Americans are dreamers – Generic statements and stereotyping in political tweets

Ines Rehbein
Data and Web Science Group
University of Mannheim
rehbein@uni-mannheim.de

Abstract

This paper addresses the question whether generics are associated with stereotyping, a claim often found in the literature. Most of the work, however, has been based on a small number of hand-crafted examples. We contribute to the discussion by looking at the use of generics in political tweets, written by members of the U.S. Congress, and present a framework for comparing the usage patterns of generics to other linguistic constructions that have been discussed as devices for stereotyping and *othering*. Our initial findings suggest that generics are not strongly associated with stereotypes while framing might play a more important role for stereotyping in political tweets.

1 Introduction

The construction of in-groups and out-groups, also refered to as *othering*, describes the categorisation of individuals as belonging to a specific group that is perceived as different from (and often inferior to) the social norm. In doing so, individuals considered as the *other* are usually depicted as a coherent and homogenuous group, based on stereotypes and overgeneralisation, making them an easy victim of discrimination (Staszak, 2009).

Othering is a common rhetorical move in political communication, and in particular in populist rhetoric (see, e.g., Fielder and Catalano (2017)). It can be strategically used to create a sense of shared identity among party supporters and mobilise them against a perceived enemy. While *othering* can have positive effects on the in-group, such as increasing group cohesiveness and cooperation, it also comes at the cost of prejudice, conflict and discrimination for the out-group.[1]

The creation of in- and out-groups plays an important role in political communication and is often associated with stereotyping. By *stereotype*, we refer to "a set of cognitive generalizations (e.g., beliefs, expectations) about the qualities and characteristics of the members of a group or social category", following the definition of the APA Dictionary of Psychology.[2]

Several linguistic devices have been discussed as a means for othering and stereotyping, among them generic statements (Novoa et al., 2023; Bosse, 2024; Davani et al., 2024), the use of pronouns (Íñigo-Mora, 2004; Bull and Fetzer, 2006; Proctor and Su, 2011; Tyrkkö, 2016; Alavidze, 2017), framing by word choice (Sheshadri et al., 2021; van den Berg et al., 2020; Dreier et al., 2022), or entity framing (Entman, 1993; Mahmoud et al., 2025).

In the paper, we look at othering and stereotyping in political tweets written by members of the U.S. Congress. The focus of our work, however, is not so much on investigating stereotyping of in- and out-groups from a social science perspective but, instead, on studying the linguistic devices used for stereotyping in political tweets. Our work contributes to the linguistic discussion on whether generics are a common means to express stereotypes, as has often been argued (Geurts, 1985; Leslie, 2014; Radden, 2009; Novoa et al., 2023; Bosse, 2024; Ralston, 2024) but also refuted (Krifka et al., 1995).[3] In particular, we examine the role of generics, universals, and quantified statements based on real data from the political domain, where previous studies have mostly relied on a small number of constructed examples. We introduce a framework for extracting real-world generic statements about social groups and present a qualitative analysis of our findings.

[1] Early work investigating the effects of *othering* on human behaviour is the classic Robbers Case Study (Sherif et al., 1961) where young boys have been divided into groups to study how they behave when having to compete or cooperate with members of the out-group.

[2] https://dictionary.apa.org/stereotype.

[3] For a more detailed discussion, see Section 2.

Proceedings of the 21st Conference on Natural Language Processing (KONVENS), Volume 2: Workshops, pages 128–137, 2025

2 Related work

Generic statements are often described as a powerful tool to control the perception of groups as *in-groups* and *out-groups* by ascribing them certain characteristics and attributes and by creating an *otherness* (Staszak, 2009). One property that makes the use of generics so compelling is their tolerance toward exceptions. In contrast to universal statements like "all birds fly", which would be rejected by everybody with at least minimal knowledge of ornithology, the generic sentence "birds fly" will be accepted as true even though there are birds that do not have this capacity. Interestingly, this is also the case for statements that are true only for a small subset of its group members, such as "mosquitoes transfer the West Nile Virus" which holds for only a small percentage of mosquitoes.

In addition, Novoa et al. (2023) point out another striking feature of generics, namely that "despite expressing general claims that gloss over exceptions, they are endorsed in the face of variable or even minimal evidence". This tolerance to exceptions and ignorance towards a lack of evidence seems to make generics especially suited to transport stereotypes. Bosse (2024, p.3878), for example, describes generics as "mental states that associate properties with social groups".[4] However, there are also other means to associate properties with groups, such as explicitly quantified statements ("most birds can fly") and universals ("all birds can fly").

Othering and generics have both been discussed in relation to polarisation (Erdoğan and Uyan-Semerci, 2025; Iyengar and Westwood, 2015). Novoa et al. (2023) investigate the role of generics for partisan polarisation in the US and show that not only generics but also explicitly quantified statements can be interpreted by humans as universally true, pointing at a tendency of the human brain to convert more nuanced, quantified statements into generic, categorising ones. In three experiments, the authors show that generic expressions and, to a lesser extent, explicitly quantified statements can evoke polarised political judgments.

Some works in NLP have focussed on the detection of stereotypes (Blodgett et al., 2020; Sap et al., 2020), often using artificial data based on *(identity term, attribute)* tuples (Jha et al., 2023). One recent example is the Multilingual SeeGULL dataset of Bhutani et al. (2024) which includes LLM generated tuples, validated by human coders. Davani et al. (2024) focus on the use of generic sentences for stereotyping and present a multilingual dataset that combines specific group mentions with attributes, annotated for genericity. Their schema also distinguishes mentions of generalisations (e.g., reporting what others have said) from promoting it. The authors show that the co-occurrence of group mentions and attributes are not a reliable incdicator of generalisation, thus questioning the widespread approach of using group, attribute pairs for stereotype detection. They present a classifier that detects generalisations, reporting a PR-AUC (Area Under the Precision Recall (PR) Curve) of 58.7.

Figure 1: Example tweet including a numerically modified statement about Americans.

More linguistically inspired work has focussed on the annotation and detection of generics in language (Mitchell et al., 2003; Doddington et al., 2004; Walker et al., 2006; Suh, 2006; Herbelot and Copestake, 2008; Zirn et al., 2008; Mathew and Katz, 2009; Reiter and Frank, 2010; Friedrich et al., 2015b; Friedrich and Pinkal, 2015; Friedrich et al., 2016). We take this as the starting point for our framework to extract generic and generalising statements from real-world political text.

3 Framework

We now describe the different steps in our framework for extracting instances of generic, universal and quantified statements about groups, in order to investigate their function in discourse (Figure 2).

Data The data we use in our case study are daily tweets of both houses of congress, from over 1,000 campaign, office, committee and party accounts.[5] The data has been collected over a time period from June 2017 to October 2020 and amounts to ca. 9.8 mio tweets. For an example, see Figure 1.

[4] Others, however, have questioned the idea that generics express stereotypes. See, e.g., Krifka et al. (1995, p.48).

[5] The data is available from `https://github.com/alexlitel/congresstweets`.

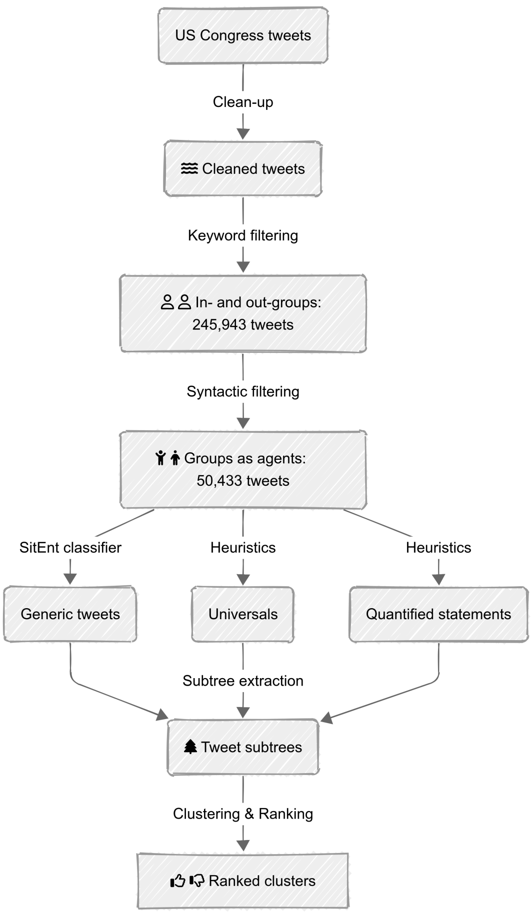

Figure 2: Framework for the extraction of real-world linguistic examples from U.S. Congress tweets.

Preprocessing We do some basic clean-up during preprocessing where we extract the tweet messages and filter out noise. Preprocessing includes the removal of non-ascii characters, empty lines and punctuation at the beginning of a tweet. We remove duplicates and tweets with no lowercase letters and delete URLs, at-mentions and hashtags at the beginning or end of the tweet.

Keyword-based filtering In the next step, we collect a list of keywords representing in- and out-groups in the migration discourse, i.e., *alien, american, asylum, migrant, refugee*. We then do a string-based keyword search for these group terms, ignoring case, and extract all tweets that include (at least) one of the terms. Please note that this also extracts tweets that include collocations like *American dream, American economy* or *immigrant rights* which are not in the center of our interest. We remove those in the next step, syntactic filtering. Keyword-based filtering reduces the overall amount of tweets from 9.8 mio to a much smaller set of 245,943 tweets (see Table 1) that can be processed more efficiently.

	group	is-subj	generic
alien	2,593	224	92
american	205,986	46,500	8,525
asylum	5,781	362	166
migrant	26,996	2,895	1,565
refugee	4,587	452	214
total	245,943	50,433	10,562

Table 1: Number of tweets after the different filtering steps (group: keyword-based filtering; is-subj: tweets where the group term is in subject position; generic: generic and generalising statements, according to the predicted Situation Entities classifier).

Syntactic filtering Next, we use a dependency parser[6] to extract tweets where the group terms occur in subject position. This filters out adjectival uses of the terms and extracts plausible candidates for generic statements about groups. We specify two exceptions, *American people, asylum seeker*, that we include in the data. This syntactic filtering step further reduces the data from roughly 246,000 messages to 50,433 tweets where one of the group terms appears in subject position in the tweet.

3.1 Classifying genericity

To identify generic and generalising statements, we train a supervised text classifier on the Situation Entities (SitEnt) dataset (Friedrich and Palmer, 2014; Friedrich et al., 2015b). Situation Entities are clause-level semantic annotations that introduce *situations* to the discourse (Smith, 2003), such as EVENTS, STATES, GENERICS, GENERALIZING SENTENCES, REPORTS or QUESTIONS (see A.1). We use a simple text classifier, based on a pre-trained RoBERTa model (Liu et al., 2019). For more training details, see §A.2 in the appendix.

Table 2 shows results on the SitEnt test set for all classes. Overall, results are high with an accuracy of 0.8 and a weighted F1 of 0.79. Not surprisingly, we get higher scores for classes with more training instances and also high results for Situation Entity types that have some easily identifiable features (such as questions, reports or speech acts). Results for GENERIC are substantially lower than for EVENT, even though both classes have roughly the same number of instances, and generalising sentences seem even harder to detect.

Our results are in the same range as the state-of-the-art (Rezaee et al., 2021) which combines a BERT model (Devlin et al., 2019) with GPT-2 (Radford et al., 2019) and reports an F1 of 0.79

[6]We use the spaCy v2.3.9 dependency parser with the English model `en_core_web_sm`.

	Prec	Rec	F1	Support
STATE	0.82	0.89	0.85	3,174
EVENT	0.85	0.83	0.84	1,524
GENERIC	0.80	0.72	0.76	1,429
UNDECIDED	0.67	0.63	0.65	465
GENERALISING	0.54	0.47	0.50	294
REPORT	0.86	0.81	0.84	274
QUESTION	0.89	0.92	0.91	132
GENERALSTATIVE	0.00	0.00	0.00	86
IMPERATIVE	0.51	0.90	0.65	51
SPEECHACT	1.00	1.00	1.00	1
Accuracy			0.80	7,430
F1(weighted)	0.79	0.80	0.79	7,430

Table 2: Results for the RoBERTa-based Situation Entity classifier on the test set. Support shows the number of instances for each class in the test set.

(acc: 0.81) on the SitEnt corpus.[7] The lower results for generic and generalising statements shows the difficulty of the task, which is in line with previous work (e.g., Davani et al. (2024), see §2 above). For an error analysis, see §A.3.

We apply our classifier to the Twitter dataset and extract all messages that have been predicted as either generic or generalising. As a result, we obtain a dataset with 10,562 tweets (see Table 1).

3.2 Extracting clusters of similar sentences

To find out whether there is an association between form and function for the different linguistic devices (i.e., generic, universal and quantified statements), we need to compare those constructions in context, i.e., in sentences with similar meaning. To extract those, we proceed as follows.

Subtree extraction As input, we use the set of tweets where instances of each group term appear in subject position. This includes all different statement types. Next, we identify the verb governing the group term and extract its syntactic subtree. This can be done, using the spaCy subtree function which extracts a token and all its syntactic dependents. For the example sentence below, the extracted subtree is the clause that includes the group term in subject position (Ex 1).

Ex 1. "Climate change is expected to create hundreds of millions of refugees, but right now climate refugees have no legal rights."
Subtree: "refugees have no legal rights"

Clustering We encode the subtree strings, using sentence embeddings, and cluster the embeddings to obtain semantically similar sentences.[8] For additional details on the clustering process, see A.4. We use a high cosine similarity threshold for clustering. As a result, we loose many instances that are below the threshold. However, for our use case, we are not interested in high coverage (i.e., being able to assign every instance to a cluster), but rather in identifying usage patterns of similar sentences in order to examine the variation in the use of different constructions. Table 3 shows an example cluster of statements expressing the americans' support for law and order. All but one sentence in this cluster use generic statements while the remaining one uses a quantified statement.

Ranking Finally, we rank the resulting clusters to identify similar sentences that show a high preference for specific constructions (i.e., a strong *form-to-meaning association*), providing us with linguistic evidence for further qualitative analysis. For ranking, we use the ratio of instances for the respective construction (i.e. generic, universal, etc.), relative to the total number of instances in the cluster.[9] To identify generic and generalising statements, we rely on the predictions of the SitEnt classifier (see §3.1). Universal and quantified statements can be easily extracted, based on heuristics. For universal statements, we extract sentences where the group term in subject position is modified by either *all* or *every* while for quantification, we search for the modifiers *most, many, some, few, several*.

4 Exploration of generics and stereotypes in political tweets

To explore to what extend generics are used to stereotype social groups in political tweets, we apply our framework to the U.S. Congress Tweets dataset described above. We present a qualitative analysis, focussing on clusters of similar sentences that show a strong preference for one particular construction.

[7]Please note that we do not report averaged results over different model initialisations, as we are not interested in benchmarking our models on the Friedrich et al. (2015b) data but only in identifying and extracting generic and generalising statements from social media.

[8]We use the Fast Clustering algorithm https://sbert.net/examples/sentence_transformer/applications/clustering/README.html#fast-clustering from the Sentence Transformer (Reimers and Gurevych, 2019) library.

[9]There are many more sophisticated measures that could be used, however, for the purpose of exploration our simple ranking metric already works well in identifying semantic clusters with a strong preference for one particular construction.

ID	**Cluster 1112**, 10 elements ratio: 0.9
1	Americans want LAW & ORDER!
2	The American people stand with law enforcement.
3	Americans support law enforcement and want law & order.
4	Americans support our # police officers and law and order in our cities.
	...
10	**Most Americans** support law & order, the fair application of it, and pray for our police.

Table 3: Example cluster for the group term "american", including 9 generic sentences (ratio: 0.9) and one explicitly quantified statement (**Most** Americans).

Cluster 267, 23 elements ratio: 1.0
ALL Americans have the constitutional right to marry
ALL Americans have the right to marry no matter whom they love
Cluster 813, 11 elements ratio: 1.0
all Americans have the opportunity to achieve success
every American has the opportunity to succeed
Cluster 838, 11 elements ratio: 1.0
all Americans can enjoy freedom & equality
every American is equal

Table 4: Examples for clusters with a high ratio of universal statements.

Generics Generic statements about the in-group in our data mostly express political demands (see A_{1-5} below).

Americans

(A_1) deserve better

(A_2) want the truth / want transparency

(A_3) should know / need to know

(A_4) are tired of waiting

(A_5) need help

For the out-group, clusters that show a strong preference for generics include mostly positive statements about the group (see B_{1-4} below).

Immigrants / refugees / asylum seekers

(B_1) enhance our economy

(B_2) make America stronger

(B_3) are not criminals / animals

(B_4) deserve protection

We did not find instances of generics used for negative stereotyping of social groups.

Universals Table 4 shows examples for semantic clusters that exclusively included universals. It is obvious that those instances do not contain stereotypes but express normative generalisations about how the world *should be*, rather than how it is. All of the clusters with a high ratio of universal sentences in our analysis are of that type.

Aliens	224 mentions in subject position:
	illegal (155), criminal (30), deported (4), convicted (3), dangerous (3), charged (2)
Immigrants	2,917 mentions in subject position:
	illegal (223), undocumented (154), young (62), legal (30), working (14)
Asylum s.	365 mentions in subject position:
	vulnerable (5), fleeing (5), cuban (3), transgender (3), legitimate (3)
Refugees	457 mentions in subject position:
	welcome (15), fleeing (14), rohingya (9), syrian (9), american (4), somali (4), liberian (4), clean (3)

Table 5: Entity framing of social groups in tweets.

Quantified statements When comparing generics to sentences that include explicit quantification, we did not see similarly strong associations between form and meaning (also see Table 9 in the apendix).

Entity framing While we did not see strong evidence for an assiciation between generics and stereotypes in our data, we found numerous instances of entity framing. Table 5 shows the most frequent modifiers for each group term, extracted from the spaCy dependency parse. It is obvious that the terms *aliens, immigrants* are mostly framed in negative terms. *Aliens* in particular are portrayed as dangerous and criminal, which is reflected in framing through word choice. Politicians with a more positive stance towards immigration did not use the term *alien* but preferred *refugee*, shown by the lack of negative attributes for this term.

5 Conclusion

In the paper, we present a framework for extracting generic and generalising statements about social groups and apply it to tweets written by members of the U.S. Congress. Our framework allows us to extract linguistic evidence for the various constructions while controlling for sentence meaning. This enables us to study usage patterns for generics, universals and quantified statements in a controlled setting with real-world data. While generics are often considered as an important device for stereotyping, our initial findings suggest that the association is less strong as expected and that framing might play a more important role for stereotyping social groups in political tweets.

Limitations

We would like to point out some limitations of our work. First, we would like to emphasize once again that the framework is not suitable for comparing the use of stereotypes across parties or across time. In the paper, we only consider three constructions that have the potential to be used for stereotyping but do not aim to measure the total amount of stereotypical statements overall.

Second, while this short paper introduces a framework for extracting real-world examples in context for comparative linguistic analysis, it does not (yet) present a validation of the framework. Third, while the clustering step manages to produce clusters of sentences with similar meaning, it is conceivable that the surface form of the different constructions also has some impact on the clustering process and that this results in artificially "clean" groups of statements where the different constructions end up in separate clusters. One way to address this issue is to mask the markers for the various constructions so that the input to the clustering process is normalised (see example below) while keeping the information on the original statement type for further analysis.

Orig: *All|Most|∅*	Americans support the law.
Cluster input →	Americans support the law.

Acknowledgments

The work presented in this paper is funded by the German Research Foundation (DFG) under the UNCOVER project (RE3536/3-1). We would also like to thank the anonymous reviewers for their constructive feedback.

References

Maia Alavidze. 2017. The use of pronouns in political discourse. *International Journal of Arts & Sciences*, 9(4):349–356.

Mukul Bhutani, Kevin Robinson, Vinodkumar Prabhakaran, Shachi Dave, and Sunipa Dev. 2024. SeeGULL multilingual: a dataset of geo-culturally situated stereotypes. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 842–854, Bangkok, Thailand. Association for Computational Linguistics.

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.

Anne Bosse. 2024. Stereotyping and generics. *Inquiry: An Interdisciplinary Journal of Philosophy*, 67(10):3876–3892.

Peter Bull and Anita Fetzer. 2006. Who are we and who are you? the strategic use of forms of address in political interviews. *Text and Talk*, 26.

Aida Mostafazadeh Davani, Sagar Gubbi, Sunipa Dev, Shachi Dave, and Vinodkumar Prabhakaran. 2024. Genil: A multilingual dataset on generalizing language. *Preprint*, arXiv:2404.05866.

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. Association for Computational Linguistics.

George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. 2004. The automatic content extraction (ACE) program – tasks, data, and evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).

Sarah K Dreier, Emily K Gade Dallas Card, and Noah Smith. 2022. Patterns of bias: How mainstream media operationalize links between mass shootings and terrorism. *Political Communication*.

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

Emre Erdoğan and Pinar Uyan-Semerci. 2025. Othering in politics: how affective polarization undermines democratic philia? *Frontiers in Political Science*, pages 1–17.

Grace E. Fielder and Theresa Catalano. 2017. Othering others: Right-wing populism in uk media discourse on "new" immigration. In Jan Chovanec and Katarzyna Molek-Kozakowska, editors, *Representing the Other in European Media Discourses*, pages 207–234. Amsterdam: John Benjamins.

Annemarie Friedrich, Kleio-Isidora Mavridou, and Alexis Palmer. 2015a. Situation entity types (annotation manual). Version 1.1.

Annemarie Friedrich and Alexis Palmer. 2014. Situation entity annotation. In *Proceedings of LAW VIII - The 8th Linguistic Annotation Workshop*, pages 149–158, Dublin, Ireland. Association for Computational Linguistics and Dublin City University.

Annemarie Friedrich, Alexis Palmer, Melissa Peate Sørensen, and Manfred Pinkal. 2015b. Annotating genericity: a survey, a scheme, and a corpus. In *Proceedings of the 9th Linguistic Annotation Workshop*, pages 21–30, Denver, Colorado, USA. Association for Computational Linguistics.

Annemarie Friedrich, Alexis Palmer, and Manfred Pinkal. 2016. Situation entity types: automatic classification of clause-level aspect. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1757–1768, Berlin, Germany. Association for Computational Linguistics.

Annemarie Friedrich and Manfred Pinkal. 2015. Discourse-sensitive automatic identification of generic expressions. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1272–1281, Beijing, China. Association for Computational Linguistics.

Barts Geurts. 1985. Generics. *Journal of Semantics*, 3(4):247–255.

Aurelie Herbelot and Ann Copestake. 2008. Annotating genericity: How do humans decide? (a case study in ontology extraction). In Sam Featherston and Susanne Winkler, editors, *The Fruits of Empirical Linguistics*, volume 1. Berlin: de Gruyter.

Isabel Íñigo-Mora. 2004. On the use of the personal pronoun we in communities. *Journal of Language and Politics*, 3(1):27–52.

S Iyengar and S Westwood. 2015. Fear and loathing across party lines: new evidence on group polarization. *American Journal of Political Science*, 59:690–707.

Akshita Jha, Aida Mostafazadeh Davani, Chandan K Reddy, Shachi Dave, Vinodkumar Prabhakaran, and Sunipa Dev. 2023. SeeGULL: A stereotype benchmark with broad geo-cultural coverage leveraging generative models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9851–9870, Toronto, Canada. Association for Computational Linguistics.

Manfred Krifka, Francis J Pelletier, Gregory N Carlson, Gennaro Chierchia, Godehard Link, and Alice ter Meulen. 1995. Genericity: An introduction. In *The Generic Book*, pages 1–24. University of Chicago Press, Chicago and London.

Sarah-Jane Leslie. 2014. Carving up the social world with generics. *Oxford studies in experimental philosophy*, 1:208–232.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Tarek Mahmoud, Zhuohan Xie, Dimitar Dimitrov, Nikolaos Nikolaidis, Purificação Silvano, Roman Yangarber, Shivam Sharma, Elisa Sartori, Nicolas Stefanovitch, Giovanni Da San Martino, Jakub Piskorski, and Preslav Nakov. 2025. Entity framing and role portrayal in the news. *Preprint*, arXiv:2502.14718.

Thomas A. Mathew and Graham Katz. 2009. Supervised categorization for habitual versus episodic sentences. In *Sixth Midwest Computational Linguistics Colloquium*. Bloomington, Indiana: Indiana University.

Alexis Mitchell, Stephanie Strassel, Mark Przybocki, JK Davis, George R. Doddington, Ralph Grishman, Adam Meyers, Ada Brunstein, Lisa Ferro, and Beth Sundheim. 2003. Ace-2 version 1.0. LDC2003T11.

Gustavo Novoa, Margaret Echelbarger, Andrew Gelman, and Susan A. Gelman. 2023. Generically partisan: Polarization in political communication. *Proceedings of the National Academy of Sciences*, 120(47):e2309361120.

Katarzyna Proctor and Lily I-Wen Su. 2011. The 1st person plural in political discourse – American politicians in interviews and in a debate. *Journal of Pragmatics*, 43(13):3251–3266.

Günter Radden. 2009. Generic reference in english: A metonymic and conceptual blending analysis. In Klaus-Uwe Panther, Linda L. Thornburg, and Antonio Barcelona, editors, *Metonymy and Metaphor in Grammar*, pages 199–228. John Benjamins Publishing Company.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI*.

Tom Ralston. 2024. Reconceptualising the psychological theory of generics. *Philosophical Studies*, 181:2973–2995.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Nils Reiter and Anette Frank. 2010. Identifying generic noun phrases. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 40–49, Uppsala, Sweden. Association for Computational Linguistics.

Mehdi Rezaee, Kasra Darvish, Gaoussou Youssouf Kebe, and Francis Ferraro. 2021. Discriminative and generative transformer-based models for situation entity classification. *CoRR*, abs/2109.07434.

Maarten Sap, Saadia Gabriel, Lianhui Qin, Dan Jurafsky, Noah A. Smith, and Yejin Choi. 2020. Social bias frames: Reasoning about social and power implications of language. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5477–5490, Online. Association for Computational Linguistics.

Muzafer Sherif, O.J. Harvey, Jack B. White, William R. Hood, and Carolyn W. Sherif. 1961. *Intergroup Conflict and Cooperation. The Robbers Cave Experiment [1954]*. University of Oklahoma Book Exchange, Norman.

Karthik Sheshadri, Chaitanya Shivade, and Munindar P. Singh. 2021. Detecting framing changes in topical news. *IEEE Transactions on Computational Social Systems*, 8(3):780–791.

Carlota S Smith. 2003. *Modes of discourse: The local structure of texts*. Cambridge University Press.

Jean-François Staszak. 2009. Other/otherness. In Rob Kitchin and Nigel Thrift, editors, *International Encyclopedia of Human Geography*, pages 43–47. Elsevier, Oxford.

Sangweon Suh. 2006. Extracting generic statements for the semantic web. Master's thesis, University of Edinburgh. Master's thesis.

Jukka Tyrkkö. 2016. Looking for rhetorical thresholds: Pronoun frequencies in political speeches. *Studies in Variation, Contacts and Change in English*, 17.

Esther van den Berg, Katharina Korfhage, Josef Ruppenhofer, Michael Wiegand, and Katja Markert. 2020. Doctor who? framing through names and titles in German. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4924–4932, Marseille, France. European Language Resources Association.

Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda. 2006. Ace 2005 multilingual training corpus. LDC2006T06.

Cäcilia Zirn, Vivi Nastase, and Michael Strube. 2008. Distinguishing between instances and classes in the wikipedia taxonomy. In *Proceedings of the 5th European Semantic Web Conference on The Semantic Web: Research and Applications*, ESWC'08, page 376–387, Berlin, Heidelberg. Springer-Verlag.

A Appendix

A.1 Situation Entity types

Table 6 illustrates the different Situation Entity types. Most relevant for our work are the classes GENERIC and GENERALISING. GENERIC sentences have a generic main referent while GENERALISING sentences make statements about individuals that generalise over situations and report regularities.

A.2 Classification setup

We follow the suggested train-test splits,[10] resulting in 30,859 instances for training and 7,430 instances for testing. We further split the training data into a training and development set (29,316 and 1,543 instances).

Preprocessing The original dataset has been automatically segmented into clauses. As the segmentation step introduces noise, we do not rely on the provided clause boundaries but extract the left and right context within the sentence for each clause and mark the relevant clause information by inserting an underscore before the verbal head of the clause. This means that the same sentence can be included multiple times, but with different verbs highlighted within the sentence. This enables the model to learn which clause to focus on while, at the same time, providing the necessary contextual information.

Hyperparameter tuning We do hyperparameter tuning on the development set and obtain best results of 0.86 F1(micro) and 0.63 (macro) across all 10 labels. We use the following hyperparameter configuration:

- learning rate: 3.134214402336582e-05
- weight decay: 3.215675091397119e-05
- batch size: 16
- warmup ration: 0.0936967038422192
- number of train epochs: 2

For implementation, we use the huggingface transformers library: `https://huggingface.co/docs/transformers`.

A.3 Error analysis

The training data for the Situation Entity classifier includes texts from 13 different genres and text types, including blogs, emails, essays, fiction, letters, news, and wikipedia articles. To get an idea about the error rate for predicting generics and generalising sentences, we randomly sample 20 instances for each group and manually inspect the classifier's predictions.

Our error analysis shows that the overall accuracy of the classifier on the twitter data is quite high with 0.84. For the term *aliens*, most instances are predicted as GENERALISING as this term is mostly

[10] `https://github.com/annefried/sitent/tree/master`.

SitEnt type	Definition	Example
STATE	States introduce specific properties of specific individuals to the discourse.	Joseph loves cake.
EVENT	Events introduce a specific event to the discourse: things that happen or happened.	Mira won the race.
GENERIC	Generic sentences make statements about kinds.	Lions are carnivores.
GENERALISING	Generalising sentences report regularities related to specific individuals.	Rob often feeds my cat.
REPORT	SitEnt introduced by verbs of speech.	"Let's go home" said Tim.
SPEECH ACTS		
IMPERATIVE	Commands, requests, advice.	Stay calm.
QUESTION	Functional speech act to elicit information.	Are you tired?

Table 6: Description and examples for the different Situation Entity types, adapted from Friedrich et al. (2015a) (available from https://www.coli.uni-saarland.de/projects/sitent/page.php?id=resources).

group term	correct	acc.
alien	15	0.75
american	18	0.90
asylum	16	0.80
migrant	18	0.90
refugee	17	0.85
all groups	84	0.84

Table 7: Number of correctly predicted labels (predicted as either GENERIC or GENERALISING by the classifier) for a random sample of 100 instances (20 instances per group) and accuracy for each social group.

used in combination with the adjective *illegal*, thus making this a reference not to a kind but to a specific subgroup of the kind (here: all aliens that are illegal). This is different from the other group terms where the main referent is generic at least for half of the instances (*Americans*) or for the majority of the messages (*asylum seekers, migrants, refugees*).

We observe the highest error rate for *asylum seekers* where our filtering heuristic does not seem to be accurate enough. Here we find 4 instances where the term *asylum* does not refer to the group, thus making the extracted instances irrelevant. Some other errors are due to mistakes made by the dependency parser failing to identify the correct subject.

A.4 Clustering similar clauses

We use the Fast Clustering algorithm[11] from the Sentence Transformer (Reimers and Gurevych, 2019) library. Cluster similarity is determined, based on cosine similarity. We set the minimum cluster size to 10 and the similarity threshold to 0.8, meaning that sentences need to have a similarity of ≥ 0.8 to be considered as similar. Sentences that do not meet this criterium are ignored.

group term	no. clusters	no. instances
alien	0	0
american	1,129	29,027
asylum	4	215
migrant	53	1,402
refugee	8	229

Table 8: Number of clusters and clustered instances for each social group term.

We cluster the tweets for each group term separately and use the `all-MiniLM-L6-v2` model[12] to map the extracted clauses to a 384 dimensional dense vector space, as input for clustering. Table 8 shows the number of clusters and clustered instances for each group.

For the term *alien*, we do not get any clusters, due to the small number of instances and rather high similarity threshold. However, for our use case we are not interested in high coverage (i.e., being able to assign each instance to a cluster) but rather in identifying usage patterns of similar sentences to study the variation in the use of the different constructions.

Table 9 shows examples for the clusters with the highest ratio of quantified statements for the term *American*. We only find two clusters where the ratio of quantified statements is higher than 0.5, indicating that explicit quantification does not seem to have a similar function in the discourse as universals.

[11] https://sbert.net/examples/sentence_transformer/applications/clustering/README.html#fast-clustering.

[12] https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

Cluster 551, 14 elements	ratio: 0.86
Americans are struggling to pay their rent and mortgages.	
Many Americans are struggling to pay their rent or mortgage mortgage.	
Cluster 725, 12 elements	ratio: 0.73
The American people overwhelmingly support this measure .	
most Americans from both parties are in support	
Cluster 1072, 10 elements	ratio: 0.30
Americans still face too many barriers at the ballot box	
too many Americans still face barriers at the ballot box with voter ID	

Table 9: Examples for clusters with a high ratio of quantified statements.