A Machine Learning Perspective on the Pragmatics of Indirect Commands

Matthew Lamm*
Stanford Linguistics
Stanford NLP Group
mlamm@stanford.edu

Mihail Eric*
Stanford Computer Science
Stanford NLP Group
meric@cs.stanford.edu

A major goal of computational linguistics research is to enable organic, language-mediated interaction between humans and artificial agents. In a common scenario of such interaction, a human issues a command in the imperative mood—e.g. *Pick up the box*—and a robot acts in turn [14, 15]. While this utterance-action paradigm presents its own set of challenges [13], it greatly simplifies the diversity of ways in which natural language can be used to elicit action of an agent, be it human or artificial [2, 10, 7, 3, 6]. Computational semantics must leverage the key insight from early speech act theory that most clause types, even vanilla declaratives, instantiate as performative requests in certain contexts [1, 12, 9].

To this end, we employ machine learning to study the use of performative commands in the Cards corpus, a set of transcripts from a web-based game that is constrained so as to require a high degree of linguistic and strategic collaboration [4, 5, 11]. For example, players are tasked with navigating a maze-like gameboard in search of six cards of the same suit, but since a player can hold at most three cards at a time, they must coordinate their efforts to win the game.

The Cards corpus is particularly well-suited to studying the pragmatics of commands because it records both utterances made as well as the actions taken during the course of a game. Where commands are concerned, we can observe who acts in response to an utterance, and test hypotheses about the discourse conditions surrounding an utterance-action exchange.

We focus on a subclass of performative commands that are ubiquitous in the Cards corpus: Non-agentive declaratives about the locations of objects, e.g. "The five of hearts is in the top left corner," hereafter referred to as locatives. Despite that their semantics makes no reference to either an agent or an action—thus distinguishing them from conventional imperatives—locatives can be interpreted as commands when embedded in particular discourse contexts. In the Cards game, it is frequently the case that an addressee will respond to such an utterance by fetching the card mentioned.

We hypothesize that the illocutionary effect of a locative utterance is a function of contextual features that variably constrain the actions of discourse participants, e.g. the card

* Authors contributed equally

Model	F_1
Random	23.5
Bigram	58.9
Explicit Goal	76.2
Full Hand	**82.3**

Table 1: F_1 performance as reported on the test set. Note our baselines are italicized.

mentioned is relevant to a winning collaborative strategy, but the speaker cannot act by picking it up as such.

To test this idea, we manually annotate a dataset of 94 locative utterances in the Cards transcripts that we deem to be truly ambiguous, out of context, between informative and command readings. We also annotate their respective transcripts for a simplified representation of the tabular common ground model described in [8]. Here, we identify the common ground with the state of a game as reflected by the utterances made by both players up to a specific point in time. Finally, we train machine learning classifiers on features of the common ground to predict whether the addressee will act on the information provided by the speaker.

We train standard logistic regression classifiers on the following features.

- **Explicit Goal**: This binary feature is triggered in two cases: 1) When the suit of card mentioned matches the agreed-upon suit strategy in the common ground and 2) When the card mentioned appears in the set of cards the addressee claims to need. This models the prediction that locative utterances are more likely to elicit follow-up action of an addressee when they are relevant to a common goal.

- **Full Hands**: This binary feature is triggered when the speaker has three cards of the same suit as the card mentioned, and which are associated with some winning six-card straight, but the addressee does not. This models the prediction that locative utterances are likely to be indirect commands when they provide information relevant to winning, but only the addressee can act as such.

Two baseline classifiers benchmark our predictive task. Our random baseline predicts the addressee follow-up using a Bernoulli distribution weighted according to the class priors of the training data. The second baseline incorporates surface-level dialogue context via bigram features of all the utterances exchanged between players up to and including the locative utterance in question. We also experimented with a unigram baseline, but found it inferior to that trained on bigrams.

We report the results of our experiments using an F_1 measure and a 0.8/0.2 train/test split of our data in Table 1. We find that of the two baselines, the bigram model performs better. This bigram model also uses 2,916 distinct lexical features which makes it a highly overspecified model for our moderate data size. In contrast, we test our two context-sensitive features one at a time with our logistic regression model, as we are interested in seeing how successfully they encode agents' pragmatic inferences. We find that both of our single-feature context-sensitive models significantly outperform our baselines, thus confirming the hypothesis that discourse context plays an important role in the interpretation of non-agentive declaratives as indirect commands.

References

[1] Austin, J. L. *How to do things with words*. Oxford University Press, 1975.

[2] Clark, H. H. *Using language*. Cambridge University Press, Cambridge, 1996.

[3] Condoravdi, C., and Lauer, S. Imperatives: Meaning and illocutionary force. *Empirical issues in syntax and semantics 9* (2012), 37–58.

[4] Djalali, A., Clausen, D., Lauer, S., Schultz, K., and Potts, C. Modeling expert effects and common ground using Questions Under Discussion. In *Proceedings of the AAAI Workshop on Building Representations of Common Ground with Intelligent Agents* (Washington, DC, November 2011), Association for the Advancement of Artificial Intelligence.

[5] Djalali, A., Lauer, S., and Potts, C. Corpus evidence for preference-driven interpretation. In *Proceedings of the 18th Amsterdam Colloquium: Revised Selected Papers* (Berlin, 2012), M. Aloni, V. Kimmelman, F. Roelofsen, G. W. Sassoon, K. Schulz, and M. Westera, Eds., Springer, pp. 150–159.

[6] Kaufmann, M. Fine-tuning natural language imperatives. *Journal of Logic and Computation* (2016), exw009.

[7] Kaufmann, S., and Schwager, M. A unified analysis of conditional imperatives. In *Semantics and Linguistic Theory* (2009), vol. 19, pp. 239–256.

[8] Malamud, S. A., and Stephenson, T. Three ways to avoid commitments: Declarative force modifiers in the conversational scoreboard. *Journal of Semantics 32*, 2 (2015), 275–311.

[9] Perrault, C. R., and Allen, J. F. A plan-based analysis of indirect speech acts. *Computational Linguistics 6*, 3-4 (1980), 167–182.

[10] Portner, P. Imperatives and modals. *Natural Language Semantics 15*, 4 (2007), 351–383.

[11] Potts, C. Goal-driven answers in the Cards dialogue corpus. In *Proceedings of the 30th West Coast Conference on Formal Linguistics* (Somerville, MA, 2012), N. Arnett and R. Bennett, Eds., Cascadilla Press, pp. 1–20.

[12] Searle, J. R. How performatives work. *Linguistics and philosophy 12*, 5 (1989), 535–558.

[13] Tellex, S., Thaker, P., Deits, R., Kollar, T., and Roy, N. Toward information theoretic human-robot dialog. In *Robotics: Science and Systems* (2012), vol. 2, p. 3.

[14] Tellex, S. A., Kollar, T. F., Dickerson, S. R., Walter, M. R., Banerjee, A., Teller, S., and Roy, N. Understanding natural language commands for robotic navigation and mobile manipulation.

[15] Walter, M. R., Antone, M., Chuangsuwanich, E., Correa, A., Davis, R., Fletcher, L., Frazzoli, E., Friedman, Y., Glass, J., and How, J. P. A situationally aware voice-commandable robotic forklift working alongside people in unstructured outdoor environments. *Journal of Field Robotics 32*, 4 (2015), 590–628.