The Pragmatics of Indirect Commands in Collaborative Discourse

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

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

Abstract

Today’s artificial assistants are typically prompted to perform tasks through direct, imperative commands such as Set a timer or Pick up the box. However, to progress toward more natural exchanges between humans and these assistants, it is important to understand the way non-imperative utterances can indirectly elicit action of an addressee. In this paper, we investigate command types in the setting of a grounded, collaborative game. We focus on a less understood family of utterances for eliciting agent action, locatives like The chair is in the other room, and demonstrate how these utterances indirectly command in specific game state contexts. Our work shows that models with domain-specific grounding can effectively realize the pragmatic reasoning that is necessary for more robust natural language interaction with robots.

1 Introduction

A major goal of robotics 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 (Tellex et al., 2011; Walter et al., 2015). While this utterance-action paradigm presents its own set of challenges (Tellex et al., 2012), it greatly simplifies the diversity of ways in which natural language can be used to elicit action of an artificial agent (Clark, 1996; Portner, 2007; Kaufmann and Schwager, 2009; Condoravdi and Lauer, 2012; Kaufmann, 2016).

In this work, we explore and model a less well-understood type of command: Locative utterances like The dog is outside, whose semantics makes no reference to which agent must act or which action should be taken, but which can be interpreted as commands when embedded in particular discourse contexts.

Consider for example a hypothetical scenario in which two people are setting up a room for a conference and must find chairs elsewhere in the building. One walks into the room carrying two chairs and, before putting them down says to her empty-handed partner, “There is a chair in the room next door.” In one possible scenario, the addressee realizes he has the capacity to act on this information and goes to fetch the chair in question. In another, he simply stands where he is, and in response the speaker puts down the chairs she is carrying, and exasperatedly fetches the chair she had previously mentioned.

Neither the case where the addressee infers that the speaker wants him to act, nor the case where the speaker gets exasperated that he failed to do so, would be unreasonable in the course of a natural, cooperative dialogue.

We hypothesize that the semantic representations of locative utterances alone do not realize the interpretation of these utterances as commands. Rather, when they are embedded within a discourse where agents’ actions are subject to situational constraints, as in our motivating chair example, pragmatic inferences resolve them as commands.

We explore this question using the Cards Corpus, a corpus of transcripts from a game in which pairs of players must collaboratively collect cards scattered on a gameboard (Djalali et al., 2011, 2012; Potts, 2012). In a first study, we examine the distribution of three classes of commands—imperative, non-locative performa-
tive, and locative—surrounding a fixed set of common actions taken in the Cards game. We find that for most actions the imperative strategy is the most prominent, but that the locative command strategy predominates for eliciting card pickups of an addressee.

In a follow-up experiment, we devise a predictive task to study contextual features associated with the interpretation of locative constructions as commands. We manually annotate a set 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 common ground: the state of a game as reflected by the utterances made by both players up to a specific point in time. Using these annotations, we derive a few very powerful contextual features that predict when a locative utterance will be interpreted as a command.

The contributions of our work are therefore as follows: 1) an extensive exploration of command types in the Cards Corpus, 2) an annotated dataset of command types and fine-grained game state in the corpus which we make publicly available, and 3) a predictive model demonstrating contextual features are highly effective in pragmatically resolving command interpretations of locative utterances.

2 Related Work

There is a long formal literature on the semantics of commands. In Portner’s (2007) influential model, imperatives are utterances whose conventional effect updates an abstract “to-do list” of an addressee. More recent debate has asked whether this effect is in fact built into the semantics of imperatives, or if their directive force is resolved by pragmatic reasoning in context (Condoravdi and Lauer, 2012; Kaufmann, 2016).

The subject of indirect commands, of which the locative utterances we describe are an example, has been extensively analyzed in terms of speech act and decision theory (Austin, 1975; Clark, 1979; Perrault and Allen, 1980; Allen and Perrault, 1980; Searle, 1989). A common thread in this line of work is the idea that language understanding is influenced by the beliefs, goals, social positions, reasoning patterns, and physical surroundings of interlocutors, and therefore that a significant component of meaning emerges directly from context.

The present work synthesizes intuitions from the theory of commands with recent computational work on natural language pragmatics (Vogel et al., 2013, 2014; Degen et al., 2013) and collaborative dialogue (Chai et al., 2014). We are particularly influenced by previous work demonstrating the complexity of pragmatic phenomena in the Cards corpus (Djalali et al., 2011, 2012; Potts, 2012).

3 The Cards corpus

The Cards corpus is a set of 1,266 transcripts from a two-player, collaborative, web-based game. It was designed with the explicit purpose of studying natural language use as it pertains to the goals of participants in a discourse (Potts, 2012).

The Cards Corpus is 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.

At the start of the game, player locations are randomly initialized on a two-dimensional, grid-style game board, as shown in Figure 1. Cards from a conventional 52-card deck are scattered randomly throughout the board. Upon signing on, players are prompted with the following task:

Gather six consecutive cards of a particular suit (decide which suit together), or determine that this is impossible. Each of you can hold only three cards at a time, so youll have to coordinate your efforts. You can talk all you want, but
you can make only a limited number of moves.

In addition to the fact that players can only hold three cards at a time, the game is further-constrained in ways that stimulate highly collaborative talk. In particular, while players can see their own location, they cannot see the locations of their partners and so must inquire about them. Players can only see cards within a small neighborhood around their respective locations, and so must explore the board to find relevant cards. Moreover, while some walls are visible, others are invisible and so lead to surprise perturbations in the course of exploring the gameboard.

4 Command types in the Cards Corpus

Players in the game can perform simple actions which are built into the interface, such as picking up or dropping a card, moving around the board, and clicking “Task Complete.” In addition to these are more complex actions like searching, meeting, strategizing, finding, and telling, which entail synthesizing knowledge about the current world state. Instances of commands concerning these actions are ubiquitous in the Cards corpus.

4.1 Imperative and performative commands

Many of the commands in the Cards data are imperatives. For example, with respect to picking cards up:

pick up the 9
pick it up!
and get the 5
or hell, grab the 234 of D
ok when you get here pick up the 8H
grab the 2c 6c and 2d

With respect to dropping (or not dropping) cards:

get rid of 6d. i found 7h
drop the 2
no dont drop it.)
keep the 3
ok drop the 7 i guess

With respect to conversational actions (some of these utterances are shortened for convenience):

tell me where it is
tell me if you see 5 or 6
don’t just say “a lot of cards” [...] 
talk to me dude [...] 
awesome let me know once you have it.

Beyond imperatives, conversational language makes productive use of so-called performative commands. These are utterances which act as commands in context but whose clause type is not conventionally associated with the effect of commanding (Clark, 1979; Searle, 1989; Wierzbicka, 1991). For example, the non-imperative utterances “You need to pick up that box up”, “It would be great if you could pick up that box”, and “Could you pick up that box?” can all be interpreted as indirect commands.

The Cards corpus exhibits a diverse set of performatives, such as the following conditional utterances:

so if you can find 5S,6S,7S that would be great
so if you come and drop the 8h and pick up the 6h we are good
If you will, find that K for me [...] 

Commands involving modal verbs like should, as well as the auxiliary verbs need and have to, are also quite common:

you need to find an A or 6
you should just keep moving [...] 
i think you should pick up the 3h, 5h, and 8h
you will have to go all around to get there

There are subtle distinctions to be made in the semantics of these performative command strategies. However, they are similar to imperatives insofar as they all make reference to an action with a second person agent.

4.2 Locative indirect commands

Imperatives and performatives that mention agents contrast with the lesser understood subclass of performative commands that are the focus of this work. Utterances like “there is a chair in the other room” do not even encode an action with respect to the object mentioned, let alone an agent, but can nevertheless be used to elicit action of an addressee in certain contexts.

As a motivating example, consider the following exchanges between two players describing their respective hands:

P1: 3h, 4h and ks
P2: i have a queen of diamonds and ace
of club
P2: we have a mess lol

Despite Player 2’s concerns, a strategy emerges shortly thereafter when Player 1 finds an additional hearts card:

P1: i have 3h,4h,6h
P2: ok so we need to collect hearts then

At this point in the transcript, all that has been committed to the common ground is that Player 1 has a full hand of three proximal hearts cards that could be relevant to a winning strategy, and Player 2 has two non-hearts cards. This is the very next utterance in the exchange:

P1: there is a 5h in the very top left corner

Player 2 is seen immediately hereafter to navigate to the top left corner, pick up the five of hearts, and confirm:

P2: ok i got it :)

In this exchange, Player 2 appears to understand not only that the five of hearts is relevant to the winning strategy of six consecutive hearts, but also that it makes more sense for her to act on information about its location than it does for Player 1 to do so.

This discourse encapsulates the collaborative reasoning pattern described in (Perrault and Allen, 1980). The speaker assumes that the addressee is a cooperative agent. Thus, the addressee shares in the goal of attaining a winning game state and will act in a way so as to realize that goal. Recognizing the fact that the speaker would have to drop cards relevant to the goal to pick up the card at issue, a cooperative addressee will infer that she should act by picking it up instead. In this way, locative utterances can be indirectly used as commands.

4.3 The breakdown of command-types per action

In a corpus-level examination of the Cards data, we find that that the indirect, locative command strategy is used extensively to elicit pickup actions of an addressee. We establish this fact by manually annotating the first two hundred instances of utterances that are discernibly command-like and analyzing their distribution.

The main challenge faced in annotating commands are instances where a speaker pursues a mixed command strategy to elicit action of an addressee. For example, when a speaker mentions the location of a card and then uses the imperative to tell one’s game partner to go fetch it. As a general rule, where mixed strategies are observed we label them with the “most direct” command strategy involved. Imperative constructions are treated as the most direct, non-locative performatives as the second most direct, and locatives as the least direct.

The distribution of command types for a subset of actions (pickup, drop, conversation, and search) tracked in our annotations is displayed in Figure 2. As depicted, for the majority of actions compelled by a speaker and taken by an addressee, the imperative is the predominant command strategy, followed by non-locative performatives. However locative commands appear to be the dominant strategy for eliciting card pickups in the corpus, constituting nearly half of all such commands observed.

This pattern demonstrates that for certain kinds of actions, it is quite natural to use the least direct, most context-dependent command strategy to elicit action of an addressee.
5 Common Ground Effects on Resolving Locative Utterances

We seek to understand how the discourse context of a game can influence the interpretation of locative utterances as commands. We therefore construct a binary classification task whereby we test how the role of a locative utterance can be resolved in context, evidenced by the actions that are taken as follow-ups to the utterance. In our task, one label denotes addressee follow-up in the form of acting to pick up the card in question, signaling her intention to act, or asking a clarifying question about its whereabouts. The second label denotes that either the speaker acts on their own utterance or neither agent does.

5.1 Annotation Details

Using a random sample of 200 transcripts from the corpus, we identify instances where a locative utterance is made and we annotate the common ground up to this utterance. This gives us 55 distinct transcripts constituting 94 utterances with this particular phenomenon.

Our common ground annotations include the following information about the game state as indicated by players’ utterances: cards in the players’ hands, player location, known information about the existence or location of cards, strategic statements made by players about needed cards, and whether a player is able to act with respect to an at-issue card.

We also include information about the location of the at-issue card and include a reference to the line in the transcript file where the speaker or addressee acts on a given locative utterance. If no agent acts, this field is marked NONE.

5.2 Model Details

Our aim in devising this task is to investigate the extent to which common ground knowledge determines the effect of ambiguous locative utterances. We therefore train a standard logistic regression classifier and experiment with a few carefully designed features that encode constraints on player action, and which should hypothetically trigger the interpretation of locative utterances as indirect commands. We experiment with the following features:

- **Edit Distance**: We use the minimal number of edits for an optimal solution as a feature. Given the cards in the players’ hands at a given point in the game, we can define an optimal solution based on the number of edits that must be made to the hands to achieve that optimal solution. An edit is defined as either picking up or dropping a card, and each such action has a cost of 1. An optimal solution is defined as the one that requires the minimal number of edits given the current hands. For example, if player 1 has a 2H, 3H, and 4H and player 2 has a 6H and a 7H, the optimal solution is the 2H, 3H, 4H, 5H, 6H, and 7H. Such a solution requires a single edit because player 2 simply has to pick up a 5H. This feature seeks to capture the intuition that an addressee should tend to act with respect to a card when the edit distance is not particularly high and hence the game is near a winning state.

- **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 be indirect commands when they are relevant to a well-defined 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.

5.3 Baselines

We provide a number of baselines to help benchmark our predictive task. Our first baseline, which is context-agnostic, seeks to capture the intuition that the role of a locative utterance is entirely ambiguous when considered in isolation. This baseline predicts the agent 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 utterance exchanged between players up to and in-
Model | $F_1$
---|---
Random | 23.5
Bigram | 58.9
Edit Distance | 62.5
Explicit Goal | 76.2
Full Hand | 82.3

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

including the locative utterance. We also experimented with a unigram baseline but found that its performance was inferior to that of the bigram.

5.4 Results

We test our three 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 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 see that of our 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.

We find that our single-feature context-sensitive models all outperform our baselines. Our Edit Distance feature is the weakest for predicting resolution of locative utterances, which may be due to data sparsity as well as the fact that it encodes a more complex form of reasoning than players typically employ in the game. Our Explicit Goal feature outperforms the Edit Distance feature by almost 14%, which indicates that locative utterances are often interpreted as commands in the presence of an explicit, common goal. The Full Hands feature outperforms the Explicit Goal feature by about 6%, which suggests that players are also methodical about incorporating each other’s card-constraints in collaboratively working toward a winning solution. In other words, an addressee of a locative utterance will tend to act accordingly when their partner cannot pick up a card, but they seem to be close to a winning state.

6 Conclusion and Future Work

In this work, we have performed an extensive study of command types as present in the Cards corpus. Using the corpus as a test bed for grounded natural language interaction among agents with a shared goal, we describe a variety of utterances that may function as indirect commands when regarded in context. In particular, locative utterances, which are not conventionally associated with command interpretations, are shown to operate as commands when considered in relation to situational constraints in the course of collaborative interaction. We develop a predictive task to show that models with carefully-designed features incorporating game state information can help agents effectively perform such pragmatic inferences. Future work will look to incorporate such pragmatic reasoning in real-world, open-ended settings as a means of augmenting the natural language understanding abilities of modern-day robots.

References

James F Allen and C Raymond Perrault. 1980. Analyzing intention in utterances. Artificial intelligence 15(3):143–178.

John Langshaw Austin. 1975. How to do things with words. Oxford University Press.

Joyce Y Chai, Lanbo She, Rui Fang, Spencer Ottarson, Cody Littley, Changsong Liu, and Kenneth Hanson. 2014. Collaborative effort towards common ground in situated human-robot dialogue. In Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction. ACM, pages 33–40.

Herbert H Clark. 1979. Responding to indirect speech acts. Cognitive psychology 11(4):430–477.

Herbert H. Clark. 1996. Using language. Cambridge University Press, Cambridge.

Cleo Condoravdi and Sven Lauer. 2012. Imperatives: Meaning and illocutionary force. Empirical issues in syntax and semantics 9:37–58.

Judith Degen, Michael Franke, and Gerhard Jäger. 2013. Cost-based pragmatic inference about referential expressions. In CogSci.

Alex Djalali, David Clausen, Sven Lauer, Karl Schultz, and Christopher Potts. 2011. 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. Association for the Advancement of Artificial Intelligence, Washington, DC.

Alex Djalali, Sven Lauer, and Christopher Potts. 2012. Corpus evidence for preference-driven interpretation. In Maria Aloni, Vadim Kimmelman, Floris Roelofsen, Galit Weidman Sasson, Katrin Schulz, and Matthijs Westera, editors, Proceedings of the 18th Amsterdam Colloquium: Revised Selected Papers. Springer, Berlin, pages 150–159.
Magdalena Kaufmann. 2016. Fine-tuning natural language imperatives. *Journal of Logic and Computation* page exw009.

Stefan Kaufmann and Magdalena Schwager. 2009. A unified analysis of conditional imperatives. In *Semantics and Linguistic Theory*. volume 19, pages 239–256.

C Raymond Perrault and James F Allen. 1980. A plan-based analysis of indirect speech acts. *Computational Linguistics* 6(3-4):167–182.

Paul Portner. 2007. Imperatives and modals. *Natural Language Semantics* 15(4):351–383.

Christopher Potts. 2012. Goal-driven answers in the Cards dialogue corpus. In Nathan Arnett and Ryan Bennett, editors, *Proceedings of the 30th West Coast Conference on Formal Linguistics*. Cascadilla Press, Somerville, MA, pages 1–20.

John R Searle. 1989. How performatives work. *Linguistics and philosophy* 12(5):535–558.

Stefanie Tellex, Pratiksha Thaker, Robin Deits, Thomas Kollar, and Nicholas Roy. 2012. Toward information theoretic human-robot dialog. In *Robotics: Science and Systems*. volume 2, page 3.

Stefanie A Tellex, Thomas Fleming Kollar, Steven R Dickerson, Matthew R Walter, Ashis Banerjee, Seth Teller, and Nicholas Roy. 2011. Understanding natural language commands for robotic navigation and mobile manipulation.

Adam Vogel, Andrés Gómez Emilsson, Michael C. Frank, Dan Jurafsky, and Christopher Potts. 2014. Learning to reason pragmatically with cognitive limitations. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*. Cognitive Science Society, Wheat Ridge, CO, pages 3055–3060.

Adam Vogel, Christopher Potts, and Dan Jurafsky. 2013. Implicatures and nested beliefs in approximate Decentralized-POMDPs. In *Proceedings of the 2013 Annual Conference of the Association for Computational Linguistics*. Association for Computational Linguistics, Stroudsburg, PA, pages 74–80.

Matthew R Walter, Matthew Antone, Ekapol Chuangsawanich, Andrew Correa, Randall Davis, Luke Fletcher, Emilio Frazzoli, Yuli Friedman, James Glass, Jonathan P How, et al. 2015. A situationally aware voice-commandable robotic forklift working alongside people in unstructured outdoor environments. *Journal of Field Robotics* 32(4):590–628.

A Wierzbicka. 1991. *Cross-cultural pragmatics: the semantics of human interaction*. Mouton de Gruyter, Berlin.