

The Climate Change Debate and Natural Language Processing

Manfred Stede

Applied Computational Linguistics
University of Potsdam
Potsdam/Germany
stede@uni-potsdam.de

Ronny Patz

Hertie School
Berlin/Germany
patz@hertie-school.de

Abstract

The debate around climate change (CC)—its extent, its causes, and the necessary responses—is intense and of global importance. Yet, in the natural language processing (NLP) community, this domain has so far received little attention. In contrast, it is of enormous prominence in various social science disciplines, and some of that work follows the “text-as-data” paradigm, seeking to employ quantitative methods for analyzing large amounts of CC-related text. Other research is qualitative in nature and studies details, nuances, actors, and motivations within CC discourses. Coming from both NLP and Political Science, and reviewing key works in both disciplines, we discuss how social science approaches to CC debates can inform advances in text-mining/NLP, and how, in return, NLP can support policy-makers and activists in making sense of large-scale and complex CC discourses across multiple genres, channels, topics, and communities. This is paramount for their ability to make rapid and meaningful impact on the discourse, and for shaping the necessary policy change.

1 Introduction

Anthropogenic climate change (CC) has become a central topic of global, national, and local debates across multiple arenas and channels that involve virtually all branches of society. From private talk to public social media exchanges, from scientific papers to journalistic articles in traditional mass media, from statements by stakeholders (industry, civil society groups, etc.) to political deliberations in national parliaments or in international organizations—no sphere is without references to climate change. While climate scientists have reached a consensus that climate change is real, that it is caused by human activity on the planet, and that it has and will have adverse effects

for humanity and the biosphere around the planet (Cook et al., 2016), public debates on CC and on the policy implications remain highly controversial (see, e.g., (Hulme, 2009)).

Natural Language Processing (NLP) is well-positioned to help study the dynamics of the large-scale and complex discourse on CC. Activists and policy-makers need NLP tools through which they can filter, order, and make sense of the vast amount of textual data produced on CC. However, within the NLP community, the amount of work done so far on CC remains limited. In the words of Luo et al. (2020, p. 3296), the topic of climate change “has received little attention in NLP despite its real world urgency”. This is in contrast to the attention that CC discourses receive in climate and environmental science and in various social sciences.

We argue in this paper that the research questions, insights and methods applied in these disciplines can provide useful orientation for NLP practitioners. And conversely, the general advances in NLP can provide more reliable and valid tools to actors aiming at shaping policy and influencing individual behavior. Such tools for monitoring the discourses across the multitude of channels, genres, speakers, and topics can enable policy-makers and activists to more rapidly respond to discourse shifts, which is of huge importance given the speed of the ongoing climate change.

To set the stage, in Section 2, we explain what we mean by CC “discourses” and we delineate the different readings of the term. Next, Section 3 takes the viewpoint of the NLP community and summarizes work that has been done in the field so far. Section 4 describes key studies taken from the social science literature, which study CC discourse in different ways and to different ends. Our emphasis here is on the methodological choices that are being made. Section 5 provides a comparative analysis and proposes points of synergy that we regard as

recommendations for NLP work. Our conclusions on the potential positive impact of NLP for making sense of the CC debate are presented in Section 6.

2 Climate Change "Discourses"

The term *discourse* is both polysemous and vague. In NLP and its branch of 'discourse processing', its default reading refers to a single text or a single dialogical interaction that becomes an object of study, involving phenomena that cross sentence boundaries (anaphoric reference, coherence relations, and so on). That reading is largely irrelevant for our purposes here.

In the social sciences, theories and definitions of *discourse(s)* and methods of *discourse analysis* are highly diverse. In the context of environmental policy, [Hajer and Versteeg \(2005, p. 175\)](#) define a discourse as the "ensemble of ideas, concepts and categories through which meaning is given to social and physical phenomena, and which is produced and reproduced through an identifiable set of practices". Thus, when we refer to the *climate change discourse*, we refer to the ensemble of practices of writing about or debating CC-related matters by one or multiple actors in various physical or digital arenas.

In much of the empirical literature on CC debates that we review below, this results in a focus on one of two dimensions of discourse:

- Discourse₁: Focus on exchanges on different technical media ("channels") and in different genres:
 - Traditional news media
 - Social media
 - Scientific exchange
 - Parliamentary debate
 - ...
- Discourse₂: Focus on social communities engaged in the topic-specific interaction, possibly using multiple channels (but studies often focus on single channels):
 - Grouped by role in the social constellation:
 - * Politicians
 - * Scientists
 - * Industrial stakeholders
 - * Interest groups (environmental, etc.)
 - * Individuals

- * ...
- Grouped by stance toward the topic:
 - * Climate change believers/accepters
 - * Climate change sceptics/deniers
 - * ...

Once one zooms in on the stances on CC more closely, further dimensions of Discourse₂ become visible. For example, [Anshelm and Hultman \(2015\)](#) develop a more fine-grained stance classification distinguishing between "industrial fatalism", "Green Keynesian", "eco-socialist" and "climate-sceptic" discourses.¹

Whether studies on CC detect a divided debate or a relatively unified conversation ([Wetts, 2020](#)) will depend on the types of discourse dimensions studied as well as on the level of analysis. This should be important also for NLP practitioners when they select a set of data for their work, as certain differences in nuances on stances may remain inconsequential in a social media debate between individuals, but can have significant policy implications when uttered by political leaders in a parliamentary debate.

3 CC discourse: Research in the NLP community

The difference between this and the following section is one of scientific community: In the present section, we briefly summarize work that has been done on CC-related data and was presented at NLP/Computational Linguistics or AI meetings. The number of such publications is small, so we mention them here in chronological order. Henceforth, we use lowercased "cc" and "gw" as shorthand for "climate change" and "global warming", respectively, as a search bigram employed by researchers for retrieving their data.

[Diakopoulos et al. \(2014\)](#) crawled 1.5 mio posts from 3,000 blogs, found by the query term cc or one from a short list of other terms, and manually coded a selection of blogs as belonging to sceptic or accepter discourse. 133 topical terms of CC discourse are taken from previous work, and for each term, correlations with "virtue" and "vice" words

¹"Industrial fatalism": apocalyptic scenarios are to be answered by technological solutions; "green Keynesian": CC is but one symptom of an institutional ecological crisis that requires redistribution of global resources; "eco-socialist": CC is a result of the pathological growth ideology of industrial capitalist society; "climate-sceptic": emissions with anthropogenic causes are not responsible for climate change, and no huge interventions are necessary.

(from the General Inquirer lexicon) are computed for both groups of blogs. Then visual analytics are applied to manually compare the discourses. Differences between blogs are found to be mainly in the framing of "climate science" and "quality of life". In continuation of this work, [Salway et al. \(2016\)](#) built a corpus of CC blog posts in three languages. They applied network analysis to the graph of blog linkages and detected four prominent communities of bloggers.

The CC topic became more visible in the NLP community when ([Mohammad et al., 2016](#)) introduced the new SemEval task "stance detection of tweets", where "Climate change is a real concern" was one of five statements for which a dataset was built. Beyond this, however, CC was not addressed in any more specific way.

[Pathak et al. \(2017\)](#) collected tweets around the 2015 UN CC conference in Paris, using about 20 search keywords and a similar number of hashtags, as well as three Twitter accounts dedicated to the conference. Term lists for CC subtopics are constructed by extending seed words with similar words gathered by a word2vec model. Then, opinion and emotion analysis tools are applied. Results are plotted in particular for correlations of emotions and topics and the role of "influencers" versus less prominent accounts.

[Jiang et al. \(2017\)](#) gathered 11,000 newspaper articles from four British broadsheets over the years 2007-2016. The search criterion was that cc has to occur at least three times. They use LDA to find sentiment targets in the texts, and by employing SentiWordNet to label keywords in the associated topics, they found some differences between newspapers in their topic-sentiment association.

Recently, [Luo et al. \(2020\)](#) were the first to apply a broad range of current NLP techniques to the CC domain. They introduce a corpus of 2,000 CC sentences from 63 US news sources (2000-2020), which were labeled by crowdworkers for stance toward "climate change is a real concern" (cf. ([Mohammad et al., 2016](#)) above). The base corpus of 56,000 articles was built with four bigram and two unigram query terms. Dependency parsing and coreference resolution are applied to enable extraction of opinion statements using a set of hand-coded patterns. These statements allow to distinguish self-affirming versus opponent-doubting frames in quoting sources of information. A BERT model is employed for stance classification, allow-

ing to identify acceptor and sceptic media.

Recently, [Koenecke and Feliu-Fabà \(2020\)](#) study whether CC sentiment in tweets changed in response to five natural disasters occurring in the US in 2018. Tweets had to contain one of the terms cc or gw, plus at least one instance of a set of natural-disaster terms. This yielded 800 pre-event and 6,000 post-event tweets. An array of standard ML tools were tested for classifying acceptor versus sceptic tweets. RNNs with GloVe embeddings performed best, yielding an accuracy of 75%. A cohort-level analysis then shows that the 2018 hurricanes yielded a statistically significant increase in average tweet sentiment affirming CC, while other disasters did not.

Summary In the absence of any "standard CC dataset", the NLP research so far has been scattered. Types of target texts (Discourse₁) were limited to news ([Jiang et al., 2017](#); [Luo et al., 2020](#)), blogs ([Diakopoulos et al., 2014](#); [Salway et al., 2016](#)) and Twitter ([Pathak et al., 2017](#); [Koenecke and Feliu-Fabà, 2020](#)); no comparisons across genres or channels were made, and there was no attention on political arenas or on statements by individuals and interest groups that are meant to directly influence policy-making. In terms of methods and goals we found network analysis for detecting communities ([Salway et al., 2016](#); [Pathak et al., 2017](#)), sentiment/stance classification for Discourse₂ grouping ([Diakopoulos et al., 2014](#); [Pathak et al., 2017](#); [Jiang et al., 2017](#); [Luo et al., 2020](#); [Koenecke and Feliu-Fabà, 2020](#)), topic modeling for computing topic/sentiment correlations ([Jiang et al., 2017](#)), and fine-grained framing distinction ([Luo et al., 2020](#)).

4 CC discourse: Research in the social sciences

In the following we provide a synthesis of a subjective selection of papers from journals in communication science, political science, and climate/environmental science that address CC discourse. All selected contributions take a "text-as-data" approach ([Grimmer and Stewart, 2013](#)) and use either semi-automatic methods such as corpus-linguistic collocation analysis or fully-automatic text mining methods. The papers we chose are either frequently cited or representative for widespread methodological approaches; a few are selected because they are innovative, either in terms

of method or in terms of the text genre(s) being addressed.

We group the discussion along the targeted text genres or media (i.e., our Discourse₁ dimension), to illustrate the range of underlying social science research questions and the data used to answer them. Then, in the second subsection, we summarize and assess the methods used, and we close the section with remarks on the relation between qualitative and quantitative research.

4.1 Genres and research questions

News media News text has for a long time been a highly prominent object of study in quantitative text analysis in the social sciences. In an early paper on CC, [Trumbo \(1996\)](#) determined how much coverage the topic received in 5 US newspapers, and he manually coded texts for using frames in the sense of [Entman \(1993\)](#) (see Sect. 5). Frames were also studied intensively by [Hoffman \(2011\)](#), who hand-coded 800 newspaper op-eds for (i) overall stance (convinced, sceptical, neutral, unclear); (ii) topical frame categories (science, risk, technology, economics, religion, political ideology, national security); and (iii) whether arguments used diagnostic, prognostic or motivational frames ([Entman, 1993](#)). Findings included that in the press, acceptor articles usually come from journalists, while sceptical texts tend to be letters to the editor. Yet another conception of frames was recently used by [Stecula and Merkley \(2019\)](#) who employed supervised classification to obtain 14,000 articles on the CC topic. The authors found that frames of "economic decline as a result of mitigating CC" are on the decline, and that frames highlighting scientific uncertainty (rather than CC consensus) are in sharp decline.

A different question was investigated by [Boykoff and Boykoff \(2007\)](#), who studied CC coverage on TV and in newspapers to determine whether adherence to the "journalistic norms" of personalization, drama, novelty, authority-order and balance contributed to impediments in covering anthropogenic CC. They found that the goals of balance and drama lead to fringe scientists getting more attention than would be proportionally warranted.

A different, in some sense more "modest", line of work is interested in the amount of coverage of CC in the press, and possible correlations with important events. [Lyytimäki and Tapio \(2009\)](#) studied 4,000 texts from the Finnish press, with man-

ual coding of topical relevance following an automatic retrieval. Other work in this vein added the aspect of cross-country comparison: [Grundmann and Krishnamurthy \(2010\)](#), for example, worked with newspapers from four countries. Besides comparing attention to CC across the countries, they offered observations on the basis of word frequencies and collocation lists. [Schmidt et al. \(2013\)](#) extended the comparison of attention to an impressive list of 27 countries with a corpus spanning 15 years. In contrast, [O'Neill et al. \(2015\)](#) focused specifically on the coverage of newly-released IPCC reports in newspapers, and also on TV and in Twitter. Studying the frames used in reporting about specific IPCC working groups, the authors proposed some recommendations on how to communicate particular kinds of information in future climate science reports.

Topic modeling is generally a popular tool in "text-as-data" research. Applying it to a corpus of 78,000 CC articles from 52 US newspapers, [Bohr \(2020\)](#) identified 28 themes related to climate change, whose prevalence (according to his interpretation) partly depends on the political orientation of the respective editorial boards.

Social media Key questions in research on CC discourse in social media concern how discursive networks and "discursive landscapes" ([Schoenfeld et al., 2018](#)) form, and what drives the polarization in CC debates. For example, [Elgesem et al. \(2014\)](#) aimed to "chart the entire structure of the climate change blogosphere". They crawled 1.3 mio posts from 3,000 blogs and ran community detection algorithms. Blogs were manually classified as sceptic, acceptor, or neutral; after running LDA, certain associations between blogger subcommunities and topics were found. Similarly, [Pérez-González \(2020\)](#) used concordance and visualization tools on 450,000 tokens from five blogs and show that terms such as "bias", "dogma" or "peer review" are framed with different motifs depending on the bloggers ideological orientation.

Many studies are performed on Twitter data. As an example of a largely descriptive analysis, [Dahal et al. \(2019\)](#) collected 360,000 tweets with five CC-related bigrams, and plotted distributions over topics (via LDA), countries and time. [Veltri and Atanasova \(2015\)](#) collected 60,000 tweets representing a random week (using the bigrams cc and gw), built cooccurrence networks over weighted terms and used centrality measures to determine

the salient topics. Further, using an emotion lexicon revealed that emotionally arousing text was more likely to be shared. [Samantray and Pin \(2019\)](#) worked with 14 mio Tweets from 3.5 mio users, written over 10 years (also found with the bigrams cc, gw). They classified Tweets and users for stance believer/denier/neutral, and with sentiment and emotion lexicons they computed correlations between polarizing language and the degree of interaction between people with similar versus antagonistic viewpoints.

Parliamentary debate and political speech

Though the amount of available data from CC-dedicated political debate is small, the research perspectives taken here show that attention to different genres is crucial for moving beyond the foci on measuring coverage and polarization. For instance, by working with a speech corpus of 100,000 words from the UK parliament's debate on the 2008 Climate Change Bill, [Willis \(2017\)](#) found that climate change is presented through "strongly scientific, technical and economic language", and he thus derived a tendency to de-politicize CC in parliament, and to frame it as a technical issue that is amenable to straightforward policy action.

More advanced research questions at the intersection of social science and linguistics also come with somewhat more elaborate computational methods. [Majdik \(2019\)](#) worked with US congressional records from 1994 to 2016 and retrieved 30,000 instances of speech mentioning cc or gw. After POS tagging and extracting bigrams, regular expressions are employed to analyze the context of selected combinations of cc/gw and verbs, which lead to a comparison of "active-agentive" to "passive-agentive" mentions in the speeches. On a related genre, [Calderwood, 2020](#) took a random sample of presidential speeches, ranging from Georg H. Bush to Obama, querying with "climat*" and "warm*". One resulting observation showed certain patterns of invoking CC when the speech is given in specific geographical locations.

Institutional text and reports Documents from specific institutions play an important role for many social science research questions. When [Barke-meyer et al. \(2016\)](#) compared the "summary for policymakers" of IPCC reports to other scientific communication, they found that the summaries have a low readability and differ notably in terms of "optimism scores" as derived with a sentiment

dictionary. Other types of documents reveal a shift in the CC discourse from prevention to mitigation: [Jaworska \(2018\)](#) studied corporate social responsibility and environmental reports that were produced by major oil companies from 2000 to 2013. Using corpus-linguistic tools she found a trend toward highlighting the risks of CC. This suggests that future research may find a new divide, not between deniers and accepters but between the attitudes "we can do something" and "CC is an unpredictable risk". A different trend was found by [Wetts \(2020\)](#) in a corpus of 1,700 institutional press releases (1985 to 2013). With topic modeling and cluster analysis she found the discourse among interest groups to become "post-political", i.e., less polarized, over time.

Looking specifically at CC denial, [Boussalis and Coan \(2016\)](#) used LDA on 16,000 documents from 19 organizations to find typical topics that contrarian actors link to CC. Going a significant step further, [Farrell \(2019\)](#) turned to intentional misinformation. Using the Stanford NER system he detected 28,000 different names of individuals and organizations connected to the American "Philanthropy Roundtable" organization (in magazines, almanacs and other online sources). Similarly he built a list of people known to be associated with deliberate misinformation, and then he computed the intersection with an approximate string matching algorithm.

Other genres Finally, we mention two examples of work on corpora from other sources. [Hulme et al. \(2018\)](#) built a CC subcorpus of the editorials of the *Nature* and *Science* journals, ranging from the mid 1960s to 2017. Eight frame categories, similar to those mentioned above for [\(Hoffman, 2011\)](#), were manually assigned to the texts. Observing the shifting frames over time and the differences between Europe and North America underscores that scientific communication around the CC discourse is not homogeneous and deserves continued attention.

Citizens' voices on CC can be found not only on social media. [Devaney et al. \(2020\)](#) compiled a small corpus of 1,885 citizen submissions to the Irish Citizens' Assembly on climate change. Combining LDA with a qualitative analysis of a 10 per cent sample, they drew lessons "for enhancing environmental literacy by improving climate crisis communication and engagement strategies". Beyond the polarization question, the submissions show what issues citizens care about when they

talk about climate change—which in turn can advise policy-makers in shaping policy solutions.

4.2 Methods applied

We briefly summarize the text mining/NLP methods that have been used in the work mentioned above (and in some other social science research), vaguely in the order of increasing complexity or sophistication.

- perform bigram matching for finding texts about climate change (often just the two bigrams *cc* and *gw*; sometimes more extensive Boolean queries, as in (Schmidt et al., 2013)), occasionally followed by manual filtering (e.g., (Lyytimäki and Tapio, 2009; Hulme et al., 2018))
- run straightforward term frequency and collocation analysis as a preparation for manual corpus inspection (e.g., (Willis, 2017)); sometimes with sophisticated visualisation (Pérez-González, 2020)
- compute bigram frequencies, or combine POS tagging with regex search to find verb usage patterns (Majdik, 2019)
- apply lexicons (sentiment, emotions, LIWC, etc.) ”out of the box” (e.g., (Barkemeyer et al., 2016))
- apply supervised classification to find CC texts and detect the presence of frames (economy, ideology, uncertainty) (Stecula and Merkle, 2019)
- apply topic modeling, usually LDA, without much further interpretation (e.g., (Dahal et al., 2019)) or with extensive subsequent interpretation (e.g., (Boussalis and Coan, 2016))
- apply topic modeling and combine this with other methods, such as network analysis (Elgesem et al., 2014) or cluster analysis (Wetts, 2020), in order to study a dedicated research question
- combine multiple techniques (sentiment, emotion, network analysis) to arrive at a fairly complex concept like ”credibility of a tweet” (Samantray and Pin, 2019)

4.3 Qualitative and quantitative research

We wish to point out that in the social sciences, the body of qualitative research on CC-related discourse is hardly smaller or less diverse than that of the quantitative work. Qualitatively-oriented studies show, for example, that effective communication on CC policy can result in citizen assemblies supporting specific policy proposals (Muradova et al., 2020). Carpenter (2002) traced how shifts in interest group discourses impacted negotiations of states at the COP-6. And studies on public opinion demonstrated that the quantity of media coverage on CC did not impact public opinion as much as ”elite cues” represented through partisan press releases or voting. A common theme, in any case, is that one needs to study CC discourse *across* channels and communities in order to understand the (lack of) impact on opinion or policy.

5 Analyzing the CC debate: Goals and methods

In the social sciences, three criteria are often used to assess the quality of research (see, e.g., (Kantner and Overbeck, 2020)):

- **Reliability:** Are analyses stable over time and can they be reproduced by other researchers?
- **Representativeness:** Does the selected data represent the variability in the underlying textual population?
- **Validity:** Do the analyses on the data actually measure the theoretically-derived (or underlying) concepts, i.e., are they helpful for the research question?

The first point corresponds quite clearly to the goal of reproducibility in NLP and does not require further comment here. In this section, we will thus reflect on the other two points. At the end, we summarize the takeaway messages that we propose for NLP.

5.1 Representativeness

Unless a certain dataset trivially represents the totality of a target discourse (e.g., all CC submissions to the Irish Citizens’ Assembly; (Devaney et al., 2020)), the work starts with assembling the subcorpus of texts that are relevant for the research question. As we pointed out in Section 4, the majority of studies employ just two bigrams (*cc*, *gw*), while a few use longer flat lists of terms (Pathak et al.,

2017) or combine terms into elaborate Boolean queries (Schmidt et al., 2013). In comparison, climate change is a relatively "friendly" domain in this respect, as the cc bigram intuitively promises relatively good quality in terms of both precision and recall. Nonetheless, one has to be aware of pitfalls, for instance when working with older text, where "global warming" and "greenhouse effect" in many discourses were the central representative terms. These questions have consequences for comparing the results and insights of different studies, for example on polarization; as noted by Calderwood (2020): "climate change" and "global warming" can be used as politically-sensitive terms, while others like "carbon emission" are more neutral.

A follow-up question concerns the "degree of topicality" of texts. The vast majority of work discussed above ran algorithms on the retrieved set of documents under the assumption that they are of equal relevance. However, in our own (ongoing) work on building a CC subcorpus of newspaper articles, we noted that querying the cc bigram also yields plenty of wine discussions and restaurant reviews. Depending on the size of the dataset, either noise is to be tolerated, or a step of manual filtering can be undertaken to improve precision, as also noted for news text by Lyytimäki and Tapio (2009) and for *Science/Nature* editorials by Hulme et al. (2018). On the latter corpus, ongoing work in our group found that supervised topic-frame classification works better for those texts that have a higher degree of "climate topicality", in comparison to texts that only mention CC in passing.

In general, supervised classification has not yet received a great deal of attention in the social science work, the exception in our survey being the study by Stecula and Merkley (2019), who used it both for finding topical texts in a large corpus and for identifying framing categories within the texts. They did not provide any evaluation of these steps, though; this is a point where established NLP research routines could inform the social science methodology.

5.2 Validity

Grimmer and Stewart (2013) stressed the danger of applying automatic tools to a text corpus without thorough reflection on what they actually measure. In the studies discussed in the previous sections, we find different attitudes toward this caution. Some-

times, the output of topic modeling or sentiment analysis is rather straightforwardly used to plot correlations with media types, time, or geographical regions. Stipulating such correlations based on NLP measures becomes much more critical when people or communities are directly affected, for example when Farrell (2019) relies on out-of-the-box NER to find out which people or organizations are associated both with philanthropy and with misinformation campaigns. Awareness of the risks of noisy or imprecise tool behavior is important for social scientists. The NLP community thus needs to consider its responsibility for making quality measures and domain or genre dependencies for their tools transparent, so that they are not used where their **validity** is low. One example of this discussion is the realm of sentiment lexicons, where the political science community found "one of their own" domain-specific tools (Young and Soroka, 2012) to be more trustworthy than so-called general-purpose lexicons.

Notwithstanding this note of caution, we believe that social science research should be open to embracing NLP tools that move beyond the well-established bag of words models and lexicon matching, especially where it increases validity. We agree with Grimmer and Stewart (2013) that NLP starts when the analysis goes beyond bags and "digs deeper" into the linear order of words and sentences for the purpose of extracting information. We think that, for example, word embeddings could receive more attention in social science in contexts where the meaning of CC terms is complex or shifting. Similarly, dependency parsing as a preparatory step to deeper content analysis can be highly relevant (also in conjunction with manual rules), as demonstrated for CC texts by Luo et al. (2020).

The "deeper analysis" concerns in particular the notion of *framing*, which is well-known to be highly ambiguous and vague (Scheufele and Iyengar, 2014, p. 6). This problem directly concerns the axiom of validity in quantitative research: what is, actually, being analyzed or measured? The majority of work discussed in Section 4 refers to Entman (1993), who stated that "to frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation". However, while much research

refers to Entman, we noted only one paper that actually uses his categories for annotation and analysis (Trumbo, 1996). Most other frame sets are, essentially, *topic perspectives*, as the list by Hoffman (2011) (quoted above) illustrates. Similar lists have been defined by, inter alia, Hulme et al. (2018) and O’Neill et al. (2015).

Whether frames are conceived as topics or as epistemic categories (e.g., (Entman, 1993; Luo et al., 2020)) makes a huge difference for validity of measurement in different research questions: The mere presence of a topic-frame in a text is to be distinguished from the stipulation that an intentional communicative act of selecting or emphasizing has been performed. The computational identification of subtle and purposeful framing requires approaches that most certainly have to go beyond bags of words. Linguistically-inspired NLP researchers can help in sorting out these phenomena, e.g., by systematically relating forms of framing to types of subjectivity analysis that are established in the NLP community, such as stance, aspect-based sentiment or argument mining.

Our final remark is that many interesting phenomena in discourse analysis are simply too subtle for automatic mining and instead require human analysis to increase validity. Here, NLP has an important role in preparing and annotating the corpora, and also in making them available to analysts in effective and comfortable ways.

5.3 Key takeaways for NLP

Considering the discussion in the previous sections, we summarize our main recommendations for how the NLP community can contribute to sense-making of the CC debate and of similar debates that are being studied in the social sciences.

- Given the importance of subcorpus building to the interdisciplinary study of the CC discourse, NLP can provide advanced and effective methods of finding topic-relevant cc texts without relying on a few predefined bigrams.
- By studying ”smaller” genres such as political speech or citizen voices on CC, NLP can increase its relevance for policy debates even where it does not deal with ”big data”, viz. by increasing efficiency and reliability/reproducibility of analyses.
- NLP can contribute to tools that provide for valid cross-channel and cross-genre analyses

to understand how CC discourses travel across communities, genres, and time.

- NLP tools regularly need to be adapted to domains and genres that are relevant for social science questions on CC discourses, as opposed to just using them ”out of the box”. This includes clarifying in what way a tool depends on its training data or other sources and how well it can be expected to perform elsewhere.
- While social scientists studying CC may have the domain expertise, the linguistic expertise from the NLP community can help understanding how notions of ”framing” correspond to established NLP tasks in subjectivity analysis and topic classification, so that social science can adopt tools that are relevant for such tasks.
- More attention can be given to the connections between network analysis (actors and their social relations) and NLP analyses, for example to extend multiplex community detection or to trace CC-related frame diffusion in online and offline social networks.
- For phenomena that eschew fully-automatic analysis, NLP and social sciences can collaborate on developing tools that support the human analyst and/or annotator in tracing CC discourses, for example by easy corpus filtering or visual analytics of frames, speaker-topic networks and the like.

6 Conclusions: Climate change, NLP, and the impact for social good

In this contribution, we have argued that NLP and social science can enrich each other to more comprehensively study the complex discourse(s) on climate change across channels, genres, communities, and topics. This is important because the CC debate is unfolding among three large and diverse actor communities:

- the general public,
- the policy-making communities (governments, public administrations, interest groups) at national or international levels, and
- the scientific communities.

Each community uses different genres, registers, and terminologies to communicate with each other and with other communities about CC. These communities shape individual and collective ideas, frames, and, ultimately, the behavior that is consequential for the future evolution of anthropogenic climate change. While social scientists explore this complex discourse in qualitative and quantitative research, they lack the full toolbox to do so at scale. And while NLP researchers are continuously expanding the general NLP toolbox, they have so far been selective in the channels and questions they focus on when it comes to CC, more or less choosing "the usual suspects".

The positive impact of combining both perspectives is not guaranteed, but possible. As societies increase their ability of "making sense" of the CC discourse, they get better at understanding and evaluating the politics and discourse landscape: Who is trying to frame CC discussions, on what channel, in what way, and for what interests? Is the CC debate polarized, controversial, fragmented into echo chambers or simply nuanced in an attempt to find socially and politically accepted solutions? Which frames are intentionally placed, and which are taken over, consciously and subconsciously, in traditional and new media? Why are some frames more successful and thus more likely to shape ideas that define public policy or collective behavior in relation to CC?

Where NLP can help answer these questions in reliable/reproducible, representative, and valid ways, it can have a positive impact for the social good beyond enriching the social sciences: Ultimately, it may provide each of the three communities mentioned above with the ability to judge in what direction one of the most important debates of our time—the climate change discourse—is evolving, and to respond accordingly.

Acknowledgments

We thank the anonymous reviewers for detailed constructive suggestions for improving the previous version of the paper.

References

Jonas Anshelm and Martin Hultman. 2015. *Discourses of Global Climate Change – Apocalyptic framing and political antagonisms*. Cambridge UP, Cambridge.

Ralf Barkemeyer, Suraje Dessai, Beatriz Monge-Sanz, Barbara Gabriella Renzi, and Giulio Napolitano. 2016. Linguistic analysis of IPCC summaries for policymakers and associated coverage. *nature climate change*, 6:311–317.

Jeremiah Bohr. 2020. Reporting on climate change: A computational analysis of U.S. newspapers and sources of bias, 1997–2017. *Global Environmental Change*, 61(102038).

Constantine Boussalis and Travis G. Coan. 2016. Text-mining the signals of climate change doubt. *Global Environmental Change*, 36:89–100.

Maxwell Boykoff and Jules Boykoff. 2007. Climate Change and Journalistic Norms: A Case-Study of US Mass-Media Coverage. *Geoforum*, 38(6):1190–2004.

Kevin J. Calderwood. 2020. Going Global: Climate Change Discourse in Presidential Communications. *Environmental Communication*, 14(1):52–67.

Chad Carpenter. 2002. [Businesses, Green Groups and The Media: The Role of Non-Governmental Organizations in the Climate Change Debate](#). *International Affairs*, 77(2):313–328.

John Cook, Naomi Oreskes, Peter T Doran, William R L Anderegg, Bart Verheggen, Ed W Maibach, J Stuart Carlton, Stephan Lewandowsky, Andrew G Skuce, Sarah A Green, Dana Nuccitelli, Peter Jacobs, Mark Richardson, Bärbel Winkler, Rob Painting, and Ken Rice. 2016. Consensus on consensus: a synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*, 11(4).

Biraj Dahal, Sathish A. P. Kumar, and Zhenlong Li. 2019. Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis and Mining*, 9(24).

Laura Devaney, Pat Brereton, Diarmuid Torney, Martha Coleman, Constantine Boussalis, and Travis G. Coan. 2020. Environmental literacy and deliberative democracy: a content analysis of written submissions to the Irish Citizens' Assembly on climate change. *Climatic Change*, 162:1965–1984.

Nicholas Diakopoulos, Amy X. Zhang, Dag Elgesem, and Andrew Salway. 2014. Identifying and analyzing moral evaluation frames in climate change blog discourse. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, pages 583–586.

Dag Elgesem, Lubos Steskal, and Nicholas Diakopoulos. 2014. [Structure and Content of the Discourse on Climate Change in the Blogosphere: The Big Picture](#). *Environmental Communication*, 9(2).

R.M. Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4):51–58.

- Justin Farrell. 2019. The growth of climate change misinformation in US philanthropy: evidence from natural language processing. *Environmental Research Letters*, 14:034013.
- Justin Grimmer and Brandon M. Stewart. 2013. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21:267–297.
- Reiner Grundmann and Ramesh Krishnamurthy. 2010. The Discourse of Climate Change: A Corpus-based Approach. *Critical Approaches to Discourse Analysis across Disciplines*, 4(2):113–133.
- Maarten Hajer and Wytse Versteeg. 2005. A decade of discourse analysis of environmental politics: Achievements, challenges, perspectives. *Journal of Environmental Policy & Planning*, 7(3):175–184.
- Andrew J. Hoffman. 2011. Talking Past Each Other? Cultural Framing of Skeptical and Convinced Logics in the Climate Change Debate. *Organization & Environment*, 24(1):3–33.
- Mike Hulme. 2009. *Why we disagree about climate change: Understanding controversy, inaction and opportunity*. Cambridge UP, Cambridge.
- Mike Hulme, Noam Obermeister, Samuel Randalls, and Maud Borie. 2018. Framing the challenge of climate change in Nature and Science editorials. *nature climate change*, 8:515–521.
- Sylvia Jaworska. 2018. Change But no Climate Change: Discourses of Climate Change in Corporate Social Responsibility Reporting in the Oil Industry. *Int'l Journal of Business Communication*, 55:194–219.
- Ye Jiang, Xingyi Song, Jackie Harrison, Shaun Quegan, and Diana Maynard. 2017. Comparing attitudes to climate change in the media using sentiment analysis based on Latent Dirichlet Allocation. In *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, pages 25–30, Copenhagen, Denmark.
- Cathleen Kantner and Maximilian Overbeck. 2020. Exploring soft concepts with hard corpus-analytic methods. In Nils Reiter, Axel Pichler, and Jonas Kuhn, editors, *Reflektierte algorithmische Textanalyse*. De Gruyter, Berlin.
- Allison Koenecke and Jordi Feliu-Fabà. 2020. Learning twitter user sentiments on climate change with limited labeled data. In *Proceedings of the 14th International AAI Conference on Web and Social Media*.
- Yiwei Luo, Dallas Card, and Dan Jurafsky. 2020. Detecting stance in media on global warming. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3296–3315, Online.
- J. Lyytimäki and P. Tapio. 2009. Climate change as reported in the press of Finland: From screaming headlines to penetrating background noise. *International Journal of Environmental Studies*, 66(6):723–735.
- Zoltan P. Majdik. 2019. A Computational Approach to Assessing Rhetorical Effectiveness: Agentive Framing of Climate Change in the Congressional Record, 1994–2016. *Technical Communication Quarterly*, 28(3):207–222.
- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, California.
- Lala Muradova, Hayley Walker, and Francesca Colli. 2020. Climate change communication and public engagement in interpersonal deliberative settings: evidence from the irish citizens’ assembly. *Climate Policy*, 20(10):1322–1335.
- Saffron O’Neill, Hywel T. P. Williams, Tim Kurz, Bouke Wiersma, and Maxwell Boykoff. 2015. Dominant frames in legacy and social media coverage of the IPCC Fifth Assessment Report. *nature climate change*, 5:380–385.
- N. Pathak, M.J. Henry, and S. Volkova. 2017. Understanding social media’s take on climate change through large-scale analysis of targeted opinions and emotions. In *AAAI Spring Symposia: AI for the Social Good*.
- Luis Pérez-González. 2020. ‘Is climate science taking over the science?’: A corpus-based study of competing stances on bias, dogma and expertise in the blogosphere. *Humanities and Social Sciences Communications*, 7(92).
- Andrew Salway, Dag Elgesem, Knut Hofland, Øystein Reigem, and Lubos Steskal. 2016. Topically-focused blog corpora for multiple languages. In *Proceedings of the 10th Web as Corpus Workshop*, pages 17–26, Berlin.
- Abhishek Samantray and Paolo Pin. 2019. Credibility of climate change denial in social media. *palgrave communications*, 5(127).
- D.A. Scheufele and S. Iyengar. 2014. The state of framing research. In Kate Kenski and Kathleen Hall Jamieson, editors, *The Oxford Handbook of Political Communication Vol 1*. Oxford University Press, Oxford.
- Andreas Schmidt, Ana Ivanova, and Mike S. Schäfer. 2013. Media Attention for Climate Change around the World: A Comparative Analysis of Newspaper Coverage in 27 Countries. *Global Environmental Change*, 23(5):1233–1248.

- Mirco Schoenfeld, Steffen Eckhard, Ronny Patz, and Hilde van Meegdenburg. 2018. [Discursive Landscapes and Unsupervised Topic Modeling in IR: A Validation of Text-As-Data Approaches through a New Corpus of UN Security Council Speeches on Afghanistan](#). *Computing Research Repository*, arxiv:1810.05572.
- Dominik A. Stecula and Eric Merkley. 2019. Framing Climate Change: Economics, Ideology, and Uncertainty in American News Media Content From 1988 to 2014. *Frontiers in Communication*, 4(6).
- Craig Trumbo. 1996. Constructing climate change: claims and frames in US news coverage of an environmental issue. *Publ. Underst. Science*, 5:269–283.
- Giuseppe A. Veltri and Dimitrinka Atanasova. 2015. [Climate change on Twitter: Content, media ecology and information sharing behaviour](#). *Public Understanding of Science*.
- Rachel Wetts. 2020. Models and Morals: Elite-Oriented and Value-Neutral Discourse Dominates American Organizations’ Framings of Climate Change. *Social Forces*, 98(3):1339–1369.
- Rebecca Willis. 2017. Taming the Climate? Corpus analysis of politicians’ speech on climate change. *Environmental Politics*, 26(2):212–231.
- L. Young and S. Soroka. 2012. [Affective News: The Automated Coding of Sentiment in Political Texts](#). *Political Communication*, 29(2):205–231.