.

Research on bilingual lexical processing has demonstrated that lexical access in bilinguals is language nonselective (van Heuven & Dijkstra, 2010; Schwartz & Kroll, 2006). In other words, lexical representations in both languages are activated in parallel regardless of which language is being processed. This is evidenced by the cross-language priming paradigm in which a prime word in one language facilitates a target word in another language. Particularly interesting is the well-known finding that primes in L1 obviously facilitate targets in L2, but L2 primes do not reliably facilitate L1 targets (e.g., Jiang & Forster, 2001; Schwartz & Kroll, 2006). This asymmetry effect may be able to be explained by the multilingual DSM proposed in this paper. One reasonable way to do this is to employ the rank of the target word under the ordering imposed by the cosine similarity to the prime word as a measure

for the degree of its priming effects (e.g., Griffiths, Steyvers, & Tenenbaum, 2007). The rationale behind this assumption is that the target word which ranks higher by the cosine similarity to the prime word is more activated and accessible by the prime word. For example, if among all the words in the semantic space the target word ranks first by the cosine similarity to the prime word, it seems to suggest that the target is activated most preferably by the prime. On the other hand, if the target word ranks very low even though the cosine value does not differ from the above case, it is less likely to be activated by the prime. Figure 2 shows the median rank of the target words obtained by applying this methodology to the 181 word pairs used in the evaluation experiment. The result is consistent with the asymmetry effect of cross-language priming. The Wilcoxin signed-rank test indicated that the median rank in the case of L1 prime and L2 target (i.e., J $\to$ E for the DSM with the Japanese pivot and E $\to$ J for the DSM with the English pivot) is significantly higher than that of L2 prime and L1 target.

Another well-known finding on multilingual lexical processing is that bilinguals generally perform more poorly on lexical tasks in both languages than monolinguals (Bialystok, 2009; Bialystok et al., 2012). This disadvantage of bilinguals is considered to be due to the interference from the other language. This interference effect can also be possibly explained by comparing the median rank of word pairs in the same language between the multilingual and monolingual DSMs. For example, we computed the median rank over 163 Japanese word association pairs (chosen from the Japanese word association norm "Renso Kijunhyo") by means of multilingual and monolingual DSMs. The result is that, as predicted, the median rank of the monolingual DSM (38.0) is higher than those of the multilingual DSMs (46.0 for the English pivot, $p < .001$; 56.0 for the Japanese pivot, $p < .01$).

From the above discussion, it is clear that the multilingual DSM proposed in this paper may have the potential to simulate several empirical findings on bilingual lexical processing. In addition, the proposed DSM framework may be able to simulate the behaviors of a variety of bilinguals with different degrees of language proficiency and with different developmental patterns. This may be realized, for example, by controlling context words (e.g., reducing context words into basic ones according to their age of acquisition) and/or by using multiple pivot languages (e.g., concatenating multilingual semantic spaces with different pivots). It would be interesting and vital for further research to explore these issues.

Acknowledgments

References

Allen, D., & Conklin, K. (2014). Cross-linguistic similarity norms for Japanese-English translation equivalents. *Behavior Research Methods*, *46*, 540–563.

Bader, B. W., & Chew, P. A. (2008). Enhancing multilingual latent semantic analysis with term alignment information. In *Proceedings of the 22nd international conference on computational linguistics (COLING-2008)* (pp. 49–56).

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

Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Science*, *16*, 240–250.

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*(3), 510–526.

Dijkstra, T. (2007). The multilingual lexicon. In M. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 251–265). Cambridge: Oxford University Press.

French, R. M., & Jacquet, M. (2004). Understanding bilingual memory: Models and data. *Trends in Cognitive Science*, *8*, 87–93.

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

Jiang, N., & Forster, K. I. (2001). Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of Memory and Language*, *44*, 32–51.

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

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174.

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.

Recchia, G., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis. *Behavior Research Methods*, *41*, 647–656.

Schwartz, A. I., & Kroll, J. F. (2006). Language processing in bilingual speakers. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics, 2nd edition* (pp. 967–999). Academic Press.

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188.

van Heuven, W. J., & Dijkstra, T. (2010). Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain Research Review*, *64*, 104–122.

Wei, C.-P., Yang, C. C., & Lin, C.-M. (2008). A latent semantic indexing-based approach to multilingual document clustering. *Decision Support System*, *45*, 606–620.

Widdows, D. (2004). *Geometry and meaning*. CSLI Publications.