Multilingual Connotation Frames: A Case Study on Social Media for Targeted Sentiment Analysis and Forecast

Hannah Rashkin[†] **Eric Bell**[‡] **Yejin Choi**[†] **Svitlana Volkova**[‡]

[†]Paul G. Allen School of Computer Science & Engineering, University of Washington
{hrashkin,yejin}@cs.washington.edu
[‡]Data Sciences and Analytics, Pacific Northwest National Laboratory
{eric.bell,svitlana.volkova}@pnnl.gov

Abstract

People around the globe respond to major real world events through social media. To study targeted public sentiments across many languages and geographic locations, we introduce *multilingual connotation frames*: an extension from English connotation frames of Rashkin et al. (2016) with 10 additional European languages, focusing on the implied sentiments among event participants engaged in a frame. As a case study, we present large scale analysis on targeted public sentiments toward salient events and entities using 1.2 million multilingual connotation frames extracted from Twitter.

1 Introduction

People around the globe use social media to express their reflections and opinions on major real world events (Atefeh and Khreich, 2015; Radinsky and Horvitz, 2013). In order to facilitate multilingual public sentiment tracking on social media, we introduce *multilingual connotation frames*,[1] a multilingual extension from English connotation frames of Rashkin et al. (2016) with 10 additional European languages, including low-resource languages such as Polish, Finnish, and Russian.

Definition 1.1. Connotation Frames: A framework for encoding predicate-specific connotative relationships implied by a predicate towards its arguments.

Figure 1 shows a selected subset of the connotation frames that is relevant in our study. See Rashkin et al. (2016) for the full description of the connotation frames.

---

[1]Publicly available at homes.cs.washington.edu/~hrashkin/multicf.html.

English Verb: survive
Other languages: survivre, sobrevivir, überleben…

Example Tweets

"US teenager … also survived Boston Marathon bombing"

"19-jähriger Missionar überlebt drei Terroranschläge"

"Este joven ha sobrevivido a los atentados de Boston, de París y de Bruselas"
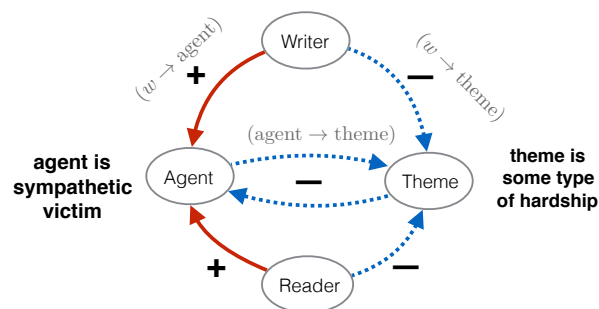
Connotation Frame for *surviving* verbs:

Figure 1: The connotation frame of *"survive"* with respect to directional sentiments among *"writer"*, *"agent"*, *"theme"*, and *"reader"*. The tweet examples show automatically induced multilingual connotation frames.

There are two important benefits to develop multilingual connotation frames. First, they serve as a unique lexical resource to enable *targeted* sentiment analysis, which rarely exists for most languages.

Definition 1.2. Targeted Sentiment: A sentiment label indicating how a source entity feels about a target entity.

In the example shown in Figure 1, *"teenager survived Boston Marathon bombing"*, the connotation frame allows us to correctly interpret (implied) targeted sentiments including:

1. sentiment(teenager → bombing) = −
2. sentiment(writer → bombing) = −
3. sentiment(writer → teenager) = +

Second, they allow us to study a broad spectrum of sentiments including nuanced ones; in the ex-

Lang	# Tuples	Examples
EN	643,004	(korea, fires, missile)
ES	305,310	(acuerdo, vulnera, derecho)
FR	85,286	(obama, quitte, cuba)
PT	76,849	(valentino, renova, contrato)
RU	28,511	(путин, обсуди, моста)
DE	23,197	(seehofer, lösen, flüchtlingskrise)
NL	14,091	(artiesten, bevestigd, komst)
IT	13,586	(conte, lascia, nazionale)
FI	2,859	(hans, tekemään, valintansa)
SV	2,229	(rubio, avslutar, kampanj)
PL	2,226	(papież, wyrazi, zgodę)

Table 1: The number and the example of (agent, verb, theme) tuples extracted from tweets across languages.

amples discussed above, connotation frames allow us to infer (1) the likely sentiment among the event participants (e.g., a surviving teenager is likely to be negative toward the Boston bombing), and (2) the likely sentiment of the author towards events and entities (e.g., the writer is likely to be sympathetic toward the teenager while negative toward the incident), even though none of these sentiment implications are overtly expressed.

To validate the empirical utility of the new multilingual connotation lexicon, we present a successful case study of large scale connotation analysis (Section 4.1) and forecast (Section 4.2) based on connotation frames extracted from 1.2 million tweets in 10 different European languages spanning over a 15 day period.

2 Multilingual Twitter Dataset

We obtained multilingual geo-located tweets spanning Mar 15 – Mar 29, 2016. This 15 day duration covers Brussels attacks on Mar 22 as well as one whole week before and after, allowing us to study the public sentiment dynamics in response to a major terrorist event. We focus on tweets that are likely to be about "news-worthy" topics by selecting tweets that came from trusted sources such as twitter-verified accounts or known news accounts, or contained hashtags `#breaking` or `#news`.[2] We used SyntaxNet dependency parser (Andor et al., 2016) and trained additional SyntaxNet models for 10 non-English languages using Universal Dependencies annotations.[3] We extracted 1.2 million agent-verb-theme tuples as listed in Table 1.

---

[2]We experimented with other automatic ways of finding news sources by identifying popular hashtags, but this method did not translate well to other languages.
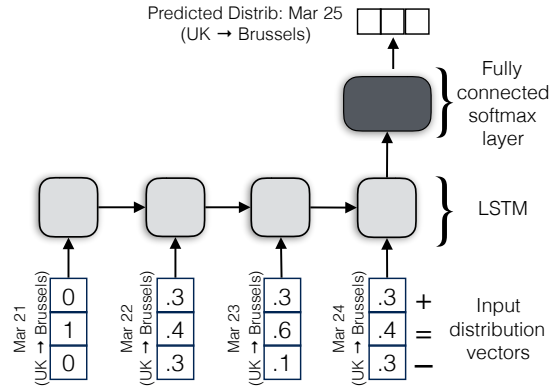
[3]http://universaldependencies.org/

Figure 2: Diagram of LSTM model for predicting the distribution of perspectives from a location (e.g., UK) towards an entity (e.g., Brussels) on a given day (e.g., March 25), based on the previous days.

3 Methods

3.1 Multilingual Connotation Frames

We perform context-based projection of English connotation frames to 10 additional European languages using large parallel corpora. Since connotation of a word arises from the context in which the word is used, we want to ensure the translated connotation frames are used in similar contexts. We use existing parallel corpora with automatic word-alignment: the Opus Corpus (Tiedemann, 2012) using Multi-UN parallel data (Eisele and Chen, 2010) for Russian and EuroParl parallel data (Koehn, 2005) for all other languages.

More concretely, for each non-English verb, v' (e.g., *assassiner* in French), we compute the probability of it being translated to English verb v by counting the alignments.

We then define the connotation frame of v', $\mathcal{F}(v')$, by transferring the connotation frame of the English verb $v*$, $\mathcal{F}(v*)$, that has the highest translation probability:

$$v* = \text{argmax}_v p(v|v')$$

$$\mathcal{F}(v') = \mathcal{F}(v*)$$

For example, the connotation frame for *assassiner* is propagated from *murder*, the English word that it is aligned with the most.

3.2 Extracting Targeted Sentiments

Using the connotation frame lexicon, we compute the distribution of targeted sentiments towards most-frequently-mentioned named entities. We also compute sentiments expressed by each
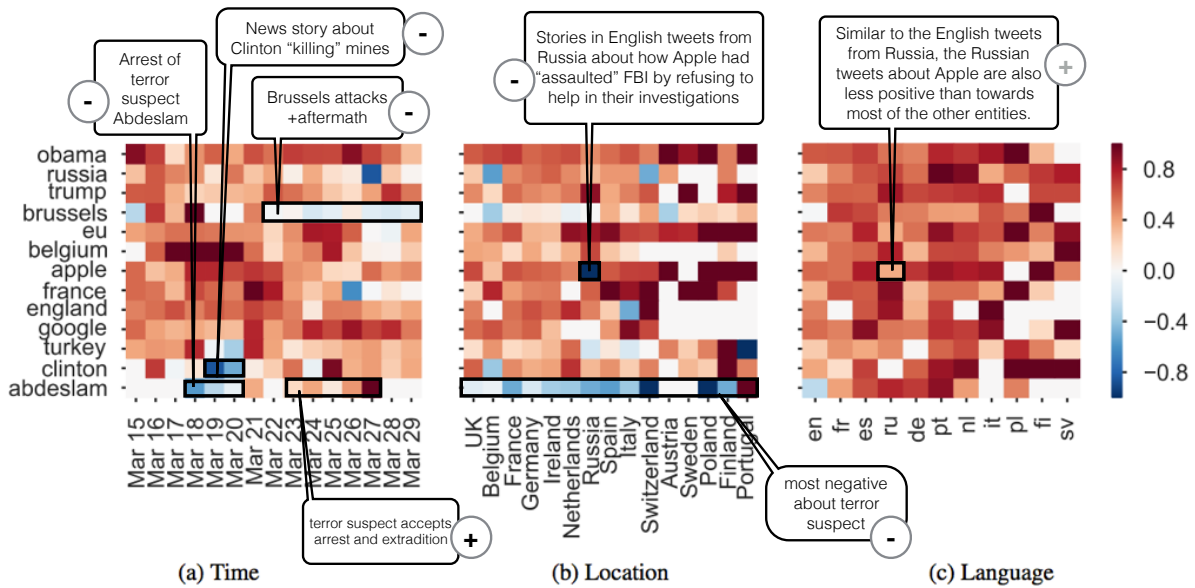
Figure 3: Heatmap of expected perspectives towards 13 named entities over 2 week period using only English tweets from European countries. Red is more positive, blue is more negative.

country by aggregating all sentiments of the writers located in that country (e.g., the distribution of positive, neutral, and negative perspectives expressed towards Obama in British tweets). The aggregated polarities can be represented as a 3-dimensional probability vector, $p = [p^+ p^= p^-]$, as will be used in the sentiment forecast task below. For other analysis, we summarize this polarity distribution as a scalar score by taking the expected value of the polarity: $E[p] = p^+ - p^-$.

3.3 Forecasting Sentiment Dynamics

We also study forecasting sentiment dynamics: predicting the sentiment distribution of the next day given the sentiment trend of previous days. For this task, we track the distribution of directional sentiments from each country towards the hundred most-frequently-mentioned named entities. At test-time, each model is given directional sentiment distributions for the previous 4 days as input and predicts tomorrow's distribution (e.g., forecasting 1 day ahead). We also train models for predicting the distribution half a week later (forecasting 4 days ahead).

We performed an additional experiment for English (EN_J) where the perspectives of all countries are aggregated together in order to predict the global perspective. For all experiments, we use 10-fold cross-validation and measure the symmetric Kullback-Leibler (KL) divergence between the true distribution and the predicted one.

We experiment with Long Short-term Memory

models (LSTMs) (Hochreiter and Schmidhuber, 1997) to integrate the dynamic contextual information from the past, as depicted in 2. Hidden dimension is 16, and we use ADAM optimization with KL divergence as the objective. For implementation, we use Keras[4] on top of Theano.[5]

Baselines We use two baselines. The first is MEAN, the average distribution seen in the training data. The second are SVMs with linear kernels, which worked well in predicting influenza activity in a similar set-up (Santillana et al., 2015). For the baselines, we encode the distributions from the 4 previous days as a flattened 12-dim. vector, and each portion of the distribution is predicted separately.

4 Results

4.1 Connotation Analysis

For the most frequently mentioned named entities, we compute heatmaps of the expected perspective being expressed towards that entity.

In Figure 3A, we use the English tweets from European countries to plot the change in connotative polarity towards these entities over the course of the 15 day period. Generally, the changes in polarity from day-to-day seem to be gradual and frequently are similar to the day before. There are a couple of exceptions e.g., the polarity towards

---

[4] https://keras.io
[5] https://deeplearning.net/software/theano

Brussels changes abruptly on March 22 (the day of the Brussels attacks), reflecting the change in tone of all tweets related to Brussels at that time.

Overall, there are mostly positive polarities expressed. This may reflect people's tendency to avoid phrasing stories too harshly, choosing to be more euphemistic even when they discuss bad news.

In Figure 3B, we aggregated the polarities of these tweets by country of origin. While most of the polarities are positive-strongly positive, the tweets about Brussels and Belgium are more neutral or even slightly negative.

Lastly, in Figure 3C, we used all of the tweets from European countries to aggregate expected polarities in 11 different languages. Non-English languages show a much higher tendency towards positive scores, particularly the languages with less tweets (Polish, Finnish, Swedish).
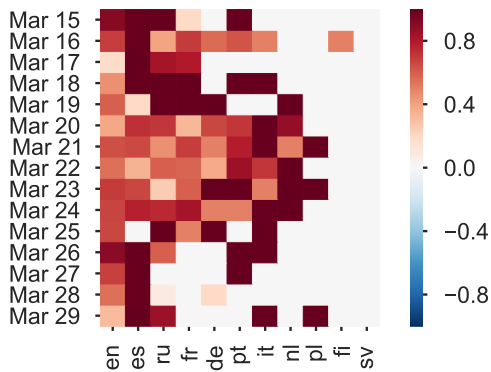

Figure 4: Heatmap of perspectives towards Obama over time in 11 different languages.

As a more detailed analysis, Figure 4 shows a heatmap of how the connotation expressed towards Obama shifts over time across different languages. Obama was not discussed much in Finnish or Swedish, whereas he was discussed everyday in English, Spanish, and Russian. In the middle of the two week period, the perspective towards Obama drops slightly, most notably in Spanish, which overlaps with his controversial trip to Cuba (March 20 – 22).

4.2 Sentiment Dynamics

In Table 2, we summarize the results of our experiments for predicting targetted sentiment dynamics. For each language, we report the average Kullback-Leibler divergence for the baselines and the LSTM model (higher scores are worse). We show prediction results in two settings: predicting

Data	Forecast + 1 day			Forecast + 4 days		
	Mean	SVM	LSTM	Mean	SVM	LSTM
EN_J	1.96	1.76	1.67	2.03	2.14	2.02
EN	4.88	1.94	1.79	5.12	3.91	3.91
RU	4.27	1.72	1.34	4.50	3.11	2.74
FR	2.23	1.80	1.76	4.33	3.24	3.12
ES	3.57	2.02	1.82	3.74	3.22	2.98
DE	4.43	2.24	1.77	4.73	3.61	3.55
NL	3.69	2.62	2.05	3.83	3.71	3.41
PT	3.84	1.97	2.04	3.94	3.03	3.11
IT	4.56	2.08	1.97	4.67	3.48	3.28
PL	4.01	1.67	1.46	4.06	3.37	3.32
SV	4.01	1.92	0.76	4.17	2.70	2.23
FI	4.90	2.04	1.84	4.99	4.14	4.20

Table 2: Average Kullback-Leibler divergence of output of the LSTM in predicting the distribution of writers' perspectives per entity for 11 different languages. In the first row, the perspectives from all countries are aggregated together.

the distribution one day ahead vs. four days ahead.

The LSTM outperforms the baselines in most languages with a few exceptions, such as Portuguese. All models perform worse at forecasting 4 days into the future than one day ahead, demonstrating how much connotation can vary over time as news events change, even in a small time period. On average, the LSTM achieves KL-divergence of 1.7 when predicting one day ahead and 3.26 when predicting 4 days ahead, lower than any of the baselines.

4.3 Error Analysis

For error analysis, we removed entities from Figure 3 from the training data and used them as a small test set for an LSTM trained on the remaining data in English with aggregation over all countries. In Figure 5, we have plotted the predicted marginal probabilities of four entities with the positive portion of the distribution (blue line) on the top half of the y-axis and the negative portion (red line) flipped onto the negative half of the axis.

The LSTM follows the general shape of the true curves, but frequently misses sudden spikes (e.g., the spike in negative polarity towards Russia on March 27th). In Table 3, we also report the KL divergences on predictions towards these entities. The model tends to perform less well at predicting sentiment towards entities where there were sudden spikes in sentiment based on news stories.

5 Related Work

There have been substantial studies for sentiment analysis on twitter (Agarwal et al., 2011; Kouloumpis et al., 2011; Pak and Paroubek, 2010;
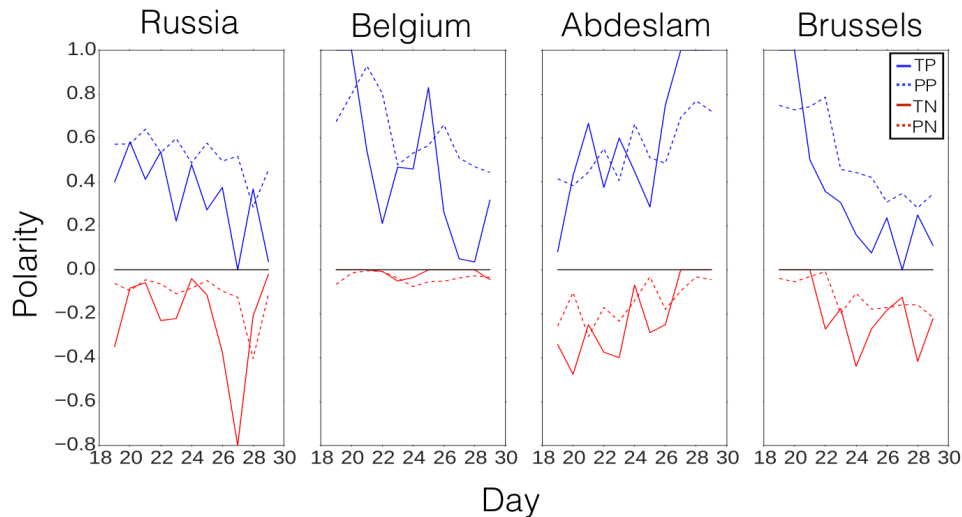
Figure 5: True versus predicted polarity distributions over time on three specific entities (TP: True Positive, PP: Predicted Positive, TN: True Negative, PN: Predicted Negative).

Entity	KL	Entity	KL	Entity	KL
Trump	0.12	England	0.14	EU	0.22
Obama	0.25	Turkey	0.30	Google	0.32
apple	0.37	Russia	0.76	Belgium	0.95
Clinton	1.13	Brussels	1.14	Abdeslam	1.79

Table 3: Error analysis on held-out entities.

Liu and Zhang, 2012), as well as targetted sentiment (Deng and Wiebe, 2015), implicit sentiment (Deng and Wiebe, 2014; Feng et al., 2013; Greene and Resnik, 2009) and specific aspects of subjective language (Mohammad and Turney, 2010; Choi and Wiebe, 2014) in other domains. Previous investigations include using targeted sentiment to predict international relations (Chambers et al., 2015), analyzing stylistic elements to predict tweet popularity (Tan et al., 2014), and exploring the re-phrasing of social media posts referencing specific news articles (Tan et al., 2016). Compared to most prior studies that focused on overt sentiment in English-only tweets, our work aims to study targeted implied sentiments across temporal, spatial, and linguistic borders.

Some work (Tsytsarau et al., 2014; O'Connor et al., 2010; De et al., 2016) has analyzed the transition of overt sentiment over a period of time and related the shifts in sentiment to news events. A body of work has also used predictive signals in Twitter to track and sense upcoming unrest and protests in specific countries (Ramakrishnan et al., 2014; Goode et al., 2015), and the future progression of flu activity based on multiple text sources (Santillana et al., 2015). In contrast, we focus on predicting the sentiment dynamics in social media based on previous trends.

6   Conclusions

When reporting news, people write with their own implicit and explicit biases and judgments. An author's choice of language reveals connotations towards entities, which can be captured within the connotation frames that we have extended to 10 European languages.

This work is one of the first to present a large scale analysis of multilingual connotation dynamics, and helps explore multiple perspectives on diverse issues across languages, time and countries – a critical piece in understanding journalistic portrayal and biases.

7   Acknowledgments

References

Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. Sentiment analysis of twitter data. In *Proceedings of the workshop on languages in social media*. ACL, pages 30–38.

Daniel Andor, Chris Alberti, David Weiss, Aliaksei Severyn, Alessandro Presta, Kuzman Ganchev, Slav Petrov, and Michael Collins. 2016. Globally normalized transition-based neural networks. In *Proceedings of ACL*.

Farzindar Atefeh and Wael Khreich. 2015. A survey of techniques for event detection in twitter. *Computational Intelligence* 31(1):132–164.

Nathanael Chambers, Victor Bowen, Ethan Genco, Xisen Tian, Eric Young, Ganesh Harihara, and Eugene Yang. 2015. Identifying political sentiment between nation states with social media. In *Proceedings of EMNLP*. pages 65–75.

Yoonjung Choi and Janyce Wiebe. 2014. +/-effectwordnet: Sense-level lexicon acquisition for opinion inference. In *EMNLP*. pages 1181–1191.

Abir De, Isabel Valera, Niloy Ganguly, Sourangshu Bhattacharya, and Manuel Gomez Rodriguez. 2016. Learning and forecasting opinion dynamics in social networks. In *Proceedings of NIPS*.

Lingjia Deng and Janyce Wiebe. 2014. Sentiment propagation via implicature constraints. In *EACL*. pages 377–385.

Lingjia Deng and Janyce Wiebe. 2015. Mpqa 3.0: An entity/event-level sentiment corpus. In *HLT-NAACL*. pages 1323–1328.

Andreas Eisele and Yu Chen. 2010. Multiun: A multilingual corpus from united nation documents. In *Proceedings of LREC*. pages 2868–2872.

Song Feng, Jun Seok Kang, Polina Kuznetsova, and Yejin Choi. 2013. Connotation lexicon: A dash of sentiment beneath the surface meaning. In *ACL*. pages 1774–1784.

Brian J Goode, Siddharth Krishnan, Michael Roan, and Naren Ramakrishnan. 2015. Pricing a protest: Forecasting the dynamics of civil unrest activity in social media. *PloS one* 10(10):e0139911.

Stephan Greene and Philip Resnik. 2009. More than words: Syntactic packaging and implicit sentiment. In *Proceedings of NAACL-HLT*. ACL, pages 503–511.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9(8):1735–1780.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT summit*. volume 5, pages 79–86.

Efthymios Kouloumpis, Theresa Wilson, and Johanna D Moore. 2011. Twitter sentiment analysis: The good the bad and the omg! In *Proceedings of ICWSM*. AAAI.

Bing Liu and Lei Zhang. 2012. A survey of opinion mining and sentiment analysis. In *Mining text data*, Springer, pages 415–463.

Saif M Mohammad and Peter D Turney. 2010. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*. Association for Computational Linguistics, pages 26–34.

Brendan O'Connor, Ramnath Balasubramanyan, Bryan Routledge, and Noah Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of ICWSM*. 122-129, pages 1–2.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of LREC*. volume 10, pages 1320–1326.

Kira Radinsky and Eric Horvitz. 2013. Mining the web to predict future events. In *Proceedings of WSDM*. ACM, pages 255–264.

Naren Ramakrishnan, Patrick Butler, Sathappan Muthiah, Nathan Self, Rupinder Khandpur, Parang Saraf, Wei Wang, Jose Cadena, Anil Vullikanti, Gizem Korkmaz, et al. 2014. 'beating the news' with embers: forecasting civil unrest using open source indicators. In *Proceedings of KDD*. ACM, pages 1799–1808.

Hannah Rashkin, Sameer Singh, and Yejin Choi. 2016. Connotation frames: A data-driven investigation. In *Proceedings of ACL*.

Mauricio Santillana, André T Nguyen, Mark Dredze, Michael J Paul, Elaine O Nsoesie, and John S Brownstein. 2015. Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS Comput Biol* 11(10):e1004513.

Chenhao Tan, Adrien Friggeri, Menlo Park, and Menlo Park. 2016. Lost in Propagation? Unfolding News Cycles from the Source. In *Proceedings of AAAI*.

Chenhao Tan, Lillian Lee, and Bo Pang. 2014. The effect of wording on message propagation: Topic- and author-controlled. In *Proceedings of ACL*.

Jörg Tiedemann. 2012. Parallel data, tools and interfaces in opus. In *LREC*. pages 2214–2218.

Mikalai Tsytsarau, Themis Palpanas, and Malu Castellanos. 2014. Dynamics of news events and social media reaction. In *Proceedings of KDD*. ACM, pages 901–910.