From Shakespeare to Twitter: What are Language Styles all about?

Wei Xu
Department of Computer Science and Engineering
The Ohio State University
weixu@cse.ohio-state.edu

Abstract

As natural language processing research is growing and largely driven by the availability of data, we expanded research from news and small-scale dialog corpora to web and social media. User-generated data and crowdsourcing opened the door for investigating human language of various styles with more statistical power and real-world applications. In this position/survey paper, I will review and discuss seven language styles that I believe to be important and interesting to study: influential work in the past, challenges at the present, and potential impact for the future.

1 Top Three Problems

The top three problems for studying language styles are data, data and data. More specifically, they are data shortage, data fusion, and data annotation problems. The data shortage problem has been improving, which is the main reason that there is surge in the number of research studies on language styles. The data fusion problem is more specific to the area, due to the subtle and often subjective nature of linguistic styles. For instance, while men and women talk in different ways (note this is not the same as talking about different things), they also talk about a lot of things in an indistinguishable way. Moreover, there is also a huge variance from one man to another, one woman to another. The styles are often fused together in the data and not easy to separate out or make black-and-white judgements on. This also leads to challenges in data annotation or data collection, comparing to other NLP tasks (e.g. question answering). Throughout the rest of this paper, we shall see many creative solutions, interesting work, and promising potential.

2 Seven Styles of Language

Disclaimers: (i) We discuss primarily in the context of natural language processing research; (ii) There are certainly more than seven language styles as there are more than seven wonders in the world.

2.1 Simple and Short

Text simplification is one of the earliest topics in computational linguistics that directly deals with language styles, rewriting regular texts into simpler versions for people with limited reading capabilities. The major transition from rule-based to machine learning approach for automatic sentence simplification did not happen until 2010 after Simple English Wikipedia became available. It is worth noting that the Simple Wikipedia data has some issues on the quality and degree of simplicity (Xu et al., 2015b). The shortage of high quality data is becoming gradually alleviated as the Newsela corpus (Xu et al., 2015b) of professionally edited 1000+ articles is released, and as more and more attention and appreciation are given by the research community to data construction (Brunato et al., 2016; Hwang et al., 2015). Multiple studies have shown crowcourcing workers can produce high quality simplifications (Xu et al., 2016; Amancio and Specia, 2014; Pellow and Eskenazi, 2014), though it is costly to scale up. Data will remain a central problem[1] as the data-hungry neural generation models (Nisioi et al., 2017) are a promising direction for future work.

Besides data, another severe problem is evaluation. In fact, one common human evaluation that uses a five point Likert scale on grammaticality, meaning and simplicity should be considered

[1]Lexical simplification as a subtask can utilize or bypass the need of parallel data (Glavaš and Štajner, 2015; Paetzold and Specia, 2016; Pavlick and Callison-Burch, 2016).

Proceedings of the Workshop on Stylistic Variation, pages 1–9
Copenhagen, Denmark, September 7–11, 2017.

unacceptable when deletion is involved, as it unfairly biases towards deletions over paraphrasing. There has been some progress on creating automatic evaluation metrics (Xu et al., 2016) and exploring new human evaluation methodologies (Xu et al., 2016; Nisioi et al., 2017; Siddharthan and Mandya, 2014). We are going to need more data, clever ideas and careful evaluation designs.

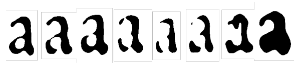
For the record, everything about sentence simplification is much harder than sentence compression[2] primarily due to the interactions between deletion and paraphrasing. Like simplification, previously, sentence compression also use human evaluation with Likert scale on grammaticality and meaning. However, it is shown to be problematic without controlling for compression ratio (Napoles et al., 2011). Now sentence compression systems are mostly compared at the same compression ratio. It is also worth noting that neural compression is similarly lacking in large-scale parallel data (Toutanova et al., 2016) and currently relies on news headline data which results in headline-like outputs (Filippova et al., 2015; Rush et al., 2015).

2.2 Instructional and Robotic

Despite the fact that instructional language is important in our everyday lives, there have been relatively limited efforts to design automated algorithms that link language to action in real world applications. Largely because of the limited availability of annotated datasets which are much-needed for training and evaluating machine learning models, existing works are primarily on cooking recipes (Tasse and Smith, 2008), airline booking conversations (Zettlemoyer and Collins, 2007), software help documents (Branavan et al., 2009) and robot navigation commands (Chen and Mooney, 2011). In particular, cooking recipe has sprouted a rich line of research as a proxy to robotic instructions (Bollini et al., 2013; Jermsurawong and Habash, 2015; Kiddon et al., 2015). Recent efforts aim to study natural language instructions for biology lab experiments (Kulkarni et al., 2017). Two closely relevant research areas, semantic parsing and dialog, have also both made major advances in recent years to utilize large-scale data via weak supervision (Cai and Yates, 2013; Artzi and Zettlemoyer, 2013) and neural network models (Lee et al., 2016; Misra and Artzi, 2016). The 1st Workshop on Language Grounding for Robotics (RoboNLP) will be held at ACL 2017. We shall expect research on instructional language become more and more fruitful in the near future.

2.3 Historical and Evolving

The rise of digital humanities certainly helps to provide more digitized materials for leaning techniques. Historical documents are proven fun (in the other word, hard) to work with. Garrette and Alpert-Abrams (2016) used the following example to present the challenges of having multiple unknown fonts and inking on a single page of a book in the Primeros Libros corpus:



A series of work (Berg-Kirkpatrick et al., 2013; Berg-Kirkpatrick and Klein, 2014; Garrette et al., 2015) have been conducted on this and other corpora to develop historical document optical character recognition (OCR) better handle fonts, offsets, etc, together with language models through unsupervised learning. Unsupervised domain adaptation to historic text was also attempted by Yang and Eisenstein (2015) using feature embedding on the part-of-speech tagging task.

Shakespeare plays in contrast are perfect for investigating a consistent writing style from a single author. Even with a relatively small amount of parallel training data, it is possible to learn paraphrase models which capture stylistic phenomena and can transform the line in the Star Wars *"If you will not be turned, you will be destroyed!"* to Shakespearean style *"If you will not be turn'd, you will be undone!"* (Xu et al., 2012b; Xu, 2014). One can image such stylistic paraphrasing, as it continues to improve, would possibly help preserve privacy and anonymity (Brennan et al., 2012). This is one thing about research on language styles, it often involves a sense of social justice and for social good (e.g. simplification for children, robotics for repetitive wet lab experiments).

Being able to handle evolving language is crucial in natural language processing applications. As the most high-performance systems often utilize fully supervised or weakly supervised learning, the time elapsed from training data to new test data will cause performance deteriorating

[2] which is closely related to, sometimes used interchangeably with, though different from, abstractive summarization, headline generation, sentence fusion.

(Plank, 2018). The most apparent case is out-of-vocabulary (OOV) words (van der Wees et al., 2015; Seraj et al., 2015), especially new emerging named entities and newly coined words (e.g. *"selfie"*, *"Brexiteers"*). This problem will become more pressing and more feasible to study as more and more time-sensitive online text data is accumulating. Learning up-to-date paraphrases (Lan et al., 2017), vector semantics (Cherry and Guo, 2015) and character-based neural models (Ling et al., 2015; Rei et al., 2016) from online data streams could be plausible solutions that connect unseen data with known expressions.

2.4 Colloquial and Internet

As social media started booming, especially after Twitter released the streaming API for free in 2010 that provides real-time tweets as posted, there is a huge explosion on social media research. Multiple workshops are dedicated to this special type of text including the Workshop on Noisy User-generated Text (WNUT) and Workshop on Making Sense of Microposts (#microposts) that hold annual shared tasks. Before that, most unedited text data (vs. well-edited such as news) is from web forums and blogs, while short message service (SMS) and email data are limited to rather small amounts due to privacy reasons (Baldwin et al., 2013). Interesting research falls into two camps: normalize lexical variants to standard form (Han and Baldwin, 2011; Xu et al., 2013) or develop domain adapted NLP systems (Ritter et al., 2011; Gimpel et al., 2011; Kong et al., 2014; Tabassum et al., 2016). The iconic opinion paper *What to do about bad language on the Internet* by Jacob Eisenstein (2013) highlighted this divide.

There is a third point we have often missed. Besides the noisy hard-to-understand Internet language, many users also use rather standard language on social networks, formal or colloquial. Don't forget that all the traditional news agencies also have Twitter accounts (Hu et al., 2013). Can we make the connections between the formal and colloquial languages as they are heavily mixed on social media? I think the answer is yes, and the twin research topics of paraphrasing and semantic similarity could be part of the solution as many language styles are heavily mixed on social media. For example, in the SemEval shared task PIT-2015 corpus (Xu et al., 2015a), the figurative meaning of the phrase *"on fire"* is captured by the sentential paraphrase of *"Aaaaaaaaand stephen curry is on fire"* and *"What a incredible performance from Stephen Curry"*. Semantic equivalences, as formal as *"fetuses"* and *"fetal tissue"* (Lan et al., 2017) or as informal as *"gets the boot from"* and *"has been sacked by"* (Xu et al., 2014; Xu, 2014), can also be learned automatically from Twitter data. Not to mention that there are also studies that focus on multiword expressions (Schneider and Smith, 2015), idioms (Muzny and Zettlemoyer, 2013), and slang.

2.5 Gendered and Personalized

One unique and exciting opportunity offered by social media data is to learn about the users authoring the texts. Much interesting research on gender difference[3] in language styles appeared in the past few years. Besides gender (Verhoeven et al., 2016; Bamman et al., 2014), other user attributes such as age (Sap et al., 2014), race (Jørgensen et al., 2015) and personality (Schwartz et al., 2013; Ruan et al., 2016; Plank and Hovy, 2015) are also commonly studied for social science and strongly motivated by commercial usages of profiling users and personalized services. Leveraging user demographic factors also shows benefits on improving natural language processing applications such as sentiment analysis (Volkova et al., 2013) and sarcasm detection (Bamman and Smith, 2015).

One particularly interesting challenge is how to handle the situation that stylistic differences (e.g. female users more likely use *"wonderful"* while male users use *"superb"*) are much more subtle than topical preferences (e.g. using word *"husband"* is a strong indicator of female user). Our recent work (Preoţiuc-Pietro et al., 2016) isolated stylistic differences from topic bias by using paraphrase pairs and clusters, and showed their predictive power in user profiling and potential for future work. We also found crowdsourcing workers are surprisingly good at perceiving gender from lexical choices when aggregating their judgments – an infamous phenomenon of so-called *The Wisdom of Crowds* (Surowiecki, 2005). Beyond lexical choice, Johannsen et al. (2015) further showed demographic differences in syntactic variances using multilingual data of online customer reviews and universal dependency parsing.

[3]Although unrelated to linguistic styles, the readers may find *He Said, She Said: Gender in the ACL Anthology* (Vogel and Jurafsky, 2012), a paper on gender-based statistics of NLP researchers, interesting.

Another subsequent challenge is how to transfer the subtle style differences into natural language generation and dialog systems. While we were able to transform contemporary texts into Shakespeare style (Xu et al., 2012b), we found gendered language style much harder to impose. It is possibly that because we have not found the right data for evaluation, for instance, it is hard to expect a randomly drawn sentence to be possible to take on a feminine or masculine style. It could also be the case that it is easier for finer-grained language style to show distinctions. One evident example is author recognition based on an individual's frequent word choices (Clark and Hannon, 2007). Another example is persona-based dialog system that not only captures background knowledge of a user (Li et al., 2016) but also speaking style (Mizukami et al., 2015). It is not a coincident that the later work (Mizukami et al., 2015) is on spoken Japanese, which exhibits extensive gender differences as well as honorifics (not as much in written Japanese).

2.6 Pervasive and Framing

The increasing availability of data also make feasible to study the textural characteristics of persuasion, argumentation and framing in realistic (not laboratory) settings and quantitatively. Besides movie quotes, political speeches, and tweets (Guerini et al., 2015), many interesting data are created and discovered, leading to a growing number of studies. Online discussion platforms provide almost ideal real world data with users stating, reasoning and contesting opinions (Somasundaran and Wiebe, 2009), and sometimes even with explicitly marked successful arguments such as ChangeMyView on Reddit. One recent work (Tan et al., 2016) found that in the ChangeMyView data, after controlled for similar arguments, stylistic choices in how the opinion is expressed carry more predictive power on how likely a user to be persuaded than how likely an argument is persuasive. However, predicting pervasiveness turns out to a difficulty task with about 60-65% accuracy using bag-of-words and linguistic features, in constrast of 75-85% accuracy for predicting politeness). Another interesting work (Recasens et al., 2013) utilized Wikipedia edit history to study biased language (e.g. *"stated"* vs. *"claimed"*) as well as framing (e.g. *"pro-life"* vs. *"anti-abortion"*). The recent construction of the Media Frames Corpus (Card et al., 2015)[4] presents another encouraging opportunity to study framing. The legal domain, such as supreme court documents, is another common place for arguments (Sim et al., 2015) and would possibly be used for studying linguistic styles.

2.7 Polite and Abusive

Another angle that has been looked at is the politeness conveyed in language. Unlike many other styles that come in close pairs (e.g. formal vs. informal, feminine vs. masculine), the polite language does not necessarily have an impolite counterpart. In addition, politeness is expressed more through function words. For example, showing gratitude by *"I appreciate that"* or apologizing by *"Sorry to bother you"*. In fact, the phrase *"in fact"* can be negative as *"in fact you did ..."*. Many other cues are identified and annotated (Danescu-Niculescu-Mizil et al., 2013) on the online interchanges of Wikipedia editors and StackExchange QA users, which can train classifier to predict politeness at about 80% accuracy. A recent study (Voigt et al., 2017) also used automatic methods to examine the respectfulness of police officers toward white and black people from transcripts of body-worn camera footage.

In other words, abusive language is closely related to politeness but not the reverse. The targets could vary from one swear word to multi-sentences, such as the mean tweet Barack Obama read on Jimmy Kimmel's show: *"Obama's hair is looking grayer these days. Can't imagine why since he doesn't seem to be one bit worried about all that's going on."* The context-dependent nature makes it challenging to collect data or design experiments. Moreoever, although bullying traces are abundant, it is a tiny fraction out of random samples which is estimated to 0.02∼0.73% of a 95% confidence internal on 2011 TREC Microblog track corpus (Xu et al., 2012a). The compromise is to look at tweets that include keywords *"bully"*, *"bullied"*, *"bullying"* instead, which is inspiring and an important first step, but far from satisfying. Another representative solution is a carefully designed crowdsourcing experiment which reveals patterns of Internet trolling behavior using user comments on CNN.com news website (Cheng et al., 2017). Perhaps, the 1st

[4] which is a great example why data resource papers even without learning results should be considered *acceptable* in ACL/EMNLP/NAACL/EACL main conferences.

Workshop on Abusive Language Online (ALW) at ACL 2017 will spark more ideas. I would like to quote an anonymous source who raised a thoughtful question: *"Under what circumstances is language use considered to be an abuse? For example, in many states when a women criticizes her husband in public, this might be considered there as abuse of language or hate speech"*, as a reminder of being aware and mindful of the great social factors and impacts embedded in the research of language styles.

3 Conclusion

At this point of the development, natural language processing research ranges a wide variety of genre, domain, register or type of data. I think the term *style* is an all-in-one umbrella concept to bring researchers and scattered attentions in various NLP subareas to a common place. There are certainly many nuances in language styles besides those mentioned in this paper. For example, connotation (e.g. *"childlike"* vs. *"childish"* vs. *"youthful"*) (Rashkin et al., 2016; Carpuat, 2015) and geographical lexical variations from regional (e.g. *"sode"* vs. *"coke"* vs. *"pop"*) to cross-country (e.g. Austrilian vs. American English) (Eisenstein et al., 2010; Garimella et al., 2016; Han et al., 2016). There are also certainly many other relevant works besides those mentioned in this paper. Last but not least, we would like to point out Dan Jurafsky's recent book *The Language of Food* (2014) and one more paper: *Do Linguistic Style and Readability of Scientific Abstracts Affect their Virality?* (Guerini et al., 2012).

Acknowledgments

I would like to thank three anonymous reviewers for helpful feedback, Julian Brooke, Thamar Solario and Moshe Koppel for organizing the EMNLP 2017 Workshop on Stylistic Variation.

References

Marcelo Amancio and Lucia Specia. 2014. An analysis of crowdsourced text simplifications. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*.

Yoav Artzi and Luke Zettlemoyer. 2013. Weakly supervised learning of semantic parsers for mapping instructions to actions. *Transactions of the Association for Computational Linguistics (TACL)* 1:49–62.

Timothy Baldwin, Paul Cook, Marco Lui, Andrew MacKinlay, and Li Wang. 2013. How noisy social media text, how diffrnt social media sources? In *Proceedings of the Sixth International Joint Conference on Natural Language Processing (IJCNLP)*.

David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. 2014. Gender identity and lexical variation in social media. *Journal of Sociolinguistics* 18(2):135–160.

David Bamman and Noah A. Smith. 2015. Contextualized sarcasm detection on Twitter. In *ICWSM*.

Taylor Berg-Kirkpatrick, Greg Durrett, and Dan Klein. 2013. Unsupervised transcription of historical documents. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*.

Taylor Berg-Kirkpatrick and Dan Klein. 2014. Improved typesetting models for historical ocr. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*.

Mario Bollini, Stefanie Tellex, Tyler Thompson, Nicholas Roy, and Daniela Rus. 2013. Interpreting and executing recipes with a cooking robot. In *Experimental Robotics*. Springer, pages 481–495.

Satchuthananthavale RK Branavan, Harr Chen, Luke S Zettlemoyer, and Regina Barzilay. 2009. Reinforcement learning for mapping instructions to actions. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*.

Michael Brennan, Sadia Afroz, and Rachel Greenstadt. 2012. Adversarial stylometry: Circumventing authorship recognition to preserve privacy and anonymity. *ACM Transactions on Information and Systems Security* 15(3):12:1–12:22.

Dominique Brunato, Andrea Cimino, Felice Dell'Orletta, and Giulia Venturi. 2016. PaCCSS-IT: A parallel corpus of complex-simple sentences for automatic text simplification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Qingqing Cai and Alexander Yates. 2013. Large-scale semantic parsing via schema matching and lexicon extension. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and 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 and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*.

Marine Carpuat. 2015. Connotation in translation. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*.

David L Chen and Raymond J Mooney. 2011. Learning to interpret natural language navigation instructions from observations. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI)*.

Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone can become a troll: Causes of trolling behavior in online discussions. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*.

Colin Cherry and Hongyu Guo. 2015. The unreasonable effectiveness of word representations for Twitter named entity recognition. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Jonathan Clark and Charles Hannon. 2007. A classifier system for author recognition using synonym-based features. *MICAI 2007: Advances in Artificial Intelligence* pages 839–849.

Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*.

Jacob Eisenstein. 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 (NAACL-HLT)*.

Jacob Eisenstein, Brendan O'Connor, Noah A. Smith, and Eric P. Xing. 2010. A latent variable model for geographic lexical variation. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*.

Katja Filippova, Enrique Alfonseca, Carlos A. Colmenares, Lukasz Kaiser, and Oriol Vinyals. 2015. Sentence compression by deletion with LSTMs. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Aparna Garimella, Rada Mihalcea, and James Pennebaker. 2016. Identifying cross-cultural differences in word usage. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING)*.

Dan Garrette and Hannah Alpert-Abrams. 2016. An unsupervised model of orthographic variation for historical document transcription. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Dan Garrette, Hannah Alpert-Abrams, Taylor Berg-Kirkpatrick, and Dan Klein. 2015. Unsupervised code-switching for multilingual historical document transcription. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. 2011. Part-of-speech tagging for Twitter: Annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)*.

Goran Glavaš and Sanja Štajner. 2015. Simplifying lexical simplification: Do we need simplified corpora? In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*.

Marco Guerini, Gözde Özbal, and Carlo Strapparava. 2015. Echoes of persuasion: The effect of euphony in persuasive communication. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)*.

Marco Guerini, Alberto Pepe, and Bruno Lepri. 2012. Do linguistic style and readability of scientific abstracts affect their virality? In *Proceedings of the 6th International Conference on Weblogs and Social Media (ICWSM)*.

Bo Han and Timothy Baldwin. 2011. Lexical normalisation of short text messages: Makn sens a #twitter. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)*.

Bo Han, Afshin Rahimi, Leon Derczynski, and Timothy Baldwin. 2016. Twitter geolocation prediction shared task of the 2016 Workshop on Noisy User-generated Text. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*.

Yuheng Hu, Kartik Talamadupula, and Subbarao Kambhampati. 2013. Dude, srsly?: The surprisingly formal nature of Twitter's language. In *Proceedings of the 7th International Conference on Weblogs and Social Media (ICWSM)*.

William Hwang, Hannaneh Hajishirzi, Mari Ostendorf, and Wei Wu. 2015. Aligning sentences from Standard Wikipedia to Simple Wikipedia. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*.

Jermsak Jermsurawong and Nizar Habash. 2015. Predicting the structure of cooking recipes. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Anders Johannsen, Dirk Hovy, and Anders Søgaard. 2015. Cross-lingual syntactic variation over age and gender. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning (ACL)*.

Anna Jørgensen, Dirk Hovy, and Anders Søgaard. 2015. Challenges of studying and processing dialects in social media. In *Proceedings of the Workshop on Noisy User-generated Text (WNUT)*.

Dan Jurafsky. 2014. *The Language of Food*. W. W. Norton Company, Inc.

Chloé Kiddon, Ganesa Thandavam Ponnuraj, Luke Zettlemoyer, and Yejin Choi. 2015. Mise en Place: Unsupervised interpretation of instructional recipes. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Lingpeng Kong, Nathan Schneider, Swabha Swayamdipta, Archna Bhatia, Chris Dyer, and Noah A. Smith. 2014. A dependency parser for Tweets. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Chaitanya Kulkarni, Wei Xu, Alan Ritter, and Raghu Machiraju. 2017. Taking the first essential steps in automating the wet laboratory: Annotating a corpus of protocols for reproducibility. In *Submission*.

Wuwei Lan, Siyu Qiu, Hua He, and Wei Xu. 2017. A continuously growing dataset of sentential paraphrases from Twitter. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2016. Global neural CCG parsing with optimality guarantees. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and Bill Dolan. 2016. A persona-based neural conversation model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Wang Ling, Chris Dyer, Alan W Black, Isabel Trancoso, Ramon Fermandez, Silvio Amir, Luis Marujo, and Tiago Luis. 2015. Finding function in form: Compositional character models for open vocabulary word representation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Dipendra Kumar Misra and Yoav Artzi. 2016. Neural shift-reduce CCG semantic parsing. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Masahiro Mizukami, Graham Neubig, Sakriani Sakti, and Tomoki Toda. 2015. Linguistic individuality transformation for spoken language. In *Proceedings of the 6th International Workshop On Spoken Dialogue Systems*.

Grace Muzny and Luke Zettlemoyer. 2013. Automatic idiom identification in Wiktionary. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Courtney Napoles, Benjamin Van Durme, and Chris Callison-Burch. 2011. Evaluating sentence compression: Pitfalls and suggested remedies. In *Proceedings of the Workshop on Monolingual Text-To-Text Generation*.

Sergiu Nisioi, Sanja Štajner, Simone Paolo Ponzetto, and Liviu P. Dinu. 2017. Exploring neural text simplification models. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Gustavo Paetzold and Lucia Specia. 2016. Benchmarking lexical simplification systems. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*.

Ellie Pavlick and Chris Callison-Burch. 2016. Simple PPDB: A paraphrase database for simplification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*.

David Pellow and Maxine Eskenazi. 2014. An open corpus of everyday documents for simplification tasks. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*.

Barbara Plank. 2018. What to do about non-standard (or non-canonical) language in nlp. In *Proceedings of the 13th Conference on Natural Language Processing (KONVENS)*.

Barbara Plank and Dirk Hovy. 2015. Personality traits on Twitter -or- how to get 1,500 personality tests in a week. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*.

Daniel Preoţiuc-Pietro, Wei Xu, and Lyle Ungar. 2016. Discovering user attribute stylistic differences via paraphrasing. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*.

Hannah Rashkin, Sameer Singh, and Yejin Choi. 2016. Connotation frames: A data-driven investigation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*.

Marek Rei, Gamal Crichton, and Sampo Pyysalo. 2016. Attending to characters in neural sequence labeling models. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING)*.

Alan Ritter, Sam Clark, Mausam, and Oren Etzioni. 2011. Named entity recognition in Tweets: An experimental study. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Xianzhi Ruan, Steven Wilson, and Rada Mihalcea. 2016. Finding optimists and pessimists on Twitter. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Alexander M. Rush, Sumit Chopra, and Jason Weston. 2015. A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Maarten Sap, Gregory Park, Johannes Eichstaedt, Margaret Kern, David Stillwell, Michal Kosinski, Lyle Ungar, and Hansen Andrew Schwartz. 2014. Developing age and gender predictive lexica over social media. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Nathan Schneider and Noah A. Smith. 2015. A corpus and model integrating multiword expressions and supersenses. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

H. Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Stephanie M. Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin E. P. Seligman, and Lyle H. Ungar. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS One* 8:1–16.

Ramtin Mehdizadeh Seraj, Maryam Siahbani, and Anoop Sarkar. 2015. Improving statistical machine translation with a multilingual paraphrase database. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Advaith Siddharthan and Angrosh Mandya. 2014. Hybrid text simplification using synchronous dependency grammars with hand-written and automatically harvested rules. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*.

Yanchuan Sim, Bryan R. Routledge, and Noah A. Smith. 2015. The utility of text: The case of amicus briefs and the supreme court. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI)*.

Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing stances in online debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing (ACL-AFNLP)*.

James Surowiecki. 2005. *The Wisdom of Crowds*. Anchor.

Jeniya Tabassum, Alan Ritter, and Wei Xu. 2016. A minimally supervised method for recognizing and normalizing time expressions in Twitter. In *Proceedings of The 2016 Conference on Empirical Methods on Natural Language Processing (EMNLP)*.

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th International Conference on World Wide Web (WWW)*.

Dan Tasse and Noah A Smith. 2008. SOUR CREAM: Toward semantic processing of recipes. *Carnegie Mellon University, Pittsburgh, Tech. Rep. CMU-LTI-08-005* .

Kristina Toutanova, Chris Brockett, Ke M. Tran, and Saleema Amershi. 2016. A dataset and evaluation metrics for abstractive compression of sentences and short paragraphs. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Marlies van der Wees, Arianna Bisazza, and Christof Monz. 2015. Five shades of noise: Analyzing machine translation errors in user-generated text. In *Proceedings of the Workshop on Noisy User-generated Text (WNUT)*.

Ben Verhoeven, Walter Daelemans, and Barbara Plank. 2016. TwiSty: A multilingual Twitter stylometry corpus for gender and personality profiling. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*.

Adam Vogel and Dan Jurafsky. 2012. He said, she said: Gender in the ACL Anthology. In *Proceedings of the ACL 2012 Special Workshop on Rediscovering 50 Years of Discoveries*.

Rob Voigt, Nicholas P. Camp, Vinodkumar Prabhakaran, William L. Hamilton, Rebecca C. Hetey, Camilla M. Griffiths, David Jurgens, Dan Jurafsky, and Jennifer L. Eberhardt. 2017. Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences* 114(25):6521–6526.

Svitlana Volkova, Theresa Wilson, and 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 (EMNLP)*.

Jun-Ming Xu, Kwang-Sung Jun, Xiaojin Zhu, and Amy Bellmore. 2012a. Learning from bullying traces in social media. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Wei Xu. 2014. *Data-Drive Approaches for Paraphrasing Across Language Variations*. Ph.D. thesis, Department of Computer Science, New York University.

Wei Xu, Chris Callison-Burch, and William B. Dolan. 2015a. SemEval-2015 Task 1: Paraphrase and semantic similarity in Twitter (PIT). In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval)*.

Wei Xu, Chris Callison-Burch, and Courtney Napoles. 2015b. Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics (TACL)* 3:283–297.

Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. 2016. Optimizing statistical machine translation for text simplification. *Transactions of the Association for Computational Linguistics (TACL)* 4:401–415.

Wei Xu, Alan Ritter, Chris Callison-Burch, William B Dolan, and Yangfeng Ji. 2014. Extracting lexically divergent paraphrases from Twitter. *Transactions of the Association for Computational Linguistics (TACL)* 2:435–448.

Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman, and Cherry Colin. 2012b. Paraphrasing for style. In *Proceedings of the 28th International Conference on Computational Linguistics (COLING)*.

Wei Xu, Alan Ritter, and Ralph Grishman. 2013. Gathering and generating paraphrases from Twitter with application to normalization. In *Proceedings of the Sixth Workshop on Building and Using Comparable Corpora (BUCC)*.

Yi Yang and Jacob Eisenstein. 2015. Unsupervised multi-domain adaptation with feature embeddings. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Luke Zettlemoyer and Michael Collins. 2007. Online learning of relaxed CCG grammars for parsing to logical form. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*.