Cross-lingual Flames Detection in News Discussions

Josef Steinberger[1,2] , **Tomáš Brychcín**[1,2] , **Tomáš Hercig**[1,2] , and **Peter Krejzl**[2]

[1]NTIS – New Technologies for the Information Society,
Faculty of Applied Sciences, University of West Bohemia, Czech Republic
[2]Department of Computer Science and Engineering,
Faculty of Applied Sciences, University of West Bohemia, Czech Republic
{jstein,brychcin,tigi,krejzl}@kiv.zcu.cz
http://mediagist.kiv.zcu.cz/flames

Abstract

We introduce Flames Detector, an online system for measuring flames, i.e. strong negative feelings or emotions, insults or other verbal offences, in news commentaries across five languages. It is designed to assist journalists, public institutions or discussion moderators to detect news topics which evoke flames. We propose a machine learning approach to flames detection and calculate an aggregated score for a set of comment threads. The demo application shows the most flaming topics of the current period in several language variants. The search functionality gives a possibility to measure flames in any topic specified by a query. The evaluation shows that the flame detection in discussions is a difficult task, however, the application can already reveal interesting information about the actual news discussions.

1 Introduction

News portals are highly busy online places where people express their opinions. Journalists write about controversial topics because these attract the readers. The large number of commentaries is a sign that the topic was read by a lot of readers and it could be viewed as a sign of success of the article. Besides journalists who need to identify such topics, political institutions have to know the current trends of the society to react accordingly. International institutions (e.g. European Commission) find useful cross-lingually organized news and commentaries, as they can quickly find and understand different views on controversial topics in different countries.

There are many news aggregators and analysers. Google News[1] aggregates headlines and displays the stories according to each reader's interests. IBM Watson News Explorer[2] gives a more analytical way to read news through linked data visualizations. Europe Media Monitor (EMM)[3] produces a summary of news stories clustered near real-time in various languages and compares how the same events have been reported in the media written in different languages. MediaGist[4] exploits together with news articles another source of information: the commentaries. Including comments opens many above-mentioned use cases.

Natural language processing (NLP) technology can help to make sense out of this data, in particular, the field of argumentation mining (Habernal and Gurevych, 2017). Sentiment analysis, in its basic definition, reveals the polarity of the texts (positive, negative, neutral). Habernal et al. (2014) state that posts in social media often contain sarcasm and irony, they later experiment with sarcasm detection in (Ptáček et al., 2014).

Stance detection (Mohammad et al., 2016) is a related field which tries to infer whether the author of the comment is in favor or against a predefined topic. The migration crisis in Europe raised attention to the *hate speech*, which is defined as "abusive speech targeting specific group characteristics, such as ethnic origin, religion, gender, or sexual orientation" (Djuric et al., 2015). Detecting offensive language was studied by Razavi et al. (2010) or Chen et al. (2012), who analyzed the use case of protecting adolescent online safety by detecting offensive language in social media.

Here, we study *flames*, which were defined as

[1] https://news.google.com/

[2] http://news-explorer.mybluemix.net/

[3] Europe Media Monitor is developed at Joint Research Centre, European Commission: http://emm.newsbrief.eu (Atkinson and van der Goot, 2009).

[4] http://mediagist.eu (Steinberger, 2016).

Proceedings of Recent Advances in Natural Language Processing, pages 694–700,
Varna, Bulgaria, Sep 4–6 2017.
https://doi.org/10.26615/978-954-452-049-6_089

non constructive, aggressive posts, which do not contribute to the discussion, where users attack each other at a personal level instead of contrasting the discussion partner for his/her approach, ideas, contribution or argumentation (Pazienza and Tudorache, 2011). We consider *flaming* those comments which carry strong negative feelings or emotions, insults or other verbal offences regardless of agreement or disagreement with the given topic. Flames detection can help discussion moderators to manage and filter the discussion.

We propose a machine learning approach to detect flames in news discussions, and an aggregation metric. Another contribution is the corpus we developed. It contains flames annotations in news discussions for 5 languages (English, Czech, German, French and Italian). The developed application, running at `http://mediagist.kiv.zcu.cz/flames`, is unique as near real-time flames detection in online discussions has not been realized yet. It displays the most flaming topic of the current week and the search functionality provides a way to measure flaming of the topic defined by the query. Because the topics are linked across languages, the topic can be easily studied and compared across the languages. The trend figures provide a historical view on the flaming of the query topic.

In the next Section we introduce the functionality of the web application for detecting flames in news discussions. Section 3 reveals how a flaming dialogue interaction can be recognized. It is followed by a definition of the aggregation measure which calculates the intensity of a set of comment threads (Section 4). In Section 5, we describe the corpus we developed in order to train and test the flame classifier. In the rest, we discuss the performance and future extensions.

2 Proposed System

The proposed system uses data feed from MediaGist (Steinberger, 2016) and brings flame detection. MediaGist processing starts with a crawler. It gathers articles and their comments from predefined news sites[5]. It creates an RSS file for each article, which goes through the NLP pipeline. The pipeline first recognizes entities, in both the article and its comments, and assigns a cross-lingual id to each mention. The named entity recognizer is based on JRC-Names[6], which is a highly multilingual named entity resource for person and organization names (Steinberger et al., 2011c). The coreference resolver (Steinberger et al., 2011a) then enriches the list of entity mentions by name part references and definite descriptions. Sentiment analyzer (Steinberger et al., 2011b) assigns to each article, comment and entity mention a sentiment score. Here comes the place to assign a flame label (FLAME/NO-FLAME) to each interaction (i.e. to every response). Except the posts which start a new thread, each comment receives a flame annotation.

Article comments are then summarized according to (Kabadjov et al., 2013). These fully annotated article RSS files enter the clustering phase. Every four hours, for each language, the clustering takes the articles published during the current week and creates monolingual clusters. After this step, RSS files contain information about all articles in the cluster. The cross-lingual linker then connects the most similar clusters across languages. Cross-lingual linking uses two kinds of features: entities and descriptors from EuroVoc[7]. Using Eurovoc features ensures that the linked clusters share the same topic. If at the same time the clusters share the same entities[8], it is very likely that the clusters are about the same story (Steinberger, 2013). The last step is creating a summary of clustered articles and a summary of cluster's comments. The RSS now contains all information needed by the presentation layer, the MediaGist website.

The clustered data are then indexed in ElasticSearch[9]. During displaying the front page, Elasticsearch runs a query which returns the cluster, which has the largest sum of flame scores (defined in Section 4) across languages. The cross-lingual links are used to calculate flame scores of the other language variants of the cluster. As the sum is used, highly cross-lingual topics have more chances to be selected.

The core of the flames detector is the search functionality. The language of the query is currently set to English. After submitting a query ElasticSearch finds relevant English news clusters

[5] Currently, it gathers data from 8 sources in 5 languages: English (theguardian.com), Czech (idnes.cz, ihned.cz, novinky.cz), Italian (corriere.it, repubblica.it), French (lemonde.fr) and German (spiegel.de).

[6] `https://ec.europa.eu/jrc/en/language-technologies/jrc-names`

[7] `http://eurovoc.europa.eu`

[8] The entity ids are unified across languages.

[9] `https://www.elastic.co/`.

in the last 10 weeks. The fields used for searching are: article titles, descriptions (initial sentence/paragraph), the summary of articles and the summary of comments. The cross-lingual links are then used to receive the relevant clusters in other languages. It can thus find, e.g., French topics relevant to the English query. For every language (a set of clusters), it measures the flame score and its confidence interval. It uses all comment threads in all retrieved clusters for the calculation. It also detects the most flaming cluster, the most flaming article and also the post which triggered the most flaming discussion. In the case of the post, the ratio of flame replies and all replies is taken as a score. For all the flame extremes (cluster, article, post) it requires a minimum number of replies (set to 10) unless there is no post having such an amount of replies. A search result can be seen at Figure 1. Each row shows information about a language. On the left, statistic about the whole retrieved set of clusters, the most flaming cluster, article and flame trigger post can be found. On the right, a historical trend is drawn. The trend is based on weekly flame scores because week is the clustering unit.

The extracted clusters are linked with MediaGist, which can be further used to analyze the results. The articles are linked to the source site from both ends (Flames detector and MediaGist).

3 Flames Detector

Distinguishing between flaming/not-flaming utterances is in fact a binary classification task. We employ Maximum Entropy (ME) classifier (Berger et al., 1996) implemented in the Brainy machine learning library (Konkol, 2014). The following feature functions were used:

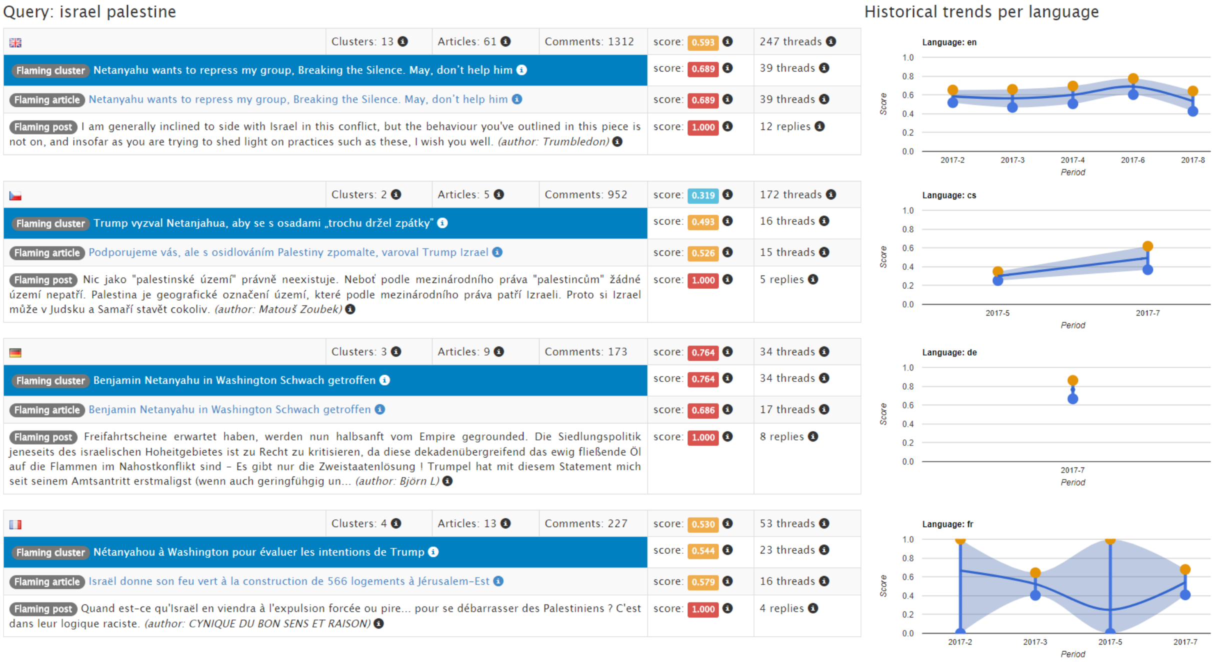
- **GloVe**: We express the meaning of an utterance as the real-valued vector. Each value in the vector is then used as a separate feature. We use semantic composition approach. It is based on *Frege's principle of compositionality* (Pelletier, 1994), which states that the meaning of a complex expression is determined as a composition of its parts, i.e. words. We use linear combination of word vectors, where the weights are represented by the inverse-document-frequency (IDF) values of words. We use Global Vectors (GloVe) (Pennington et al., 2014) for word vector representation. We trained the word vectors on data from MediaGist gathered during the last 17 month (approximately 1M comments for each language). Brychcín and Svoboda (2016) showed that this approach leads to very good sentence representation.
- **BoW**: We use bag-of-words (BoW) representation of an utterance, i.e. separate binary feature representing the occurrence of a word in the utterance.
- **DisSize**: The number of utterances in the discourse. The larger discourses are assumed to be about the more controversial topic or at least the topic which worth discussing.
- **Level**: The number expressing how many preceding utterances occur in the sequence of replies. The larger number means the people are paying attention to the discussed topic.
- **CharNgrams**: Separate binary feature for each character n-gram in the utterance text. We do it separately for different orders $n \in \{1, 2, 3\}$.
- **WordShape**: We tried to improve pattern features by using word-shape classes for words. We assign words into one of 24 classes[10] similar to the function specified in (Bikel et al., 1997). Each class is separate binary feature.

GloVe and *BoW* features use preprocessed text. We lowercase the text, remove stop-words, and apply stemming for remaining words. We use HPS unsupervised stemmer (Brychcín and Konopík, 2015). We trained HPS on the same data as GloVe word vectors. The surface features (*CharNgrams* and *WordShape*) work with the text in the form it was originally written. When the current comment starts with a repetition of a part of the previous comment, we remove the repetition.

4 Flaming Intensity Metric

A not nested post (i.e. a post on the top level of the discussion) with its replies (and replies on replies, etc.), which evolve the discussion, make a tree of utterances. The tree is in the following text denoted simply as discourse. Assume we have a set of n discourses and we want to determine the

[10] We use edu.stanford.nlp.process.WordShapeClassifier with the WORDSHAPECHRIS1 setting available in Standford CoreNLP library (Manning et al., 2014).



Query: israel palestine

	Clusters: 13	Articles: 61	Comments: 1312	score: 0.593	247 threads
Flaming cluster Netanyahu wants to repress my group, Breaking the Silence. May, don't help him				score: 0.689	39 threads
Flaming article Netanyahu wants to repress my group, Breaking the Silence. May, don't help him				score: 0.689	39 threads
Flaming post I am generally inclined to side with Israel in this conflict, but the behaviour you've outlined in this piece is not on, and insofar as you are trying to shed light on practices such as these, I wish you well. *(author: Trumbledon)*				score: 1.000	12 replies

	Clusters: 2	Articles: 5	Comments: 952	score: 0.319	172 threads
Flaming cluster Trump vyzval Netanjahua, aby se s osadami „trochu držel zpátky"				score: 0.493	16 threads
Flaming article Podporujeme vás, ale s osidlováním Palestiny zpomalte, varoval Trump Izrael				score: 0.526	15 threads
Flaming post Nic jako "palestinské území" právně neexistuje. Neboť podle mezinárodního práva "palestincům" žádné území nepatří. Palestina je geografické označení území, které podle mezinárodního práva patří Izraeli. Proto si Izrael může v Judsku a Samaří stavět cokoliv. *(author: Matouš Zoubek)*				score: 1.000	5 replies

	Clusters: 3	Articles: 9	Comments: 173	score: 0.764	34 threads
Flaming cluster Benjamin Netanyahu in Washington Schwach getroffen				score: 0.764	34 threads
Flaming article Benjamin Netanyahu in Washington Schwach getroffen				score: 0.686	17 threads
Flaming post Freifahrtscheine erwartet haben, werden nun halbsanft vom Empire gegrounded. Die Siedlungspolitik jeneseits des israelischen Hoheitgebietes ist zu Recht zu kritisieren, da diese dekadenübergreifend das ewig fließende Öl auf die Flammen im Nahostkonflikt sind – Es gibt nur die Zweistaatenlösung ! Trumpel hat mit diesem Statement mich seit seinem Amtsantritt erstmaligst (wenn auch geringfühgig un... *(author: Björn L)*				score: 1.000	8 replies

	Clusters: 4	Articles: 13	Comments: 227	score: 0.530	53 threads
Flaming cluster Nétanyahou à Washington pour évaluer les intentions de Trump				score: 0.544	23 threads
Flaming article Israël donne son feu vert à la construction de 566 logements à Jérusalem-Est				score: 0.579	16 threads
Flaming post Quand est-ce qu'Israël en viendra à l'expulsion forcée ou pire... pour se débarrasser des Palestiniens ? C'est dans leur logique raciste. *(author: CYNIQUE DU BON SENS ET RAISON)*				score: 1.000	4 replies

Historical trends per language

Figure 1: The search results for query "israel, palestine".

flaming intensity. For i-th discourse we firstly calculate the ratio of flaming utterances and the discourse size. We denote this ratio as w_i. The flaming intensity is then simply a mean, $\bar{w}$. In addition, we calculate t-statistics confidence interval for the mean $\bar{w}$. In our system we use 90% confidence level. Flaming intensities across less and highly discussed topics can be then better compared.

5 Proposed Dataset

We created a corpus for five languages (Czech, English, French, Italian, and German), in which we annotated flames. The data comes from noisy user-generated news article discussions. The statistics of the corpus are shown in Table 1.

Language	Size	Flame ratio
CS	1812	702 (38.7%)
DE	1122	149 (13.3%)
EN	1007	250 (24.8%)
FR	487	144 (29.6%)
IT	649	249 (38.4%)

Table 1: Data statistics.

For every language, we used the same annotation scheme and we included topics from multiple domains. The annotators were given instructions to label a comment as flame when it contained strong negative feelings, negative emotions, insults or other verbal offenses regardless of agreement or disagreement with the given topic. See the examples of flame comments bellow:

- `You mean like Bullsh!t?`
- `Rubbish.`
- `Thats a lovely collection of random thought bubbles.`
- `you really think you speak for everyone eh? Says a lot about you.`
- `When do you leave school, Tim?`
- `Cool story, bro. But you still seem to be suffering from the view that walking and chewing gum is impossible. Try it.`

Czech and English data were annotated by three experts. The annotators each labeled approximately a third of the data and a small sample was annotated by all to assess their agreement. The majority voting scheme was applied on the gold label selection. Table 2 shows the inter-annotator agreement metrics for the Czech corpus (100 comments) and the English corpus (76 comments). Only one annotator labeled French, Italian and German data.

Metric	EN	CS
AVG Accuracy	79.83%	86.67%
Fleiss' κ	0.590	0.578
AVG Cohen's κ	0.591	0.580
Krippendorff's α	0.592	0.579

Table 2: Inter-annotator agreement.

The corpus will be available for research purposes at `http://nlp.kiv.zcu.cz/projects/flame`.

6 Results and Discussion

We used the same settings and features for experiments in all languages. The proposed datasets for different languages have different ratios of flaming/not flaming posts. Due to the small size, the datasets are not assumed to cover all discussion topics. We apply Maximum Entropy principle to achieve the best estimate for previously unknown topic, i.e. we balance the training data to contain the same number of flaming/not flaming posts.

Model	CS	DE	EN	FR	IT
ME BoW	53.5%	62.4%	53.2%	51.7%	55.6%
ME GloVe	56.4%	64.4%	68.0%	54.2%	54.0%
ME all	62.8%	67.8%	72.2%	60.1%	60.2%

Table 3: Accuracy of flame detection across all five languages.

The results reported in Table 3 are achieved by 20-fold cross-validation on balanced datasets. The difficulty of flames detection is similar to polarity or stance detection (accuracy 60%–70% in the case of commentary data). The best accuracy was for English (72.2%), the worst for French (60.1%). The size of the training data (Table 1) clearly affects the classification accuracy. Note the best contributing feature was *GloVe*. Also *DisSize* and *Level* were proved to be very useful.

To illustrate the search functionality we show results of query "climate change" in Table 4. We can observe that the supplied query actually defines the topic. As the cross-lingual linking is based on topical words and entities it gives the answer wider than just climate change. The top entities mentioned in the clusters relevant to climate change were: *Donald Trump*, *Barrack Obama*, *United Nations*, and *European Commission*. These play an important role in cross-lingual linking and thus the topics related to them are retrieved for the other languages.

7 Conclusion and Future Work

This is the first study for detecting flames in news commentaries across languages. The flame classification performance very much depends on the size of training data. An extension of the corpus will directly lead to an improved performance.

The application measures the flames in discussions in near real-time and extracts the flaming topics which can be further analyzed in the news aggregator (MediaGist) or at the source news site itself. Future plans include increasing the data volume on both vertical (sources) and horizontal (historical data) axes. This will allow us to study the evolution of flames on a larger scale.

The system currently consumes raw commentaries. Measuring the flames among real Internet users will require to fight trolls and filter the conversations (Mihaylov et al., 2015). On the other hand, without fighting the trolls, the system can identify those discussions in which trolls are active because such comments many times raise negative emotions and thus help to identify the trolls.

Flame ratio depends very much on the source. We will search for a way how to remove the source bias. This will make the flaming more comparable across languages.

Acknowledgements

This publication was supported by the project LO1506 of the Czech Ministry of Education, Youth and Sports under the program NPU I., by Grant No. SGS-2016-018 Data and Software Engineering for Advanced Applications and by project MediaGist, EUs FP7 People Programme (Marie Curie Actions), no. 630786.

References

Martin Atkinson and Erik van der Goot. 2009. Near real time information mining in multilingual news. In *Proceedings of the 18th International World Wide Web Conference (WWW 2009)*. Madrid, Spain, pages 1153–1154.

Adam L. Berger, Vincent J. D. Pietra, and Stephen A. D. Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics* 22:39–71.

Daniel M. Bikel, Scott Miller, Richard Schwartz, and Ralph Weischedel. 1997. Nymble: a high-performance learning name-finder. In *Proceedings of the fifth conference on Applied natural language processing*. Association for Computational Linguistics, pages 194–201.

Tomáš Brychcín and Miloslav Konopík. 2015. Hps: High precision stemmer. *Information Processing & Management* 51(1):68–91.

Tomáš Brychcín and Lukáš Svoboda. 2016. UWB at SemEval-2016 Task 1: Semantic Textual Similarity using Lexical, Syntactic, and Semantic Information. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, pages 588–594.

Language	Clusters	Articles	Comments	Flame score	Threads	Flame cluster
CS	7	30	4556	0.307	691	Lidé mohou domů. Přehrada Oroville je podle šerifa na bouři připravena
DE	7	23	265	0.552	153	Rede zur Amtseinführung Das hat Donald Trump versprochen
EN	110	388	186,640	0.461	30,301	Trump can save his presidency with a great deal to save the climate
FR	8	23	237	0.492	44	Téhéran teste la détermination de Washington
IT	9	34	947	0.479	39	Brexit, governo sconfitto a Camera Lord su emendamento che protegge diritti cittadini Ue

Table 4: An example of queries and their results.

Ying Chen, Yilu Zhou, Sencun Zhu, and Heng Xu. 2012. Detecting offensive language in social media to protect adolescent online safety. In *Proceedings of the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust*. IEEE Computer Society, Washington, DC, USA, SOCIALCOM-PASSAT '12, pages 71–80.

Nemanja Djuric, Jing Zhou, Robin Morris, Mihajlo Grbovic, Vladan Radosavljevic, and Narayan Bhamidipati. 2015. Hate speech detection with comment embeddings. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, New York, NY, USA, WWW '15 Companion, pages 29–30.

Ivan Habernal and Iryna Gurevych. 2017. Argumentation mining in user-generated web discourse. *Computational Linguistics* 43(1):(in press).

Ivan Habernal, Tomáš Ptáček, and Josef Steinberger. 2014. Supervised sentiment analysis in czech social media. *Information Processing & Management* 50(5):693–707.

Mijail Kabadjov, Josef Steinberger, and Ralf Steinberger. 2013. Multilingual statistical news summarization. In *Multilingual Information Extraction and Summarization*, Springer, volume 2013 of *Theory and Applications of Natural Language Processing*, pages 229–252.

Michal Konkol. 2014. Brainy: A machine learning library. In Leszek Rutkowski, Marcin Korytkowski, Rafa Scherer, Ryszard Tadeusiewicz, Lotfi A. Zadeh, and Jacek M. Zurada, editors, *Artificial Intelligence and Soft Computing*, Springer-Verlag, Berlin, volume 8468 of *Lecture Notes in Computer Science*.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*. pages 55–60.

Todor Mihaylov, Georgi Georgiev, and Preslav Nakov. 2015. Finding opinion manipulation trolls in news community forums. In *Proceedings of the 19th CoNLL*. ACL, pages 310–314.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of SemEval-2016*. Association for Computational Linguistics, San Diego, California, pages 31–41.

Maria Teresa Pazienza and Alexandra Gabriela Tudorache. 2011. Interdisciplinary contributions to flame modeling. In *Congress of the Italian Association for Artificial Intelligence*. Springer, pages 213–224.

Francis Jeffry Pelletier. 1994. The principle of semantic compositionality. *Topoi* 13(1):11–24.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, pages 1532–1543.

Tomáš Ptáček, Ivan Habernal, and Jun Hong. 2014. Sarcasm detection on czech and english twitter. In *Proceedings of COLING 2014*. Dublin City University and Association for Computational Linguistics, Dublin, Ireland, pages 213–223.

Amir H. Razavi, Diana Inkpen, Sasha Uritsky, and Stan Matwin. 2010. Offensive language detection using multi-level classification. In Atefeh Farzindar and Vlado Kešelj, editors, *Advances in Artificial Intelligence: 23rd Canadian Conference on Artificial Intelligence*. Springer Berlin Heidelberg, Berlin, Heidelberg, pages 16–27.

Josef Steinberger. 2016. Mediagist: A cross-lingual analyser of aggregated news and commentaries. In *Proceedings of ACL-2016 System Demonstrations*. Association for Computational Linguistics, Berlin, Germany, pages 145–150.

Josef Steinberger, Jenya Belyaeva, Jonathan Crawley, Leonida Della-Rocca, Mohamed Ebrahim, Maut Ehrmann, Mijail Kabadjov, Ralf Steinberger, and Erik Van der Goot. 2011a. Highly multilingual coreference resolution exploiting a mature entity repository. In *Proceedings of the 8th RANLP Conference*. Incoma Ltd., pages 254–260.

Josef Steinberger, Polina Lenkova, Mijail Kabadjov, Ralf Steinberger, and Erik Van der Goot.

2011b. Multilingual entity-centered sentiment analysis evaluated by parallel corpora. In *Proceedings of the 8th RANLP Conference*. pages 770–775.

Ralf Steinberger. 2013. Multilingual and cross-lingual news analysis in the europe media monitor (emm). In *Multidisciplinary Information Retrieval*, Springer, volume 8201 of *LNCS*, pages 1–4.

Ralf Steinberger, Bruno Pouliquen, Mijail Kabadjov, Jenya Belyaeva, and Erik Van der Goot. 2011c. Jrc-names: A freely available, highly multilingual named entity resource. In *Proceedings of the International RANLP Conference*. Incoma Ltd.