Event Detection from Blogs using Large Scale Analysis of Metaphorical Usage

Brian J. Goode[1,2], Juan Ignacio Reyes M.[3], Daniela R. Pardo-Yepez[3], Gabriel L. Canale[3], Richard M. Tong[4], David Mares[3], Michael Roan[2], and Naren Ramakrishnan[1]

[1] Discovery Analytics Center (Dept. of Computer Science), Virginia Tech
900 N. Glebe Rd., Arlington, VA, United States
{bjgoode,naren}@vt.edu
http://dac.cs.vt.edu

[2] Virginia Tech, Dept. of Mechanical Engineering
445 Goodwin Hall, Blacksburg, VA, United States
{mroan}@vt.edu
http://www.me.vt.edu

[3] Center for Iberian and Latin American Studies (CILAS), UC San Diego
9500 Gilman Dr., La Jolla, CA, United States
{jireyes,dpardoye,gcanale,dmares}@ucsd.edu
https://cilas.ucsd.edu

[4] Tarragon Consulting Corporation
1563 Solano Avenue, #350, Berkeley, CA, United States
{rtong}@tgncorp.com
http://www.tgncorp.com

Abstract. Metaphors shape the way people think, decide, and act. We hypothesize that large-scale variations in metaphor usage in blogs can be used as an indicator of societal events. To this end, we use metaphor analysis on a massive scale to study blogs in Latin America over a period ranging from 2000-2015, with most of our data occurring within a nine-year period. Using co-clustering, we form groups of similar behaving metaphors for Argentina, Ecuador, Mexico, and Venezuela and characterize overrepresented as well as underrepresented metaphors for specific locations. We then focus on the metaphor's potential relation to events by studying the tobacco tax increase in Mexico from 2009-2011. We study correspondences between changes in metaphor frequency with event occurrences, as well as the effect of temporal scaling of data windows on the frequency relationship between metaphors and events.

Keywords: metaphors, blogs, open source indicators, event detection

1 Introduction and Related Work

Conceptual metaphors associate an abstract target concept with a concrete source concept. So, we say things like "life is a journey" or "poverty is a disease". Individual mappings of source-target concept pairs in language are linguistic metaphors. Such constructions facilitate knowledge dissemination, bridge concepts over a variety of subject areas, and unwittingly affect human behavior through intrinsic associations and world-views [4, 7, 10]. There have been a number of attempts to quantify these effects in an experimental framework [14–16]. Despite the confirming evidence, the effects of metaphors in everyday interactions and mental constructs remains unclear [5]. Metaphor exposure and adoption alone is not necessarily responsible for action, because anchoring and time also dictate decision outcomes [14]. The impact of metaphor on behavior may largely depend on its age or relative conventionality (i.e., becomes a definition) in society [6, 8]. Applied to complex social interactions and events in everyday scenarios, capturing changes in metaphor usage can become an untenable undertaking. At present, we adopt the assumption that conceptual mappings governing action within a population are present to some degree in language, and can be used for event detection.

Prior effort has focused on the use of Open Source Indicators (OSI) as a tool for improving the forecasting of events [13]. In such analyses, models with keywords relevant to events such as protests and disease epidemics steer algorithms to predict the probability of occurrence [11, 3]. However, we are not blind to the operational modeling challenges that exist when incorporating metaphors [2]. The view taken in this paper is that we can re-purpose the idea of the keyword dictionary to extract relevant text that now consists of a tuple, $M_L := (source, target)$. This tuple is one of many linguistic metaphors mapped to a given conceptual metaphor. Extracting these linguistic metaphors in documents is a subject of continued research (e.g., [1, 12]).

This is the first study to present a method of large scale analysis of metaphorical usage in blogs. Our data derives from 327 Latin American blogs spanning the years 2000-2015. We extracted 589,089 documents from Argentina, Ecuador, Mexico, and Venezuela. Blogs from these countries were chosen for their rich use of metaphorical language and discussion of significant political events. We performed spectral co-clustering over the time-series correlation matrix to identify groups of metaphors that appeared in the text in close temporal proximity and linear proportionality. Through the course of the analysis, we investigate:

1. What do concepts mapped to linguistic metaphors reveal about differences in blogger populations?
2. Do prevailing conceptual references of linguistic metaphors change over the course of specific events?

In general, identifying equilibrium behavior is highly subjective. We can detect clusters of different slants on particular target concepts like taxation and government, but locally prevailing source concepts can be subject to variations that are dependent on the size of the data window (scaling effects).

2 Methods

Blog selection. Blogs are studied, because initial investigations revealed more metaphorical language than similarly timed news articles containing more factual description of events. Candidate blogs were required to be located in the country of interest, express political opinion, and reference events published in a major news outlet. These were found using either a directory (e.g., `http://blogsdemexico.com.mx/`) or through Google searches. The main search format was: [Political blog, name of the country/topic]. Examples of searches include: "Blog politica, Argentina, Venezuela", "blogs problemas sociales", "blogs politicos", etc. Specific references to hosting sites were also used such as "blogspot Pena Nieto". Topics of the blogs include societal issues, the economy, and criticism of current or former government and politicians. Most of the blogs were written in Spanish; however, some of the Venezuelan blogs were written in English for a number of possible reasons. This is an unavoidable source of bias resulting from inconsistencies across languages in metaphor usage. Therefore, we only consider blogs written in Spanish. Summary details of the corpus are given in Table 1. The ratio of documents containing linguistic metaphors (LMs) remains fairly consistent across countries. The LMs included are only those mapped to seven target concepts discussed below, not all LMs present in the data.

Table 1: Summary details of the blog corpus by country.

Country	# blogs	# docs	# LMs	LM-doc ratio
Argentina	89	111437	53283	0.48
Ecuador	59	13899	5278	0.38
Mexico	97	306101	157553	0.51
Venezuela	82	157652	75869	0.48
Total	327	589089	291983	0.49

Metaphor extraction. Linguistic metaphors were extracted using software, the Metaphor Detection System (MDS), developed on the IARPA Metaphor program.[5] The goal of this program was to develop automated systems to identify the beliefs and world-views of different cultures by examining their use of metaphorical language. The MDS automatically detects linguistic metaphors in the original Spanish text and then maps them to pairs of pre-defined conceptual constructs. The MDS uses a mixture of techniques, including lexico-syntactic pattern matching and classification using semantic and psycholinguistic features. Internal program testing showed the MDS capable of achieving an LM detection F-score of 0.74 (precision 0.82, recall 0.68) in Spanish. As stated earlier, the goal

[5] See: `http://www.iarpa.gov/index.php/research-programs/metaphor`.

of this paper is not to focus on the mechanics of MDS itself (covered elsewhere) but to present the first large-scale analysis of metaphors in blogs.

Both linguistic metaphors and their mappings to source-target concept pairs are used in this analysis.[6] The conceptual targets we consider in this analysis are $m_t \in \{$ BUREAUCRACY, DEMOCRACY, ELECTIONS, GOVERNMENT, POVERTY, TAXATION, and WEALTH$\}$. We analyze 84 conceptual sources $m_s \in \{$ CRIME, MACHINE, MOVEMENT, $\ldots\}$. The linguistic sources and targets are a dictionary of political keywords, such as *"government"*, that map to a conceptual source or target domain. Given the translated text from our corpus, *"the usurper government"*, *usurper* is the linguistic source and *government* is the linguistic target. Similarly, *usurper* maps to the source concept CRIME, and *government* maps to the target concept GOVERNMENT. These source and target concepts form the tuple, $M := (m_s, m_t) \in \mathcal{M}$. The set $\mathcal{M}$ consists of all source-target pairs, and has cardinality ($|\mathcal{M}| = 672$). Our dataset is $X \subseteq T \times \mathcal{M} \times \mathcal{B}$ for daily timestamps $t \in T$ and blogs $B \in \mathcal{B}$. Time-series are generated for each blog and concept pair, $f_{M,B}(t) = |\{x \in X | x = (t, M, B)\}|$, that indicate the frequency of the metaphors appearing at time, t.

3 Results and Discussion

Blogs are heterogeneous in nature, because they can range in publication frequency, content length, and production effort. The prototypical blog post may not exist on all dimensions, because they display characteristics of both academic articles and Twitter posts. Although germane to the discussion, we do not directly address the qualities composing the blogs, and focus the analysis to aggregating a variety of different blogs by target concept through co-clustering.

Clustering of concept pairs. Clusters of related metaphors are formed over summed times series of metaphor frequencies,

$$F_M(t) = \sum_{\mathcal{B}} w_B f_{B,M}(t) \tag{1}$$

where we assume $w_B = 1$. Setting $w_B = 1$ assumes that the blogs contributing to the proportional response are of equal weighting. This is perhaps sub-optimal, but it is a maximum-entropy approach for reducing noise fluctuations in individual blogs given no prior information. We could reduce the argument further by considering the relative contribution per writer (multiple per blog) given certain types of events, but then lose the information inherent in the collection. For now, w_B is a parameter for investigation in future work.

The spectral co-clustering algorithm [9] forms clusters of similar source concepts, m_s for a given target concept m_t. Co-clustering is performed over the

[6] Target and source concepts will be represented in all caps: e.g., ELECTIONS.

Pearson correlation matrix C, to find evidence of similar linear proportional responses. The correlation matrix is $N \times N$, where N is the number of source concepts m_s for a given target concept,

$$c_{ij} = \frac{\sum_1^K \hat{F}_{M_i}(t_k)\hat{F}_{M_j}(t_k) - K\bar{\hat{F}}_{M_i}\bar{\hat{F}}_{M_j}}{\sqrt{\sum_1^K \hat{F}^2_{M_i}(t_k) - K\bar{\hat{F}}^2_{M_i}}\sqrt{\sum_1^K \hat{F}^2_{M_j}(t_k) - K\bar{\hat{F}}^2_{M_j}}} \tag{2}$$

where $\hat{F}_M = (F_M * W)(t)$ is a windowed time-series from $t \in \{1, K\}$ for metaphor M_i and M_j, and $\bar{\hat{F}}$ is the mean of the time-series. The Hamming window is applied in the convolution of F_M,

$$W(\gamma) = 0.54 - 0.46\cos\left(\frac{2\pi\gamma}{\Gamma - 1}\right), 0 \leq \gamma \leq \Gamma - 1 \tag{3}$$

where Γ is a parameter that controls the length of the window. The Hamming window places more weight on the current time step while the future and past extrema are minimized. The convolution is phase centered to preserve temporal accuracy and assume a symmetric lead-lag duration between blog and event dates. This choice of filter (as opposed to cross-correlation) reduces the effects of inconsistent lead-lag responses of publication dates around events for small Γ. For large Γ, this window compensates for unknown lead-lag relationships to frequencies across multiple events.

The number of clusters were chosen to minimize the objective function,

$$\beta^* = \arg\min_\beta \phi(C(\beta)),\ \phi(C) = \sum_{i=1,j>i} ijc_{ij} \tag{4}$$

where $\phi(C)$ minimizes entries appearing on the off-diagonal, and β is the optimal number of clusters. Minimizing this objective discourages highly correlated time-series from appearing in separate clusters. This increases discrimination of clusters by forcing smaller highly correlated clusters, when possible. The co-clustering algorithm was initialized randomly for 20 trials at each choice of β.

The source concept clusters grouped by target for $\Gamma = 365$ are shown for Mexico[7] in Fig. 1. Source concepts with higher frequency counts are in the center of the cluster along the y-axis. For each target concept, different countries have different groupings of source concepts. However, there are some common trends in all of our datasets. ELECTIONS, as a target concept, tend to have a thin cluster having COMPETITION as the most frequent source concept. In our data for both Mexico and Argentina, we tend to see increased metaphor usage around the time of elections. The target concepts, POVERTY and GOVERNMENT, tend to have one large cluster and one small cluster of source concepts. The large cluster tends to be the source BUILDING, and the smaller cluster varies. POVERTY, however,

[7] Argentina, Ecuador, and Venezuela are not shown to conserve space. Qualitative results from these countries are discussed in terms of their similarities to our Mexico dataset. Differences are explicitly highlighted in Table 2.

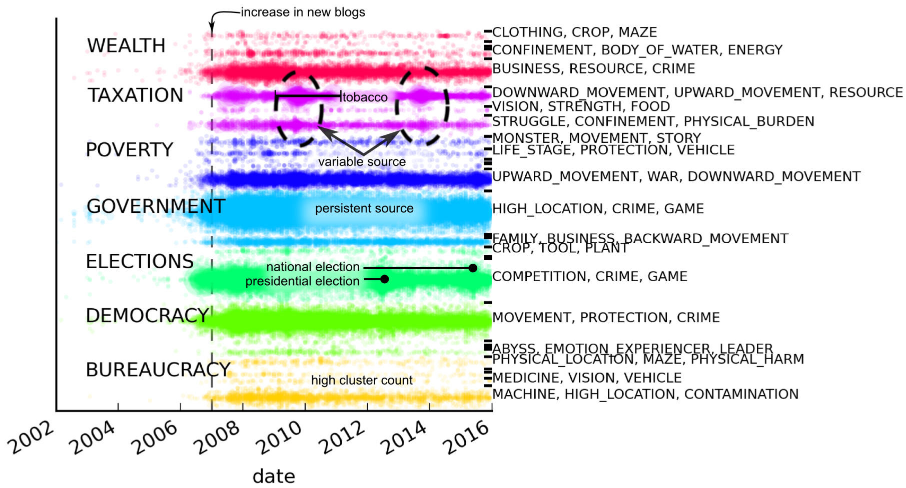

Fig. 1: Source clusters of linguistic metaphors are shown for blogs in Mexico, $\Gamma = 365$. Each horizontal position corresponds to a particular source concept, M. Clusters are identified by the partitions on the right y-axis. Each time-series is color-coded to a target, and the opaqueness shows increased metaphor frequency per blog at that timestamp, and marker size indicates the frequency of usage on a given date by a particular blog. Larger clusters show source names.

has a different set of source clusters for each country indicating the possibility of regional variations. Lastly, BUREAUCRACY tends to have many smaller source clusters, which indicates more variety in the descriptions of this target.

Quantitatively, with this amount of data, it is likely that the distributions of source concepts associated with targets are going to be different[8], if not by chance alone. Therefore, we consider the table of standard residuals to find the relative target and source pairs that are markedly different between countries as shown in Table 2. This table shows results for summed blog data, $F_M(t)$ over the entire window of time from 2000-2015. The subtables are grouped by target, and the source concepts showing the most differences are shown in bold. Ecuador is not included for having too few data relative to Argentina, Mexico, and Venezuela to make a proper comparison. Of the highest relative standard residuals, highlighted in bold, Mexico tends to deviate from the distributions of Argentina and Venezuela for BUREAUCRACY as a HIGH_LOCATION and for the increased usage of ELECTIONS as both a CRIME and BUSINESS, without as much ELECTIONS as a GAME (i.e., something won). Mexican blogs also the DEMOCRACY as BUILDING tends to be underrepresented as well. However, target concept WEALTH shows more diverse differences where compared to Argentina, Mexico shows increased usage of BUSINESS, CRIME, RESOURCE, and LOW_LOCATION sources as opposed to MOVEMENT and BUILDING.

[8] χ^2 test ($\Gamma = 365$), $p < 0.001$ for all targets in all countries

Table 2: χ^2 standardized residuals for target-source concepts; larger (± 10) differences in source concepts relative to sample countries are shown in bold.

BUREAUCRACY	Argentina	Mexico	Venezuela
CRIME	-2.19	-5.05	7.70
UPWARD_MOVEMENT	3.20	-4.65	2.56
RESOURCE	4.00	-4.17	1.31
BARRIER	6.56	-5.33	0.42
PHYSICAL_BURDEN	-0.52	-1.02	1.62
PROTECTION	1.03	2.16	-3.37
MORAL_DUTY	-2.57	5.10	-3.63
HUMAN_BODY	3.98	-4.04	1.18
HIGH_LOCATION	**-10.35**	**18.89**	**-12.69**
MACHINE	-1.26	-0.81	2.03
WAR	**4.75**	**-10.07**	**7.43**
BUILDING	3.08	-1.48	-0.97
CONTAMINATION	-2.37	1.59	0.24

DEMOCRACY	Argentina	Mexico	Venezuela
CRIME	-2.47	3.81	-2.37
MOVEMENT	-0.07	6.00	-6.39
PROTECTION	-7.70	5.08	-0.08
BUILDING	**15.70**	**-19.52**	**10.01**

ELECTIONS	Argentina	Mexico	Venezuela
BUSINESS	**-11.30**	**21.26**	**-14.85**
CRIME	**-40.19**	**50.06**	**-23.98**
GAME	**36.45**	**-52.60**	**29.88**
COMPETITION	8.87	-7.41	1.19
STRUGGLE	-3.94	7.34	-5.10

POVERTY	Argentina	Mexico	Venezuela
FORCEFUL_EXTRACTION	-3.94	-6.02	10.41
CRIME	-1.64	3.45	-2.46
RESOURCE	1.25	1.06	-2.33
CONFINEMENT	4.25	-7.11	4.28
PHYSICAL_LOCATION	1.19	-1.38	0.51
DISEASE	-2.97	3.15	-0.92
STRUGGLE	-1.54	-2.24	3.94
WAR	**-12.04**	**15.33**	**-6.66**
SIZE	0.77	3.93	-5.18
BUILDING	**22.49**	**-15.62**	**-2.41**
DOWNWARD_MOVEMENT	**8.64**	**-11.48**	**5.32**
PHYSICAL_HARM	-4.39	6.35	-3.29
UPWARD_MOVEMENT	-5.40	6.27	-2.30

GOVERNMENT	Argentina	Mexico	Venezuela
FURNISHINGS	**-11.85**	**18.04**	**-10.34**
BUSINESS	**-5.11**	**11.73**	**-8.93**
CRIME	**-18.78**	**27.48**	**-15.12**
STRENGTH	-0.14	3.61	-3.97
GAME	-4.50	1.72	1.89
PROTECTION	2.13	0.03	-1.84
RESOURCE	6.66	-5.40	0.45
PHYSICAL_BURDEN	0.11	2.90	-3.37
SHAPE	4.72	-9.12	6.30
COMPETITION	1.23	7.86	-9.95
MORAL_DUTY	-0.85	3.78	-3.55
CONFINEMENT	6.80	-4.79	-0.37
ENSLAVEMENT	1.00	-5.62	5.51
THEFT	**-13.79**	**27.67**	**-19.57**
HUMAN_BODY	4.34	-9.88	7.49
HIGH_LOCATION	**-20.94**	**19.83**	**-4.62**
LEADER	-4.36	6.93	-4.13
MOVEMENT	**3.86**	**-11.09**	**9.26**
STORY	**12.49**	**-13.67**	**4.85**
PHYSICAL_LOCATION	-1.39	0.04	1.14
MACHINE	**5.40**	**-15.14**	**12.54**
STRUGGLE	13.01	-6.38	-3.85
WAR	**6.02**	**-25.65**	**23.91**
BUILDING	6.68	-9.49	5.06
WEAKNESS	2.36	-0.66	-1.25
PHYSICAL_HARM	6.72	-6.82	2.00
CONTAMINATION	-4.86	7.95	-4.86
UPWARD_MOVEMENT	8.75	-4.70	-2.13

TAXATION	Argentina	Mexico	Venezuela
RESOURCE	1.91	-2.98	2.15
BUILDING	-1.56	2.94	-2.60
DOWNWARD_MOVEMENT	**13.47**	**-16.59**	**7.68**
CONFINEMENT	1.81	-1.53	-0.15
UPWARD_MOVEMENT	**-10.51**	**12.30**	**-4.89**

WEALTH	Argentina	Mexico	Venezuela
BUSINESS	**-15.72**	**21.21**	**-7.93**
CRIME	**-18.70**	**20.35**	**-2.79**
GAME	-6.67	7.07	-0.75
RESOURCE	**-19.83**	**16.99**	**3.28**
MOVEMENT	**40.14**	**-39.03**	**-0.34**
LOW_LOCATION	**-6.82**	**11.28**	**-6.28**
BUILDING	**17.91**	**-23.49**	**8.12**

Cluster variation around events. One of the challenges of using blogs and metaphors to detect events is that there is no one specific type of event that corresponds to a particular conceptual metaphor. For example, POVERTY as a MONSTER can refer to a number of different events in the realm of government, economics, civil unrest, and politics. Modeling an individual interpretation of an event through metaphor is likely not to converge to a particular expected outcome. However, given enough data points of correlated trends in a cluster surrounding a particular event may have insight into this complex relationship.

As an example, we study the metaphors associated with the target concept, TAXATION in Mexico. In Fig. 1, we see that there a number of intervals of time, particularly between 2009-2011, and just after 2014 where there is an increase in both the number of blogs and source concepts referenced in our data. These are noted by the more opaque patches and increased path width, respectively. During the 2009-2011 interval, Mexico had an increase in the tobacco tax. Figure 2 shows the time-series produced by the top three sources of both time-varying TAXATION metaphor clusters shown in Fig. 1. The clusters are labeled “1” and “2”, respectively. The black time-series is the frequency of occurrence of the words “tabaco” and “cigarrillo” in our Mexican blog corpus.

These metaphors tend to vary proportionally with respect to the appearance of the tobacco keywords. However, the CONFINEMENT, STRUGGLE, and PHYSICAL_BURDEN source concepts tend to focus around the same interval as the tobacco keywords peak. In these time-series 82% of the tobacco references and

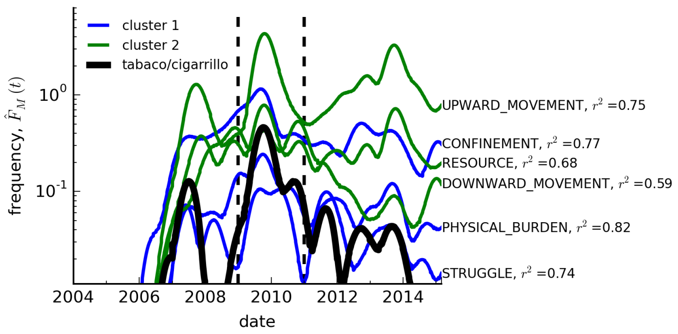

Fig. 2: $\hat{F}_M(t)$ ($\Gamma = 365$) for the TAXATION target concept in Mexico. The vertical lines indicate the years during the period of increased tobacco tax.

taxation metaphors occurred within the same sentence, indicating a strong relationship to the metaphors referring to the tax increase during this time period.

The association of concept pairs to events is limited by the mapping of linguistic metaphors. In this example, the TAXATION as an UPWARD_MOVEMENT is commonly referring to a tax increase or "impuesto" and "elevar". The linguistic metaphors supporting TAXATION as a CONFINEMENT conceptual metaphor are referring to evasion or being captive. There is no guarantee that any of these metaphors is strictly referring to the tobacco tax, but in aggregate we see mutual linearly correlated fluctuations in the times series. The least correlated is TAXATION as a DOWNWARD_MOVEMENT, and the most correlated is PHYSICAL_BURDEN. Therefore, we can extract a linearly proportional relationship between metaphors and events, but is currently on a case-by-case basis.

Effects of varying the data window, Γ. Inconsistencies in publication times, metaphor usage and events affect the temporal relationship between metaphors and events. Our results up to this point have focused on clustering using a data window of size, $\Gamma = 365$. We assume that applying a Hamming window of length Γ over the data helps to disperse signal energy symmetrically in time to smooth variations in blog publications around relevant events. However, in larger data windows, we expect a decrease in the number of clusters, and certain metaphors may exhibit more similar behaviors over different temporal windows.

If the metaphors had similar temporal characteristics at every windowing parameter, Γ, we would expect to see the clusters merging as the window length increases. However, the network in Figure 3 shows a different path of cluster formation. Each edge weight corresponds to the number of same sources in each node (cluster). Initially, there are many small clusters containing few sources. As the window length increases, many of these smaller clusters are absorbed as would be expected. These are identified by the downward arrow showing similar temporal trends. There are also a number of crossings between clusters, and often these crossings contain groups of 10 metaphors. As the window approaches the length of the time-series (4096 time steps), the clusters converge to two nodes with one containing most of the sources. Although there is a trajectory of thick edges connecting the windows of increasing size, the cross-overs indicate the possibility of different temporal similarities between source-target concept pairs, and the possibility of no "characteristic" time window for analysis of events.

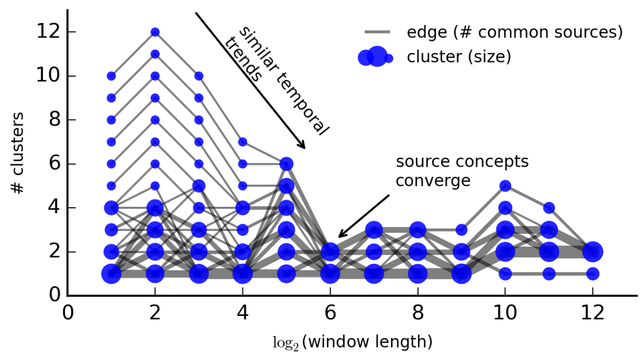

Fig. 3: The path of source concept clusters for TAXATION in the Mexico dataset reveals different cluster configurations, and the "right" time-scale is less obvious.

4 Conclusions

This is the first large scale analysis of metaphors in Latin American blogs for event detection. We discuss qualitative similarities of cluster formation when grouped according to correlation, but source concepts will vary when aggregated by country. Using a co-clustering approach, we identify times of interest and show an example of a linear proportional relationship of metaphor usage regarding taxation around a period of time when Mexico enacted a tobacco tax. However, when generalizing the approach to multiple scales, the clusters do not necessarily show a characteristic time-scale. This indicates that different relationships could be realized over different data windows. Future work will consist of a two-pronged approach. First, we aim to better understand what the different window lengths may come to represent. Second, aggregating heterogeneous blogs is a non-trivial process. To refine our approach, we aim to better understand the uncertainty of blog temporal behavior to better estimate prior probabilities of publication style and metaphor usage parameters in this diverse data source.

Acknowledgments. Supported by the Intelligence Advanced Research Projects Activity (IARPA) via DoI/NBC contract number D12PC000337, the US Government is authorized to reproduce and distribute reprints of this work for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the US Government.

References

1. Bracewell, D., Tomlinson, M., Mohler, M., Rink, B.: A tiered approach to the recognition of metaphor. In: Gelbukh, A. (ed.) Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science, vol. 8403, pp. 403–414. Springer Berlin Heidelberg (2014)
2. Carroll, J.M., Mack, R.L.: Metaphor, computing systems, and active learning. Int. J. Hum.-Comput. Stud. 51(2), 385–403 (Aug 1999)
3. Chakraborty, P. et al.: Forecasting a moving target: Ensemble models for ili case count predictions. In: Proceedings of the 2014 SIAM International Conference on Data Mining. pp. 262–270 (2014)
4. Geary, J.: I Is an Other: The Secret Life of Metaphor and How it Shapes the Way We See the World. HarperCollins Publishers, New York, NY (2011)
5. Gibbs Jr., R.W.: The real complexities of psycholinguistic research on metaphor. Language Sciences 40, 45 – 52 (2013)
6. Golshaie, R., Golfam, A.: Processing conventional conceptual metaphors in persian: A corpus-based psycholinguistic study. Journal of Psycholinguistic Research 44(5), 495–518 (2014)
7. Johnson, M.: Metaphorical reasoning. The Southern Journal of Philosophy 21(3), 371–389 (1983)
8. Keysar, B., Shen, Y., Glucksberg, S., Horton, W.S.: Conventional language: How metaphorical is it? Journal of Memory and Language 43(4), 576 – 593 (2000)
9. Kluger, Y., Basri, R., Chang, J.T., Gerstein, M.: Spectral biclustering of microarray data: Coclustering genes and conditions. Genome Research 13(4), 703–716 (2003)
10. Lakoff, G., Johnson, M.: Metaphors We Live By. University of Chicago Press, Chicago, IL (2008)
11. Muthiah, S. et al.: Planned protest modeling in news and social media. In: Innovative Applications of Artificial Intelligence (2015)
12. Neuman, Y. et al.: Metaphor identification in large texts corpora. PLoS ONE 8(4), e62343 (04 2013)
13. Ramakrishnan, N. et al.: 'Beating the News' with EMBERS: Forecasting civil unrest using open source indicators. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1799–1808. KDD '14, ACM, New York, NY, USA (2014)
14. Thibodeau, P.H., Boroditsky, L.: Metaphors we think with: The role of metaphor in reasoning. PLoS ONE 6(2), e16782 (02 2011)
15. Thibodeau, P.H., Boroditsky, L.: Natural language metaphors covertly influence reasoning. PLoS ONE 8(1), e52961 (01 2013)
16. Thibodeau, P.H., Boroditsky, L.: Measuring effects of metaphor in a dynamic opinion landscape. PLoS ONE 10(7), e0133939 (07 2015)