The Chinese Route Argument: Predicting the Longitude and Latitude of Cities in China and the Middle East Using Statistical Linguistic Frequencies

Max M. Louwerse (mlouwerse@memphis.edu)
Department of Psychology / Institute for Intelligent Systems, University of Memphis
365 Innovation Drive, Memphis, TN 38152 USA

Sterling Hutchinson (schtchns@memphis.edu)
Department of Psychology / Institute for Intelligent Systems, University of Memphis
365 Innovation Drive, Memphis, TN 38152 USA

Zhiqiang Cai (zcai@memphis.edu)
Institute for Intelligent Systems, University of Memphis
365 Innovation Drive, Memphis, TN 38152 USA

Abstract

The Chinese Room argument describes a thought experiment that suggests that for symbols to become meaningful, they must be grounded in perceptual experiences. Embodied cognition theorists frequently use this argument to claim that cognition requires perceptual simulation. We shed light on the symbol grounding problem by arguing that the structure of natural language provides language users with cues that allow them to bootstrap meaning from non-grounded symbolic co-occurrences, such that the statistical linguistic structure can bootstrap meaning with minimal grounding. Two studies show that co-occurrences of both Chinese and Arabic city names can reliably predict their longitude and latitude in China and in the Middle East. Using the statistical linguistic technique Latent Semantic Analysis, similarity ratings were obtained for Chinese city names (Study 1) and for Arabic city names (Study 2). Multidimensional scaling (MDS) coordinates of these similarity ratings correlated with the actual longitude and latitude of these cities, showing that cities that are located together share similar semantic contexts. These results suggest that the Chinese Room argument might be substituted with a Chinese Route argument: statistical linguistic frequencies of word co-occurrences provide language users with implicit cues about how to form perceptual representations.

Keywords: symbol grounding problem; geography; spatial cognition; latent semantic analysis; symbolic cognition; embodied cognition

Chinese Room Argument

A monolingual English speaker sits in a room; all he has is a Chinese newspaper. Even though there is a wealth of Chinese language at his disposal, few would argue that he understands Chinese. In fact, even if he can successfully find a specific Chinese word, e.g., 上海, in his newspaper, and the collocations of that word, say, 北京, and 香港, there is little evidence he knows the meaning of those words. This Chinese room argument has been used by Searle (1980) to illustrate the symbol grounding problem in cognition (Harnad, 1990), which questions a computational account of meaning acquisition.

Many cognitive scientists place a strong theoretical emphasis upon how symbols become grounded (Barsalou, 1999; Glenberg, 1997; Harnad, 1990; Pulvermüller, 1999; Searle, 1980). These researchers express an increased concern regarding symbolic representations of meaning, and do not endorse analogies between computational and human approaches towards deriving meaning (Pecher & Zwaan, 2005; Semin & Smith, 2008). Embodiment theorists state that meaning cannot lie within arbitrary amodal symbol systems; instead, meaning extraction continuously involves the activation of perceptual experiences. Indeed, learning Chinese as a monolingual English speaker with only a Chinese-Chinese dictionary would lead to a symbolic merry-go-round (Harnard, 1990). Computationally translating symbols into other symbols is however what computer models do, and computational simulations are therefore fundamentally different from human cognitive processes (Glenberg & Robertson, 2001). One computational model of meaning extraction that is frequently used is latent semantic analysis (LSA; Landauer, McNamara, Dennis, & Kintsch, 2007). LSA establishes meaning representation based on co-occurrences of words in same contexts. The words *Beijing* and *Shanghai* therefore have a high similarity in (computational) meaning. However, because *Beijing* is not grounded in perceptual experiences, it does not have human-like meaning (Glenberg & Robertson, 2001). For instance, when humans process the word *Beijing* they might see a map of China and are able to 'see' the city in the northeast of the country. Estimating the location of *Beijing* using a corpus-based computational model would seem impossible, because of the Chinese Room argument. If the location of *Beijing* were to be estimated using amodal symbol systems, explicit spatial cues, such as prepositions and cardinal directions are needed (e.g., *Beijing is north of Shanghai*). Such spatial cues would lead to a combinatorial explosion with each city being added (*Beijing is north of Shanghai, Chongqing is southwest of Beijing, Chongqing is west of Shanghai,* etc.*)*.

Returning to the Chinese room with the monolingual English speaker described earlier, how could our English speaker possibly extract geographical locations from Chinese newspapers without seeing a map of China and seeing the cities marked on the map? Louwerse and Zwaan (2009) concluded that language encodes geographical information. Louwerse and Zwaan took the 50 largest cities in the United States and computed their co-occurrence frequencies in the New York Times, Wall Street Journal, and Los Angeles Post. None of these newspaper corpora necessarily described the spatial locations of the American cities. Yet, using LSA, Louwerse and Zwaan (2009) were able to estimate the longitude and latitude of the 50 cities using statistical linguistic frequencies for each of the three corpora (for a detailed description of the method being used, see below). Moreover, the population size of these cities could be estimated using frequency in the newspapers. The computational estimates were on par with human performance. The findings in this study showed that statistical linguistic frequencies can be used to estimate the location and the size of cities. Determining the semantic associations between cities in a corpus allows for estimating physical distance between cities. In fact, a heuristic like this might be used during cognitive map construction (Goldstein & Gigerenzer, 2002).

Harnad (1990) suggested that ungrounded symbolic representations can inherit meaning from grounded words related to them; we similarly propose that with minimal grounding of some symbols, the meaning of all symbols can be bootstrapped. If language encodes geographical information, we will be at least one step closer towards a bootstrapping solution. That is, the vacuum of the Chinese room now becomes an opportunity for a Chinese route: with very limited symbol grounding, the native speaker of English can bootstrap the geography of China.

Latent Semantic Analysis

To test the possibility of a Chinese Route Argument, i.e., the Chinese language encodes geographical locations in China, we used LSA on a sample of texts segmented into paragraphs as input. Mathematical transformations created a large term-document matrix from the input. For example, if there are m terms in n paragraphs, a matrix of $A = (f_{ij} \mathrm{x} \mathrm{G}(j) \mathrm{x} \mathrm{L}(i,j))_{mxn}$ is obtained. The value of f_{ij} is a function of the integer that represents the number of times term i appears in document j, $L(i; j)$ is a local weighting of term i in document j, and $G(j)$ is the global weighting for term j. Such a weighting function is used to differentially treat terms and documents to reflect knowledge that is beyond the collection of the documents. As in most LSA studies (Dumais, 2007; Martin & Berry, 2007), we used natural log as the local weight and log entropy as the global weight in the current analyses. The large matrix of A has, however, much redundant information. Singular Value Decomposition decomposes the matrix A into three matrices $A = \mathrm{U}\Sigma\mathrm{V}'$; where U is an m by m square matrix and V is an n by n square matrix, with Σ being an m by n diagonal matrix with singular values on the diagonal. Removing dimensions corresponding to smaller singular values and keeping the dimensions corresponding to larger singular values reduces the representation of each word to a low dimensional vector. Although the new representation for the words (the reduced U matrix) is no longer orthogonal, each word now becomes a weighted vector on 300 dimensions, with only the most important dimensions that correspond to larger singular values being kept. The number of dimensions can be determined *ad hoc*, but we followed the trend set by most LSA studies and used 300 factors (Landauer & Dumais, 1997). The semantic relationship between words can be estimated by taking the cosine between two vectors. With LSA the semantic relatedness is not only determined by the relation between words, but also by the words that accompany a word (Landauer & Dumais, 1997).

Two studies each tested two hypotheses: 1) Cities that are located together are debated together. That is, cities that are in close geographical proximity are also in close proximity in the text, so that language structure itself provides cues to derive perceptual-semantic information. 2) Cities that are populated more are debated more. That is, larger cities are talked about more, so that city word frequency provides cues about the importance of the city. In Study 1 we tested these hypotheses with city names in China in a Chinese text corpus, in Study 2 with city names in the Middle East in an Arabic corpus.

Study 1: China

In Study 1 we used a Chinese corpus collected online, consisting of 4 of the most popular classic fiction books, 29 popular modern fiction books, 26 history books, 49 philosophy books, 34 economy books, 15 politics books, and 8 military books. These books provided 86MB of text in 14768 documents (paragraphs) and 47,226 word types. In terms of text size, 33.4% texts are in history, 24.6% in philosophy, 10.4% in economics, 9.3% in modern fiction, 9.3% in politics, 7% in military and 6% in classic fiction. Note that the texts did not explicitly describe geographical relations between Chinese cities, and that the corpus was very heterogeneous.

The standard procedure was used when creating the LSA space, whereby each word was a weighted vector on 300 dimensions. The 50 largest cities in China were selected, and their latitude and longitude were determined using census data. All cities had a population size of more than one million (M = 2,393,188, SD = 2,340,707.88) (Table 1).

Cosine values were computed for each of the city pairs. Two cities resulted in cosine errors and were removed from the analysis, resulting in a 48 x 48 cosine matrix. This matrix was submitted to an MDS analysis using the ALSCAL algorithm. A Euclidean distance measure transformed the semantic similarities into dissimilarities, such that the higher the value, the longer the distance. Default MDS criteria were used with an S-stress convergence of .001, a minimum stress value of .005, and a maximum of 30 iterations.

We chose a low-dimensionality to rule out over-fitting the data. The fitting on a two-dimensional scale was moderate, with a Stress value = .33 and an R^2 = .59. The LSA estimated coordinates of the 48 cities were compared with the actual coordinates of the cities.

The loadings of the 48 cities on the two dimensions generated by the MDS analysis correlated with the longitude and latitude of the cities, latitude – dimension 1, r = .64, p < .001, n = 48; longitude – dimension 2, r = .33, p = .02, n = 48.

To do justice to the geometry of the 2D variables, we used bi-dimensional regression analyses to compare the computational estimates with the actual coordinates of the 50 cities. Tobler (1964) and Friedman & Kohler (2004) introduced bi-dimensional regressions in order to compute the mapping of any two planes under consideration. Whereas in a uni-dimensional regression each data point is shifted by intercept and slope, each actual and predicted value of the dependent variable are presented by a point in space, whereby vectors represent intercept and slope.

A bi-dimensional regression yielded a significant correlation between the LSA estimates and the actual city coordinates, r = .57, p < .001, n = 48. This result supported the hypothesis that Chinese cities that are located together in China are debated together in the Chinese language.

The question can be raised whether the bi-dimensional regressions not always yield significant correlations. To answer this question we conducted 1000 Monte Carlo simulations on the 48 x and y pairs. The average bi-dimensional regression of these simulations yielded no significant result, average r = .13 (SD = .06), p = .37, n = 48.

In addition, we tested the hypothesis that cities that are populated more are debated more by comparing the frequency of the 50 cities in the Chinese corpus with their actual population size. A Pearson correlation was significant, r = .47, p < .001, n = 50.

Table 1: Chinese Cities

City	Lat.	Long.	Dim.1	Dim.2	City	Lat.	Long.	Dim.1	Dim.2
上海	31.23	121.40	-0.38	-1.42	苏州	31.30	120.60	-0.83	-1.34
北京	39.93	116.40	-0.13	-1.24	汕头	23.37	116.60	-1.04	-0.69
重庆	29.57	106.50	-1.07	0.81	荣成	23.54	116.30	1.05	-0.17
西安	34.27	108.90	-0.92	1.28	兰州	36.05	103.60	-0.68	1.06
武汉	30.58	114.20	-1.25	-0.93	**合肥**	31.85	117.20	-0.27	1.40
成都	30.67	104.00	-1.24	0.45	抚顺	41.87	123.80	1.28	-0.50
天津	39.13	117.20	0.28	-1.26	洛阳	34.68	112.40	0.83	1.32
沈阳	41.80	123.40	1.96	-0.64	邯郸	36.58	114.40	0.88	1.09
哈尔滨	45.75	126.60	1.36	-0.19	包头	40.60	110.00	-0.98	0.68
南京	32.05	118.70	-0.47	-1.48	香港	22.27	114.10	-0.68	0.64
广州	23.12	113.20	-0.63	-1.31	苏州	34.27	117.10	-0.84	-1.31
太原	37.87	112.50	0.16	1.40	深圳	22.53	114.10	-0.74	-0.01
长春	43.87	125.30	1.98	-0.55	**福州**	26.08	119.30	-1.09	-0.99
石家庄	38.05	114.40	1.63	0.07	无锡	31.58	120.30	-0.35	1.57
长沙	28.20	112.90	-1.30	-0.30	淮南	32.63	116.90	-1.16	0.46
济南	36.67	117.00	0.86	-0.91	贵阳	26.58	106.70	-1.27	-0.57
大连	38.92	121.60	0.56	-0.89	**鞍山**	41.12	122.90	1.61	-0.55
吉林	43.85	126.50	1.70	0.11	**保定**	38.87	115.40	-0.17	0.99
南昌	28.68	115.80	-1.02	-0.94	咸阳	34.37	108.70	-0.01	1.42
郑州	34.75	113.60	1.25	0.28	**昆明**	25.05	102.70	0.01	-0.83
九龙	22.32	114.10	-0.04	1.73	大同	40.08	113.30	0.66	0.86
杭州	30.25	120.10	-0.92	-1.18	**本溪**	41.33	123.70	1.82	-0.30
青岛	36.07	120.30	1.03	-0.23	淮北	33.95	116.70	-0.75	1.08
唐山	39.62	118.10	0.37	1.53	**常州**	31.78	119.90	-1.06	0.50

Study 2: Middle East

In order to determine whether the findings could be generalized beyond China and the Chinese language, we used a different language (Arabic) and a different geography (the Middle East) in the second study.

An LSA space was created using an Arabic corpus collected online, consisting of books and news on history (49%), fiction (42%), politics (3%), philosophy (2%), economy (1%) and other unknown types of texts (3%). The total size of the corpus was 71.8 MB, including 27,937 paragraphs and 147,535 word types. Again, the texts did not specifically discuss the geography of the Middle East. Instead, the corpus covered many topics and was, again, very heterogeneous in nature.

Similar to the previous analysis, 50 of the largest cities across the Arabic speaking countries in the Middle East were selected (M = 1,304,154, SD = 1,451,579). These cities were located in Egypt, Iraq, Jordan, Kuwait, Lebanon, Oman, Syria, United Arab Emirates, and Yemen (Table 2).

Some countries were not included because the Arabic notation of cities for those countries was unavailable (Saudi Arabia, Ethiopia, Eritrea, Somalia, Djibouti).

As in the Chinese analysis, geographical location (longitude and latitude) as well as population size for these 50 cities were determined. A 50 x 50 cosine matrix was submitted to an MDS ALSCAL analysis, and the MDS coordinates were compared with the actual coordinates.

Again, the fitting on a two-dimensional scale was moderate, with Stress = .35, R^2 =.69. The LSA estimated coordinates of the 50 cities were compared with the actual coordinates of the cities.

The loadings of the 50 cities on the two dimensions generated by the MDS analysis correlated with the longitude and latitude of the cities, latitude – dimension 1, r = .41, p < .001, n = 50; longitude – dimension 2, r = .57, p < .001, n = 50.

A bi-dimensional regression also yielded a significant correlation between the LSA estimates and the actual city coordinates (r = .53, p < .001, n = 50). These results again supported the hypothesis that cities in the Middle East that share geographical context, share textual context (cities that are located together are debated together).

As in Study 1, we ran 1000 Monte Carlo simulations to rule out the possibility that the significant bi-dimensional regressions could be obtained from an accidental pairing of the estimates. The average regression coefficient again ruled out that the findings could be obtained by chance, average r =.13 (SD =.07), p = .37, n = 50.

Finally, as before, we compared the frequency of the 50 cities in the Arabic corpus with their actual population size. A Pearson correlation was significant (r = .61, p < .001, n = 50), providing evidence for the hypothesis that cities in the Middle East that have a higher population, are talked about more frequently.

Table 2: Middle Eastern Cities

	Lat.	Long.	Dim1.	Dim.2		Lat.	Long.	Dim. 1	Dim. 2
تهران	35.67	51.43	0.2169	1.5093	كرمانشاه	34.38	47.06	-0.075	2.0242
بغداد	33.33	44.44	0.6485	-1.2606	السليمانية	35.56	45.43	1.0587	-1.0911
الرياض	24.65	46.77	-0.9637	-0.9586	اروميه	37.53	45.00	0.3161	1.2087
جدة	21.50	39.17	-1.1027	-0.7991	زاهدان	29.50	60.83	0.4181	1.1594
الموصل	36.34	43.14	1.007	-1.0505	رشت	37.30	49.63	0.3014	1.2215
مشهد	36.27	59.57	-0.2017	1.0937	الطائف	21.26	40.38	-1.044	-0.7242
كابل	34.53	69.17	1.2753	-0.0876	كرمان	30.30	57.08	-0.04	2.0397
بيروت	33.89	35.50	-0.6608	-0.9994	حماة	35.15	36.73	-1.3177	-0.7392
البصرة	30.53	47.82	0.0922	-1.4058	الحلة	32.48	44.43	1.034	-0.7171
حلب	36.23	37.17	-1.2136	-0.7176	تبوك	28.39	36.57	-1.1946	-0.7448
اصفهان	32.68	51.68	-0.4703	1.6961	كربلاء	32.61	44.03	1.0672	-1.0522
دمشق	33.50	36.32	-0.7743	-1.0202	همدان	34.77	48.58	-0.0679	2.01
كرج	35.80	50.97	0.2241	1.2555	العمارة	31.84	47.15	1.0166	-0.8805
مكه	21.43	39.82	-1.0349	-0.7787	الزرقاء	32.07	36.10	1.0995	-0.6294
تبريز	38.08	46.30	-0.9509	0.7826	اراك	34.08	49.70	0.578	0.9256
شيراز	29.63	52.57	-0.8681	0.8734	الديوانية	31.99	44.93	0.973	-0.9436
اربل	36.18	44.01	1.2833	-0.3946	خميس مشيط	18.31	42.73	-0.9944	-0.6579
عمان	31.95	35.93	-0.5422	-1.022	يزد	31.92	54.37	-0.0563	1.81
المدينه	24.48	39.59	-0.7522	-1.0729	بريده	26.37	43.97	-1.1197	-0.2557
اهواز	31.28	48.72	0.1934	1.3174	اردبيل	38.25	48.30	-0.0501	2.002
قم	34.65	50.95	-0.1317	1.2444	بعقوبة	33.74	44.65	1.1115	-0.6968
الدمام	26.43	50.10	-1.1988	-0.7561	بندرعباس	27.25	56.25	0.6984	0.9249
حمص	34.73	36.72	-1.3472	-0.7446	هرات	34.35	62.18	1.2645	-0.043
كركوك	35.47	44.39	1.0518	-1.1014	اسلام شهر	35.54	51.20	1.233	0.0334
النجف	32.00	44.34	1.0731	-1.0775	اللاذقية	35.54	35.78	-1.0628	-0.7091

General Discussion

This study showed that statistical linguistic frequencies can be used to estimate the location and population size of cities. In the first study we estimated the location and size of cities in China using Chinese text, in the second the location and size of cities in the Middle East using Arabic texts. These findings show that the results reported by Louwerse, Cai, Hu, Ventura, & Jeuniaux (2006) for France, those reported by Louwerse and Zwaan (2009) for the United States, and by Louwerse and Benesh (in press) for (the fictional) Middle Earth can be extended to China and the Middle East. Moreover, the current study has demonstrated that geographical locations are not only encoded in English (Louwerse et al., 2006; Louwerse & Benesh, in press; Louwerse & Zwaan, 2009), but also in Chinese and Arabic.

There are several questions that should be addressed with regards to the findings reported in this study. First, we should address the question whether the findings reported in this study should be attributed to LSA or to statistical linguistic frequencies. Louwerse and Zwaan (2009) addressed this question by demonstrating that geographical locations of cities in the United States could be predicted using higher-order co-occurrences (using LSA), but also by first-order co-occurrences. Louwerse (2011), however, pointed out that for first-order co-occurrences a corpus needs to be approximately 25,000 times larger than a corpus that is the appropriate size for an LSA analysis. A second question concerns the explicit spatial cues potentially present in the corpus. After all, the argument could be made that the Chinese and Arabic corpora we used for our semantic spaces consisted of explicit spatial cues (e.g., cardinal directions, prepositions) that explained our findings rather than implicit semantic relationships. This seems extremely unlikely for two reasons. First, the LSA algorithm shows minimal sensitivity to explicit cues because it uses higher-order co-occurrences (see Landauer, McNamara, Dennis, & Kintsch, 2007). Secondly, the corpora were so diverse in nature that the results can better be explained by statistical linguistic frequencies than by the specifics of the texts.

We began this paper with the Chinese Room Argument, which suggests that meaning cannot be extracted from symbols unless a referent is perceptually activated (Searle, 1980). Even though this study did not compare the computational results with experimental data (see Louwerse and Zwaan, 2009 and Louwerse and Benesh, in press, for such a comparison), it does provide some insight in the Chinese Room argument. The current study puts forward that, with minimal grounding of some symbols (city names), the meaning of all symbols (city names) can be bootstrapped, because of the organization of the symbolic network. The language system has many built-in regularities that are utilized during cognitive processing (Louwerse, 2011; Louwerse & Jeuniaux, 2010). To illustrate this further, the current study has shown that if a language user knows the location of the city 乌鲁木齐, and knows only that the other Chinese words are Chinese city names, the language user can bootstrap the geographical locations of these other cities on a country map of China. Moreover, they can make estimates about the size of each city, because frequency correlates with population size.

Obviously, we do not deny the essence of the symbol grounding problem: the language user must ground at least one symbol and must also have partial meaning with regards to the other words (i.e., know that they are city names). Moreover, the geographical estimates are relative estimates, rather than a specific longitude and latitude. However, findings like these do challenge an extreme view of symbol grounding that dismisses the possibility of statistical linguistic frequencies playing a significant role in cognition. Experimental evidence has shown that statistical linguistic frequencies often explain experimental findings better than perceptual simulations account do (Louwerse, 2008; Louwerse, 2011), yet whether humans rely more on statistical linguistic frequencies or perceptual simulations depends on at least the cognitive task and the stimulus (Louwerse & Jeuniaux, 2010). We therefore advocate the pursuit of a unified account in which both statistical linguistic frequencies and perceptual simulation help establishing meaning. In line of this research agenda, this study has shown that with minimal grounding the symbolic vacuum of the Chinese Room can become a guiding Chinese route.

Acknowledgments

This research was in part supported by grant NSF BCS-0904909. The usual exculpations apply.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577 660.

Dumais, S. T. (2007). LSA and information retrieval: Getting back to basics. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.

Friedman, A., & Kohler, B. (2003). Bidimensional regression: A method for assessing the configural similarity of cognitive maps and other two-dimensional data. *Psychological Methods, 8*, 468–491.

Glenberg, A. M. (1997). What memory is for: Creating meaning in the service of action. *Behavioral and Brain Sciences*, *20*, 41-50.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory & Language, 43,* 379-401.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109*, 75–90.

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335–346.

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

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.) (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.

Louwerse, M. M. (2008). Embodied relations are encoded in language. *Psychonomic Bulletin & Review*, *15*, 838-844.

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *TopiCS in Cognitive Science, 3,* 273-302.

Louwerse, M. M. & Benesh, N. (in press). Representing spatial structure through maps and language: Lord of the Rings encodes the spatial structure of Middle Earth. *Cognitive Science*.

Louwerse, M. M., Cai, Z., Hu, X., Ventura, M., & Jeuniaux, P. (2006). Cognitively inspired natural-language based knowledge representations: Further explorations of Latent Semantic Analysis. *International Journal of Artificial Intelligence Tools, 15,* 1021-1039.

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*, 96-104.

Louwerse, M.M. & Zwaan, R.A. (2009). Language encodes geographical information. *Cognitive Science*, *33*, 51-73.

Martin, D. I., & Berry, M. W. (2007). Mathematical foundations behind latent semantic analysis. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.

Pecher, D., & Zwaan, R. A. (Eds.) (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. New York, NY: Cambridge University Press.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences, 22*, 253-270.

Searle, J.R. (1980). Minds, brains, and programs. *Behavioral & Brain Sciences, 3,* 417- 424.

Semin, G. R. & Smith, E. R. (Eds.) (2008). *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches*. New York, NY: Cambridge University Press.

Tobler, W. R. (1964). Bidimensional regression. *Geographical Analysis, 26,* 187–212.