A TOPIC-BASED DIACHRONIC ACCOUNT OF THE POLYSEMY OF THE ENGLISH VERB '*RUN*'[1]

NAOKI KIYAMA
University of Kitakyushu
n-kiyama@kitakyu-u.ac.jp

YOSHIKATA SHIBUYA
Kanazawa University
y.shibuya@staff.kanazawa-u.ac.jp

Abstract
This study discusses the polysemy and diachronic semantic frequency shifts of the English verb *run* using a topic model. The increase in corpus-based research in cognitive linguistics has yielded many empirical findings on various aspects of language. However, there are still areas that have not received sufficient attention. One such area is the study of word meaning in terms of related topics and social interest. In this paper, we analyze diachronic data collected from the NOW corpus (News on Web) and argue that the polysemy of *run* can be described in terms of topic, and that the change in the use of the meanings that this verb has can be described in terms of a change in social interest.

Keywords: topic models, polysemy, semantic frequency shifts, meaning in society, the NOW corpus

1. Introduction

In cognitive linguistics, it is assumed that the meaning of a word is encyclopedic, that what one knows about a concept is part of its meaning (Croft, 1993), and that in actual settings of language use, one "perspectivizes" (Taylor, 1995, p. 90) an aspect of the relevant knowledge in order to understand the word appropriately. Based on these premises, corpus-based research in cognitive linguistics has provided many important insights into this framework. However, there are some areas that have been somewhat overlooked. One such area concerns the study of the changes in the frequency with which a meaning is used. The lack of research in this area could be attributed to the idealization that is likely to be made in corpus-based studies (Hilpert, 2012). To elaborate on this point from the perspective of the interest of the present study, many previous corpus-based

1 We would like to thank Hisashi Iwane (Osaka University) for useful discussions. We also thank two anonymous reviewers of *Research in Language* for their comments. Remaining inadequacies are of course ours. This work was supported by JSPS KAKENHI Grant Number 20K00667.

diachronic semantic studies have tended to forgo examining changes in the frequency of individual meanings of polysemous words. However, as with various levels of language, the frequency of each meaning that a word has should also always be in a constant state of flux. For example, while the verb *run* has several meanings (see e.g., Gries, 2006), it is not necessarily the case that all of the meanings will undergo the same frequency changes over time. Instead, it is conceivable that each meaning will undergo its own unique frequency changes. This implies that it is important for diachronic semantic studies to analyze the frequency change of individual meanings. Against this background, this study examines the polysemy and semantic frequency shifts of the English verb *run*, using it as a case study. We offer a diachronic semantic account using a topic model, which is still not widely used in semantic research, arguing that the polysemy and the semantic frequency shifts of *run* can be characterized in terms of topic and the changes in social interest.

This paper is structured as follows. In section 2, previous corpus-based studies on polysemy are described, followed by an overview of the research design of the present study. Section 3 describes the relationship between word meaning, topic and social interest. It will be argued that quantifying context in terms of topic is useful as a means of exploratory investigation into the relationship between social interest and language use. Finally, section 4 concludes with a brief discussion of the implications of our findings for the study of linguistic knowledge and semantics.

2. Theoretical background

2.1. Previous studies and the hypotheses of the present study

In corpus-based cognitive linguistic lexical research, analysis tends to be conducted mainly on words that co-occur with the target word under investigation. For example, in the study based on 'behavioral profiles', various features of the target word in question, including, for instance, whether the subject or object in the clause is animate or not, are (semi-)manually coded by the analyst(s) (e.g. Gries, 2006), after which the 'behavior' of the target word is explored by statistically measuring the similarities (and dissimilarities) of the usages of the target word (for other approaches, see e.g., Peirsman et al., 2010).

The approach based on the behavior of the co-occurring words of the word under study is an excellent method from a descriptive point of view and has so far provided many insights into lexical semantic research. However, there are certain perspectives that have been under-explored. One of them is concerned with the concept of 'topic' (see section 2.2 for details). In this study, we argue that topic can be an important notion in conducting polysemy research.

In this study, we hypothesize that in understanding a polysemous word, one understands it in terms of what is being talked about in the text, i.e., in terms of 'topic' (the definition will be detailed in the next section). For the sake of illustration, let us consider the following two simple fictious situations.

For example, if you were watching the Olympic 100-meter race on TV at home and a family member asked you, 'Who's running?', you would probably think that the meaning of the verb *run* that should be foregrounded here is related to meaning of 'someone moving at a speed faster than walking'. On the other hand, if a family member asked you 'Who's running?' while watching a news report on an election, you would probably think that the meaning of *run* in this case is related to 'someone who announced candidacy for the election' (see also Deignan, 2005, p. 27ff). The difference between these two situations seems to be amenable to explanation in terms of the differences in the topics involved. Based on the considerations as stated above, it is hypothesized here that the meaning of a word is rooted in the topic on which it is used (hypothesis 1).

There is one more aspect that appears lacking in many previous corpus-based studies: the investigation of the relationship between word meaning and social interest. Although various diachronic studies have been conducted, there has been little research on how differences in word meaning relate to changes in social interest (with the exceptions of Hilpert, 2020 and Peirsman et al., 2010). In cognitive linguistics, language is assumed to be a communicative tool used in social behavior (Croft, 2009). Based on this conception, we hypothesize here that changes in social interest are closely related to the usage and frequency of words (hypothesis 2).

In this paper, the above two hypotheses are examined through the analysis of topics. Today, with the development of computer technology, methods have been devised to identify topics in documents/texts in probabilistic terms. One such method is the topic model, which will be described in the next section.

2.2 Topic models

Topic modeling is a method of identifying the topic(s) of a text/document. For example, consider the following.

(1) First baseman Anthony Rizzo is a slugger for the Yankees who owns outstanding home-run power and run-producing abilities. Rizzo creates great power due to his big swing and frame. (CBS Sports, Jul. 16, 2022)

Those who are familiar with American baseball will probably interpret this text as being about Major League Baseball. However, note that the phrase *Major League Baseball* is not found in it. Those who understood the above passage to be about American baseball would have understood the topic based on such words as *first baseman, the Yankees,* and *home-run* found therein.

As seen in the example above, it seems that people interpret the topic of a text through a particular set of words found in it, even if the topic is not directly observable in it. A topic model is a method developed to classify texts by statistically analyzing the distribution of words in them to determine what topics they are about. Simply put, a 'topic' in a topic model can be paraphrased as "what is being talked about" (Sagi et al., 2012, p. 163fn).

In the example above, the topic was discussed based on an observation of the entire text, but in the present study, in order to align with its purpose, which is to discuss the polysemy of *run*, the input data used for the analysis are concordance lines including the target word. This means that it is necessary to use topic modeling methods suitable for analyzing short texts. To this end, here we opted to use a topic model called a biterm topic model (Yan et al., 2013, hereafter BTM).

2.3. Corpus used and the procedures for the BTM analysis

In this study, we used the NOW corpus (News on the Web) (Davies, 2016-), which contains news articles from 20 English-speaking countries. This corpus contains approximately 15 billion words (as of May 2022). In order to keep the amount of data manageable, data were collected exclusively from the U.S. data, covering the period from January 2015 to January 2017. Data were collected in two steps. First, for the range of co-occurring words of *run* obtained, 20 words each were extracted from L/R, for a total of 40 words. Then, stop words, numbers, etc., which were considered irrelevant for identifying the meanings of *run*, were removed from the data set. The verb *run* is a "highly polysemous English verb" as described by Gries (2006, p. 58), and has been the subject of numerous studies (e.g. Glynn, 2014; Gries, 2006). By utilizing a topic model to examine the polysemy of *run*, the present study aims to show that this method of analysis opens a new avenue for the study of word meaning.

3. Results of the BTM analysis

3.1. Determining the optimal number of topics

First, note that, as with other probabilistic topic modeling methods, when performing BTM the number of topics must be determined at the discretion of the analyst. In BTM, to control the arbitrariness in determining the number of topics, the optimal number of topics can be mathematically determined using an index called 'exclusivity'. A higher value of exclusivity is considered to indicate a more appropriate number of topics, although the value with the highest exclusivity is not necessarily easy for humans to interpret (Chang et al., 2009). Therefore, while the values of exclusivity are used as reference values, the analyst needs to determine the number of topics that are considered qualitatively most appropriate.

Figure 1 shows the results of estimating the exclusivity values when the number of topics was set from 5 to 40. As shown in the figure, exclusivity reaches its highest value when the number of topics is around 30 and 40. A qualitative interpretation based on the classification of the topics yielded by the BTM analysis indicated that when the number of topics is set to 40, the topics become too specific to interpret compared to when the number of topics is set to 30. We therefore decided to set the number of topics in this study to 30. That said, BTM is a method of concentrating extremely frequent words in Topic 1 thereby improving classification accuracy of the remaining topics. This means that Topic 1 is treated as an 'anything goes' topic and is hence not subject to analysis. Ultimately, 29 topics were left for analysis in this study.

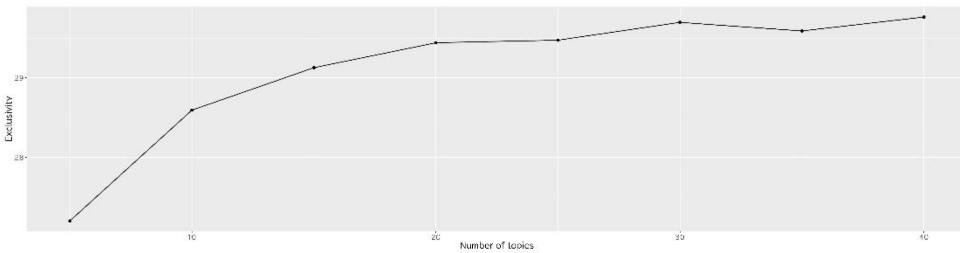

Figure 1: Change in exclusivity when the number of topics is set from 5 to 40

3.2. Hypothesis 1: Topics and the meanings of *run*

Due to limitations of space, it is not possible to look at all of the 29 topics analyzed here. Instead, this section focuses on three topics (Topics 2, 3 and 4) that were deemed particularly important. Table 1 shows the results for these topics. The ID column indicates the respective topic number. The Keywords column shows the top 10 words that have the highest probability of occurrence among the words making up each topic. The Topic column shows the topic names assigned by the analysts based on the keywords. The Sense column indicates the meanings of *run* as identified by the analysts based on the concordance lines.

Table 1: Topics identified by the BTM analysis (excerpts)

ID	Key words	Topic	Sense
2	android, devices, windows, version, system, new, software, operating, ios, device	Smartphones	To execute
3	president, campaign, trump, party, presidential, candidate, republican, election, democratic, house	Election Trump	To become a candidate
4	black, wearing, restaurant, store, food, white, man, men, blood, big	Physical characteristics	To escape

Let us start with Topic 2 (ID number 2) to examine the results of the analysis. Looking at the keywords that make up this topic, we see many words related to smartphones including *Android, device(s),* and *iOS*. From this observation, we can assume that this topic is about smartphones and can therefore name it 'Smart phones' (see the Topic column). To determine the meaning of *run* in this topic, we need to look at the concordance lines. Below is a concordance line containing the keywords in Topic 2.

(2) … the Android operating system used to ***run*** almost all Android phones continues to dominate the global mobile OS market with ...

In the concordance line (2), *run* is used to mean something like 'one system makes another system work,' and therefore the meaning of *run* here can be roughly. defined as 'to operate' (see the Sense column).

Moving on to Topic 3, which contains election-related vocabulary such as *president(ial), campaign, candidate,* etc., it is particularly noteworthy that the word *Trump* is included in the Keywords column. Taking into account the time period covered in this study, it can be inferred that the topic here pertains to the 2016 election in which former U.S. President Donald Trump ran for office. Based on this observation, we labeled Topic 3 as 'Election Trump'. The next example is a concordance line related to Topic 3.

(3) … he wants Trump to unify "all wings of the Republican Party and the conservative movement" and then ***run*** a campaign that will allow Americans to "have something that they're proud to support and proud to be a part of."

Here, *run* is used to mean 'to run for president'. Hence, we can identify the meaning of *run* when used in this topic as 'to become a candidate'.

Topic 4 contains a collection of words related to people, such as *men* and *women,* and words related to clothing, such as *shirts* and *shoes*. Therefore, in this study, we label this topic as 'Physical characteristics'. Example (4) is an instance containing the keywords of this topic.

(4) After he ***ran***, witnesses said they saw Jones, who was wearing only shorts and a black baseball hat, turn on Jefferson.

The above example sentence is an eyewitness account of the crime scene, describing how the perpetrator was dressed. When presenting the characteristics of a suspect in newspapers, the clothing features of the suspect are often described. Since *run* here can be paraphrased by verbs such as *escape*, we can assume that *run* is used in this sentence in the sense of 'to escape'.

We have so far described the relationship between the topics and the meanings of *run*. Although the BTM analysis did yield some topics for which we could not find any consistency in the meanings of *run,* we were for the most part able to confirm the correspondence in 27 of the 29 topics, meaning that hypothesis 1 described in section 2.1 generally holds true.

Before we consider hypothesis 2, let us examine one more aspect of word meaning. Let us consider the relationship between the meanings of *run* in the 27 topics where a correspondence between a topic and the meaning of *run* was found. Given hypothesis 1, we assume that the relationship between the meanings of *run* can be characterized on the basis of the similarities of the topics involved. To substantiate this assumption, we computed cosine similarity between the topics. The results are shown in Figure 2 using multidimensional scaling.

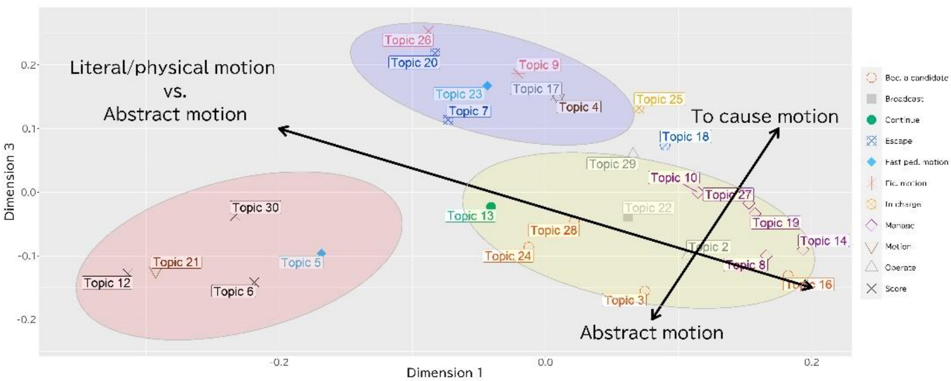

Figure 2: Classification of meanings of run by MDS (Dimensions 1 and 3)

In the lower left corner of the figure, the literal meanings such as 'to score' (red oval), and 'fast pedestrian motion' (blue oval), which are often used in sports, are clustered together. The yellow oval in the lower right is a cluster representing the abstract meanings of *run* such as 'to become a candidate' and 'to broadcast'. The yellow cluster can be further split into two groups: to the upper right of the arrow, the meanings of 'to cause motion' (Gries, 2006) such as 'to manage' are found, and to the lower left of the arrow, the meanings of 'abstract motion' (Gries, 2006) such as 'to become a candidate' are clustered together.

3.3. Hypothesis 2: Semantic frequency shifts and social interest

Another purpose of this paper is to discuss the relationship between semantic frequency shifts and social interest. Let us first look at the change in frequency of *run*, indicated by the number of hits per million words in Figure 3. The gray dashed line illustrates the adjusted frequency, while the blue solid line illustrates the values smoothed by the data for each month (LOWESS, locally weighted scatter plot smooth). In order to capture the changes in frequency of *run*, we have here used a method called variability-based neighbor clustering (VNC) (Gries & Hilpert, 2008).[2]

2 Although this study investigates a very short period of only two years, it still qualifies as a diachronic approach since it examines changes in language use over time.

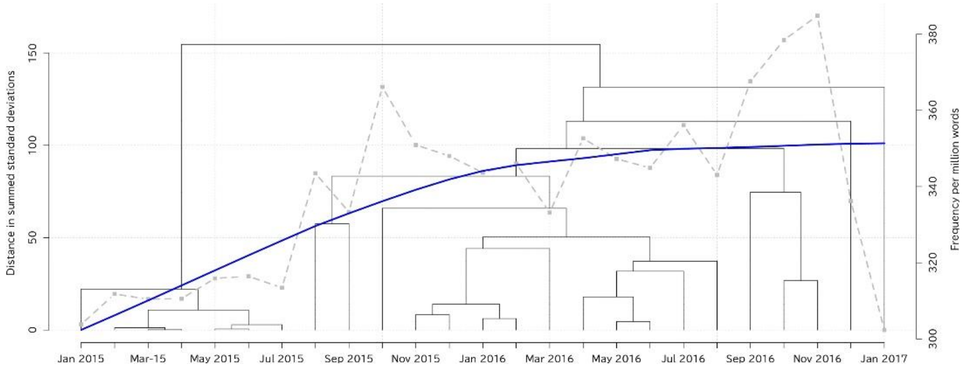

Figure 3: Frequency changes of run in the NOW data

The figure shows that the frequency of *run* was on an upward trend (although there were periods of decrease) until November 2016, with a significant change in the upward trend after July and August 2015.

Peirsman et al. (2010) used Word Space models to investigate how language users' impressions of specific words changed after 9/11 (see also, e.g., Jansegers & Gries, 2017). While the approach taken by Peirsman et al. is very powerful for semantic research, it requires prior knowledge of the specific dates on which the meaning of a word is thought to have changed. In contrast, BTM analysis allows one to examine the changes in the frequency with which a meaning is used without setting a priori a date when the change is thought to have occurred.

In the interest of space limitations, in the following we will take as an example the meaning "to become a candidate" (topics 3, 16, 24, and 28) among those discussed in section 3.2 and examine how the probability of occurrence of each topic has changed.

Before we get into the details of the discussion, an explanation is in order with respect to the term 'probability of occurrence'. Since topics are not directly observable (see section 2.2), the frequency of topic occurrence cannot be represented either. However, BTM can be applied to infer the topic(s) contained in each document by means of computing the *theta* values. The *theta* values can then be combined with the NOW corpus data to indicate how likely each topic is to appear on a particular day. The concordance lines in the data allow us to observe changes on a daily basis, but since the present study does not seek detailed day-to-day changes, we computed the average of *theta* on a monthly basis. Based on the assumption that there is a relationship between topic and word meaning, changes in the probability of occurrence of each topic can be viewed as reflecting the changes in the probability of use of the meaning of *run*.

Before presenting the results, let us examine each topic. Table 2 lists the topics related to elections (for Topic 3, see Table 1).

Table 2: The BTM analysis of Topics 16, 24, and 28

ID	Keywords	Topic
16	people, office, political, campaign, country, government, trump, women, president, business	Election Trump
24	people, going, time, run, way, really, things, good, president, business	Policy
28	state, county, seat, city, district, former, office, election, senate, board	Senate election

Let us start with Topic 16, which is considered to be a topic related to ‘Election Trump’ because of the appearance of words such as *president*, *Trump*, and *campaign*. It appears that Topic 16 is similar to Topic 3. However, note that while Topic 3 expresses expectations for Trump, Topic 16 questions whether Trump will actually run for office, as in (5).

(5) Did you know that Donald Trump might actually ***run*** this time, instead of using our nation’s highest office to promote his reality-TV show?

Thus, it can be said that Topic 16, despite its semantic similarities, brings to the fore a slightly different aspect than Topic 3.

Turning now to Topic 24, a closer inspection of the keywords suggests that it contains many words of a generic nature, such as *way* and *things*. Looking at the instances related to this topic (see example 6 below), there are many words related to candidates running for presidential primaries, etc.

(6) It wasn’t until John McCain ***ran*** for president that I had serious reason to think otherwise.

Topic 28 is also slightly different from the election topics described so far. Although the topics discussed in this section are all related to the presidential election, neither the word *president* nor its derivatives are found in this topic, as shown in the following example.

(7) Johnson County Commissioner Bill Novotny, who […] ***ran*** former Nebraska Treasurer Shane Osborn’s unsuccessful campaign for Senate last year …

The example above talks about someone who ran for the Senate election. In other words, Topic 28 is not about a presidential election, but about running for Congress.

As described above, Topic 3, 16, 24, and 28 are election-related topics, but the meanings of *run* in these topics are slightly different from each other.[3] How have

3 In the encyclopedic semantics on which this paper draws, the notion ‘meaning’ is defined as follows: “[a] word meaning is […] *a perspective on our knowledge of the world…*” (Croft & Cruse,

the frequencies of these four meanings changed over time? Figure 4 is a VNC dendrogram overlaid on the monthly *theta* values. The horizontal axis represents the time series data points from January 2015, and the vertical axis the average monthly *theta* values. The dotted line indicates the transition of the average, with the solid blue line showing the smoothed LOWESS of the average.

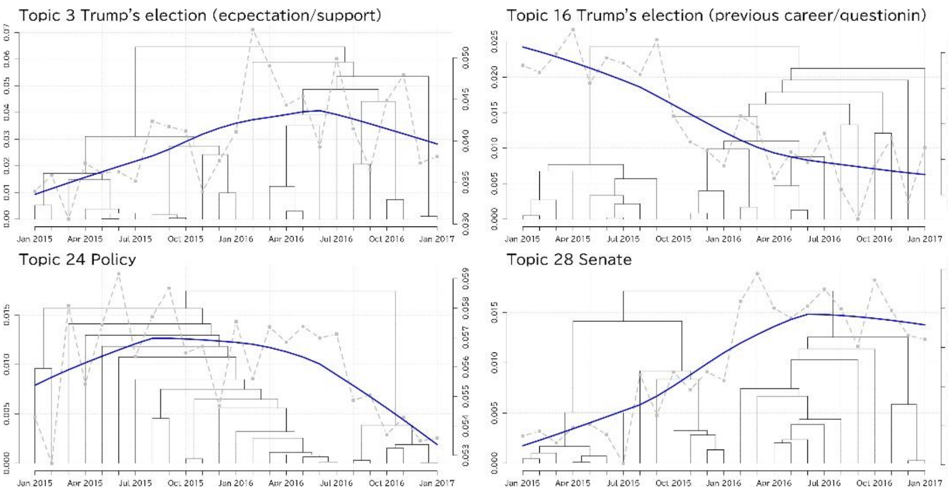

Figure 4: Changes in the probability of occurrence

From this figure, it can be seen that the changes in the frequency of use of each meaning of *run* show variously different patterns, in contrast to the pattern seen in Figure 3. The probability of occurrence of Topic 3 and Topic 28 are both on the increase, and their patterns of change are similar. It is possible that these similar patterns are due to the fact that the presidential and congressional elections were held on the same day, November 8, 2016. On the other hand, Topic 16 and Topic 24 both show a downward trend, but there are differences in the way the probability of occurrence of these topics changed depending on the meaning: as shown by the VNC dendrogram, the former changed significantly after September 2015. This could be taken as suggesting that Topic 16 is related to the question of whether Trump is really running for office, and may be because the peak in the probability of topic occurrence came around the time of the official announcement of his candidacy (June 2015). In contrast, Topic 24 represents a single cohort from 2015 to July 2016. This likely reflects the fact that policy-related topics appeared

2004, p. 30, emphasis by the authors, see also Fillmore, 1982). Given this view, the four meanings represented by each topic are considered to be used in slightly different ways in that they foreground different aspects of knowledge involved: knowledge related to endorsement and support, knowledge related to the qualifications of the electorate, knowledge related to policy, and knowledge related to senate elections.

numerous times during the election period, but the amount of coverage declined significantly after the election ended in September 2016.

4. Discussion and conclusion

In this article, the relationship between the meanings of *run* and the respective topics was discussed based on the results of the BTM analysis. Also discussed here was the relationship between the meanings of *run* and the changes in social interest. The results of the analysis indicate the importance of discussing word meaning in terms of topic and social interest. The findings beg two fundamental questions. What implications does the relationship between word meaning and topic have for the study of linguistic knowledge? And what are the implications of the study of word meaning and social interest for the study of meaning?

As for the former question, a possible answer seems to lie in the way linguistic knowledge is viewed in exemplar-based models of language, where it is argued that "memory storage for linguistic experience includes detailed information about the tokens that have been processed, *including their form and the contexts in which they were used*" (Bybee, 2013, p. 52; emphasis by the authors; see also Divjak, 2019, p. 53). The notion 'context' encompasses a variety of elements. If topic constitutes the elements of context, the implication is that linguistic experience with topic is also stored in the mind, i.e., it may also be part of the linguistic knowledge of language users.

Regarding the second question, that the change in the use of meanings of a word can be described in terms of a change in social interest lends support to the social turn in cognitive linguistics, which argues for the need to define the meaning of words in the broader context of "meaning-in-society" (Harder, 2010, p. 3. For a discussion of the relation between frame semantic knowledge and culture, see Kövecses, 2006). In other words, the implication of this study is that semantic research requires consideration from a social perspective.

Although BTM is a relatively new method and needs some improvements (section 3.2), its strength is in its ability to substantiate the relationship between word meanings and the topics in which they are used from a statistical standpoint. BTM also allows us to address the relationship between the respective meanings of a word and the social interest involved. The topic-based semantic analysis presented in this study can be useful in that it allows for methodologically rigorous analysis while largely eliminating the subjectivity of the analyst that tends to creep into semantic research.

References

Bybee, Joan. 2013. Usage-Based Theory and Exemplar Representations of Constructions. In T. Hoffmann and G. Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 49-69. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195396683.013.0004

Chang, Jonathan, Boyd-Graber, Jordan, Gerrish, Sean, Wang, Chong, and Blei, David M. 2009. Reading Tea Leaves: How Humans Interpret Topic Models. Proceedings of the 22nd International Conference on Neural Information Processing Systems (NIPS'09), 288-296.

Croft, William. 1993. The Role of Domains in the Interpretation of Metaphors and Metonymies. Cognitive Linguistics, 4(4), 335-370. https://doi.org/10.1515/cogl.1993.4.4.335

Croft, William. 2009. Toward a Social Cognitive Linguistics. In V. Evans and S. Pourcel (eds.), *New Directions in Cognitive Linguistics,* 395-420. John Benjamins. https://doi.org/10.1075/hcp.24.25cro

Croft, William and Cruse, Alan. 2004. *Cognitive Linguistics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511803864

Davies, Mark. 2016-. Corpus of News on the Web (NOW). Available online at https://corpus.byu.edu/now/.

Deignan, Alice. 2005. *Metaphor and Corpus Linguistics*. John Benjamins. https://doi.org/10.1075/celcr.6

Divjak, Dagmar. 2019. *Frequency in Language*. Cambridge University Press. https://doi.org/10.1017/9781316084410

Fillmore, Charles J. 1982. Frame Semantics. In The Linguistics Society of Korea (ed.), *Linguistics in the Morning Calm*, 111–137. Hanshin Publishing Company.

Glynn, Dylan. 2014. The Many Uses of *Run*: Corpus Methods and Soci-Cognitive Semantics. In D. Glynn and J. A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, 117-144. John Benjamins. https://doi.org/10.1075/hcp.43.05gly

Gries, Stefan Th. 2006. Corpus-Based Methods and Cognitive Semantics: The Many Senses of to *Run*. In S. T. Gries and A. Stefanowitsch (eds.), *Corpora in Cognitive Linguistics Corpus-Based Approaches to Syntax and Lexis,* 57-99. Mouton de Gruyter. https://doi.org/10.1515/9783110197709.57

Gries, Stefan Th. and Hilpert, Martin. 2008. The Identification of Stages in Diachronic Data: Variability-Based Neighbor Clustering. Corpora, 3(1), 59-81. https://doi.org/10.3366/E1749503208000075

Harder, Peter. 2010. *Meaning in Mind and Society: A Functional Contribution to the Social Turn in Cognitive Linguistics*. Mouton de Gruyter. https://doi.org/10.1515/9783110216059

Hilpert, Martin. 2012. Diachronic Collostructional Analysis: How to Use it and How to Deal with Confounding Factors. In K. Allan and J. A. Robinson (eds.), *Current Methods in Historical Semantics*, 133-160. Mouton de Gruyter. https://doi.org/10.1515/9783110252903.133

Hilpert, Martin. 2020. The Great Temptation: What Diachronic Corpus Data Do and Do Not Reveal about Social Change. In A. Nurmi, P. Rautionaho and J. Klemola (eds.), *Corpora and the Changing Society: Studies in the Evolution of English*, 3-28. John Benjamins. https://doi.org/10.1075/scl.96.01hil

Jansegers, Marlies and Gries, Stefan Th. 2017. Towards a Dynamic Behavioral Profile: A Diachronic Study of Polysemous *Sentir* in Spanish. Corpus Linguistics and Linguistic Theory, 16(1), 145-187. https://doi.org/10.1515/cllt-2016-0080

Kövecses, Zoltán. 2006. *Language, Mind, and Culture: A Practical Introduction*. Oxford University Press.

Peirsman, Yves, Heylen, Kris, and Geeraerts, Dirk. 2010. Applying Word Space Models to Sociolinguistics. Religion Names Before and After 9/11. In D. Geeraerts, G. Kristiansen, and Y. Peirsman (eds.), *Advances in Cognitive Sociolinguistics*, 111-137. Mouton de Gruyter. https://doi.org/10.1515/9783110226461.111

Sagi, Eyal, Kaufmann, Stefan, and Clark, Brady. 2012. Tracing Semantic Change with Latent Semantic Analysis. In K. Allan and J. A. Robinson (eds.), *Current Methods in Historical Semantics*, 161-183. Mouton de Gruyter. https://doi.org/10.1515/9783110252903.161

Taylor, John R. 1995. *Linguistics Categorization*. Oxford University Press.

Yan, Xiaohui, Guo, Jiafeng, Lan, Yanyan, and Cheng, Xueqi. 2013. A Biterm Topic Model for Short Texts. Proceedings of the 22nd International Conference on World Wide Web, 1445-1456. https://doi.org/10.1145/2488388.2488514