Speech Act Prediction in Conversation

Lena Warnke[a] *& Jan P. de Ruiter*[a,b]

[a] *Department of Psychology, Tufts Universty, Medford, MA, USA*
[b] *Department of Computer Science, Tufts Universty, Medford, MA, USA*

Keywords: Speech Acts, Prediction, Turn-Taking.

Introduction

In conversation, every utterance performs a speech act, such as a question, a complaint, or an invitation (Austin, 1962; Searle, 1969). Speech act comprehension is a critical skill for conversation: since different speech acts constrain for different responses (Schegloff, 2007), it is only once the action of an utterance has been recognized that the listener can begin planning a response. Speech act recognition, however, presents a complex cognitive challenge. There is no one-to-one mapping between an utterance's linguistic form and speech act. For example, "*Are you kidding?"* is formally interrogative, yet does not always request information. Speech acts also need to be recognized quickly in order to respond within the 200 ms gap that is typical in conversation (Stivers et al., 2009). Therefore, listeners must extract the underlying speech act and begin preparing their response before the incoming utterance has completely unfolded.

In the face of these challenges, cognitive prediction provides an explanation for the ease of listeners' speech act comprehension. According to this view, listeners predict the speech act of the upcoming turn from the preceding context. Evidence suggests that comprehenders anticipate language at many levels of linguistic representation, including semantic (Federmeier & Kutas, 1999), syntactic (Strijkers et al., 2019), and orthographic levels (Laszlo & Federmeier, 2009), yet to date only limited research has explored prediction at the speech act level (e.g. Gisladottir et al., 2018). The present study investigates whether listeners form expectations about upcoming speech acts based on the preceding turn in conversation.

Methods

Participants (n=60) listened to naturalistic conversations consisting of two turns – a context utterance followed by the critical turn. The critical turn was either spoken by the same speaker or a different speaker as the context utterance and fell into one of three conditions. In *congruent* trials, the critical turn confirmed speech act expectations given the preceding context utterance. In *speech act violation* trials, the critical turn violated speech act expectations because the speech act was incongruent with the preceding context utterance. We created *speech act violation* trials by switching the speaker (same or different) of the critical turn in the *congruent* trials such that the speech act of the second turn became implausible, but the words remained identical. This allowed us to control for lexico-semantic content. Lastly, in *speaker-independent violation* trials, the critical turn violated both speech act and lexico-semantic expectations (see Table 1). All stimuli were fully counterbalanced across conditions and normed for plausibility. This experimental design allowed us to measure any effects of anticipation purely at the speech act level.

We used the same behavioral paradigm as De Ruiter et al. (2006). For each trial, participants were asked to anticipate the precise ending of the second turn with a button press. We measured the duration of time between the end of the turn and the button press (called *bias*). Prior work shows that listeners are more accurate at predicting *when* a turn will end if they predict *how* the turn ends

(Magyari & de Ruiter, 2012). We hypothesized that if listeners are predicting speech acts, they would be more accurate at estimating the end of the critical turn in the *congruent* compared to both the *speech act violation* and the *speaker-independent violation* conditions, reflected by a shorter *bias*.

Table 1

Conditions of the experiment, 3 (congruency: congruent, speech act violation, speaker-independent violation) x 2 (speaker: same speaker, different speaker)

Context Utterance (TCU 1): "*We just moved into a new house.*" Target Utterance (TCU 2):			
	Congruent	**Speech act violation**	**Speaker-independent violation**
Different speaker	"*Where?*"	"*Come by*"	"*I'm lost*"
Same speaker	"*Come by*"	"*Where?*"	"*You sure?*"

Results

The mean *bias* was 307 ms for the *congruent* condition, 332 ms for the *speech act violation* condition, and 368 ms for the *speaker-independent violation* condition (Figure 1). The data were analyzed with Bayesian mixed effects models. Analyses were performed in R (R Core Team, 2021) using the rstanarm (Goodrich et al., 2020) and bridgesampling (Gronau et al., 2020) packages. We added fixed and random effects incrementally to a minimal model and tested if the additional term was justified by comparing the likelihood of the data under each model. The data were most likely under the model that contained condition as a fixed factor, and random intercepts for both participants and items. The model estimated beta coefficients show a the three-way dissociation in *bias* between the three experimental conditions. The Bayes factor for this model was 2770, providing decisive evidence for this model over the null model (intercept only). The data were 24 times more likely under this model than under a model that also included speaker switch as a factor, and 2212 times more likely under this model than under a model that included speaker switch and the interaction between speaker switch and condition. This provides decisive evidence that speech act congruency predicts turn-end estimation in conversation, and very strong evidence that whether the critical turn was spoken by the same or a different speaker (a factor that was not part of our hypothesis but part of our experimental design) did not affect *bias*.

Discussion

Our results show that listeners are more accurate at estimating the ends of turns that confirm speech act expectations compared to turns that violate speech act expectations, even when we controlled for lexico-semantic content. This suggests that listeners draw on conversation context to anticipate the speech of the upcoming turn. Interestingly, our results also show that listeners were more accurate at estimating the end of the turn when the lexico-semantic content was congruent despite violating speech act expectations. We conclude that comprehenders draw on their pragmatic speech act knowledge while comprehending unfolding utterances, but that lexico-semantic predictions also plays a role.

The results from this study provide support for the theory that listeners anticipate speech acts in natural conversation. Listeners draw on the context of the prior turn to interpret the speech act of the subsequent turn, which helps them plan their response early when they are in a conversation themselves. These results not only shed light ono the fact that we have a cognitive architecture oriented to speech acts, but also extend linguistic prediction to social interaction.

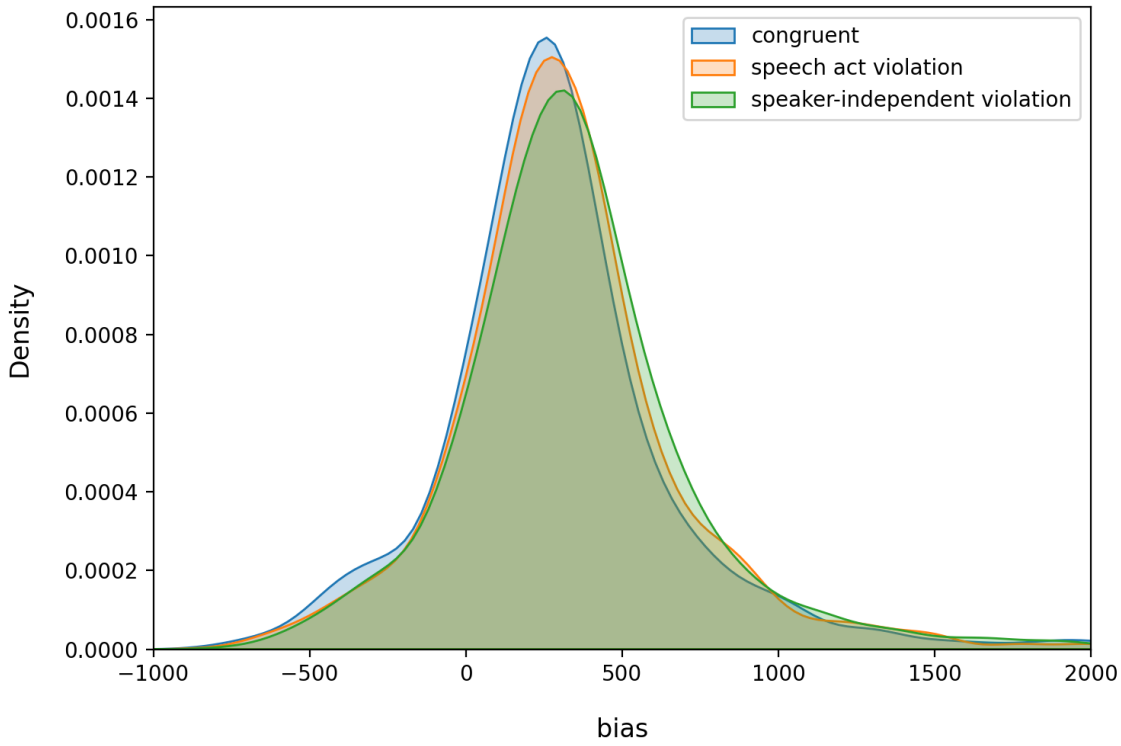

Figure 1. Probability density of bias separated by the three experimental conditions

References

Austin, J. L. (1962). *How to do things with words*. Oxford University Press.

de Ruiter, J. P., Mitterer, Holger., & Enfield, N. J. (2006). Projecting the End of a Speaker's Turn: A Cognitive Cornerstone of Conversation. *Language*, *82*(3), 515–535. https://doi.org/10.1353/lan.2006.0130

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *4*(41), 469–495. https://doi.org/10.1006/jmla.1999.2660

Gisladottir, R. S., Bögels, S., & Levinson, S. C. (2018). Oscillatory Brain Responses Reflect Anticipation during Comprehension of Speech Acts in Spoken Dialog. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00034

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2020). *Rstanarm: Bayesian Applied Regression Modeling via Stan. R Package v. 2.19. 2*.

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2020). bridgesampling: An R Package for Estimating Normalizing Constants. *Journal of Statistical Software*, *92*(10), 1–29. https://doi.org/10.18637/jss.v092.i10

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326–338. https://doi.org/10.1016/j.jml.2009.06.004

Magyari, L., & de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front Psychol*, *3*, 376. https://doi.org/10.3389/fpsyg.2012.00376

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis I* (Vol. 1). Cambridge University Press.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge university press.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106

Strijkers, K., Chanoine, V., Munding, D., Dubarry, A.-S., Trébuchon, A., Badier, J.-M., & Alario, F.-X. (2019). Grammatical class modulates the (left) inferior frontal gyrus within 100 milliseconds when syntactic context is predictive. *Scientific Reports*, *9*(1), 4830. https://doi.org/10.1038/s41598-019-41376-x

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, *30*(6), 648–672. https://doi.org/10.1080/23273798.2014.995679