3.1 Code and title

A5 – Dog Whistles and Gender Biases in a Rational Model of Language Interpretation

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

The reference systems of participants in a discourse can rarely, if ever, be assumed to perfectly align, which creates room for ambiguity and misunderstandings. This is particularly the case with social meaning, i.e. meaning that informs us about where interlocutors position themselves in the social coordinate system and which arises implicitly and fluidly from how we express ourselves. The goal of this project is to use empirical, experimental, and computationally methods to investigate two topical instances of social meaning, linguistic gender biases and dog whistle expressions, and to explain how they naturally arise from the assumption of miscalibrated reference systems when viewed through the lens of rational pragmatic inference.

Gender bias is exemplified by a speaker referring to the future president as "he" even though the president is expected to be female, or by a listener experiencing processing difficult upon hearing "she" in reference to a likely female future president. So-called dog whistles are expressions that encode a literal, typically benign meaning, while also carrying a hidden, more controversial, or even offensive non-literal social meaning that only a subset of the audience is receptive to. An example is the use of "inner city crime" in the United States which is argued to carry racial overtones by implying crime committed predominantly by black people. These two phenomena are superficially different in that gender biases are assumed to be unintended and unconscious whereas dog whistles are believed to reflect a conscious strategy that exploits blind spots in the audience's reference system. 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" when referring to organized crime conducted by extended Arabic families).

This project contends that both phenomena, gender biases and dog whistles, can be understood as products of the same underlying dynamics, namely the interaction of rational pragmatic inference with 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 (**FrankGoodman2012**) can explain gender biases and dog whistles as naturally arising from the framework's independently motivated assumptions about efficient communication. Gender biases and especially dog whistles have been predominantly 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.

3.3.2 Current state of research and preliminary work

A contentious issue in German is the generic masculine, e.g., the use of the masculine noun "die Lehrer" when talking about a set of male and female teachers. Many studies have demonstrated that the generic masculine is interpreted with a bias toward male referents (**GygaxEtAl2008**; **Kusterle2011**), an observation that has inspired the introduction of novel gender-neutral forms such as those containing the so-called gender star ("Lehrer*innen"). However, there are two key issue in this area research: 1. There is currently no clearly articulated theory that explains how these biases arise and where in the system they are rooted. 2. The debate does not stand on solid empirical footing. Experiments typically expose participants to many instances of the generic masculine along with gender-neutral forms and feminine forms. It is practically certain that participants develop some awareness of what is being tested and this may alter their behavior. For instance, a salient gender-neutral alternative may lead participants to interpret the generic masculine as more male-oriented than they usually would; or participants might interpret the generic more neutral due to social pressure. Further, most studies tested local undergraduate populations and results may therefore not generalize. With regard to gender biases, this project therefore identifies two desiderata: 1. A theoretical account that explains the use and interpretation of the generic masculine and its alternatives. 2. An improved empirical basis for understanding the generic masculine.

The starting point with regard to dog whistles is different: There is a rich 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 signaling games (**Burnett2017**) and contents that dog whistles are used to signal one's persona and ideologies to a subset of the audience while keeping others in the dark (**HendersonMcCready2018**; **HendersonMcCready2019**). While theoretically attractive and precisely formulated, this proposal has not yet been thoroughly evaluated. It is also not clear to us that it captures the full breadth of the 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 speakers' choice of language rather displays their biases much like the choice of masculine expressions does in the case of gender biases especially when explicit (e.g., racist) language is off limits. In this project we will formally implement this hypothesis as a variant of the RSA model and compare it to the **HendersonMcCready2019** proposal using corpus data and experimental data combined with Bayesian model comparison.

Both PIs contribute crucial experience to this project. TODO Tonhauser is an expert .... Von der Malsburg used large-scale online experimentation to study gender biases during the 2016 US presidential campaign and found that speakers avoided feminine pronouns to refer to the future president even when they believed that the female candidate (Clinton) was going to win (**MalsburgEtAl2020** see also **PoppelsEtAl2021CUNY**; **PozniakBurnett2021**). This study further found that the strength of these biases varies as a function of age, political alignment, and education, suggesting that data from just undergraduate populations may indeed be misleading. **BoyceEtAl2019LSA** generalized these findings and demonstrated that feminine pronouns are systematically under-used and that comprehenders systematically under-infer female gender both relative to perceived real-world statistics. Von der Malsburg is also an expert for Bayesian inference (**MorganEtAl2020**; **MeziereEtAl2021**) which puts us in an ideal the position to formally implement our theoretical ideas and to evaluate them using state-of-the-art Bayesian model comparison tools.

References to work outside the CRC

3.3.3 Work programme and role within the proposed Collaborative Research Centre

Ignore the text below. Just a scratch pad for text blocks.

Specifically, we will embed this issue in the RSA model of rational pragmatic inference and evaluate it using corpus data and experimental comprehension and production data. Within this framework we will also investigate the alternatives to the generic masculine since there is no strong a priori reason to believe that they are necessarily neutral (e.g., "Lehrer*innen" may still be male-biased, or female-biased).

- Explain - Explain online methods.

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 stimulis item and hence cannot develop an experiment-specific strategy. Point 3. will be adressed 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 empricial 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 use 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

b) Other publications

3.5 Project funding

(To be filled in at a later date.)