NLP und Journalismus
Journalismus und NLP

Michael Strube

`michael.strube ät h-its.org`

February 1, 2018

Inhalt

Computerlinguisten arbeiten mit Texten. Journalisten arbeiten mit Texten. Zwischen den beiden Gebieten besteht eine natürliche Nähe, die erst seit kurzem thematisiert wird. Im Seminar werden wir ausloten, welche Anwendungsmöglichkeiten es für NLP-Technologie im Bereich Journalismus gibt. Das betrifft sowohl 'alte' Themen wie die Unterstützung von Journalisten bei der Recherche durch Textmining und Textanalyse als auch hochaktuelle wie das Verifizieren der Authentizität von Meldungen (oder das Erkennen von 'Fake News'), das Moderieren von Kommentaren, die Erkennung von 'Hate Speech', usw.

Journalismus hat aber noch eine zweite Bedeutung für uns. Computerlinguisten sehen sich ebenso wie andere Wissenschaftler zunehmend der Erwartung ausgesetzt, ihre Arbeit persönlich und durch Medien vermittelt der Öffentlichkeit vorzustellen. Im Seminar werden wir deshalb statt oder in Ergänzung zu langweiligen Referaten verschiedene Formen der Kommunikation und der Wissenschaftskommunikation kennenlernen und einüben:

- Tages-/Wochenzeitung

- Wissenschaftsjournal

- Radio

- Fernsehen

- soziale Medien

- Podiumsdiskussion

- öffentliche Rede - . . .

Wie können wir sicherstellen, daß unsere Aussagen unverfälscht und korrekt beim Rezipienten ankommen? Wie können wir für die Öffentlichkeit verständlich sein, ohne trivial zu werden?

Warnung: Thema und Form sind experimentell! Und: Die Benutzung von Smartphones, Tablets, Notebooks und anderen elektronischen Geräten ist im Seminar nicht gestattet.

Literatur

- Proceedings of the 2016 IJCAI Workshop on Natural Language Processing meets Journalism, *New York, N.Y., 10 July 2016*

- Proceedings of the First Workshop on Abusive Language Online, *Vancouver, B.C., Canada, 4 August 2017*

- Proceedings of the Events and Stories in the News Workshop, *Vancouver, B.C., Canada, 4 August 2017*

- Proceedings of the Workshop on Natural Language Processing Meets Journalsim. *Copenhagen, Denmark, 9 September, 2017.*

Termine, Themenvorschläge

24.10.2017
Einführung, Motivation, . . .

31.10.2017
fällt aus: Feiertag

07.11.2017
Podiumsdiskussion zu *Filter Bubble in Recommender Systems:* Steffen Knapp, Stefanie Kühner, Sarina Meyer, ~~Till Nocher~~, Christoph Schaller, Annalena Sailer
Überblick über *Recommender Systems:* (Koren & Bell, 2011)
Filter Bubble – Realität: (Pariser, 2011; Bozdag, 2013; Bozdag & van den Hoven, 2015; Sunstein, 2017)
Filter Bubble – Fiktion/reparierbar: (Zhang et al., 2012)
Filter Bubble, Fairness in Recommender Systems (Zhang & Hurley, 2008; Vargas & Castells, 2011; Adomavicius & Kwon, 2012; Nguyen et al., 2014; Castells et al., 2015; Burke et al., 2017; Huang & Yao, 2017)

Filter Bubble in Search: (Santos et al., 2015) . . .
Filter Bubble in Social Networks: (Garimella et al., 2017b; 2017a) . . .

zur Vorbereitung: (Bozdag & van den Hoven, 2015)
Textbeitrag: Skript für einen 3-minütigen Beitrag zu einer Podiumsdiskussion zum Thema *Filter Bubble* in Recommender Systems. Dazu gehört eine Erklärung und eine Meinung. Fragen dazu bitte an mich.

14.11.2017
Diskussion zu *Filter Bubble*: Was kann NLP leisten? Michael Strube (45 min.)
Literatur für *Recommendation Systems* 07.11.2017; und als Anregung (Carbonell & Goldstein, 1998; Santos et al., 2015; Degaetano-Ortlieb & Teich, 2017; Nema et al., 2017; Zhao et al., 2017)

Vortrag zu TV Sprache: Melanie Schindler (15 min.)

Genre *News*; Journalistische Texte als Daten in der Computerlinguistik: Jason Brockmeyer, Utaemon Toyota (30 min.)
Lead Baseline in der automatischen Zusammenfassung, Genre Zeitungstexte
(Mihalcea & Ceylan, 2007; Yang et al., 2017)

allgemein über Genre: (Swales, 1990; Kessler et al., 1997; Wolters & Kirsten, 1999; Sarkar et al., 2005; Passonneau et al., 2014) Diskurskonnektive, Diskursparsing: (Webber, 2009)

zur Vorbereitung: (Passonneau et al., 2014) (Fragen im Hinblick auf Genre News)

Textbeitrag: *Filter Bubble* – Was kann NLP zur Auflösung der Filterblase leisten? Was können wir tun? (1/2 Seite mit Fokus auf Techniken)

21.11.2017
fällt aus: Michael Strube at HITS Gruppenleiter Retreat

28.11.2017
fällt aus: Michael Strube at IJCNLP 2017

05.12.2017
fällt aus: Michael Strube at HITS Scientific Advisory Board Meeting

12.12.2018
Streitgespräch? Podiumsdiskussion? zu Überschriften in der Presse: ~~Maja Hoffmann, Viktor Zimmermann,~~ **Lasse Becker-Czarnetzki,** ~~Daniel Wachter, Stoyan Dimitrov,~~ **Utaemon Toyota (55 min.)**
(Banko et al., 2000; Dorr et al., 2003; Alfonseca et al., 2013; Chesney et al., 2017; Gatti et al., 2016; Szymanski et al., 2016; Lemke et al., 2017; Piotrkowicz et al., 2017)

zur Linguistik der Überschrift in der Presse: (Lindemann, 1990; Schaffer, 1995; Bucaria, 2004; Ifantidou, 2008; Blom & Hansen, 2015)

***Text Analytics* für Journalisten: Inhalte, Optimierung, Visualisierung: Arthur Neidlein, Uli Steinbach, Jason Brockmeyer (40 min.)**

Inhalte: (Freitas & Ji, 2016)
Unterstützung: (Bougiatiotis et al., 2016; Gravier et al., 2016; Kiomourtzis et al., 2016; Cucchiarelli et al., 2017)
Visualisierung: (Tannier, 2016; Bois et al., 2017; Laban & Hearst, 2017; van Meersbergen et al., 2017)

zur Vorbereitung: (Hardt & Rambow, 2017) oder (Freitas & Ji, 2016) oder (Chesney et al., 2017) oder (Gatti et al., 2016)

Textbeitrag (ungefähr 1 Seite): Wählen Sie eines der Papiere unter *zur Linguistik der Überschrift in der Presse* aus, sammeln zehn Überschriften aus der Presse und untersuchen Sie diese mit Hilfe des gewählten Papiers.

19.12.2017
Fortsetzung Text Analytics (15 min.)

3 Minute Madness **zu Framing: Sarina Meyer, Annalena Sailer, Christoph Schaller, Steffen Knapp, Nadja Heinzen, Daaje Meiners, Melanie Schindler, Stefanie Kühner (75 min.)**
(Dardis et al., 2008; Baumgartner et al., 2009; Boydstun, 2013; Boydstun et al., 2014; Baumer et al., 2015; Card et al., 2015; Tsur et al., 2015; Card et al., 2016; Johnson et al., 2017)

zur Vorbereitung: (Card et al., 2015)

Textbeitrag: (ungefähr 1/2-1 Seite:) Wählen Sie ein Thema, das in der Presse diskutiert wird. Wählen Sie zwei kontroverse Meinungen und beschreiben den Rahmen (?)/die Perspektive (?) (*Frame*), den der Autor einnimmt.

09.01.2018
verschoben auf den 9.1.18: Qualität von Zeitungstexten: ~~Eva Ngatchen~~ **– Michael Strube (15 min.)**
(Hardt & Rambow, 2017; Tamori et al., 2017)

Authenzität von Meldungen, *Fake News***: Daniel Schubert, Nadja Heinzen,** ~~Daniel Wachter, Eva Ngatchen~~**, Utaemon Toyota**

aktuelle Arbeiten zu *Uncertainty, Rumors, Deception* in den Medien: (Adel & Schütze, 2017; Li et al., 2017; Ma et al., 2017; Pisarevskaya, 2017; Son et al., 2017; Thorne et al., 2017; Volkova et al., 2017) und konkret zu *Fake News*: (Rashkin et al., 2017; Wang, 2017; Long et al., 2017)

(aktuelle) Arbeiten zu *Stance Detection*: (Thomas et al., 2006; Somasundaran & Wiebe, 2009; Hasan & Ng, 2013; 2014; Klenner et al., 2017; Misra et al., 2017)

Fake News Challenge Stage 1: Stance Detection (`http://www.fakenewschallenge.org/`)

SemEval-2016 Task 6: Detecting Stance in Tweets und *SemEval-2017 Task 8: RumourEval: Determining Rumour Veracity and Support for Rumours*

Textbeitrag: Beschreiben Sie kurz das *UCL Machine Reading*-System (`https://github.com/uclmr/fakenewschallenge`), das bei der *Fake News Challenge* (`http://www.fakenewschallenge.org/`) Rang 3 belegt hat? Was ist cool an

dem System? (1/2-1 Seite)

zur Vorbereitung: (Rashkin et al., 2017) oder (Wang, 2017) oder (Long et al., 2017)

16.01.2018
Fortsetzung Authenzität von Meldungen, *Fake News***: Fabian Düker, Pauline Wyderka, Jason Brockmeyer**

Gastvortrag: Angelika Kirilin – Masterarbeit zu *Fake News Detection*

Diskussion: Was können wir tun?

zur Vorbereitung: (Brandtzaeg et al., 2015) oder `https://jospang.wordpress. com/2017/05/07/fake-news-wieso-weshalb-warum-und-was-ist-das-uberhaupt/`

Textbeitrag: Sehen Sie sich an `https://www.youtube.com/watch?v=HfdWpy6mqWA& list=PLGBWUED4WiIbIw93_CCJiCr4doYo17XsE&index=16` – Vergleichen Sie die Herangehensweise von Deutscher Welle und ARD. Was können wir tun, um den Sendern zu helfen?

23.01.2018
Fake News **und Gegensteuerung**
Uli Steinbach (15-20 min)

Moderieren von Kommentaren, *Hate Speech, Abusive Language*
Nadja Heinzen, ~~Zoë Rae Zinser,~~ **Daniel Schubert,** ~~Daniel Wachter, Stoyan Dimitrov,~~ **Lasse Becker-Czarnetzki, Steffen Knapp**

(Clarke & Grieve, 2017; Gambäck & Sikdar, 2017; Kolhatkar & Taboada, 2017a; 2017b; Pavlopoulos et al., 2017a; 2017b; Serrà et al., 2017; Waseem & Hovy, 2016; Waseem, 2016; Tulkens et al., 2016)

zur Vorbereitung: (Schmidt & Wiegand, 2017) oder (Waseem et al., 2017)

Textbeitrag: Brianna Wu gab einen eingeladenen Vortrag auf "Abusive Language Workshop" nach der ACL 2017. Der Vortrag ist leider nicht als Video verfügbar. Stattdessen sehen Sie sich an: `https://www.youtube.com/watch?v=dMCN8jOlPAo`. Geben Sie die Problematik wieder, die Brianna Wu beschreibt. Was läuft falsch in der (Online-)Gesellschaft? – Optional: Wenn Sie Mut haben und es vertragen können, sehen Sie sich die Kommentare zu dem Video an. Kommentieren Sie die Kommentare.

30.01.2018
Hate Speech Detection
Arthur Neidlein: System, Reddit-Korpus (15 min)

Podiumsdiskussion zu Moderieren von Kommentaren, *Hate Speech, Abusive Language***:** ~~Stoyan Dimitrov,~~ **Christoph Schaller, Annalena Sailer, Stefanie Kühner,**

Fabian Düker, Pauline Wyderka

Textbeitrag: Sehen Sie sich an: `https://www.youtube.com/watch?v=purzZqzVPO4&list=PLGBWUED4WiIbIw93_CCJiCr4doYo17XsE&index=13`. Weshalb macht die Politik hier eine so schlechte Figur? – Ist das neue Gesetz die Lösung des Problems? – Was können wir tun?

06.02.2018
Diskussion: Daniel Schubert, Arthur Neidlein ...

zur Vorbereitung: Offene Fragen, Kommentare, Kritik, Diskussionsvorschläge zum gesamten Semester – auch: was haben wir vergessen? gibt es einen Punkt, über den wir noch sprechen müssen? (mir bitte zusenden bis spätestens 05.02.18, 13 Uhr!)

Bemerkungen:

Leistungsnachweise: Lektüre und aktive Teilnahme (1/3), mind. 3 mal Teilnahme an Sitzungsleitung, Podiumsdiskussion etc. (1/3), Textbeiträge per Email als PDF bis spätestens 13 Uhr am Montag vor der Sitzung an mich (1/3).
Regelmäßige Teilnahme (d.i. nicht mehr als einmal unentschuldigtes Fehlen) ist Voraussetzung für den Scheinerwerb. Zu jeder Sitzung müssen jeweils zwei Fragen (!) zu einem Papier abgegeben werden, das in der aktuellen Sitzung vorgestellt wird. Abgabe entweder per Email bis spätestens 13 Uhr am Montag vor der Sitzung (oder – im Notfall – schriftlich direkt vor der Sitzung). Dies geht in die Bewertung für aktive Teilnahme am Seminar ein.

Literatur: Viele Papiere können direkt aus der *ACL Anthology* kopiert werden (`http://acl.ldc.upenn.edu/`), insbesondere alle Papiere der *(E/NA)ACL-*, *Coling-* und *EMNLP-*Konferenzen, alle Workshops, die im Rahmen dieser Konferenzen veranstaltet wurden und die Zeitschrift *Computational Linguistics*. Papiere, die von der *AAAI* publiziert wurden (AAAI-Konferenz, AAAI-Workshops, AAAI-Symposia, etc.) sind in der *AAAI Digital Library* verfügbar (`http://www.aaai.org/Library`). – Die meisten weiteren Zeitschriften sind elektronisch verfügbar über die UB (`http://rzblx1.uni-regensburg.de/ezeit/search.phtml?bibid=UBHE`) – oder stehen dort im Regal.

Sprechstunde: Auf Vereinbarung (Email, Telefon) bei mir im Büro, ggf. auch im Anschluß an das Seminar.

References

Adel, Heike & Hinrich Schütze (2017). Exploring different dimensions of attention for uncertainty detection. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics,* Valencia, Spain, 3–7 April 2017, pp. 22–34.

Adomavicius, Gediminas & Young Ok Kwon (2012). Improving aggregate recommendation diversity using ranking-based techiques. *IEEE Transactions on Knowledge and Data Engineering*, 24(5):896–911.

Alfonseca, Enrique, Daniele Pighin & Guillermo Garrido (2013). HEADY: News headline abstraction through event pattern clustering. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* Sofia, Bulgaria, 4–9 August 2013, pp. 1243–1253.

Banko, Michele, Vibhu O. Mittal & Michael J. Witbrock (2000). Headline generation based on statistical translation. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics,* Hong Kong, China, 1–8 August 2000, pp. 318–325.

Baumer, Eric, Elisha Elovic, Ying Qin, Francesca Polletta & Geri Gay (2015). Testing and comparing computational approaches for identifying the language of framing in political news. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* Denver, Col., 31 May – 5 June 2015, pp. 1472–1482.

Baumgartner, Frank R., Suzanna Linn & Amber E. Boydstun (2009). The decline of the death penalty: How media framing changed capital punishment in america. In B.F. Schaffner & P. Sellers (Eds.), *Winning with Words: The Origins and Impacts of Framing*. New York, N.Y.: Routledge.

Blom, Jonas Nygaard & Kenneth Reineke Hansen (2015). Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics*, 76:87–100.

Bois, Rémi, Guillaume Gravier, Eric Jamet, Emmanuel Morin, Pascale Sébillot & Maxime Robert (2017). Language-based construction of explorable news graphs for journalists. In *Proceedings of the 2017 EMNLP Workshop on Natural Language Processing meets Journalism,* Copenhagen, Denmark, 7 September 2017, pp. 31–36.

Bougiatiotis, Konstantinos, Anastasia Krithara, Georgios Paliouras & George Giannakopoulos (2016). An exploratory analysis of news trends on Twitter. In *Proceedings of the 2016 IJCAI Workshop on Natural Language Processing meets Journalism,* New York, N.Y., 10 July 2016, pp. 80–85.

Boydstun, Amber E. (2013). *Making the News: Politics, the Media and Agenda Setting*. Chicago, Ill.: University of Chicago Press.

Boydstun, Amber E., Shaun Bevan & H.F. Thomas III (2014). The importance of attention diversity and how to measure it. *Policy Studies Journal*, 42(2):173–196.

Bozdag, Engin (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, 15(3):209–227.

Bozdag, Engin & Jeroen van den Hoven (2015). Breaking the filter bubble: Democracy and design. *Ethics and Information Technology*, 17(4):249–265.

Brandtzaeg, P.B., M. Lüders, J. Spangenberg, L. Rath-Wiggins & A. Folstad (2015). Emerging journalistic verification practices concerning social media. *Journalism Practice*, 10(3):323–342.

Bucaria, Chiara (2004). Lexical and syntactic ambiguity as a source of humor: The case of newspaper headlines. *Humor*, 17(3):279–309.

Burke, Robin, Nasim Sonboli, Masoud Mansoury & Aldo Ordoñez-Gauger (2017). Balanced neighborhoods for fairness-aware collaborative recommendation. In *RecSys Workshop about Fairness in Recommender Systems,* Como, Italy, August 2017.

Carbonell, Jaime G. & Jade Goldstein (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval,* Melbourne, Australia, 24–28 August 1998, pp. 335–336.

Card, Dallas, Amber E. Boydstun, Justin H. Gross, Philip Resnik & Noah A. Smith (2015). The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers),* Beijing, China, 26–31 July 2015, pp. 438–444.

Card, Dallas, Justin Gross, Amber Boydstun & Noah A. Smith (2016). Analyzing framing through the casts of characters in the news. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing,* Austin, Tex., 1–5 November 2016, pp. 1410–1420.

Castells, Pablo, Neil J. Hurley & Saul Vargas (2015). Novelty and diversity in recommender systems. In Ricci F., Rokach L. & Shapira B. (Eds.), *Recommender Systems Handbook*, pp. 881–918. Boston, Mass.: Springer.

Chesney, Sophie, Maria Liakata, Massimo Poesio & Matthew Purver (2017). Incongruent headlines: Yet another way to mislead your readers. In *Proceedings of the 2017 EMNLP Workshop on Natural Language Processing meets Journalism,* Copenhagen, Denmark, 7 September 2017, pp. 56–61.

Clarke, Isobelle & Dr. Jack Grieve (2017). Dimensions of abusive language on Twitter. In *Proceedings of the First Workshop on Abusive Language Online,* Vancouver, B.C., Canada, 4 August 2017, pp. 1–10.

Cucchiarelli, Alessandro, Christian Morbidoni, Giovanni Stilo & Paola Velardi (2017). What to write? A topic recommender for journalists. In *Proceedings of the 2017 EMNLP Workshop on Natural Language Processing meets Journalism,* Copenhagen, Denmark, 7 September 2017, pp. 19–24.

Dardis, Frank E., Frank R. Baumgartner, Amber E. Boydstun, Suzanna De Boef & Fuyuan Shen (2008). Media framing of capital punishment and its impact on individuals¡¯ cognitive responses. *Mass Communication And Society*, 11:115–140.

Degaetano-Ortlieb, Stefania & Elke Teich (2017). Modeling intra-textual variation with entropy and surprisal: Topical vs. stylistic patterns. In *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature,* Vancouver, B.C., Canada, 4 August 2017, pp. 68–77.

Dorr, Bonnie, David Zajic & Richard Schwartz (2003). Hedge trimmer: A parse-and-trim approach to headline generation. In *Proceedings of the Text Summarization Workshop at HLT-NAACL '03,* Edmonton, Alberta, Canada, 2003, pp. 1–8.

Freitas, Jesse & Heng Ji (2016). Identifying news from Tweets. In *Proceedings of the 1st Workshop on Natural Language Processing and Computational Social Science,* Austin, Texas, 5 November 2016, pp. 11–16.

Gambäck, Björn & Utpal Kumar Sikdar (2017). Using convolutional neural networks to classify hate-speech. In *Proceedings of the First Workshop on Abusive Language Online,* Vancouver, B.C., Canada, 4 August 2017, pp. 85–90.

Garimella, Kiran, Gianmarco De Francisci Morales, Aristides Gionis & Michael Mathioudakis (2017a). Factors in recommending contrarian content on social media. In *Proceedings of the 2017 ACM Web Science Conference,* Troy, N.Y, 25–28 June 2017, pp. 259–262.

Garimella, Kiran, Gianmarco De Francisci Morales, Aristides Gionis & Michael Mathioudakis (2017b). Reducing controversy by connecting opposing views. In *Proceedings of the 10th ACM International Conference on Web Search and Data Mining,* Cambridge, United Kingdom, 6–10 February 2017, pp. 81–90.

Gatti, Lorenzo, Gözde Özbal, Marco Guerini, Oliviero Stock & Carlo Strapparava (2016). Automatic creation of flexible catchy headlines. In *Proceedings of the 2016 IJCAI Workshop on Natural Language Processing meets Journalism,* New York, N.Y., 10 July 2016, pp. 25–29.

Gravier, Guillaume, Martin Ragot, Laurent Amsaleg, Rémi Bois, Grégoire Jadi, Eric Jamet, Laura Monceaux & Pascale Sébillot (2016). Shaping-up multimedia analytics: Needs and expectations of media professionals. In *MultiMedia Modeling - 22nd International Conference,* Miami, Florida, 4-6 January 2016, pp. 303–314.

Hardt, Daniel & Owen Rambow (2017). Predicting user views in online news. In *Proceedings of the 2017 EMNLP Workshop on Natural Language Processing meets Journalism,* Copenhagen,

Denmark, 7 September 2017, pp. 7–12.

Hasan, Kazi Saidul & Vincent Ng (2013). Stance classification of ideological debates: Data, models, features, and constraints. In *Proceedings of the 6th International Joint Conference on Natural Language Processing,* Nagoya, Japan, 14–18 October 2013, pp. 1348–1356.

Hasan, Kazi Saidul & Vincent Ng (2014). Why are you taking this stance? Identifying and classifying reasons in ideological debates. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing,* Doha, Qatar, 25–29 October 2014, pp. 751–762.

Huang, Bert & Siriu Yao (2017). New fairness metrics for recommendation that embrace differences. In *Proceedings of the Workshop on Fairness, Accountability and Transparency in Machine Learning,* Halifax, Nova Scotia, Canada, 14 August 2017.

Ifantidou, Elly (2008). Newspaper headlines and relevance: Ad hoc concepts in ad hoc contexts. *Journal of Pragmatics*, 41:699–720.

Johnson, Kristen, I-Ta Lee & Dan Goldwasser (2017). Ideological phrase indicators for classification of political discourse framing on Twitter. In *Proceedings of the 2nd Workshop on Natural Language Processing and Computational Social Science,* Vancouver, B.C., Canada, 3 August 2017, pp. 90–99.

Kessler, Brett, Geoffrey Nunberg & Hinrich Schütze (1997). Automatic detection of text genre. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and of the 8th Conference of the European Chapter of the Association for Computational Linguistics,* Madrid, Spain, 7–12 July 1997, pp. 32–38.

Kiomourtzis, George, George Giannakopoulos, Aris Kosmopoulos & Vangelis Karkaletsis (2016). A multi-lingually applicable journalist toolset for the big-data era. In *Proceedings of the 2016 IJCAI Workshop on Natural Language Processing meets Journalism,* New York, N.Y., 10 July 2016, pp. 41–46.

Klenner, Manfred, Don Tuggener & Simon Clematide (2017). Stance detection in Facebook posts of a German right-wing party. In *Proceedings of the 2nd Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics,* Valencia, Spain, 3 April 2017, pp. 31–40.

Kolhatkar, Varada & Maite Taboada (2017a). Constructive language in news comments. In *Proceedings of the First Workshop on Abusive Language Online,* Vancouver, B.C., Canada, 4 August 2017, pp. 11–17.

Kolhatkar, Varada & Maite Taboada (2017b). Using New York Times picks to identify constructive comments. In *Proceedings of the 2017 EMNLP Workshop on Natural Language Processing meets Journalism,* Copenhagen, Denmark, 7 September 2017, pp. 100–105.

Koren, Yehuda & Robert Bell (2011). Advances in collaborative filtering. In F. Ricci, L. Rokach, B. Shapira & P.B. Kantor (Eds.), *Recommender Systems Handbook*. New York, N.Y.: Springer.

Laban, Philippe & Marti Hearst (2017). newsLens: Building and visualizing long-ranging news stories. In *Proceedings of the Events and Stories in the News Workshop,* Vancouver, B.C., Canada, 4 August 2017, pp. 1–9.

Lemke, Robin, Eva Horch & Ingo Reich (2017). Optimal encoding! – Information theory constrains article omission in newspaper headlines. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Volume 2: Short Papers,* Valencia, Spain, 3–7 April 2017, pp. 131–135.

Li, Yingya, Jieke Zhang & Bei Yu (2017). An NLP analysis of exaggerated claims in science news. In *Proceedings of the 2017 EMNLP Workshop on Natural Language Processing meets Journalism,* Copenhagen, Denmark, 7 September 2017, pp. 106–111.

Lindemann, Bernhard (1990). Cheap thrills we live by: Some notes on the poetics of tabloid headlines. *Journal of Literary Semantics*, 19(1):46–59.

Long, Yunfei, Qin Lu, Rong Xiang, Minglei Li & Chu-Ren Huang (2017). Fake news detection through multi-perspective speaker profiles. In *Proceedings of the 8th International Joint Conference on Natural Language Processing,* Taipei, Taiwan, 27 November – 1 December 2017, pp. 252–256.

Ma, Jing, Wei Gao & Kam-Fai Wong (2017). Detect rumors in microblog posts using propaga-

tion structure via kernel learning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* Vancouver, B.C., Canada, 30 July –4 August 2017, pp. 708–717.

Mihalcea, Rada & Hakan Ceylan (2007). Explorations in automatic book summarization. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Language Learning,* Prague, Czech Republic, 28–30 June 2007, pp. 380–389.

Misra, Amita, Brian Ecker, Theodore Handleman, Nicolas Hahn & Marilyn A. Walker (2017). *A semi-supervised approach to detecting stance in Tweets.* arXiv preprint.

Nema, Preksha, Mitesh M. Khapra, Anirban Laha & Balaraman Ravindran (2017). Diversity driven attention model for query-based abstractive summarization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* Vancouver, B.C., Canada, 30 July –4 August 2017, pp. 1063–1072.

Nguyen, Tien T., Pik-Mai Hui, F. Maxwell Harper, Loren G. Terveen & Joseph A. Konstan (2014). Exploring the filter bubble: The effect of using recommender systems on content diversity. In *Proceedings of the 23rd World Wide Web Conference,* Seoul, Korea, 7–11 April, 2014, pp. 677–686.

Pariser, Eli (2011). *The Filter Bubble: What the Internet is Hiding from You.* New York, N.Y.: Penguin Press.

Passonneau, Rebecca J., Nancy Ide, Songqiao Su & Jesse Stuart (2014). Biber Redux: Reconsidering dimensions of variation in American English. In *Proceedings of the 25th International Conference on Computational Linguistics,* Dublin, Ireland, 23–29 August 2014, pp. 565–576.

Pavlopoulos, John, Prodromos Malakasiotis & Ion Androutsopoulos (2017a). Deep learning for user comment moderation. In *Proceedings of the First Workshop on Abusive Language Online,* Vancouver, B.C., Canada, 4 August 2017, pp. 25–35.

Pavlopoulos, John, Prodromos Malakasiotis & Ion Androutsopoulos (2017b). Deeper attention to abusive user content moderation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing,* Copenhagen, Denmark, 7–11 September 2017, pp. 1136–1146.

Piotrkowicz, Alicja, Vania Dimitrova & Katja Markert (2017). Automatic extraction of news values from headline text. In *Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics,* Valencia, Spain, 3–7 April 2017, pp. 64–74.

Pisarevskaya, Dina (2017). Deception detection in news reports in the Russian language: Lexics and discourse. In *Proceedings of the 2017 EMNLP Workshop on Natural Language Processing meets Journalism,* Copenhagen, Denmark, 7 September 2017, pp. 74–79.

Rashkin, Hannah, Eunsol Choi, Jin Yea Jang, Svitlana Volkova & Yejin Choi (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing,* Copenhagen, Denmark, 7–11 September 2017, pp. 2921–2927.

Santos, Lucio F. D., Willian D. Oliveira, Luiz O. Carvalho, Monica R. P. Ferreira, Agma J. M. Traina & Caetano Traina, Jr. (2015). Combine-and-conquer: Improving the diversity in similarity search through influence sampling. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing,* Salamanca, Spain, pp. 994–999.

Sarkar, Arvik, Anne De Roeck & Paul H. Garthwaite (2005). Term re-occurrence measures for analyzing style. In *Proceedings of the 28th Annual International ACM SIGIR Conference (SIGIR 2005) Workshop on Using Stylistic Analysis of Text for Information Access,* Salvador, Brazil, August 15-19, 2005, p. ??

Schaffer, Deborah (1995). Shocking secrets revealed! The linguistics of tabloid headlines. *ETC: A Review of General Semantics*, 52(1):27–46.

Schmidt, Anna & Michael Wiegand (2017). A survey on hate speech detection using natural language processing. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media,* Valencia, Apain, 3 April 2017, pp. 1–10.

Serrà, Joan, Ilias Leontiadis, Dimitris Spathis, Gianluca Stringhini, Jeremy Blackburn & Athena

Vakali (2017). Class-based prediction errors to detect hate speech with out-of-vocabulary words. In *Proceedings of the First Workshop on Abusive Language Online,* Vancouver, B.C., Canada, 4 August 2017, pp. 36–40.

Somasundaran, Swapna & Janyce Wiebe (2009). Recognizing stances in online debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing,* Singapore, 2–7 August 2009.

Son, Youngseo, Anneke Buffone, Joe Raso, Allegra Larche, Anthony Janocko, Kevin Zembroski, H. Andrew Schwartz & Lyle Ungar (2017). Recognizing counterfactual thinking in social media texts. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers),* Vancouver, B.C., Canada, 30 July –4 August 2017, pp. 654–658.

Sunstein, Cass R. (2017). *#Republic: Divided Democracy in the Age of Social Media.* Princeton, N.J.: Princeton University Press.

Swales, John M. (1990). *Genre Analysis: English in Academic and Research Settings.* Cambridge, U.K.: Cambridge University Press.

Szymanski, Terrence, Claudia Orellana-Rodriguez & Mark Keane (2016). Helping news editors write better headlines: A recommender to improve the keyword contents and shareability of news headlines. In *Proceedings of the 2016 IJCAI Workshop on Natural Language Processing meets Journalism,* New York, N.Y., 10 July 2016, pp. 30–34.

Tamori, Hideaki, Yuta Hitomi, Naoaki Okazaki & Kentaro Inui (2017). Analyzing the revision logs of a Japanese newspaper for article quality assessment. In *Proceedings of the 2017 EMNLP Workshop on Natural Language Processing meets Journalism,* Copenhagen, Denmark, 7 September 2017, pp. 46–50.

Tannier, Xavier (2016). NLP-driven data journalism: Time-aware mining and visualization of international alliances. In *Proceedings of the 2016 IJCAI Workshop on Natural Language Processing meets Journalism,* New York, N.Y., 10 July 2016, pp. 52–56.

Thomas, Matt, Bo Pang & Lillian Lee (2006). Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing,* Sydney, Australia, 22–23 July 2006, pp. 327–335.

Thorne, James, Mingjie Chen, Giorgos Myrianthous, Jiashu Pu, Xiaoxuan Wang & Andreas Vlachos (2017). Fake news stance detection using stacked ensemble of classifiers. In *Proceedings of the 2017 EMNLP Workshop on Natural Language Processing meets Journalism,* Copenhagen, Denmark, 7 September 2017, pp. 80–83.

Tsur, Oren, Dan Calacci & David Lazer (2015). A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* Beijing, China, 26–31 July 2015, pp. 1629–1638.

Tulkens, Stéphan, Lisa Hilte, Elise Lodewyckx, Ben Verhoeven & Daelemans Walter (2016). A dictionary-based approach to racism detection in Dutch social media. In *Proceedings of the First Workshop on Text Analytics for Cybersecurity and Online Safety,* Portorož, Slovenia, 23 May 2016.

van Meersbergen, Maarten, Piek Vossen, Janneke van der Zwaan, Antske Fokkens, Willem van Hage, Inger Leemans & Isa Maks (2017). Storyteller: Visual analytics of perspectives on rich text interpretations. In *Proceedings of the 2017 EMNLP Workshop on Natural Language Processing meets Journalism,* Copenhagen, Denmark, 7 September 2017, pp. 37–45.

Vargas, Saúl & Pablo Castells (2011). Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the 5th ACM Conference on Recommender Systems,* Chicago, Ill., 23-27 October 2011, pp. 109–116.

Volkova, Svitlana, Kyle Shaffer, Jin Yea Jang & Nathan Hodas (2017). Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on Twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers),* Vancouver, B.C., Canada, 30 July –4 August 2017, pp. 647–653.

Wang, William Yang (2017). "Liar, Liar Pants on Fire": A new benchmark dataset for fake news detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers),* Vancouver, B.C., Canada, 30 July –4 August 2017, pp. 422–426.

Waseem, Zeerak (2016). Are you a racist or am I seeing things? Annotator influence on hate speech detection on Twitter. In *Proceedings of the 1st Workshop on Natural Language Processing and Computational Social Science,* Austin, Texas, 5 November 2016, pp. 138–142.

Waseem, Zeerak, Thomas Davidson, Dana Warmsley & Ingmar Weber (2017). Understanding abuse: A typology of abusive language detection subtasks. In *Proceedings of the First Workshop on Abusive Language Online,* Vancouver, B.C., Canada, 4 August 2017, pp. 78–84.

Waseem, Zeerak & Dirk Hovy (2016). Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter. In *Proceedings of the NAACL Student Research Workshop,* San Diego, Cal., 12-17 June 2016, pp. 88–93.

Webber, Bonnie (2009). Genre distinctions for discourse in the Penn TreeBank. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing,* Singapore, 2–7 August 2009, pp. 674–682.

Wolters, Maria & Mathias Kirsten (1999). Exploring the use of linguistic features in domain and genre classification. In *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics,* Bergen, Norway, 8–12 June 1999, pp. 142–149.

Yang, Yinfei, Forrest Bao & Ani Nenkova (2017). Detecting (un)important content for single-document news summarization. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Volume 2: Short Papers,* Valencia, Spain, 3–7 April 2017, pp. 707–712.

Zhang, Mi & Neil Hurley (2008). Avoiding monotony: Improving the diversity of recommendations. In *Proceedings of the 2nd ACM Conference on Recommender Systems,* Lausanne, Switzerland, 23-25 October 2008, pp. 123–130.

Zhang, Yuan Cao, Diarmuid Ó Séaghdha, Daniele Quercia & Tamas Jambor (2012). Auralist: Introducing serendipity into music recommendation. In *Proceedings of the 5th ACM International Conference on Web Search and Data Mining,* Seattle, Wash., 8–12 February 2012, pp. 13–22.

Zhao, Tiancheng, Ran Zhao & Maxine Eskenazi (2017). Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* Vancouver, B.C., Canada, 30 July –4 August 2017, pp. 654–664.