

Chancen und Gefahren automatischer Sprachverarbeitung

Michael Strube

michael.strube at h-its.org

February 1, 2014

Inhalt

Computerlinguistische Anwendungen haben sich im Alltag durchgesetzt: Suchmaschinen, Rechtschreibkorrektur, maschinelle Übersetzung, Spracherkennung usw. stehen jedem auf Computer und Mobiltelefon zur Verfügung. Die Computerlinguistik hilft allerdings nicht nur uns Endanwendern, sondern auch dem Handel, mehr über seine Kunden zu erfahren, der Industrie, personalisierte Werbung zu platzieren, autoritären Staaten, Microblogeinträge zu zensieren, Geheimdiensten, Telefongespräche und Emails nicht mehr nur auf Stichwörter, sondern auch auf Inhalte hin zu durchsuchen.

Im Seminar sollen Methoden und Anwendungen aus den Bereichen Sentiment Analysis, maschinelle Übersetzung, Textmining, NLP und Social Media, Computational Advertising usw. untersucht und im Hinblick auf ihre gesellschaftlichen und ethischen Auswirkungen hinterfragt werden:

- Wo ist die Grenze zwischen "cooler" und gefährlicher Forschung?
- Sind wir gerade dabei die Atomphysiker des 21. Jahrhunderts zu werden?
- Sollen wir uns in der Forschung Beschränkungen auferlegen, oder geht alles?
- Wie groß ist unsere Freiheit in der Forschung angesichts des hohen Einflusses von Industrie und Militär in der Forschungsförderung?

Im Seminar werden wir uns mit computerlinguistischen Methoden und Techniken beschäftigen, und diese im Hinblick auf ihr Potential, die Gesellschaft zu beeinflussen und unsere Freiheit zu gefährden, bewerten. Es wird aber auch um Anwendungen gehen, die dem mündigen Bürger helfen, Informationen zu gewinnen, um gesellschaftliche und politische Veränderungen herbeizuführen.

Termine, Themenvorschläge

24.10.2013

Einführung, Motivation, ...

31.10.2013

Geschichte – 2. Weltkrieg, Vietnam, Förderung für KI und CL durch das Militär:

(Shapley, 1972; Glantz & Albers, 1974; Holden, 1975; Edsall, 1975; Thompson, 1986; Schuler & Jacky, 1989; Beusmans & Wieckert, 1989; Winograd, 1991; Yen, 2004; Popp et al., 2004; Coffman et al., 2004; Bradford, 2006; Rubenstein et al., 2008; Cho, 2013; Hajaj et al., 2013)

zur Vorbereitung: (Winograd, 1991)

07.11.2013

Wissenschaftsethik – andere Disziplinen: Informatik, Ingenieurwissenschaften, Biowissenschaftlern, Medizin

Wie funktioniert Google?

(Levy, 2011; Schmidt & Cohen, 2013)

zur Vorbereitung: Fragen zu ethischen Leitlinien in anderen Wissenschaften oder Fragen zu Google

14.11.2013

Microblogs – Zensur:

Referat: Chen Li – (Bamman et al., 2012b; Huang et al., 2013; Xu et al., 2013)

optional: – (Sleeper et al., 2013; Das & Kramer, 2013; Zhu et al., 2013)

zur Vorbereitung: (Bamman et al., 2012b)

21.11.2013

Entity Linking:

Referat: Anja Summa – (Guo et al., 2013b; Liu et al., 2013)

auch: [http://www.darpa.mil/Our_Work/I2O/Programs/Deep_Exploration_and_Filtering_of_Text_\(DEFT\).aspx](http://www.darpa.mil/Our_Work/I2O/Programs/Deep_Exploration_and_Filtering_of_Text_(DEFT).aspx)

und: http://www.nist.gov/tac/publications/2012/presentations/KBP2012_Entity_Linking_tasks_overview.pdf

Microblogs – Soziale Faktoren:

Referat: Raphael Schumann – (Argamon et al., 2009; Bergsma & Van Durme, 2013)

optional: (Burger et al., 2011; Bamman et al., 2012a; Ciot et al., 2013; Hasegawa et al., 2013; Eisenstein et al., 2011; Rangel et al., 2013; Nguyen et al., 2013a)

zur Vorbereitung: (Csomai & Mihalcea, 2008) oder (Milne & Witten, 2008)

28.11.2013

Microblogs – Inhaltserschließung:

Referat: Eleftherios Matios – (Diao & Jiang, 2013)

optional: (Guo et al., 2013a; Eisenstein, 2013; Chua & Asur, 2013; Grinberg et al., 2013; Kairam et al., 2013; Tsur & Rappoport, 2013)

Microblogs – Lokalisierung:

Referat: Xenia Kühling – wegen Krankheit ausgefallen – (Cheng et al., 2013)

Referat: Carolin Günzel – (Schulz et al., 2013)

optional: – (Fink et al., 2009; Gelernter & Mushegian, 2011; Varga et al., 2013; Han et al., 2013; Crooks et al., 2013; Jurgens, 2013)

zur Vorbereitung: (Diao & Jiang, 2013) oder (Cheng et al., 2010)

05.12.2013

Soziale Netzwerke und NLP – soziologische, psychologische Phänomene etc.:

Referat: Hans-Martin Ramsl – (Agarwal et al., 2013)

Referat: Danny Rehl (Fokus auf Facebook) – (Das & Kramer, 2013)

optional: – (Rao et al., 2011; Cano et al., 2013; Nitta et al., 2013; Abu-Jbara et al., 2013; Burke et al., 2013; El-Arini et al., 2013)

zur Vorbereitung: (Elson et al., 2010) oder (Sleeper et al., 2013)

12.12.2013

Sentiment Analysis – Foren:

Referat: Angela Schneider – (Qiu et al., 2013) und auch ein wenig (Qiu & Jiang, 2013; Chen et al., 2013)

Sentiment Analysis – Meinung in der Politik:

Referat: Maximilian Bacher –
(Arunachalam & Sarkar, 2013)

optional: (Mukherjee & Liu, 2013; Mukherjee et al., 2013; Lin et al., 2013; Bhosale et al., 2013; Cohen & Ruths, 2013)

Sentiment Analysis – eher “traditionell”:

Referat: Patrick Claus – (Riloff et al., 2013)

optional: (Sokolova & Lapalme, 2011; Volkova et al., 2013; Zhou et al., 2013)

zur Vorbereitung: (Qiu et al., 2013) oder (Arunachalam & Sarkar, 2013) oder (Riloff et al., 2013)

19.12.2013

Psychologie – Erkennung von Lügen, etc.:

Referat: Angelika Kirilin – (Bachenko et al., 2008)

Referat: Jasmin Schröck – (Ott et al., 2011)

Referat: Sabrina Mänz – (Takase et al., 2013)

optional: (Burgoon et al., 2003; Zhou et al., 2003; 2004; Bond & Lee, 2005; Hancock et al., 2005; Graciarena et al., 2006; Feng & Hirst, 2013; Li et al., 2013a; Resnik et al., 2013; Li et al., 2013b; 2013c; Ott et al., 2013)

zur Vorbereitung: (Bachenko et al., 2008) oder (Ott et al., 2011) oder (Takase et al., 2013)

09.01.2014

Microblogs – Soziale Faktoren:

Referat: Erwin Glockner – (Hasegawa et al., 2013)

optional: (Argamon et al., 2009; Burger et al., 2011; Bamman et al., 2012a; Ciot et al., 2013; Bergsma & Van Durme, 2013; Eisenstein et al., 2011; Rangel et al., 2013; Nguyen et al., 2013a)

Psychologie – Erkennung von Depression, etc.:

Referat: Yulia Pilkevich – (Stirman & Pennebaker, 2001; Lott et al., 2002; Rude et al., 2004; Cohn et al., 2004; Le et al., 2011; Pestian et al., 2012; Resnik et al., 2013; Lamb et al., 2013; De Choudhury et al., 2013; Nguyen et al., 2013b)

Psychologie – Macht, Einfluß, etc.:

Referat: Lyubov Nakryko – (Prabhakaran et al., 2013)

optional: – (Mayfield et al., 2013; Prabhakaran & Rambow, 2013)

zur Vorbereitung: (Hasegawa et al., 2013) oder (Prabhakaran et al., 2013)

16.01.2014

Microblogs – Autorenerkennung:

Referat: Madeline Remse – (Wang et al., 2013)

Referat: Katharina Sowa – (Schwartz et al., 2013)

optional: – (Qian & Liu, 2013)

Anonymisierung (in der medizinischen Domäne – und darüberhinaus?):

Referat: Jonas Placzek – (Narayanan & Shmatikov, 2008)

optional: (Uzuner et al., 2007; Szarvas et al., 2007; Wellner & Pustejovsky, 2007; Friedlin & McDonald, 2008; Uzuner et al., 2008; Hirschman & Aberdeen, 2010; Benitez & Malin, 2010)

zur Vorbereitung: (Narayanan & Shmatikov, 2008) oder (Schwartz et al., 2013) oder (Qian & Liu, 2013)

23.01.2014

Gesprochene Sprache und Dialogsysteme:

Julian Gerhard – (Shepstone et al., 2013)

Referat: Elisa Starke – (Johnston et al., 2013)

Referat: Leo Born – (Rakov & Rosenberg, 2013)

optional: – (Traum, 2013; Rizzo et al., 2013; Pérez-Rosas & Mihalcea, 2013; Cummins et al., 2013; Evans et al., 2013; Federico et al., 2013; Kim et al., 2013; Bigot et al., 2013; Hatmi et al., 2013)

zur Vorbereitung: (Shepstone et al., 2013) oder (Johnston et al., 2013) oder (Rakov & Rosenberg, 2013)

30.01.2014

Essay Scoring etc.:

Referat: Joachim Bingel – (Pitler & Nenkova, 2008) und evtl. (Ganjigunte Ashok et al., 2013)

optional: – (Schwarm & Ostendorf, 2005; Dikli, 2006; Burstein et al., 2010; Chen & Zechner, 2011; Chen & He, 2013; Guinaudeau & Strube, 2013)

Medizin: Informationsextraktion, Kommunikation, etc.:

Referat: Mirjam Eppinger – (Paul & Dredze, 2011)

Referat: Thomas Haider – (Ferrucci et al., 2013)

optional: – (Wallace et al., 2013; Chen, 2013; Sarioglu et al., 2013; Paul & Dredze, 2013; Rebholz-Schumann et al., 2013; Teodoro & Naaman, 2013)

zur Vorbereitung: (Paul & Dredze, 2011) oder (Pitler & Nenkova, 2008) oder (Ferrucci et al., 2013)

06.02.2014

Zusammenfassung, Diskussion

zur Vorbereitung: Offene Fragen, Kommentare, Diskussionsvorschläge – zum gesamten Semester (mir bitte zusenden bis spätestens 05.02.2014, 20 Uhr)

Weitere Themenvorschläge: Maschinelle Übersetzung – DARPA BOLT Programm: (Zbib et al., 2012)

auch: [http://www.darpa.mil/Our_Work/I20/Programs/Broad_Operational_Language_Translation_\(BOLT\).aspx](http://www.darpa.mil/Our_Work/I20/Programs/Broad_Operational_Language_Translation_(BOLT).aspx)

Bemerkungen:

Leistungsnachweise: Lektüre und aktive Teilnahme (1/3), Referat (1/3), Hausarbeit (1/3). Hausarbeit: 8-10 Seiten (Proseminar), 12-15 Seiten (Hauptseminar) inkl. Bibliographie. Die Hausarbeit kann auch per Email an mich geschickt werden, aber *nicht*

als Word-Datei sondern nur als PDF-Datei. – Ich empfehle, wissenschaftliche Texte mit Latex und Bibtex zu verfassen.

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 Tag der Sitzung oder 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.

Hausarbeiten:

Maximal 8-10 Seiten (Proseminar), 12-15 Seiten (Hauptseminar) inkl. Abbildungen, inkl. Literaturverzeichnis.

Inhalt: Fokus auf das vorgestellte Papier; NICHT *Related Work*-Kapitel referieren, wenn die entsprechenden Papiere nicht gelesen wurden; Evaluierung berichten; WICHTIG: mit eigener Meinung oder Bewertung abschließen.

Stil: Wissenschaftlichkeit drückt sich nicht durch lange, komplizierte Sätze und exzessiven Gebrauch von Fremdwörtern aus – deshalb bitte kurze Sätze, einfache Sprache; Hausarbeiten vor der Abgabe Korrektur lesen oder Korrektur lesen lassen (s. auch *Dos and donts: Hinweise zur Abfassung wissenschaftlicher Arbeiten* von Prof. Frank – http://www.cl.uni-heidelberg.de/~frank/materials/dos_and_donts.pdf). Ich schätze Wikipedia als Gegenstand meiner Forschung sehr, nicht aber als Quelle für wissenschaftliche Arbeiten. Hausarbeiten, die Wikipedia (oder auch andere allgemeine Enzyklopädien) als Beleg zitieren, werde ich zurückweisen. Bitte lesen und zitieren Sie Fachliteratur!

Seminararbeit (d.i. eine praktische Arbeit) ist auch möglich. Sollte durch 5-6 Seiten Bericht begleitet werden.

Abgabetermin: bis spätestens 6. März 2014; per Email als PDF-Datei (kein Microsoft Word!) oder ausgedruckt per Post – Matrikelnummer und Studiengang nicht vergessen!

References

- Abu-Jbara, Amjad, Ben King, Mona Diab & Dragomir Radev (2013). Identifying opinion subgroups in Arabic online discussions. In *Proceedings of the ACL 2013 Conference Short Papers*, Sofia, Bulgaria, 4–9 August 2013, pp. 829–835.
- Agarwal, Apoorv, Anup Kotalwar & Owen Rambow (2013). Automatic extraction of social networks from literary text: A case study on Alice in Wonderland. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, Nagoya, Japan, 14–18 October 2013, pp. 1202–1208.
- Argamon, Shlomo, Mosche Koppel, James W. Pennebaker & Jonathan Schler (2009). Automatically profiling the author of an anonymous text. *Communications of the ACM*, 52(2):119–123.
- Arunachalam, Ravi & Sandipan Sarkar (2013). The new eye of the government: Citizen sentiment analysis in social media. In *Proceedings of the Workshop on Natural Language Processing for Social Media (SocialNLP)*, Nagoya, Japan, 14 October 2013, pp. 23–28.
- Bachenko, Joan, Eileen Fitzpatrick & Michael Schonwetter (2008). Verification and implementation of language-based deception indicators in civil and criminal narratives. In *Proceedings of the 22nd International Conference on Computational Linguistics*, Manchester, U.K., 18–22 August 2008, pp. 41–48.
- Bamman, David, Jacob Eisenstein & Tyler Schnoebelen (2012a). *Gender in Twitter: Styles, stances, and social networks*. Published at arXiv:1210.4567.
- Bamman, David, Brendan O’Connor & Noah A. Smith (2012b). Censorship and content deletion in Chinese social media. *First Monday*, 17(3-5).
- Benitez, Kathleen & Bradley Malin (2010). Evaluating re-identification risks with respect to the HIPAA privacy rule. *Journal of the American Medical Informatics Association*, 17(2):160–177.
- Bergsma, Shane & Benjamin Van Durme (2013). Using conceptual class attributes to characterize social media users. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 4–9 August 2013, pp. 710–720.
- Beusmans, Jack & Kären Wieckert (1989). Computing, research, and war: If knowledge is power, where is responsibility? *Communications of the ACM*, 32(8):939–947.
- Bhosale, Shruti, Heath Vinicombe & Raymond Mooney (2013). Detecting promotional content in Wikipedia. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 1851–1857.
- Bigot, Benjamin, Grégory Senay, Linarès, Corinne Fredouille & Richard Dufour (2013). Combining acoustic name spotting and continuous context models to improve spoken person name recognition in speech. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 25–29 August 2013.
- Bond, Gary D. & Adrienne Y. Lee (2005). Language of lies in prison: Linguistic classification of prisoners’ truthful and deceptive natural language. *Applied Cognitive Psychology*, 19:313–329.
- Bradford, Roger B. (2006). Relationship discovery in large text collections using latent semantic indexing. In *Proceedings of the 4th Workshop on Link Analysis, Counterterrorism, and Security*, Bethesda, Md., 20–22 April 2006.
- Burger, John D., John Henderson, George Kim & Guido Zarrella (2011). Discriminating gender on Twitter. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Edinburgh, Scotland, U.K., 27–29 July 2011, pp. 1301–1309.
- Burgoon, Judee K., J.P. Blair, Tiantian Qin & Jr. Nunamaker, Jay F. (2003). Detecting deception through linguistic analysis. In *Proceedings of Intelligence and Security Informatics (ISI)*, p. 958.
- Burke, Moira, Lada A. Adamic & Karyn Wheeler Marciniak (2013). Families on Facebook. In *Proceedings of the 7th International Conference on Weblogs and Social Media*, Cambridge, Mass., 8–11 July 2013.
- Burstein, Jill, Joel Tetreault & Slava Andreyev (2010). Using entity-based features to model coherence in student essays. In *Proceedings of Human Language Technologies 2010: The*

- Conference of the North American Chapter of the Association for Computational Linguistics*, Los Angeles, Cal., 2–4 June 2010, pp. 681–684.
- Cano, Elizabeth, Yuland He, Kang Liu & Jun Zhao (2013). A weakly supervised Bayesian model for violence detection in social media. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, Nagoya, Japan, 14–18 October 2013, pp. 109–117.
- Chen, Annie (2013). Patient experience in online support forums: Modeling interpersonal interactions and medication use. In *Proceedings of the Student Research Workshop at the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 5–7 August 2013, pp. 16–22.
- Chen, Hongbo & Ben He (2013). Automated essay scoring by maximizing human-machine agreement. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 1741–1752.
- Chen, Miao & Klaus Zechner (2011). Computing and evaluating syntactic complexity features for automated scoring of spontaneous non-native speech. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, Oreg., 19–24 June 2011, pp. 722–731.
- Chen, Zhiyuan, Bing Liu, Meichun Hsu, Malu Castellanos & Riddhiman Ghosh (2013). Identifying intention posts in discussion forums. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, Georgia, 9–14 June 2013, pp. 1041–1050.
- Cheng, Zhiyuan, James Caverlee & Kyumin Lee (2010). You are where you tweet: A content-based approach to geo-locating Twitter users. In *Proceedings of the ACM 19th Conference on Information and Knowledge Management (CIKM 2010)*, Toronto, Ont., Canada, 26–30 October 2010, pp. 759–768.
- Cheng, Zhiyuan, James Caverlee & Kyumin Lee (2013). A content-driven framework for geolocating microblog users. *ACM Transactions on Intelligent Systems and Technology*, 4(1). Article 2 (27 pages).
- Cho, Adrian (2013). Network science at center of surveillance dispute. *Science*, 340(6138):1272.
- Chua, Freddy Chong Tat & Sitaram Asur (2013). Automatic summarization of events from social media. In *Proceedings of the 7th International Conference on Weblogs and Social Media*, Cambridge, Mass., 8–11 July 2013.
- Ciot, Morgane, Morgan Sonderegger & Derek Ruths (2013). Gender inference of Twitter users in non-English contexts. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 1136–1145.
- Coffman, Thayne, Seth Greenblatt & Sherry Marcus (2004). Graph-based technologies for intelligence analysis. *Communications of the ACM*, 47(3):45–47.
- Cohen, Raviv & Derek Ruths (2013). Classifying political orientation on Twitter: It’s not easy! In *Proceedings of the 7th International Conference on Weblogs and Social Media*, Cambridge, Mass., 8–11 July 2013.
- Cohn, Michael A., Matthias R. Mehl & James W. Pennebaker (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, 15(10):687–693.
- Crooks, Andrew, Arie Croitoru, Anthony Stefanidis & Jacek Radzikowski (2013). #earthquake: Twitter as a distributed sensor system. *Transactions in GIS*, 17(1):124–147.
- Csomai, Andras & Rada Mihalcea (2008). Linking documents to encyclopedic knowledge. *IEEE Intelligent Systems*, 23(5):34–41.
- Cummins, Nicholas, Julien Epps, Vidhyasaharan Sethu, Michael Breakspear & Roland Goecke (2013). Modeling spectral variability for the classification of depressed speech. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 25–29 August 2013.
- Das, Sauvik & Adam Kramer (2013). Self-censorship on Facebook. In *Proceedings of the 7th International Conference on Weblogs and Social Media*, Cambridge, Mass., 8–11 July 2013.
- De Choudhury, Munmun, Michael Gamon, Scott Counts & Eric Horvitz (2013). Predicting

- depression via social media. In *Proceedings of the 7th International Conference on Weblogs and Social Media*, Cambridge, Mass., 8–11 July 2013.
- Diao, Qiming & Jing Jiang (2013). A unified model for topics, events and users on Twitter. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 1869–1879.
- Dikli, Seimire (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment*, 5(1):??
- Edsall, John T. (1975). Scientific freedom and responsibility. report of the AAAS committee on scientific freedom and responsibility. *Science*, 188(4189):687–693.
- Eisenstein, Jacob (2013). What to do about bad language on the internet. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, Georgia, 9–14 June 2013, pp. 359–369.
- Eisenstein, Jacob, Noah A. Smith & Eric P. Xing (2011). Discovering sociolinguistic associations with structured sparsity. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, Oreg., 19–24 June 2011, pp. 1365–1374.
- El-Arini, Khalid, Min Xu, Emily Fox & Carlos Guestrin (2013). Representing documents through their readers. In *Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Chicago, Ill., 11–14 August 2013, pp. 14–22.
- Elson, David K., Nicholas Dames & Kathleen R. McKeown (2010). Extracting social networks from literary fiction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Sweden, 11–16 July 2010, pp. 138–147.
- Evans, Nicholas, Tomi Kinnunen & Junichi Yamagishi (2013). Spoofing and countermeasures for automatic speaker verification. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 25–29 August 2013.
- Federico, Alegre, Ravichander Vippera, Asmaa Amehraye & Nicholas Evans (2013). A new speaker verification spoofing countermeasure based on local binary patterns. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 25–29 August 2013.
- Feng, Vanessa Wei & Graeme Hirst (2013). Detecting deceptive opinions with profile compatibility. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, Nagoya, Japan, 14–18 October 2013, pp. 338–346.
- Ferrucci, David A., Anthony Levas, Sugato Bagchi, David Gondek & Erik T. Mueller (2013). Watson: Beyond Jeopardy! *Artificial Intelligence*, 199-200:93–105.
- Fink, Clay, Christine Piatko, James Mayfield, Tim Finin & Justin Martineau (2009). Geolocating blogs from their textual content. In *Working Notes of the AAAI Spring Symposium on Social Semantic Web: Where Web 2.0 Meets Web 3.0*, Stanford, Calif., 23–25 March 2009.
- Friedlin, F. Jeff & Clement J. McDonald (2008). A software tool for removing patient identifying information from clinical documents. *Journal of the American Medical Informatics Association*, 15(5):601–610.
- Ganjigunte Ashok, Vikas, Song Feng & Yejin Choi (2013). Success with style: Using writing style to predict the success of novels. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 1753–1764.
- Gelernter, Judith & Nikolai Mushegian (2011). Geo-parsing messages from microtext. *Transactions in GIS*, 15(6):753–773.
- Glantz, Stanton A. & Norm V. Albers (1974). Department of Defense R & D in the university. *Science*, 186(4165):706–711.
- Graciarena, Martin, Elizabeth Shriberg, Andreas Stolcke, Frank Enos, Julia Hirschberg & Sachin Kajarekar (2006). Combining prosodic, lexical and cepstral systems for deceptive speech detection. In *Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, 15–19 June 2006, pp. 1033–1036.
- Grinberg, Nir, Mor Naaman, Blake Shaw & Gilad Lotan (2013). Extracting diurnal patterns of real world activity from social media. In *Proceedings of the 7th International Conference on Weblogs and Social Media*, Cambridge, Mass., 8–11 July 2013.

- Guinaudeau, Camille & Michael Strube (2013). Graph-based local coherence modeling. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 4–9 August 2013, pp. 93–103.
- Guo, Weiwei, Hao Li, Heng Ji & Mona Diab (2013a). Linking tweets to news: A framework to enrich short text data in social media. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 4–9 August 2013, pp. 239–249.
- Guo, Yuhang, Bing Qin, Ting Liu & Sheng Li (2013b). Microblog entity linking by leveraging extra posts. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 863–868.
- Hajaj, Chen, Noam Hazon, David Sarne & Avshalom Elmalech (2013). Search more, disclose less. In *Proceedings of the 27th Conference on the Advancement of Artificial Intelligence*, Bellevue, Wash., 14–18 July 2013, p. ??
- Han, Bo, Paul Cook & Timothy Baldwin (2013). A stacking-based approach to Twitter user geolocation prediction. In *Proceedings of the ACL 2013 System Demonstrations*, Sofia, Bulgaria, 4–9 August 2013, pp. 7–12.
- Hancock, Jeffrey T., Lauren Curry, Saurabh Goorha & Michael Woodworth (2005). Automated linguistic analysis of deceptive and truthful synchronous computer-mediated communication. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*, p. 22c.
- Hasegawa, Takayuki, Nobuhiro Kaji, Naoki Yoshinaga & Masashi Toyoda (2013). Predicting and eliciting addressee’s emotion in online dialogue. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 4–9 August 2013, pp. 964–972.
- Hatmi, Mohamd, Christine Jacquin, Emmanuel Morin & Sylvain Meignier (2013). Incorporating named entity recognition into the speech transcription process. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 25–29 August 2013.
- Hirschman, Lynette & John Aberdeen (2010). Measuring risk and information preservation: Toward new metrics for de-identification of clinical texts. In *Proceedings of the 2nd Louhi Workshop on Text and Data Mining of Health Documents*, Los Angeles, Cal., 5 June 2010, pp. 72–75.
- Holden, Constance (1975). Privacy: Congressional efforts are coming to fruition. *Science*, 188(4189):713–715.
- Huang, Hongzhao, Zhen Wen, Dian Yu, Heng Ji, Yizhou Sun, Jiawei Han & He Li (2013). Resolving entity morphs in censored data. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 4–9 August 2013, pp. 1083–1093.
- Johnston, Michael, Patrick Ehlen, Frederick G. Conrad, Michael F. Schober, Christopher Antoun, Stefanie Fail, Andrew Hupp, Lucas Vickers, Huiying Yan & Chan Zhang (2013). Spoken dialog systems for automated survey interviewing. In *Proceedings of the SIGdial 2013 Conference: The 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Metz, France, 22–24 August 2013, pp. 329–333.
- Jurgens, David (2013). That’s what friends are for: Inferring location in online social media platforms based on social relationships. In *Proceedings of the 7th International Conference on Weblogs and Social Media*, Cambridge, Mass., 8–11 July 2013.
- Kairam, Sanjay Ram, Meredith Ringel Morris, Jaime Teevan, Dan Liebling & Susan Dumais (2013). Towards supporting search over trending events with social media. In *Proceedings of the 7th International Conference on Weblogs and Social Media*, Cambridge, Mass., 8–11 July 2013.
- Kim, Samuel, Fabio Valente & Alessandro Vinciarelli (2013). Annotation and detection of conflict escalation in political debates. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 25–29 August 2013.
- Lamb, Alex, Michael J. Paul & Mark Dredze (2013). Separating fact from fear: Tracking flu infections on Twitter. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, Georgia, 9–14 June 2013, pp. 789–795.

- Le, Xuan, Ian Lancashire, Graeme Hirst & Regina Jokel (2011). Longitudinal detection of dementia through lexical and syntactic changes in writing: A case study of three British novelists. *Literary and Linguistic Computing*, 26(4):435–461.
- Levy, Steven (2011). *In the plex: How Google thinks, works, and shapes our lives*. New York, N.Y.: Simon & Schuster.
- Li, Chiwei, Myle Ott & Claire Cardie (2013a). Identifying manipulated offerings on review portals. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 1933–1942.
- Li, Fangtao, Yang Gao, Shuchang Zhou, Xiance Si & Decheng Dai (2013b). Deceptive answer prediction with user preference graph. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 4–9 August 2013, pp. 1723–1732.
- Li, Jiwei, Claire Cardie & Sujian Li (2013c). TopicSpam: A topic-model based approach for spam detection. In *Proceedings of the ACL 2013 Conference Short Papers*, Sofia, Bulgaria, 4–9 August 2013, pp. 217–221.
- Lin, Ching-Sheng, Samira Shaikh, Jennifer Stromer-Galley, Jennifer Crowley, Tomek Strzalkowski & Veena Ravishankar (2013). Topical positioning: A new method for predicting opinion changes in conversation. In *Proceedings of the Workshop on Language Analysis in Social Media*, Atlanta, Georgia, 13 June 2013, pp. 41–48.
- Liu, Xiaohua, Yitong Li, Haocheng Wu, Ming Zhou, Furu Wei & Yi Lu (2013). Entity linking for tweets. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 4–9 August 2013, pp. 1304–1311.
- Lott, P.R., S. Guggenbühl, A. Schneeberger, A.E. Pulver & H.H. Stassen (2002). Linguistic analysis of the speech output of schizophrenic, bipolar, and depressive patients. *Psychopathology*, 35:220–227.
- Mayfield, Elijah, David Adamson & Carolyn Penstein Rosé (2013). Recognizing rare social phenomena in conversation: Empowerment detection in support group chatrooms. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 4–9 August 2013, pp. 104–113.
- Milne, David & Ian H. Witten (2008). Learning to link with Wikipedia. In *Proceedings of the ACM 17th Conference on Information and Knowledge Management (CIKM 2008)*, Napa Valley, Cal., USA, 26–30 October 2008, pp. 1046–1055.
- Mukherjee, Arjun & Bing Liu (2013). Discovering user interactions in ideological discussions. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 4–9 August 2013, pp. 671–681.
- Mukherjee, Arjun, Vivek Venkataraman, Bing Liu & Sharon Meraz (2013). Public dialogue: Analysis of tolerance in online discussions. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 4–9 August 2013, pp. 1680–1690.
- Narayanan, Arvind & Vitaly Shmatikov (2008). Robust de-anonymization of large sparse datasets. In *Proceedings of the IEEE Symposium on Security and Privacy*, Oakland, Calif., 18–21 May 2008, pp. 111–125.
- Nguyen, Dong, Rilana Gravel, Dolf Trieschnigg & Theo Meder (2013a). "how old do you think I am?" A study of language and age in Twitter. In *Proceedings of the 7th International Conference on Weblogs and Social Media*, Cambridge, Mass., 8–11 July 2013.
- Nguyen, Thin, Bo Dao, Dinh Phung, Svetha Venkatesh & Michael Berk (2013b). Online social capital: Mood, topical and psycholinguistic aspects. In *Proceedings of the 7th International Conference on Weblogs and Social Media*, Cambridge, Mass., 8–11 July 2013.
- Nitta, Taisei, Fumito Masui, Michal Ptaszynski, Yasutomo Kimura, Rafal Rzepka & Kenji Araki (2013). Detecting cyberbullying entries on informal school websites based on category relevance maximization. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, Nagoya, Japan, 14–18 October 2013, pp. 579–586.
- Ott, Myle, Claire Cardie & Jeffrey T. Hancock (2013). Negative deceptive opinion spam. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, Georgia, 9–14 June 2013, pp. 497–501.

- Ott, Myle, Yejin Choi, Claire Cardie & Jeffrey T. Hancock (2011). Finding deceptive opinion spam by any stretch of imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, Oreg., 19–24 June 2011, pp. 309–319.
- Paul, Michael J. & Mark Dredze (2011). You are what you tweet: Analyzing Twitter for public health. In *Proceedings of the 5th International Conference on Weblogs and Social Media*, Barcelona, Spain, 17–21 July 2011.
- Paul, Michael J. & Mark Dredze (2013). Drug extraction from the web: Summarizing drug experiences with multi-dimensional topic models. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, Georgia, 9–14 June 2013, pp. 168–178.
- Pérez-Rosas, Verónica & Rada Mihalcea (2013). Sentiment analysis of online spoken reviews. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 25–29 August 2013.
- Pestian, John P., Pawel Matykiewicz, Michelle Linn-Gust, Brett South, Ozlem Uzuner, Jan Wiebe, K. Bretonnel Cohen, John Hurdle & Christopher Brew (2012). Sentiment analysis of suicide notes: A shared task. *Biomedical Informatics Insights*, 5(Suppl.1):3–16.
- Pitler, Emily & Ani Nenkova (2008). Revisiting readability: A unified framework for predicting text quality. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, Waikiki, Honolulu, Hawaii, 25–27 October 2008, pp. 186–195.
- Popp, Robert, Thomas Armour, Ted Senator & Kristen Numrzh (2004). Countering terrorism through information technology. *Communications of the ACM*, 47(3):36–43.
- Prabhakaran, Vinodkumar, Ajita John & Dorée D. Seligman (2013). Who had the upper hand? Ranking participants of interactions based on their relative power. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, Nagoya, Japan, 14–18 October 2013, pp. 365–373.
- Prabhakaran, Vinodkumar & Owen Rambow (2013). Written dialog and social power: Manifestations of different types of power in dialog behavior. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, Nagoya, Japan, 14–18 October 2013, pp. 216–224.
- Qian, Tiejun & Bing Liu (2013). Identifying multiple userids of the same author. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 1124–1135.
- Qiu, Minghui & Jing Jiang (2013). A latent variable model for viewpoint discovery from threaded forum posts. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, Georgia, 9–14 June 2013, pp. 1031–1040.
- Qiu, Minghui, Liu Yang & Jing Jiang (2013). Mining user relations from online discussions using sentiment analysis and probabilistic matrix factorization. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, Georgia, 9–14 June 2013, pp. 401–410.
- Rakov, Rachel & Andrew Rosenberg (2013). "Sure I did the right thing": A system for sarcasm detection in speech. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 25–29 August 2013.
- Rangel, Francisco, Paolo Rosso, Moshe Koppel, Efstathios Stamatatos & Giacomo Inches (2013). Overview of the author profiling task at PAN 2013. In *Proceedings of CLEF 2013 Labs and Workshops – Notebook Papers*, Valencia, Spain, 23–26 September 2013.
- Rao, Delip, Michael Paul, Clayton Fink, David Yarowsky, Timothy Oates & Glen Coppersmith (2011). Hierarchical Bayesian models for latent attribute detection in social networks. In *Proceedings of the 5th International Conference on Weblogs and Social Media*, Barcelona, Spain, 17–21 July 2011.
- Rebbholz-Schumann, Dietrich, Simon Clematide, Fabio Rinaldi, Senay Kafkas, Erik M. van Muligen, Chinh Bui, Johannes Hellrich, Ian Lewin, David Milward, Michael Poprat, Antonio Jimeno-Yepes, Udo Hahn & Jan A. Kors (2013). Multilingual semantic resources and parallel corpora in the biomedical domain: The CLEF-ER challenge. In *Proceedings of CLEF 2013*

- Labs and Workshops – Notebook Papers*, Valencia, Spain, 23–26 September 2013.
- Resnik, Philip, Anderson Garron & Rebecca Resnik (2013). Using topic modeling to improve prediction of neuroticism and depression in college students. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 1348–1353.
- Riloff, Ellen, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert & Ruihong Huang (2013). Sarcasm as contrast between a positive sentiment and negative situation. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 704–714.
- Rizzo, Albert, Eric Forbell, Belinda Lange, John Galen Buckwalter, Josh Williams, Kenji Sagae & David Traum (2013). SimCoach: An online intelligent virtual agent system for breaking down barriers to care for service members and veterans. In R.S. Scurfield & K.T. Platoni (Eds.), *Healing War Trauma: A Handbook of Creative Approaches*, pp. 238–250. Routledge.
- Rubenstein, Ira S., Ronald D. Lee & Paul M. Schwartz (2008). Data mining and internet profiling: Emerging regulatory and technological approaches. *The University of Chicago Law Review*, 75(1):261–285.
- Rude, Stephanie S., Eva-Maria Gortner & James W. Pennebaker (2004). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, 18(8):1121–1133.
- Sarioglu, Efsun, Kabir Yadav & Hyeong-Ah Choi (2013). Topic modeling based classification of clinical reports. In *Proceedings of the Student Research Workshop at the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 5–7 August 2013, pp. 67–73.
- Schmidt, Eric & Jared Cohen (2013). *The New Digital Age: Reshaping the Future of People, Nations and Business*. New York, N.Y.: Alfred A. Knopf.
- Schuler, Douglas & Jonathan Jacky (1989). Responsibility. *Communication of the ACM*, 32(8):925–927.
- Schulz, Axel, Aristotelis Hadjakos, Heiko Paulheim, Johannes Nachtwey & Max Mühlhäuser (2013). A multi-indicator approach for geolocalization of tweets. In *Proceedings of the 7th International Conference on Weblogs and Social Media*, Cambridge, Mass., 8–11 July 2013.
- Schwarm, Sarah E. & Mari Ostendorf (2005). Reading level assessment using support vector machines and statistical language models. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, Ann Arbor, Mich., 25–30 June 2005, pp. 523–530.
- Schwartz, Roy, Oren Tsr, Ari Rappoport & Moshe Koppel (2013). Authorship attribution of micro-messages. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 1880–1891.
- Shapley, Deborah (1972). Defense research: The names are changed to protect the innocent. *Science*, 175(4024):866–868.
- Shepstone, Sven Ewan, Zheng-Hua Tan & Søren Holdt Jensen (2013). Demographic recommendation by means of group profile elicitation using speaker age and gender recognition. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 25–29 August 2013.
- Sleeper, Manya, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese & Lorie Faith Cranor (2013). The post that wasn't: Exploring self-censorship on Facebook. In *Proceedings of the Conference on Computer Supported Cooperative Work*, San Antonio, Tex., 23–27 February 2013, pp. 793–802.
- Sokolova, Marina & Guy Lapalme (2011). Learning opinions from user-generated web content. *Natural Language Engineering*, 17(4):541–567.
- Stirman, Shannon Wiltsey & James W. Pennebaker (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, 63:517–522.
- Szarvas, György, Richárd Farkas & Róbert Busa-Fekete (2007). State-of-the-art anonymization of medical records using an iterative machine learning framework. *Journal of the American Medical Informatics Association*, 14(5):574–580.
- Takase, Sho, Akiko Murakami, Miki Enoki, Naoaki Okazaki & Kentaro Inui (2013). Detecting

- chronic critics based on sentiment polarity and user's behavior in social media. In *Proceedings of the Student Research Workshop at the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 5–7 August 2013, pp. 110–116.
- Teodoro, Rannie & Mor Naaman (2013). Fitter with Twitter: Understanding personal health and fitness activity in social media. In *Proceedings of the 7th International Conference on Weblogs and Social Media*, Cambridge, Mass., 8–11 July 2013.
- Thompson, Clark (1986). Military direction of academic CS research. *Communications of the ACM*, 29(7):583–585.
- Traum, David (2013). Non-cooperative and deceptive virtual agents. *IEEE Intelligent Systems*, 27(6):66–69.
- Tsur, Oren & Ari Rappoport (2013). Efficient clustering of short messages into general domains. In *Proceedings of the 7th International Conference on Weblogs and Social Media*, Cambridge, Mass., 8–11 July 2013.
- Uzuner, Özlem, Yuan Luo & Peter Szolovits (2007). Evaluating the state-of-the-art in automatic de-identification. *Journal of the American Medical Informatics Association*, 14(5):550–563.
- Uzuner, Özlem, Tawanda C. Sibanda, Yuan Luo & Peter Szolovits (2008). A de-identifier for medical discharge summaries. *Artificial Intelligence in Medicine*, 42(1):13–35.
- Varga, István, Motoki Sano, Kentaro Torisawa, Chikara Hashimoto, Kiyonori Ohtake, Takao Kawai, Jong-Hoon Oh & Stijn De Saeger (2013). Aid is out there: Looking for help from tweets during a large scale disaster. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 4–9 August 2013, pp. 1619–1629.
- Volkova, Svitlana, Theresa Wilson & David Yarowsky (2013). Exploring demographic language variations to improve multilingual sentiment analysis in social media. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 1815–1827.
- Wallace, Byron C., Thomas A Trikalinos, M. Barton Laws, Ira B. Wilson & Eugene Charniak (2013). A generative joint, additive, sequential model of topics and speech acts in patient-doctor communication. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 1765–1775.
- Wang, Zhongqing, Shoushan Li, Fang Kong & Guodong Zhou (2013). Collective personal profile summarization with social networks. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 715–725.
- Wellner, Ben & James Pustejovsky (2007). Automatically identifying the arguments of discourse connectives. 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. 92–101.
- Winograd, Terry (1991). Strategic computing research and the universities. In C. Dunlop & R. Kling (Eds.), *Computerization and Controversy: Value Conflicts and Social Choices*, pp. 704–716. San Diego, Cal.: Academic Press Professional.
- Xu, Jun-Ming, Benjamin Burchfiel, Xiaojin Zhu & Amy Bellmore (2013). An examination of regret in bullying tweets. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, Georgia, 9–14 June 2013, pp. 697–702.
- Yen, John (2004). Emerging technologies for homeland security. *Communications of the ACM*, 47(3):33–35.
- Zbib, Rabih, Erika Malchiodi, Jacob Devlin, David Stallard, Spyros Matsoukas, Richard Schwartz, John Makhoul, Omar F. Zaidan & Chris Callison-Burch (2012). Machine translation of Arabic dialects. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Montréal, Québec, Canada, 3–8 June 2012, pp. 49–59.
- Zhou, Lina, Judee K. Burgeon & Douglas P. Twitchell (2003). A longitudinal analysis of language behavior of deception in e-mail. In *Proceedings of the First NSF/NIJ Symposium on Intelligence and Security Informatics*, Tucson, Ariz., 2003, p. ??
- Zhou, Lina, Judee K. Burgeon, Douglas P. Twitchell, Tiantian Qin & Jr. Nunamaker, Jay F.

- (2004). A comparison of classification methods for predicting deception in computer-mediated communication. *Journal of Management Information Systems*, 20(4):139–165.
- Zhou, Xinjie, Xiaojun Wan & Jianguo Xiao (2013). Collective opinion target extraction in Chinese microblogs. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 1840–1850.
- Zhu, Tao, David Phipps, Adam Pridgen, Jedidiah R. Crandall & Dan S. Wallach (2013). The velocity of censorship: High-fidelity detection of microblog post deletions. In *Proceedings of the 22nd USENIX Security Symposium*, Washington, D.C., 14–16 August 2013.