

# Integrating Speaker-Hearer Relations Into a Rational Speech Act-based Model of Politeness

Eleonore Lumer, Hendrik Buschmeier  
Digital Linguistics Lab, Bielefeld University, Germany  
eleonore.lumer@uni-bielefeld.de

Politeness, an integral aspect of communication, is a complex linguistic phenomenon that stands in opposition to other principles of conversation, such as the Gricean Maxims (Grice, 1975), as polite utterances are often neither truthful nor efficient. In our research, we mostly consider politeness strategies as ways to enhance or preserve the public self-image of the listener or the speaker (Brown & Levinson, 1987). Further research in this area is necessary, as many factors and influences on politeness are still unclear even though the phenomenon is fundamental for a tactful social interaction and thus also relevant for technical applications such as human-agent interaction. With the help of probabilistic models, influences on these phenomena can be explored. An influential Bayesian model for pragmatic language processing, ‘Rational Speech Acts’ (RSA; Frank & Goodman, 2012), has been adapted to account for politeness by considering opposing goals during language production (Yoon et al., 2016, 2020): informational (be truthful), social (be kind), and self-presentational (be considerate). A limitation of that model is that it does not consider social influences on politeness, even though it was shown that the speaker–hearer relation – expressed, for instance, in terms of power and distance (Brown & Levinson, 1987) – has an influence on the choice of politeness strategy (e.g., one tends to be more polite when talking to one’s boss than to a friend). The relationship can thus be expected to affect the utility of the opposing goals in the RSA-based politeness model.

We present work that instantiates Yoon et al. (2016)’s computational model of politeness with German data and (i) integrates speaker–hearer relations in terms of power and distance (Brown & Levinson, 1987), and (ii) additionally asks at which stage in the production process (roughly during conceptualisation or formulation; Levelt, 1989) the relationship should exert its influence. For this, we developed and tested two versions of a ‘polite Relational RSA’ (pRRSA) model. The pRRSA<sub>C</sub> model implements the relationship influence on the speaker’s goal by mapping the relationship on a parameter that determines the weighing between kindness and truthfulness (which could be considered a choice in conceptualisation). In contrast to this, the pRRSA<sub>F</sub> model maps the relationship influence on a parameter altering the degree of politeness (which could be considered a choice in formulation). Our approach differs from Yoon et al. (2016, 2020), as they learn a single parameter, while we learn one for each relationship. Results show that both pRRSA models were able to predict the average meaning of each word for each relationship, but that the probability distributions over the possible meanings diverged from the data (see Figure 1). Overall, both models achieved very similar results, that is, the choice of parameter, where the influence is modeled, appears to be less relevant than we expected. Thus the models do not allow us to answer the question on which stage in the production process the influence of politeness should occur, indicating that (Yoon et al., 2016, 2020)’s politeness models are too simplistic for mapping relevant social influences on complex linguistic phenomena.

In future work, we aim at modelling social and contextual influences on politeness in speech production and language generation and testing these models in human-agent interaction.

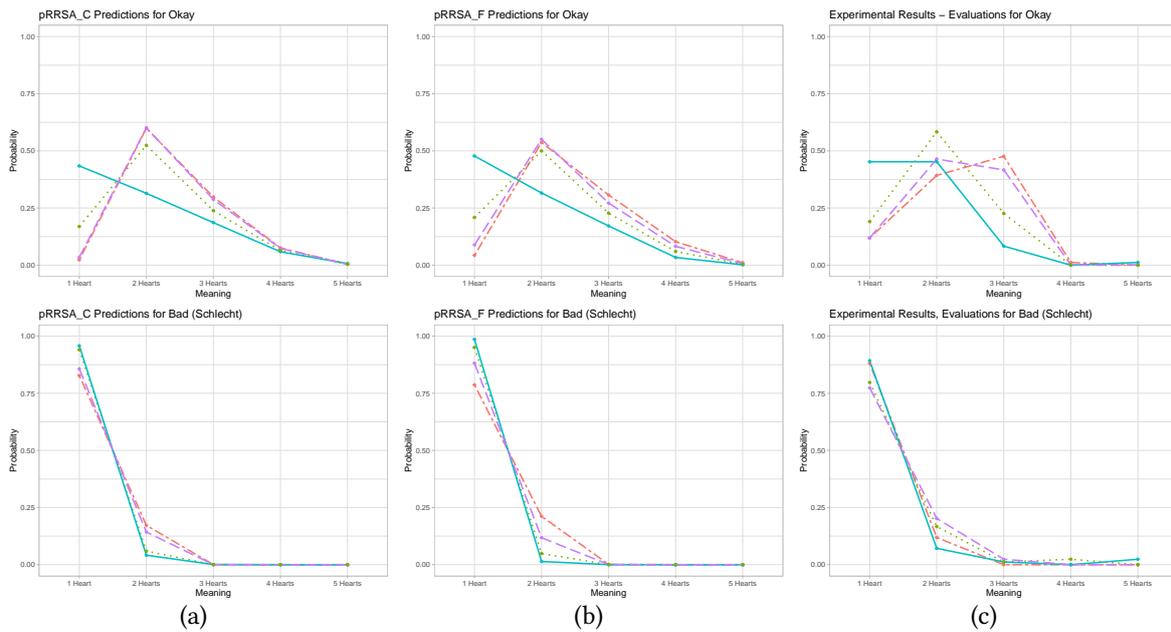


Figure 1: Model prediction results for models  $pRRSA_C$  (a) and  $pRRSA_F$  (b) and experimental results (c) showing the distribution of the meaning for the target words *okay* (top row) and *bad* (bottom row) across five states. The data in (c) was collected in an online study where participants were asked for the pragmatic meaning of words given similar scenarios and different speaker-hearer relations (dreaded boss/blue line, distant colleague/green dots, easy-going boss/violet dashes, close friend/red dash-dots) on five-point Likert-scales (♥–♥♥♥♥♥).

## References

- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.
- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336, 998–998. <https://doi.org/10.1126/science.1218633>
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). Academic Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. The MIT Press.
- Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2016). Talking with tact: Polite language as a balance between kindness and informativity. *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, 2771–2776.
- Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2020). Polite speech emerges from competing social goals. *Open Mind*, 4, 71–87. [https://doi.org/10.1162/opmi\\_a\\_00035](https://doi.org/10.1162/opmi_a_00035)