

Indirect Speech Acts Do Not Change FTOs in Conversation

Threlkeld, C.^a Warnke, L.,^b de Ruiter, J.^{ab}

^a Department of Computer Science, Tufts University, Medford, MA, USA

^b Department of Psychology, Tufts University, Medford, MA, USA

Keywords: Speech Act, Sentence Type, Floor Transfer Offset, Cognitive Processing.

Introduction

In conversation, every utterance performs a social action, known as its speech act (Austin, 1962; Searle, 1969). Speech act recognition is not trivial given that there is not a one-to-one correspondence between an utterance's speech act and its grammatical sentence type (interrogative, declarative, imperative) (Sadock & Zwicky, 1985). For example, “*Can you pass the salt?*” is interrogative, but often functions as a request rather than a question.

When an utterance's sentence type and function match, it is a direct speech act. According to the Literal Force Hypothesis (Gazdar, 1981 via Levinson, 1983), declaratives function as statements, interrogatives as questions, and imperatives as commands. When an utterance's function does not match its sentence type, it is an indirect speech act.

Two competing theories propose how we process indirect speech acts. In a traditional philosophical view, indirect speech acts flout Grice's maxim of manner (Grice, 1975). Such a model assesses the sentence type and then determines whether the corresponding speech act makes sense in the conversational context, an approach taken by e.g. Sarathy et al. (2020). Yet, recent experimental research shows we recognize speech acts early in a turn (Bögels et al., 2015; Gisladdottir et al., 2012, 2015, 2018). Indirect speech acts are then recognized not *by* their sentence type, but *in spite of* their sentence type.

These two theories predict different cognitive processes in speech act recognition. In the philosophical (i.e. Gricean) view, indirect speech act recognition is reactive: we first extract a direct speech act according to sentence type and then reinterpret the speech act given context. The alternative model does not suggest reinterpretation. Here, we investigate the *timing* of natural conversation for evidence of additional cognitive processing for indirect speech acts. We predict that if indirect speech acts require extra cognitive processing, interlocutors take longer to respond to indirect vs. direct speech acts.

We operationalize cognitive processing as the *floor transfer offset* (FTO) in conversation (Mertens & de Ruiter, 2021). FTO is the time between the end of one speaker's utterance and the start of their interlocutor's utterance (de Ruiter et al., 2006). This work builds on Roberts, Torreira, and Levinson (2015) by operationalizing sentence type in addition to FTO and speech act. If indirect speech acts

are processed reactively, we expect them to be followed by larger FTOs compared to direct speech acts.

Methods

Conversation data came from the Switchboard corpus (Godfrey & Holliman, 1993). We retrieved timing information from the MSU transcription (*Openslr.org*) and turn construction unit (TCU) segmentation and speech act tagging from the Stanford University transcriptions (Jurafsky et al., 1997). We reduced this speech act set to a schema motivated by Gazdar (1981). We assigned sentence-type probability for each TCU using fine-tuning of DistilBERT (Sanh et al., 2020). Our first analysis used only TCUs assigned at least 50% chance of some sentence type — 1930 TCUs and FTOs. Our second analysis included all 2465 available turn-end TCUs.

Results

We used R (R Core Team, 2021) with the *rstanarm* (Goodrich et al., 2020) and *bridgesampling* (Gronau et al., 2020) packages to build several linear mixed effects models. We included random intercepts for conversation ID in all models, and sentence type, speech act and the interaction between the two as fixed effects to predict FTOs. We added fixed effects incrementally to a null model, testing if the inclusion of the additional term was justified by comparing the marginal likelihood of the data under each model. The final model under which the data were most likely included speech acts as a fixed effect. The data were 522 times more likely under this model compared to the null model and 71 times more likely under this model compared to a model that included both sentence type and speech acts as fixed effects. The data under the speech act only model were also 4270 times more likely compared to a model that included the interaction between speech act and sentence type. This provides decisive evidence that speech acts predict FTOs and that the interaction between sentence type and speech act does not. We also find very strong evidence that sentence type does not predict FTOs in conversation (Wetzels, 2011).

We also analyzed the data with the probability of directness of the speech act as a continuous predictor. The probability of directness is the probability associated with the direct sentence type of the tagged speech act. We ran linear mixed effects regression models including directness probability as a fixed factor and conversation ID as a random factor to predict FTOs. The data were 5.7 times more likely under the null model, which included only random intercepts for conversation ID, than the model including the probability of directness. This constitutes moderate evidence in favor of the null model, replicating the above findings that speech act directness does not predict FTOs.

Table 1. The data in the first study involved 1930 utterances as broken down below.

	Declarative	Interrogative
Statement	1687	42
Question	43	158

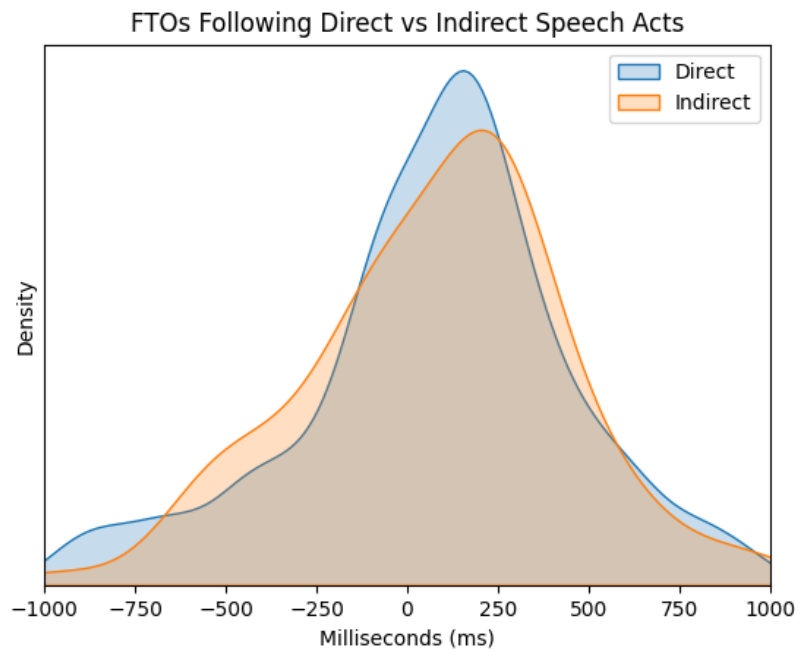


Figure 1. The data were 4270 times more likely under the speech act model than one including directness.

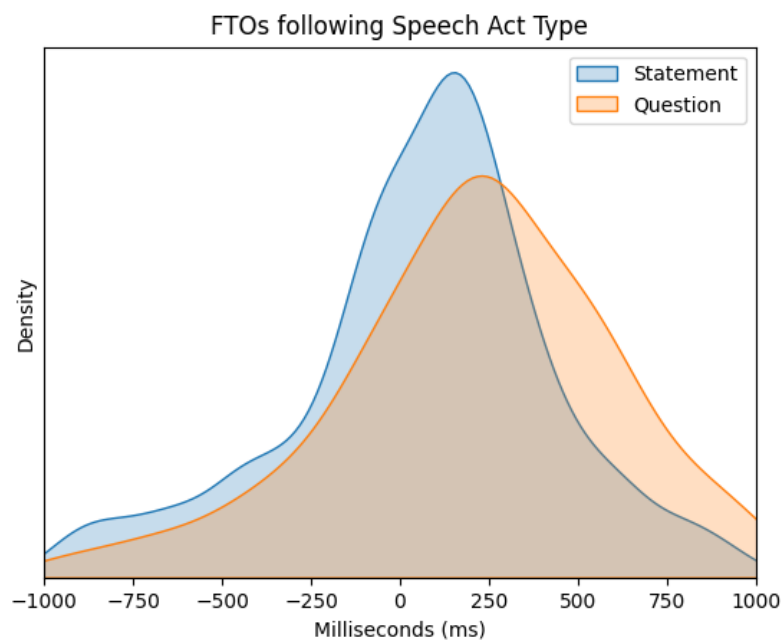


Figure 2. The data were maximally likely under a model accounting for only speech act.

References

- Austin, J. L. (1962). *How to do things with words*. Oxford University Press.
- Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Sci Rep*, 5, 12881. <https://doi.org/10.1038/srep12881>
- 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>
- Gazdar, G. (1981). Speech Act Assignment. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of Discourse Understanding*. Cambridge University Press.
- Gisladdottir, R., Chwila, D., Schriefers, H., & Levinson, S. (2012). Speech Act Recognition in Conversation: Experimental Evidence. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 34(34). <https://escholarship.org/uc/item/94n80472>
- Gisladdottir, 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>
- Gisladdottir, R. S., Chwila, D. J., & Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PLoS One*, 10(3), e0120068. <https://doi.org/10.1371/journal.pone.0120068>
- Godfrey, John J., & Holliman, Edward. (1993). *Switchboard-1 Release 2* (p. 14610176 KB) [Data set]. Linguistic Data Consortium. <https://doi.org/10.35111/SW3H-RW02>
- Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2020). *Rstanarm: Bayesian Applied Regression Modeling via Stan. R Package v. 2.19. 2*.
- Grice, H. P. (1975). Logic and Conversation. *Speech Acts*, 41–58. https://doi.org/10.1163/9789004368811_003
- 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>
- Jurafsky, D., Coccaro, N., Ries, K., Shriberg, E., Stolcke, A., Taylor, P., & Ess-Dykema, C. V. (1997). *Johns Hopkins LVCSR Workshop-97 Switchboard Discourse Language Modeling Project Final Report*. 50.
- Levinson, S. (1983). *Pragmatics*. Cambridge University Press.
- Mertens, J., de Ruiter, J. P. (2021). Cognitive and social delays in the initiation of conversational repair. In *Dialogue and Discourse*, 12(1), 21–44. <http://dx.doi.org/10.5210/dad.2021.102>
- Openslr.org. (n.d.). Retrieved March 18, 2022, from <http://www.openslr.org/5/>
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Roberts, S. G., Torreira, F., & Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: a corpus study. *Frontiers in psychology*, 6, 509.
- Sadock, J., & Zwicky, A. (1985). Speech act distinctions in syntax. In *Language typology and syntactic description* (pp. 155–196). Cambridge University Press. <http://arxiv.org/abs/1910.01108>
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. *ArXiv:1910.01108 [Cs]*. <http://arxiv.org/abs/1910.01108>
- Sarathy, V., Tsuetaki, A., Roque, A., & Scheutz, M. (2020). Reasoning Requirements for Indirect Speech Act Interpretation. *Proceedings of the 28th International Conference on Computational Linguistics*, 4937–4948. <https://doi.org/10.18653/v1/2020.coling-main.433>
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge university press.
- Wetzels, R., Matzke, D., Lee M. D., Rouder, J.N., Iverson G. J., & Wagenmakers, E.-J. (2011). Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t

Tests. *Perspectives on Psychological Science*, 6(3), 291-298.
<https://doi.org/10.1177/1745691611406923>