

3.1 Code and title

A5 – Subjective beliefs and misaligned reference systems: A unifying theory of gender biases and dog whistles

3.2 Project leaders

Jun.-Prof. Dr. Titus von der Malsburg
02.04.1977
Department of English Linguistics
Keplerstr. 17, 70174 Stuttgart
0711 / 685-8xxx
titus.vondermalsburg@ling.uni-stuttgart.de

Prof. Dr. Judith Tonhauser
19.11.1974
Department of English Linguistics
Keplerstr. 17, 70174 Stuttgart
0711 / 685-83121
judith.tonhauser@ling.uni-stuttgart.de

3.3 Project description

3.3.1 Summary and main research question

This project investigates the hypothesis that two superficially distinct cases of non-literal social meaning, gender biases and so-called dog whistles, naturally arise from mismatches in the reference systems participating in a discourse and under rational pragmatic inference.

Utterances are designed to convey intended meanings, literal and non-literal alike, accurately and efficiently when interpreted in the context of a shared reference system. Utterances can, however, also systematically fail to achieve this when the interlocutors' reference systems are miscalibrated. One example are linguistic gender biases, for example, when a speaker refers to the future president as "he" even though the future president is expected to be female. Another example are so-called dog whistles, expressions that are typically characterized as communicating a literal, typically harmless, meaning while also carrying a hidden and more controversial non-literal meaning that only a subset of the audience is receptive to. A classic example is the use of "inner city crime" in the United States which is superficially harmless but has been argued to carry racial overtones by implying crime committed predominantly by black people.

These phenomena appear dissimilar since gender biases are assumed to be unintended and unconscious whereas dog whistles are believed to be a conscious strategy that exploits the audience's blind spots. This analysis, however, may not hold up to scrutiny. Gendered language can be intentionally used to signal potentially controversial social views (e.g., German "Fräulein", roughly *missy*) and dog whistles can be unintentional expressions of unconscious biases (e.g., German "Clan-Kriminalität", organized crime conducted by extended families, often Arabic).

This project contends that both phenomena, gender biases and dog whistles, can be understood as products of the same underlying pragmatic principles when combined with the assumption of misaligned and/or miscalibrated reference systems. The overarching goal of this project is then to investigate whether an existing formal account of pragmatic inference, namely the rational speech act model (RSA model, Goodman and Stuhlmüller, 2013), can accommodate gender biases and dog whistles and explain them as naturally arising from the framework's independently motivated assumptions. Both gender biases as well as dog whistles have been primarily studied in English and in the context of US culture. A secondary goal of this project is therefore to empirically investigate both phenomena in German using corpus and experimental methods.

3.3.2 Current state of research and preliminary work

The utterances that we produce and the way we respond to them both index particular beliefs about the world, societal structures, and ideologies, but also biases and stereotypes. For instance, readers slow down when they see the gender-marked reflexive pronoun in the sentence “The tough soldier pricked *herself* with a needle” showing that readers eagerly make inferences about a referent’s gender based on stereotypical knowledge (Sturt, 2003; Kennison and Trofe, 2003). While this behavior may seem rational and in a sense unbiased (if the stereotype reflects real-world statistics), there is also evidence for intrinsic linguistic biases that are reflected in language use that does neither faithfully reflect speakers’ beliefs about the world (biased or unbiased) nor real-world statistics. In a large-scale online experiment during the 2016 US presidential campaign, von der Malsburg, Poppels, and R. P. Levy (2020) investigated how speakers refer to the future president (Clinton or Trump) and found that speakers tended to avoid feminine pronouns even when they believed that the female candidate was going to win. Similarly, readers struggled to process “she” pronouns referring to the future president. These results suggest that language users relied on reference systems that were out of touch with the current state of the world and even with their own beliefs (see also Pozniak and Burnett, 2021; Poppels, Boyce, Ajunwa, von der Malsburg, and R. Levy, 2021). Boyce, von der Malsburg, Poppels, and R. Levy (2019) generalized these results and demonstrated that feminine pronouns are systematically under-used (compared to masculine pronouns) and that comprehenders systematically under-infer female gender (both relative to perceived real-world statistics), again showing that gender information is not transparently transmitted. In German, the most contentious issue with regard to linguistic gender bias is the use and interpretation of the generic masculine, e.g., German “die Lehrer” when talking about male and female teachers. Many studies have demonstrated that these expressions are interpreted with a bias toward male referents regardless of real-world statistics (e.g., Gygax, Gabriel, Sarasin, Oakhill, and Garnham, 2008), which led to the proposal to introduce artificial gender-neutral forms, e.g., those using the so-called gender star (“Lehrer*innen”).

There are three main unresolved open problems in this area of research that this project aims to address. 1. There is a considerable amount of empirical evidence showing that gender biases exist but no clearly articulated theory that explains where in the system these biases arise and how. 2. The whole debate about linguistic gender biases and the German generic masculine in particular does not stand on solid empirical footing: Studies that assessed the interpretation of the generic masculine (and its alternatives) used repeated measurement designs in which participants likely became aware of the topic of the study and observed biases may therefore be under- or overestimates. 3. All studies known to us were conducted with small samples of participants from local student populations. Very little is therefore known about how gender biases vary as a function of age, gender, political affiliation, and other relevant factors.

The situation with regard to research on dog whistles is the opposite: There is an interesting theoretical literature but little empirical and no experimental data (none known to us). One clearly articulated proposal situates dog whistles in the framework of Bayesian signalling games (Burnett, 2017) and contents that dog whistles are used to signal one’s persona and ideologies to a subset of the audience while keeping another part of the audience in the dark (Henderson and McCready, 2018; Henderson and McCready, 2019). This hypothesis has been evaluated using a set of hand-picked examples from the US political discourse, e.g., the use of “family values” as code for *Christian values*. This proposal, while theoretically attractive and precisely formulated, has not yet been experimentally evaluated. It is also not clear to us that it necessarily captures the full breadth of the dog whistle phenomenon. For instance, when George Bush says “family values”, one could argue that he is not plausibly denying his Christian identity which is widely known. One alternative

interpretation is that the author's choice of words rather displays their biases much like the choice of masculine expressions does in the case of gender biases. A third alternative is that the use of expressions like "inner city crime" arise naturally when the speaker intends to communicate racist ideas but when overtly racist language is off limits. In this project we will formally implement these hypotheses as variants of the RSA model and compare them using corpus and experimental data combined with Bayesian model comparison.

References to work outside the CRC

- Burnett, Heather (2017). "Sociolinguistic interaction and identity construction: The view from game-theoretic pragmatics". In: *Journal of Sociolinguistics* 21.2, pp. 238–271.
- Goodman, Noah D. and Andreas Stuhlmüller (2013). "Knowledge and Implicature: Modeling Language Understanding as Social Cognition". In: *Topics in Cognitive Science* 5.1, pp. 173–184.
- Gygax, Pascal, Ute Gabriel, Oriane Sarasin, Jane Oakhill, and Alan Garnham (Apr. 2008). "Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men". In: *Language and Cognitive Processes* 23.3, pp. 464–485. ISSN: 1464-0732.
- Henderson, Robert and Elin McCready (2018). "How Dogwhistles Work". In: *New Frontiers in Artificial Intelligence*. Ed. by Sachiyo Arai, Kazuhiro Kojima, Koji Mineshima, Daisuke Bekki, Ken Satoh, and Yuiko Ohta. Springer International Publishing, pp. 231–240. ISBN: 978-3-319-93794-6.
- (2019). "Secondary Content: The Semantics and Pragmatics of Side Issues". In: Leiden, The Netherlands: Brill. Chap. Dogwhistles and the At-Issue/Non-At-Issue Distinction, pp. 222–245. ISBN: 9789004393127.
- Kennison, Shelia M. and Jessie L. Trofe (2003). "Comprehending Pronouns: A Role for Word-Specific Gender Stereotype Information". In: *Journal of Psycholinguistic Research* 32.3, pp. 355–378. ISSN: 1573-6555.
- Pozniak, Céline and Heather Burnett (2021). "Failures of Gricean reasoning and the role of stereotypes in the production of gender marking". In: *Glossa*. In press. In press.
- Sturt, Patrick (2003). "The time-course of the application of binding constraints in reference resolution". In: *Journal of Memory and Language* 48.3, pp. 542–562. ISSN: 0749-596X.

3.3.3 Work programme and role within the proposed Collaborative Research Centre

This project will address points 1. and 2. through large-scale online experiments that tap into a much wider participant population than typical lab-based studies. Online experiments also open the door to so-called one-shot designs in which each participant sees just one stimulus item and hence cannot develop an experiment-specific strategy. Point 3. will be addressed by extending the rational speech act model to see which assumptions are necessary to explain the observed data. Given the fully spelled-out mathematical nature of this model, we will be able to deploy the full arsenal of Bayesian model comparison tools for this purpose.

The goals of this project with respect to dog whistle are therefore the following: 1. We will address the thin empirical foundation for studying dog whistles in German by establishing a corpus of real-world examples from political speeches and public discourse. 2. We formulate a model explaining the use and interpretation of dog whistles. The aim is to treat dog whistles within the same model as will be used for gender biases (see above) and to contrast it with other models, again using the tools of Bayesian model comparison. 3. Finally, we will evaluate the model predictions with regard to the production and interpretation of dog whistles through psycholinguistic experiments.

3.4 Project-related publications by participating researchers

a) Peer-reviewed published or accepted articles

- de Marneffe, Marie, Mandy Simons, and Judith Tonhauser (2019). "The CommitmentBank: Investigating projection in naturally occurring discourse". In: *Sinn und Bedeutung* 23, pp. 107–124.
- Degen, Judith and Judith Tonhauser (to appear). "Managing web-experiments for psycholinguistics: An example from experimental semantics/pragmatics". In: *Open Handbook of Linguistic Data Management*. Ed. by Andrea L. Berez-Kroeker, B. McDonnell, and E. Koller. Cambridge, MA: MIT Press Open.
- Kiparsky, Paul and Judith Tonhauser (2013). "Semantics of Inflection". In: *Semantics: An International Handbook of Natural Language Meaning*. Ed. by Claudia Maienborn, Klaus von Heusinger, and Paul Portner. Vol. 3. Berlin: Mouton de Gruyter, pp. 2070–2097.
- von der Malsburg, Titus and Shravan Vasishth (2013). "Scanpaths reveal syntactic underspecification and reanalysis strategies". In: *Language and Cognitive Processes* 28.10, pp. 1545–1578.
- von der Malsburg, Titus and Bernhard Angele (2017). "False Positives and Other Statistical Errors in Standard Analyses of Eye Movements in Reading". In: *Journal of Memory and Language* 94, pp. 119–133.
- von der Malsburg, Titus, Till Poppels, and Roger P. Levy (Jan. 2020). "Implicit Gender Bias in Linguistic Descriptions for Expected Events: The Cases of the 2016 United States and 2017 United Kingdom Elections". In: *Psychological Science* 31.2, pp. 115–128. ISSN: 0956-7976.
- Stevens, Jon Scott, Marie-Catherine de Marneffe, Shari R. Speer, and Judith Tonhauser (2017). "Rational use of prosody predicts projectivity in manner adverb utterances". In: *Annual Meeting of the Cognitive Science Society* 39, pp. 1144–1149.
- Tonhauser, Judith (2012). "Diagnosing (not-)at-issue content". In: *Semantics of Under-represented Languages in the Americas* 6, pp. 239–254.
- Tonhauser, Judith, David Beaver, and Judith Degen (2018). "How projective is projective content? Gradience in projectivity and at-issueness". In: *Journal of Semantics* 35, pp. 495–542.
- Tonhauser, Judith, David Beaver, Craige Roberts, and Mandy Simons (2013). "Toward a taxonomy of projective content". In: *Language* 89, pp. 66–109.

b) Other publications

- Boyce, Veronica, Titus von der Malsburg, Till Poppels, and Roger Levy (2019). "Female gender is consistently under-expressed in pronoun production and under-inferred in comprehension". In: *93th Annual Meeting of the Linguistics Society of America*. Ed. by Roumyana Pancheva and Khalil Iskarous. New York, NY, USA: Linguistic Society of America.
- Poppels, Till, Veronica Boyce, Chelsea Ajunwa, Titus von der Malsburg, and Roger Levy (Mar. 2021). "Bias against "she" pronouns can be rapidly overcome by changing event expectations". In: *Proceedings of the 34th Annual CUNY Conference on Human Sentence Processing*. Ed. by John Trueswell, Delphine Dahan, Anna Papafragou, Gareth Roberts, Kathryn Schuler, Florian Schwarz, and Charles Yang. Philadelphia, Pennsylvania, United States: University of Pennsylvania.

3.5 Project funding

(To be filled in at a later date.)