Computational methods for sociolinguistic analysis in online discussions

The language that people use to communicate both reflects and constructs the society that they live in, both in explicit statements and implicit word choices. A person's everyday language choices can reflect where they were raised, how they accommodate to their audience, and their attitudes toward other people. These social factors have an especially strong impact on the internet, where people from a variety of backgrounds meet to share information and build social connections primarily via text communication. My work investigates the relationship between language style and social factors on the internet, using computational methods to quantify structural linguistic factors. The internet provides researchers with a bird's eye view of social interactions that is difficult to replicate in offline settings.

I focus on intuitive social factors that are understood to affect linguistic choice in everyday conversation, which include social attitudes, community dynamics, and audience expectations. My work uses natural language processing (NLP) and statistical analysis to explain the variation in language structure using these concrete social factors. Whereas prior work often investigates word frequency, I use a variety of NLP methods to characterize **structural** patterns, such as syntax, that would otherwise be ignored by typical approaches such as word frequency. For example, will a word that is more **syntactically flexible** (occurring in diverse contexts) outcompete other words in an online community? My work extends sociolinguistic theory to the context of the internet, a domain with rich linguistic diversity.

As with other work in computational social science, studying language use on the internet can extend existing social science theories and provide new methods to draw insight from large-scale text data. In my dissertation work, I have explored linguistic variation in the domains of political attitudes, online community norms, and audience expectations in public discussions of crisis events.

How do social attitudes affect a multilingual person's choice between languages in public discussions?

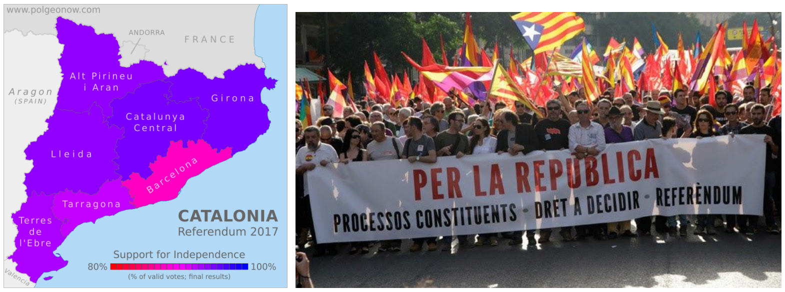


Fig. 1: High support for Catalonian independence during the 2017 referendum paralleled the use of Catalan slogans (*per la republica* "for the republic") in protests.[1]

A person's attitude toward a particular topic can result in consistent patterns in their **language choice**: in political discussions, the use of a minority language is often connected to attitudes about the status of the language's culture (Shoemark et al. 2017). In 2017, the region of Catalonia in Spain voted for

[1] https://www.polgeonow.com/2017/10/catalonia-independence-vote-2017-results-map-graph.html, https://www.rt.com/news/189072-catalonia-independence-consultation-vote/

independence (see Fig. 1), which ignited a national debate over the cultural identity of Catalonia and whether it deserved to be a separate country. Through an analysis of Twitter discussion of the independence vote, we found that bilingual activists who were pro-independence more often wrote in Catalan than Spanish, even in posts unrelated to independence discussion. This effect was stronger than a similar study of another independence referendum, which supports the idea that political identity is particularly strong in the use of Catalan and more generally that minority language use can reflect social attitudes. In follow-up work, we are investigating the influence of cultural affiliation, such as active consumption of American media, on the grammatical integration of loanwords in social media discussions. This kind of work can reveal how language reflects cultural differences among multilingual people, which is especially relevant to internet discussions where cultures clash frequently.

How well does a word's diversity of linguistic contexts predict its adoption in a community?

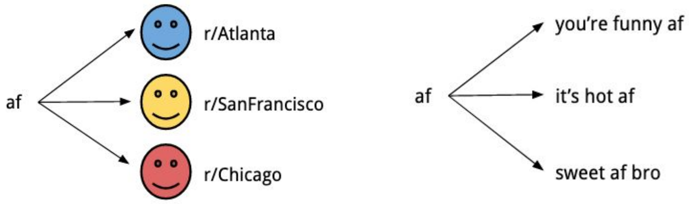
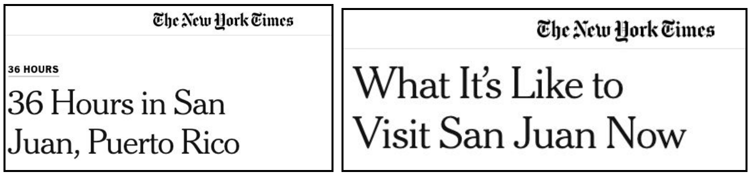


Fig. 2: A word may be socially disseminated (left) or linguistically disseminated (middle), and we find that linguistic context dissemination predicts word growth more readily than social dissemination.

In online communities, new words emerge frequently via interactions between different sub-groups and the arrival of new internet users who bring their own unique vocabulary: *haha* today can be *lol* tomorrow. While the social factors leading to word adoption are well-studied (Altmann et al. 2011), the linguistic factors remain poorly understood, such as whether a word that can apply to many contexts will be adopted more readily than a competitor word. In a study on Reddit, I investigated the role of **linguistic context** as a factor in word adoption, to test the hypothesis that the spread of a word among diverse syntactic contexts can predict the adoption of nonstandard words (e.g. if the new intensifier "af" "as fuck" can occur with a wide variety of adjectives). I developed a new metric to measure linguistic *dissemination* among different contexts and found that this metric consistently predicted word growth and decline, even when compared to the standard metric of social dissemination. In contrast to prior work in innovation diffusion that focuses on social metrics (Altmann et al. 2011), I demonstrated that the linguistic variation of nonstandard words is an important factor in the eventual adoption of nonstandard words, which should inspire further study in word adoption that tests other linguistic metrics such as topical diversity. More broadly, the study shows how variation in linguistic structure can provide insight into large-scale social dynamics in online communities.

When do audience expectations in public discussions lead people to use more descriptive information?



The New York Times

36 HOURS

36 Hours in San Juan, Puerto Rico

The New York Times

What It's Like to Visit San Juan Now

Fig. 3: Newspaper headlines that mention city “San Juan” before Hurricane Maria (2013; with descriptive information) and after Hurricane Maria (2017; no descriptor information).

When sharing information in discussion, people must determine how much *context* they need to provide for their audience. One type of linguistic context is **descriptive information** for location names, which may or may not be known to an audience: many Americans know about Puerto Rico, but they may not recognize its capital without additional description (see Fig. 3). Prior work supports a regular decrease over time in descriptive information for names in news coverage (Staliunaite et al. 2018), but it is unclear how much of that decrease is due to audience needs. To address this gap, I investigated how Twitter and Facebook users changed their use of descriptive information for location names during their discussion of the crisis events. I leveraged named entity recognition and dependency parsing to detect descriptive information, which captured the notion of descriptive information with high precision without sacrificing data diversity. I found that discussion participants decreased their use of descriptive information after the peak in collective attention, and locations that were local to a particular audience (e.g. mentioning “San Juan” to locals from the area) had fewer descriptions. This suggests that the discussion participants accommodate to their audience’s lower perceived need for information during the event, i.e. more collective attention paid toward affected locations leads to increased expectations of shared knowledge.

Future work: Detecting and evaluating linguistic polarization in online discussions

My previous work has focused on a variety of social factors that influence language style variation, including social attitudes, online community norms and audience expectations. In my future work, I will focus on developing linguistically-motivated metrics for social cohesion and division, and testing the ability of such metrics to generalize across domains. Developing such metrics will provide more accurate estimates of political division, which will guide interventions to address division in online discussion such as exposing online commenters to opposing opinions. Studying social cohesion and division will also lend itself to interdisciplinary collaborations with political and sociology researchers, who can benefit from expanding their typical methods toolkit from limited surveys to more open-domain text data.

How well can linguistically-motivated metrics for differences in opinions capture political polarization?

Political polarization is typically quantified as the split between social groups based on divergent beliefs, such as disagreements between Democrats and Republicans about policy. Online discussions about political issues often result in polarized opinions between different social groups, which can inhibit information sharing between groups. While prior computational work has quantified polarization using word count differences between predefined social groups (e.g. Demszky et al. (2019)), I am interested in developing linguistically-motivated metrics to quantify the level of polarization in online discourse. For example, knowing that Republicans tend to talk about guns more often than Democrats does not imply that their opinions are polarized, but observing a difference in the relative *valence* of gun-related discussion (positive vs. negative) can reveal polarization. Developing high-precision metrics will be useful in situations with relatively low word counts (i.e. high-variance), such as comparing individual speakers rather than large-scale political parties. I will leverage distributed representations of word and sentence semantics to measure differences in valence across discussion posts on news articles related to politics. I will experiment with several structured linguistic representations of expressed opinions using sentence structure, including restricting the scope to information connected to named entities ("Trump is wrong") and to concrete nouns ("abortion is wrong"). Ideally, such a metric will be able to determine the degree of difference between two texts even without being given *a priori* knowledge of what topics should be considered, which is an assumption made by stance detection.

How well do polarization metrics generalize across domains?

I plan to evaluate the utility of linguistically-motivated polarization metrics with both intrinsic and extrinsic comparisons to guarantee the generalizability of such metrics. Typical studies of large-scale polarization evaluate such metrics against expert judgment, which is often sparse and limited to well-understood domains such as American politics. In intrinsic evaluation, I plan to compare the estimated aggregate polarization against ground-truth data from voting: for a given newspaper in a state, the state's level of Republican versus Democrat support is available from voting records. Extrinsic evaluations will include predictive tasks, such as inferring whether a user will agree with another user's comment based on their prior computed degree of polarization, as well as descriptive tasks, such as comparing relative aggregate rates of polarization across more or less divisive topics (e.g. political elections versus sports games). The predictive tasks can also include community-level predictions such as whether a community will split into multiple sub-groups with differing opinions (e.g. when the subreddit r/News generated r/WorldNews).

As more diverse text corpora become available, quantifying polarization will become more important as a lens for understanding broad trends and changes in society. The metrics that my research develops can be extended to other domains related to social groups: the integration of immigrants into society can be better understood by comparing the relative alignment of immigrant-written texts with non-immigrant texts. Evaluating more linguistically-motivated metrics in collaboration with domain experts will encourage computational social scientists to leverage text data with a more critical lens, without relying on word frequency alone.

References

Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2011). Niche as a determinant of word fate in online groups. *PloS one*, *6*(5).

Demszky, D., Garg, N., Voigt, R., Zou, J., Shapiro, J., Gentzkow, M., & Jurafsky, D. (2019). Analyzing polarization in social media: method and application to tweets on 21 mass shootings. In *NAACL*.

Shoemark, P., Sur, D., Shrimpton, L., Murray, I., & Goldwater, S. (2017). Aye or naw, whit dae ye hink? Scottish independence and linguistic identity on social media. In *EACL*.

Stewart, I., & Eisenstein, J. (2018). Making "fetch" happen: The influence of social and linguistic context on nonstandard word growth and decline. In *EMNLP*.

Stewart, I., Pinter, Y., & Eisenstein, J. (2018). Sí o no, ¿què penses? Catalonian independence and linguistic identity on social media. In *NAACL*.

Stewart, I., Yang, Y., & Eisenstein J. (2019). Characterizing collective attention via descriptor context in public discussions of crisis events. In submission.

How do multilingual speakers modulate their language use in political discourse?

In my future work, I will continue the work in language structure variation among multilingual speakers by examining public discussions in political discourse in multilingual countries. Whereas previous work has focused on language choice (e.g. code-switching), I will focus on word choice and structure among multilingual speakers. I outline several potential projects below.

How do implicit cultural attitudes affect language choice in public discussions?

In future work, I will investigate the expression of implicit social attitudes through indirect language expressions in online discussions. Implicit social attitudes are an important aspect of everyday communication in how they reflect broader differences between social groups (Greenwald & Banaji, 1995). Even a simple phrase such as “OK boomer” can reveal a sharp generational divide between young and old (“boomer”) generations in a way that typical surveys struggle to replicate. By analyzing such language choices at a large scale, researchers can identify subtle variation in affect expressed toward social groups that can correlate with concrete consequences, including polarization. While extreme attitudes (e.g. hate speech, racism) are known to affect language choices in predictable ways (CITE), it remains to be seen whether fine-grained attitudes such as the generation gap can affect language choices in public online discussions. My future work will investigate the expression of implicit attitudes in language use and evaluate semantically-aware methods of detecting such attitudes, with the goal of augmenting more traditional opinion surveys.

Implicit attitudes in humor

While implicit attitudes may be clear in direct statements (“American food is great”), people may also make their attitudes clear through humor (“American food is great, if you hate flavor”). I will begin the study of implicit attitudes through a survey of attitudes and humor

Extending attitude-based surveys

Typical psychological studies of implicit attitudes rely on either controlled experiments (e.g. priming) or surveys of self-reported beliefs.

Effect of implicit attitudes on multilingual discussions

In my future work, I will test the ability of NLP methods to extract implicit cultural attitudes from online discussions through the lens of word choice, with the goal of improving typical surveys of cultural attitudes. By focusing on attitudes toward prominent social groups such as immigrants (Suro 2005), sociologists and political scientists can uncover insight about implicit attitudes at a larger scale and with more fine-grained nuance than previously available. I will collaborate with domain experts to identify domains worthy of analysis and to clarify the type of insight about attitudes that can be extracted with NLP techniques (e.g. subtle semantic cues available from contextual word embeddings). Beyond social media, the methods developed in future work can be extended to investigate implicit attitudes in more “official” formats including political speeches and newspaper editorial articles, which often have a broad social impact (e.g. how issues are framed in news coverage).

When studying language use, computational social science researchers often use surface-level signals such as word frequency to make claims about social processes. My work demonstrates that structural aspects of language, such as word syntax, play an important role in the linguistic correlates of social processes on the internet.

References

Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2011). Niche as a determinant of word fate in online groups. *PLoS ONE*, *6*(5), 1–12.

Greenwald, A., & Banaji, M. (1995). Implicit Social Cognition. *Psychological Review*, *102*(1), 4-27.

Labov, W. (2001). *Principles of linguistic change, Volume 2: Social Factors*. Malden, MA: Blackwell Publishers.

Shoemark, P., Sur, D., Shrimpton, L., Murray, I., & Goldwater, S. (2017). Aye or naw, whit dae ye hink? Scottish independence and linguistic identity on social media. In *EACL 2017* (Vol. 1, pp. 1239–1248).

Staliunaite, I., Rohde, H., Webber, B., & Louis, A. (2018). Getting to “Hearer-old”: Charting Referring Expressions Across Time. In *EMNLP* (pp. 4350–4359).

Suro, R. (2005). *Attitudes toward immigrants and immigration policy: Surveys among Latinos in the US and Mexico*. Washington: Pew Hispanic Center.

OLD

Assessing the social conventions of multilingual people in online discussions

- most of world’s population speaks at least 2 languages, but the default setting for most webpages is English
- some websites have tried to be accessible with e.g. automatic translation, user-provided translations
- how do multilingual social media users navigate discussions in which their primary language is not the discussion majority?
- **language learners**: do multilingual people adapt their use of L2 based on the assumed language skills of their conversation partners? (e.g. using more simple vocabulary with other multilingual speakers, trying out new words with monolingual speakers) how do early language learners rely on NLP tech (translations, pronunciation generators, auto-correct) as compared to late language learners?
- **attitudes toward automatic translations**: translations provided by platform, by other services; connection to social perception of “translation-ese” (Rabinovich et al. 2015); do multilingual speakers use these translations for their L1, and do the translations affect their responses (e.g. accommodation)?
- **adoption of language-specific hashtags**: how do multilingual speakers adapt popular hashtags from e.g. political campaigns to be more understandable? if they don’t adapt the hashtags, do they assume that the hashtags have a different meaning based on context?
- **navigating nonstandard language**: how do multilingual people approach nonstandard language from L1? word translation, specialized dictionary? do they need in-line translation or explanation, or would they prefer to receive feedback from conversation participants?
- **impact**: design socially-aware affordances for multilingual people, e.g. browser plug-ins for word lookups, feedback for automatic translations, writing style suggestions

My prior research tested social theory in the context of online discussions, using NLP techniques to detect everyday linguistic variation. One area that deserves more attention in the study of language variation is the space of multilingual conversations among second-language learners: Do multilingual speakers learn to pick up the semantic nuances of slang for their non-native language? Does a multilingual speaker's perception of another language as "cool" influence their likelihood of full acquisition of the language? My future work will advance this agenda in two ways: measuring language attitudes among bilingual speakers, evaluating the effect of such attitudes on large-scale behavior, and providing insight for ...

What can large-scale analysis of linguistic choices tell us about social attitudes and behavior in online discussion?

- the style of language that people choose can reflect their social goals and expectations
    - ex. do you say "haha" or "lol"? how does that meet your audience expectations? does it express your attitude toward other people? do you adapt to new conventions as they emerge?
- language is arbitrary! cues chosen may not directly reflect meaning but can be co-opted to express social signal
- traditional studies study language variation in spoken context (e.g. interviews): pronunciation and word choice
    - focus on dynamics within speech communities: in-group vs. out-group identities, navigating relationships, accommodating to group norms
- large scale web data lets us investigate more complicated phenomena such as lexical competition, social dimensions of semantics
    - online communities permit birds-eye view into individuals' behavior and linguistic cues, and large-scale summaries of unexpected trends
- **how can we apply NLP to extract social insight from lexical choice?**
    - (1) NLP can help extract fine-grained phenomena that would otherwise require manual examination: e.g. competition between words => detected with word embedding similarity and PMI; detecting verb/adjective use of loanwords
    - (2) NLP can address confounds that would invalidate other kinds of study, e.g. matching authors based on sentence context ("haha that's funny" vs. "lol that's funny")
    - (3) NLP can provide insight into speaker metadata: e.g. uncovering latent speaker attitudes (third-wave) that influence word choice; more fine-grained speaker demographics (e.g. gender identity)
    - (4) NLP is better at handling multilingual and multi-dialectal scenarios (without requiring direct insider knowledge)
    - pick 2-3 representative studies
- **how can we enable linguists to get the most out of NLP techniques?**
    - don't want "dual mule" problem: linguists label, programmers analyze
    - ideally: back-and-forth conversation about underlying concept that linguists want to better understand, data scientists want to analyze (e.g. Reddit stance work)
    -

- future work: big, broad
- language contains content, social ideas
- not enough work on social side -> need to separate content/social!
- systemic functional linguistics
- why do we want to study society? what is impact?
- compact research statement: include in first paragraph

Assessing cultural fit in communities of practice

- increasingly people turn to online communities for social support and advice
- many communities have their own norms and expectations for members, including explicit rules of conduct as well as implicit understanding
- members who do not fit into the communities may feel discouraged and eventually leave community
- members who fit in very well may become part of group that changes community for the better or worse
- much research on this issue has focused on predicting “user churn”, which is just one outcome of a more complicated process of group socialization
- **what constitutes a meaningful cultural fit for members of online communities?**
    - (1) in typical organizations, managers and leaders are expected to be aware of the organization members’ ability to fit in. do moderators know how to recognize members who fit in well, and if not, can NLP methods help them find such members?
    - (2) some communities have a very concrete notion of “culture” including well-defined traditions, commonly-understood vocabulary, and collective memory of specific events and people. how much of cultural “fit” requires matching the community’s *style* (e.g. shorthand conventions) as compared to the community’s *content* (e.g. ideological opinions)? what NLP methods are needed to separate style from content?
    - (3) many online communities are based in offline social organizations (e.g. r/Atlanta based on the city). how closely does the online community’s culture mirror the culture of the offline organization, and does this impact the overall well-being of the community?
    - (4) we know that more successful newcomers tend to be those that accommodate to community’s linguistic norms. how **aware** are newcomers and old-timers of linguistic norms? do they make explicit comments or ask questions about norms, or do they subconsciously “absorb” norms?
    - (5) instead of leaving newcomers to their own devices, some regulars take on the duty of acculturating the newcomers. do regulars tend to target newcomers who are struggling to fit in, and if so how are they treated differently than newcomers who may have prior experience with community norms (e.g. Linux pro joins PC community vs. computer newbie)
    - (6) cultural fit is expressed not only in language style but also in the expressed attitude toward sub-topics: e.g. members of T_D express negative attitudes toward liberals. how important is ideological expression as a marker of cultural fit for newcomers, and do regular members feel less pressure to show such ideology as they become established members?
- important next steps: verifying measures of cultural fit with perceptions (from community members and non-members), providing insight for moderators to make interventions, considering differences in community type (e.g. political culture may require more ideological expression than tech culture)

Text-as-data for understanding implicit societal norms

- in any given society, people navigate daily interactions using a combination of explicit and implicit social norms
    - explicit norms encoded in laws like “don’t murder”
    - implicit norms encoded in conventions like “greet someone when you first see them”
- often, implicit norms must be taught through observation or intervention
- implicit norms can come as a surprise (“culture shock”) to immigrants and tourists
- violating implicit norms can carry heavy social cost (e.g. loss of respect) and following such norms can provide unexpected benefits (e.g. insider knowledge)
- political norms may be reflected in non-political language use
- **how readily can we extract concrete, implicit social norms from large-scale text data (news media, travel guides, social media)?**
    - (1) students in foreign language courses often read simplified narratives to learn about life in X country. e.g. “in the morning, John eats a croissant for breakfast”. how readily can NLP generalize social norms from typical narratives, and do they match reader expectations?
    - (2) implicit norms are often invisible until they are broken, e.g. accidentally being rude to a stranger. how do people react on social media to breaking such norms (e.g. Reddit advice “am I the asshole”), and can these experiences be summarized to help others avoid breaking the norms?
    - (3) some implicit norms require repeated enforcement to be established as a part of everyday interaction, e.g. asking for a person’s pronouns when meeting them. within online communities, what kinds of people repeatedly push to establish certain norms, and what kinds of people explicitly comment on such norms?
- **impact**: providing advice to socially unaware people, comparing implicit norms across cultures and time, finding insight for future explicit norms (e.g. implicit norm of respecting gender identity later codified as law)