

On learning and representing social meaning in NLP: a sociolinguistic perspective

Nguyen, Dong; Rosseel, Laura; Grieve, Jack

Published in:

Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

DOI:

[10.18653/v1/2021.naacl-main.50](https://doi.org/10.18653/v1/2021.naacl-main.50)

Publication date:

2021

Document Version:

Final published version

[Link to publication](#)

Citation for published version (APA):

Nguyen, D., Rosseel, L., & Grieve, J. (2021). On learning and representing social meaning in NLP: a sociolinguistic perspective. In K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, & Y. Zhou (Eds.), *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 603-612). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.naacl-main.50>

Copyright

No part of this publication may be reproduced or transmitted in any form, without the prior written permission of the author(s) or other rights holders to whom publication rights have been transferred, unless permitted by a license attached to the publication (a Creative Commons license or other), or unless exceptions to copyright law apply.

Take down policy

If you believe that this document infringes your copyright or other rights, please contact openaccess@vub.be, with details of the nature of the infringement. We will investigate the claim and if justified, we will take the appropriate steps.

On learning and representing social meaning in NLP: a sociolinguistic perspective

Dong Nguyen

Utrecht University

The Netherlands

d.p.nguyen@uu.nl

Laura Rosseel

Vrije Universiteit Brussel

Belgium

Laura.Rosseel@vub.be

Jack Grieve

University of Birmingham

United Kingdom

J.Grieve@bham.ac.uk

Abstract

The field of NLP has made substantial progress in building meaning representations. However, an important aspect of linguistic meaning, social meaning, has been largely overlooked. We introduce the concept of social meaning to NLP and discuss how insights from sociolinguistics can inform work on representation learning in NLP. We also identify key challenges for this new line of research.

1 Introduction

Variation is inherent to language. Any variety of language provides its users with a multitude of linguistic forms—e.g., speech sounds, words, grammatical constructions—to express the same referential meaning. Consider, for example, the many ways of pronouncing a given word or the variety of words that can refer to a given concept.

Linguistic variation is the primary object of inquiry of **sociolinguistics**, which has a long history of describing and explaining variation in linguistic form across society and across levels of linguistic analysis (Tagliamonte, 2015). Perhaps the most basic finding of the field is that linguistic variation allows for the expression of **social meaning**, information about the social background and identity of the language user. Such sociolinguistic variation adds an additional layer of meaning onto the basic referential meaning communicated by any utterance or text. Understanding the expression of social meaning based on linguistic variation is a crucial part of the linguistic knowledge of any language user, drawn upon continuously during both the production and processing of natural language. The relationship between variation and social meaning, however, has only begun to be explored computationally (e.g., Pavalanathan et al. (2017)). Studies have shown, for example, that words, capitalisation, or the language variety used can index political identity (Shoemark et al., 2017; Stewart et al., 2018; Tatman et al., 2017).

Despite general acceptance of the link between linguistic variation and social meaning in linguistics, NLP has largely ignored this relationship. Nevertheless, the importance of linguistic variation more generally is increasingly being acknowledged in NLP (Nguyen et al., 2016). NLP tools are usually developed for standard varieties of language, and therefore tend to under-perform on texts written in varieties that diverge from the ‘standard’, including language identification (Blodgett et al., 2016), dependency parsing (Blodgett et al., 2016), and POS tagging (Hovy and Søgaard, 2015).

One approach to overcoming the challenges posed by linguistic variation is text normalisation (Han and Baldwin, 2011; Liu et al., 2011). Normalisation transforms non-standard texts into a more standardised form, which can then be analysed more accurately using NLP models trained on standard language data. Text normalisation, however, removes rich social signals encoded via sociolinguistic variation. Other approaches have also been explored to improve the robustness of NLP models across society, such as adapting them based on demographic factors (Lynn et al., 2017) or social network information (Yang and Eisenstein, 2017).

Linguistic variation, however, should not simply be seen as a problem to be overcome in NLP. Although variation poses a challenge for robust NLP, it also offers us a link to the social meaning being conveyed by any text. To build NLP models that are capable of understanding and generating natural language in the real world, sociolinguistic variation and its role in creating social meaning must be incorporated into our models. For example, over the last few years, research in NLP has been marked by substantial advancements in the area of **representation learning**, but although Bisk et al. (2020) and Hovy (2018) have recently argued that the social nature of language must be considered in representation learning, the concept of social meaning is still largely overlooked.

In this paper, we therefore introduce the concept of social meaning to NLP from a sociolinguistic perspective (Section 2). We reflect on work on representation learning in NLP and how social meaning could play a role there (Section 3), and we present example applications (Section 4). Finally, we identify key challenges moving forward for this important and new line of research in NLP for the robust processing of meaning (Section 5).

2 Social meaning

People use language to communicate a message. The same message can be packaged in various linguistic forms. For example, someone might say ‘I’m not coming, pal’. But they could also refer to their friend as ‘mate’, ‘buddy’, ‘bruv’, or ‘bro’ for instance. Or they could say ‘I am not coming’ or ‘I ain’t comin’ to express that they are not joining that friend. With each of these options, or *variants*, the language user communicates the same referential meaning, that is, they refer to the exact same entity, action or idea in the real or an imagined world. The only difference is the linguistic form used to encode that message. To put it simply: these are different ways of saying the same thing (Labov, 1972).

Although this variation in form does not change the referential meaning of the message, it is not meaningless in itself. Variation in form can also carry information about the social identity of a language user (Eckert, 2008), which sociolinguists call the *social meaning of linguistic variation*. For example, Walker et al. (2014) define social meaning as all social attributes associated with a language feature and its users. These social attributes can be highly diverse and relate to any aspect of identity a language user may want to express through their linguistic output.

Linguistic variation can express broad social attributes like national background or social class. Saying ‘I left the diapers in the trunk’ rather than ‘I left the nappies in the boot’, may suggest that the speaker is American rather than British. But linguistic variation can also be far more fine-grained and can be called upon directly by language users to construct local identities. A famous example is Labov (1972)’s groundbreaking study on Martha’s Vineyard, a small island off the northeast coast of the US. Labov found that within the small island community there were differences in the way people pronounced the diphthongs /ay/ (as in ‘right’) and /aw/ (as in ‘house’).

The study shows that a more centralised pronunciation of the diphthongs was used by local fishermen who opposed the rise in tourism from the mainland on the island. Conversely, islanders who were more oriented towards mainland culture used a more standard American pronunciation for these diphthongs. The pronunciation of these sounds was thus used in this particular community to express the local social meaning of island identity.

The social meaning of linguistic variation is not fixed. Over time, a linguistic variant can develop new meanings and lose others, while new forms can also emerge. A single linguistic feature can also be associated with multiple social meanings. Which of those meanings is activated in interaction depends on the specific context in which that interaction takes place. Campbell-Kibler’s research on the social meaning of the pronunciation of -ing in the US (e.g., ‘coming’ vs. ‘comin’) shows, for instance, that the variation can be linked to both social and regional identity. For example, velar pronunciation ‘coming’ sounds urban, while alveolar pronunciation ‘comin’ is perceived as sounding Southern (Campbell-Kibler, 2007, 2009, 2010).

Information about the speaker can also influence the social meaning attached to variation in -ing pronunciation. Experiments show that when a speaker is presented as a professor, they sound knowledgeable when using the velar pronunciation, while if the same speaker is presented as an experienced professional, they are perceived as knowledgeable when using the alveolar variant (Campbell-Kibler, 2010). The collection of social meanings a linguistic feature could potentially evoke is referred to as the *indexical field* of that feature (Eckert (2008), for a theoretical discussion of indexicality, see Silverstein (2003)).

As the above examples suggest, social meaning can be attached to various types of linguistic features. In the friend, nappy and boot examples, there is variation on the level of the lexicon, while the ‘I ain’t comin’ example shows morphosyntactic variation (‘ain’t’ vs. ‘am not’) and variation in pronunciation (‘comin’ vs. ‘coming’). A language or language variety as a whole can also carry social meaning. Think of the choice to use a standard variety or a local dialect to signal one’s regional background or the use of loans from foreign languages to come across as cosmopolitan or fashionable (Vaattovaara and Peterson, 2019).

It is also important to acknowledge that there are other types of linguistic variation—and other traditions that analyse variation within linguistics—including variation across communicative contexts, as is commonly analysed in corpus linguistics (i.e. register variation) (Biber, 1988; Biber and Conrad, 2009). For example, research on register variation has shown that texts that are intended to concisely convey detailed information, like academic writing, tend to have very complex noun phrases, as opposed to more interactive forms of communication, like face-to-face conversations, which tend to rely more on pronouns. Crucially, register variation depends on the communicative goals, affordances, and constraints associated with the context in which language is used, as opposed to the social background or identity of the individual language users, although the relationship between social and situational variation is also complicated and not fully understood (Eckert and Rickford, 2002; Finegan and Biber, 2002).

Linguistic variation and the expression of social meaning is thus a highly complex phenomenon, and one that sociolinguists are only beginning to fully grasp despite decades of research. Nevertheless, we argue that language variation and social meaning must be considered when building NLP models: not simply to create more robust tools, but to better process the rich meanings of texts in general. Moreover, we believe that methods from NLP could contribute substantially to our understanding of sociolinguistic variation.

3 Representing social meaning

Distributed representations map a word (or some other linguistic form) to a k -dimensional vector, also called an embedding. Sometimes these representations are independent of linguistic context (Mikolov et al., 2013; Pennington et al., 2014), but increasingly they are contextualised (Devlin et al., 2019; Peters et al., 2018). These representations are shown to capture a range of linguistic phenomena (Baroni et al., 2014; Conneau et al., 2018; Gladkova et al., 2016). A key question in the development of representations is *what aspects of meaning* these representations should capture. Indeed, recent reflections have drawn attention to challenges such as polysemy and hyponymy (Emerson, 2020) and construed meaning (Trott et al., 2020). However, even though Bender and Lascarides (2019, p.20) note that ‘[l]inguistic meaning includes so-

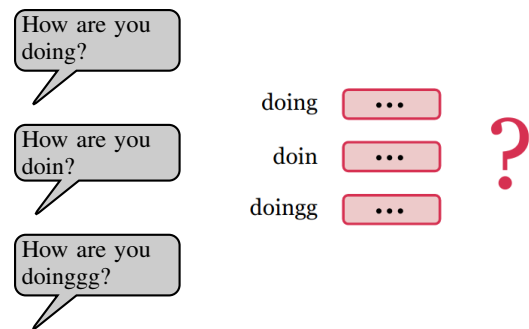


Figure 1: With all three utterances, the author asks how someone is doing, but the spelling variants carry different social meanings. For example, how should a spelling variant like *doin* be represented? Providing it the same representation as *doing* would result in a loss of social meaning associated with g-dropping.

cial meaning’, social meaning has been overlooked in the development of meaning representations, although a few recent studies have suggested that the embedding space can already exhibit patterning related to sociolinguistic variation, even when learning is based on text alone (e.g., Niu and Carpuat (2017); Nguyen and Grieve (2020); Shoemark et al. (2018)).

3.1 Example: Spelling variation

One clear example comes from spelling, where deviations from spelling conventions (e.g., *4ever*, *greattt*, *doin*) can create social meaning (Eisenstein, 2015; Herring and Zelenkauskaitė, 2009; Ilbury, 2020; Nini et al., 2020; Sebba, 2007). Androutopoulos (2000), for example, discusses how non-conventional spelling in media texts can convey social meanings of radicality or originality. Furthermore, a close textual analysis by Darics (2013) shows that letter repetition can create a relaxed style and signal ‘friendly intent’. An immediate question is therefore how to handle spelling variation when building representations (Figure 1).

Current research on representation learning that considers spelling variation is primarily motivated by making NLP systems more robust. For example, Piktus et al. (2019) modify the loss function to encourage the embeddings of misspelled words to be closer to the embeddings of the likely correct spelling. Similarly, motivated by ‘adversarial character perturbations’, Liu et al. (2020) aim to push embeddings closer together for original and perturbed words (e.g. due to swapping, substituting, deleting and inserting characters).

Although approaches to making models robust to spelling variation are useful for many applications, they necessarily result in the loss of the social meaning encoded by the spelling variants. Many of the operations (such as deleting characters) used to generate adversarial perturbations are also frequent in natural language data. In a recent study focused on a small set of selected types of spelling variation, such as g-dropping and lengthening, [Nguyen and Grieve \(2020\)](#) found that word embeddings encode patterns of spelling variation to some extent. Pushing representations of spelling variants together therefore resembles efforts to normalise texts, carrying the same risk of removing rich social information ([Eisenstein, 2013](#)).

3.2 Moving forward

So far, we have highlighted that linguistic forms (e.g., spellings, words, sentences) with different social meanings should not receive the same representation when social meaning is relevant to the task at hand. Drawing on Section 2, we now highlight key considerations for social meaning representations:

Social meaning can be attached to different types of linguistic forms Especially for evaluation, comparing representations for forms with the same referential meaning but potentially different social meanings would be the most controlled setting. However, in many cases this can be challenging. For example, paraphrases rarely have exactly the same referential meaning; to what extent we can relax this constraint remains an open question. Generally, it is easier to keep referential meaning constant when analysing spelling variation compared to other forms of variation. Spelling variation may thus be a good starting point but variation on other levels should also be considered.

Linguistic variation can index local identities

Research on linguistic variation in NLP has mainly focused on broad demographic categories (e.g., nation, sex, age) ([Nguyen et al., 2016](#)). These have often been modeled as discrete variables, although [Lynn et al. \(2017\)](#) show how treating variables as continuous can provide advantages. To represent the rich social meanings of linguistic variation, representations likely must be continuous and high dimensional. Moreover, rather than imposing static social attributes onto people, it may be more desirable to let highly localised social meanings emerge from the data itself (e.g., see [Bamman et al. \(2014b\)](#)).

Social meaning is highly contextual The same form can have different social meanings depending on context. Furthermore, variation can also occur at the semantic level ([Bamman et al., 2014a](#); [Del Tredici and Fernández, 2017](#); [Lucy and Bamman, 2021](#)). Contextual representations are therefore more suitable than static representations. Our proposed line of work also raises challenges about what should be considered context for learning representations. For learning social meaning, linguistic context alone is not sufficient. Instead, the social and communicative context in which utterances are produced must be considered as well.

4 Applications

Because the expression of social meaning is a fundamental part of language use, it should be taken into consideration throughout model development, but it is especially relevant for computational sociolinguistics ([Nguyen et al., 2016](#)) and computational social science ([Lazer et al., 2009](#); [Nguyen et al., 2020](#)). Examples where social meaning is especially important are:

Conversational systems Research on text generation has long recognised that the same message can be said in different ways, and that style choices depend on many factors, such as the conversation setting and the audience ([Hovy, 1990](#)). There is a large body of work on generating text in specific styles (e.g., [Edmonds and Hirst \(2002\)](#); [Ficler and Goldberg \(2017\)](#); [Mairesse and Walker \(2011\)](#)). An example are conversational systems that generate text in consistent speaker styles to model persona ([Li et al., 2016](#)). Rich representations of social meaning and linguistic variation could support the development of conversational systems that dynamically *adjust* their style depending on the context including the language used by interlocutors, constructing unique identities in real time, as individuals do in real world interactions ([Eckert, 2012](#)).

Abusive content detection Systems to automatically detect abusive content can contain racial biases ([Davidson et al., 2019](#); [Sap et al., 2019](#)). The task is challenging, because whether something is abusive (e.g., apparent racial slurs) depends strongly on context, such as previous posts in a conversation as well as properties of the author and audience. Considering social meaning and variation would facilitate the development of systems that are more adaptive towards the local social con-

text (going beyond developing systems for major demographic groups). This would more generally also be relevant to other tasks where interpretation is dependent on social context.

Exploration of sociolinguistic questions NLP methods can support (socio)linguistic research, e.g. methods to automatically identify words that have changed meaning (Hamilton et al., 2016) or words that exhibit geographical variation (Nguyen and Eisenstein, 2017). Likewise, if computational methods could discover forms with (likely) similar or different social meanings, these forms could then be investigated further in experimental perception studies or through qualitative analysis.

5 Challenges

5.1 Learning

Corpora such as Wikipedia and BookCorpus (Zhu et al., 2015) are often used to learn representations. However, it is likely that corpora with more colloquial language offer richer signals for learning social meaning. Text data may already allow models to pick up patterns associated with social meaning, as Bender and Lascarides (2019, p.20) note about social meaning that ‘*it is (partly) derivable from form*’. Social and communicative context can provide additional signals, for example by including information about author (Garimella et al., 2017; Li et al., 2018), geography (Bamman et al., 2014a; Cocos and Callison-Burch, 2017; Hovy and Purschke, 2018), social interaction (Li et al., 2016), or social network membership (Yang and Eisenstein, 2017). Furthermore, as argued by Bisk et al. (2020), static datasets have limitations for learning and testing NLP models on their capabilities related to the social nature of language. Instead, they argue for a ‘learning by participation’ approach, in which users interact freely with the system (Bisk et al., 2020). A key challenge is that although we know that social meaning is highly contextual, we would need to seek a balance between the richness and complexity of the context considered and computational, privacy and ethical constraints.

Another key challenge is that usually different aspects of meaning are encoded in one representation. Future work could potentially build on work on disentangling representations, such as work by Akama et al. (2018), Romanov et al. (2019) and recent work motivated by Two-Factor Semantics (Webson et al., 2020).

5.2 Evaluation

Although there are many datasets to evaluate NLP models on various linguistic phenomena (Warstadt et al., 2019, 2020; Wang et al., 2018), such datasets are missing for social meaning. Collecting evaluation data is challenging. First, relatively little is known about the link between social meaning and textual variation. Sociolinguistics has traditionally focused on the social meaning of phonetic features and to a lesser extent on grammatical and especially lexical features (Chambers, 2003). Social meaning making through spelling variation has received even less attention (exceptions include Leigh (2018)). Hence, research approaches would need to be (further) developed within sociolinguistics to allow for reliable measurement of social meanings of under-researched types of language variation such a spelling variation. One concrete avenue would be to extend and adapt traditional methods like the speaker evaluation paradigm, in which respondents indirectly evaluate accent variation, to be suitable for variation in written communication. Data generated by building on such approaches could then in turn serve as the basis for developing evaluation datasets for NLP models.

Second, collecting data is challenging due to the highly contextual nature of social meaning (Section 2). The same form can take on different social meanings and how a particular form is perceived depends on a variety of factors, including social and situational attributes of both the audience and the speaker or writer. However, carefully collected experimental data should at least be able to lay bear the social meanings that language users collectively associate with a certain linguistic form (i.e. its indexical field). This should give an overview of the social meaning potential language users have at their disposal to draw on in a specific situation.

6 Conclusion

Despite the large body of work on meaning representations in NLP, social meaning has been overlooked in the development of representations. Fully learning and representing the rich social meanings of linguistic variation will likely not be realised for years to come. Yet even small steps in this direction will already benefit a wide array of NLP applications and support new directions in social science research. With this paper, we hope to encourage researchers to work on this challenging but important aspect of linguistic meaning.

Ethical considerations

We will now discuss a few ethical considerations that are relevant to our proposed line of research. In this paper, we have discussed how language variation should be a key consideration when building and developing meaning representations. Labels such as ‘standard’, ‘bad’ and ‘noisy’ language used to describe language variation and practices can reproduce language ideologies (Blodgett et al., 2020; Eisenstein, 2013). As an example, non-standard spellings are sometimes labeled as ‘misspellings’, but in many cases they are deployed by users to communicate social meaning. A different term, such as ‘respellings’, may therefore be more appropriate (Tagg, 2009). Furthermore, even though there has been increasing attention to performance disparities in NLP systems and how to mitigate them, Blodgett et al. (2020) point out that they should be placed in the wider context of reproducing and reinforcing deep-rooted injustices. See Blodgett et al. (2020) for a discussion on different conceptualizations of ‘bias’ in NLP and the role of language variation.

Our paper also complements the discussion by Flek (2020). Recognising that language variation is inherent to language, Flek (2020) argues for personalised NLP systems to improve language understanding. The development of such systems, however, also introduces risks, such as stereotypical profiling and privacy concerns. See Flek (2020) for a discussion on ethical considerations for this line of work.

In this paper, we have argued for considering language variation and social meaning when building representations. However, such research could potentially also support the development of applications that can cause harm. Long-standing research in sociolinguistics has shown rich connections between language variation and social attributes, including sensitive attributes such as gender and ethnicity (e.g. Eckert (2012)). One may take that as a motivation to build automatic profiling systems. However, as discussed in Section 2, sociolinguists have emphasised the highly contextual nature of social meaning (the same linguistic feature can have different social meanings) and the agency of speakers (language is not just a reflection of someone’s identity, but can be actively used as a resource for identity construction). Profiling systems tend to impose categories on people based on broad stereotypical associations. They fail to recognise the rich

local identities and agency of individuals. Besides privacy concerns, misclassifications by such systems can cause severe harms.

Another ethical consideration is the training data. Data with colloquial language will likely offer richer signals for training, which could be augmented with information about the social and communicative context. Online sources such as Twitter and Reddit may be attractive given their size and availability of fine-grained social metadata. However, the use of large-scale online datasets (even though it is ‘public’) raises privacy and ethical concerns. We recommend following guidelines and discussions surrounding the use of online data in social media research—not only regarding collecting and storing data, but also how such data is shared, and how analyses based on such data are reported and disseminated (Fiesler and Proferes, 2018; Zook et al., 2017; Fiesler et al., 2020). One key step is documenting the datasets (Bender and Friedman, 2018; Gebru et al., 2018). In addition, social biases in these datasets can propagate into the learned representations (Bolukbasi et al., 2016; Caliskan et al., 2017), which may impact downstream applications that make use of these representations.

Acknowledgements

This work is part of the research programme Veni with project number VI.Veni.192.130, which is (partly) financed by the Dutch Research Council (NWO). We also like to thank Kees van Deemter for useful feedback.

References

- Reina Akama, Kento Watanabe, Sho Yokoi, Sosuke Kobayashi, and Kentaro Inui. 2018. [Unsupervised learning of style-sensitive word vectors](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 572–578, Melbourne, Australia.
- Jannis K. Androutsopoulos. 2000. [Non-standard spellings in media texts: The case of German fanzines](#). *Journal of Sociolinguistics*, 4(4):514–533.
- David Bamman, Chris Dyer, and Noah A. Smith. 2014a. [Distributed representations of geographically situated language](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 828–834, Baltimore, Maryland.
- David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. 2014b. [Gender identity and lexical varia-](#)

- tion in social media. *Journal of Sociolinguistics*, 18(2):135–160.
- Marco Baroni, Georgiana Dinu, and Germán Kruszewski. 2014. Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 238–247, Baltimore, Maryland.
- Emily M. Bender and Batya Friedman. 2018. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6:587–604.
- Emily M. Bender and Alex Lascarides. 2019. Linguistic fundamentals for natural language processing II: 100 essentials from semantics and pragmatics. *Synthesis Lectures on Human Language Technologies*, 12(3):1–268.
- Douglas Biber. 1988. *Variation across Speech and Writing*. Cambridge University Press.
- Douglas Biber and Susan Conrad. 2009. *Register, Genre, and Style*. Cambridge Textbooks in Linguistics. Cambridge University Press.
- Yonatan Bisk, Ari Holtzman, Jesse Thomason, Jacob Andreas, Yoshua Bengio, Joyce Chai, Mirella Lapata, Angeliki Lazaridou, Jonathan May, Aleksandr Nisnevich, Nicolas Pinto, and Joseph Turian. 2020. Experience grounds language. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8718–8735, Online.
- Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (technology) is power: A critical survey of “bias” in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5476, Online. Association for Computational Linguistics.
- Su Lin Blodgett, Lisa Green, and Brendan O'Connor. 2016. Demographic dialectal variation in social media: A case study of African-American English. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1119–1130, Austin, Texas.
- Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai. 2016. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In *Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16*, page 4356–4364.
- Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.
- Kathryn Campbell-Kibler. 2007. Accent, (ING), and the social logic of listener perceptions. *American Speech*, 82(1):32–64.
- Kathryn Campbell-Kibler. 2009. The nature of sociolinguistic perception. *Language Variation and Change*, 21(01):135–156.
- Kathryn Campbell-Kibler. 2010. The effect of speaker information on attitudes toward (ING). *Journal of Language and Social Psychology*, 29(2):214–223.
- Jack Chambers. 2003. *Sociolinguistic Theory: Linguistic Variation and Its Social Significance*. Oxford: Blackwell.
- Anne Cocos and Chris Callison-Burch. 2017. The language of place: Semantic value from geospatial context. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 99–104, Valencia, Spain.
- Alexis Conneau, German Kruszewski, Guillaume Lample, Loïc Barrault, and Marco Baroni. 2018. What you can cram into a single $\&\!#\&$ vector: Probing sentence embeddings for linguistic properties. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2126–2136, Melbourne, Australia.
- Erika Darics. 2013. Non-verbal signalling in digital discourse: The case of letter repetition. *Discourse, Context & Media*, 2(3):141–148.
- Thomas Davidson, Debasmita Bhattacharya, and Ingmar Weber. 2019. Racial bias in hate speech and abusive language detection datasets. In *Proceedings of the Third Workshop on Abusive Language Online*, pages 25–35, Florence, Italy.
- Marco Del Tredici and Raquel Fernández. 2017. Semantic variation in online communities of practice. In *IWCS 2017 - 12th International Conference on Computational Semantics - Long papers*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota.
- Penelope Eckert. 2008. Variation and the indexical field. *Journal of Sociolinguistics*, 12(4):453–476.
- Penelope Eckert. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, 41(1):87–100.
- Penelope Eckert and John R. Rickford, editors. 2002. *Style and Sociolinguistic Variation*. Cambridge University Press.

- Philip Edmonds and Graeme Hirst. 2002. Near-synonymy and lexical choice. *Computational linguistics*, 28(2):105–144.
- 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*, pages 359–369, Atlanta, Georgia.
- Jacob Eisenstein. 2015. [Systematic patterning in phonologically-motivated orthographic variation](#). *Journal of Sociolinguistics*, 19(2):161–188.
- Guy Emerson. 2020. [What are the goals of distributional semantics?](#) In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7436–7453, Online.
- Jessica Fidler and Yoav Goldberg. 2017. [Controlling linguistic style aspects in neural language generation](#). In *Proceedings of the Workshop on Stylistic Variation*, pages 94–104, Copenhagen, Denmark.
- Casey Fiesler, Nathan Beard, and Brian C. Keegan. 2020. [No robots, spiders, or scrapers: Legal and ethical regulation of data collection methods in social media terms of service](#). In *Proceedings of the International AAAI Conference on Web and Social Media*, pages 187–196.
- Casey Fiesler and Nicholas Proferes. 2018. [“Participant” perceptions of Twitter research ethics](#). *Social Media + Society*, 4(1).
- Edward Finegan and Douglas Biber. 2002. *Register variation and social dialect variation: the Register Axiom*, page 235–267. Cambridge University Press.
- Lucie Flek. 2020. [Returning the N to NLP: Towards contextually personalized classification models](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7828–7838, Online. Association for Computational Linguistics.
- Aparna Garimella, Carmen Banea, and Rada Mihalcea. 2017. [Demographic-aware word associations](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2285–2295, Copenhagen, Denmark.
- Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2018. [Datasheets for datasets](#). *CoRR*, abs/1803.09010.
- Anna Gladkova, Aleksandr Drozd, and Satoshi Matsuo. 2016. [Analogy-based detection of morphological and semantic relations with word embeddings: What works and what doesn’t](#). In *Proceedings of the NAACL-HLT SRW*, pages 47–54, San Diego, California, June 12–17, 2016.
- William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. [Diachronic word embeddings reveal statistical laws of semantic change](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany.
- 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*, pages 368–378, Portland, Oregon, USA.
- Susan C. Herring and Asta Zelenkauskaitė. 2009. [Symbolic capital in a virtual heterosexual market: Abbreviation and insertion in Italian iTV SMS](#). *Written Communication*, 26(1):5–31.
- Dirk Hovy. 2018. [The social and the neural network: How to make natural language processing about people again](#). In *Proceedings of the Second Workshop on Computational Modeling of People’s Opinions, Personality, and Emotions in Social Media*, pages 42–49, New Orleans, Louisiana, USA.
- Dirk Hovy and Christoph Purschke. 2018. [Capturing regional variation with distributed place representations and geographic retrofitting](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4383–4394, Brussels, Belgium.
- Dirk Hovy and Anders Søgaard. 2015. [Tagging performance correlates with author age](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 483–488, Beijing, China.
- Eduard H. Hovy. 1990. [Pragmatics and natural language generation](#). *Artificial Intelligence*, 43(2):153–197.
- Christian Ilbury. 2020. [“Sassy Queens”: Stylistic orthographic variation in Twitter and the enregisterment of AAVE](#). *Journal of Sociolinguistics*, 24(2):245–264.
- William Labov. 1972. *Sociolinguistic patterns*. 4. University of Pennsylvania Press.
- David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. [Computational social science](#). *Science*, 323(5915):721–723.
- Daisy Leigh. 2018. [Expecting a performance: Listener expectations of social meaning in social media](#). Presented at New Ways of Analyzing Variation (NWAY) 47.

- Chang Li, Aldo Porco, and Dan Goldwasser. 2018. [Structured representation learning for online debate stance prediction](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3728–3739, Santa Fe, New Mexico, USA.
- 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 (Volume 1: Long Papers)*, pages 994–1003, Berlin, Germany.
- Fei Liu, Fuliang Weng, Bingqing Wang, and Yang Liu. 2011. [Insertion, deletion, or substitution? Normalizing text messages without pre-categorization nor supervision](#). In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 71–76, Portland, Oregon, USA.
- Hui Liu, Yongzheng Zhang, Yipeng Wang, Zheng Lin, and Yige Chen. 2020. [Joint character-level word embedding and adversarial stability training to defend adversarial text](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8384–8391.
- Li Lucy and David Bamman. 2021. [Characterizing English variation across social media communities with BERT](#). In *To appear in TACL*.
- Veronica Lynn, Youngseo Son, Vivek Kulkarni, Niranjan Balasubramanian, and H. Andrew Schwartz. 2017. [Human centered NLP with user-factor adaptation](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1146–1155, Copenhagen, Denmark.
- François Mairesse and Marilyn A. Walker. 2011. [Controlling user perceptions of linguistic style: Trainable generation of personality traits](#). *Computational Linguistics*, 37(3):455–488.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. [Distributed representations of words and phrases and their compositionality](#). In *Advances in neural information processing systems*, pages 3111–3119.
- Dong Nguyen, A. Seza Doğruöz, Carolyn P. Rosé, and Franciska de Jong. 2016. [Computational sociolinguistics: A Survey](#). *Computational Linguistics*, 42(3):537–593.
- Dong Nguyen and Jacob Eisenstein. 2017. [A kernel independence test for geographical language variation](#). *Computational Linguistics*, 43(3):567–592.
- Dong Nguyen and Jack Grieve. 2020. [Do word embeddings capture spelling variation?](#) In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 870–881, Barcelona, Spain (Online).
- Dong Nguyen, Maria Liakata, Simon DeDeo, Jacob Eisenstein, David Mimno, Rebekah Tromble, and Jane Winters. 2020. [How we do things with words: Analyzing text as social and cultural data](#). *Frontiers in Artificial Intelligence*, 3:62.
- Andrea Nini, George Bailey, Diansheng Guo, and Jack Grieve. 2020. [The graphical representation of phonetic dialect features of the North of England on social media](#), pages 266–296. Edinburgh University Press, United Kingdom.
- Xing Niu and Marine Carpuat. 2017. [Discovering stylistic variations in distributional vector space models via lexical paraphrases](#). In *Proceedings of the Workshop on Stylistic Variation*, pages 20–27, Copenhagen, Denmark.
- Umashanthi Pavalanathan, Jim Fitzpatrick, Scott Kiesling, and Jacob Eisenstein. 2017. [A multidimensional lexicon for interpersonal stancetaking](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 884–895, Vancouver, Canada.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. [GloVe: Global vectors for word representation](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar.
- Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. [Deep contextualized word representations](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana.
- Aleksandra Piktus, Necati Bora Edizel, Piotr Bojanowski, Edouard Grave, Rui Ferreira, and Fabrizio Silvestri. 2019. [Misspelling oblivious word embeddings](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3226–3234, Minneapolis, Minnesota.
- Alexey Romanov, Anna Rumshisky, Anna Rogers, and David Donahue. 2019. [Adversarial decomposition of text representation](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 815–825, Minneapolis, Minnesota.
- Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. 2019. [The risk of racial bias in hate speech detection](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1668–1678, Florence, Italy.

- Mark Sebba. 2007. *Spelling and society: The culture and politics of orthography around the world*. Cambridge University Press.
- Philippa Shoemark, James Kirby, and Sharon Goldwater. 2018. [Inducing a lexicon of sociolinguistic variables from code-mixed text](#). In *Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text*, pages 1–6, Brussels, Belgium.
- Philippa Shoemark, Debnil Sur, Luke Shrimpton, Iain Murray, and Sharon Goldwater. 2017. [Aye or naw, whit dae ye hink? Scottish independence and linguistic identity on social media](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1239–1248, Valencia, Spain. Association for Computational Linguistics.
- Michael Silverstein. 2003. [Indexical order and the dialectics of sociolinguistic life](#). *Language & Communication*, 23(3):193 – 229. Words and Beyond: Linguistic and Semiotic Studies of Sociocultural Order.
- Ian Stewart, Yuval Pinter, and Jacob Eisenstein. 2018. [Si O no, que penses? Catalanian independence and linguistic identity on social media](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 136–141, New Orleans, Louisiana. Association for Computational Linguistics.
- Caroline Tagg. 2009. *A corpus linguistics study of SMS text messaging*. Ph.D. thesis, University of Birmingham.
- Sali A. Tagliamonte. 2015. *Making waves: The story of variationist sociolinguistics*. John Wiley & Sons.
- Rachael Tatman, Leo Stewart, Amandalynne Paullada, and Emma Spiro. 2017. [Non-lexical features encode political affiliation on Twitter](#). In *Proceedings of the Second Workshop on NLP and Computational Social Science*, pages 63–67, Vancouver, Canada. Association for Computational Linguistics.
- Sean Trott, Tiago Timponi Torrent, Nancy Chang, and Nathan Schneider. 2020. [\(Re\)construing meaning in NLP](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5170–5184, Online.
- Johanna Vaattovaara and Elizabeth Peterson. 2019. [Same old paska or new shit? On the stylistic boundaries and social meaning potentials of swearing loanwords in Finnish](#). *Ampersand*, 6:100057.
- Abby Walker, Christina García, Yomi Cortés, and Kathryn Campbell-Kibler. 2014. [Comparing social meanings across listener and speaker groups: The indexical field of Spanish /s/](#). *Language Variation and Change*, 26(02):169–189.
- Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. 2018. [GLUE: A multi-task benchmark and analysis platform for natural language understanding](#). In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 353–355, Brussels, Belgium.
- Alex Warstadt, Alicia Parrish, Haokun Liu, Anhad Mohananey, Wei Peng, Sheng-Fu Wang, and Samuel R. Bowman. 2020. [BLiMP: The benchmark of linguistic minimal pairs for English](#). *Transactions of the Association for Computational Linguistics*, 8:377–392.
- Alex Warstadt, Amanpreet Singh, and Samuel R. Bowman. 2019. [Neural network acceptability judgments](#). *Transactions of the Association for Computational Linguistics*, 7:625–641.
- Albert Webson, Zhizhong Chen, Carsten Eickhoff, and Ellie Pavlick. 2020. [Do “undocumented workers” == “illegal aliens”? Differentiating denotation and connotation in vector spaces](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4090–4105, Online.
- Yi Yang and Jacob Eisenstein. 2017. [Overcoming language variation in sentiment analysis with social attention](#). *Transactions of the Association for Computational Linguistics*, 5:295–307.
- Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. [Aligning books and movies: Towards story-like visual explanations by watching movies and reading books](#). In *The IEEE International Conference on Computer Vision (ICCV)*, pages 19–27.
- Matthew Zook, Solon Barocas, danah boyd, Kate Crawford, Emily Keller, Seeta Peña Gangadharan, Alyssa Goodman, Rachel Hollander, Barbara A. Koenig, Jacob Metcalf, Arvind Narayanan, Alondra Nelson, and Frank Pasquale. 2017. [Ten simple rules for responsible big data research](#). *PLOS Computational Biology*, 13(3):1–10.