A Machine Learning Perspective on the Pragmatics of Indirect Commands

Matthew Lamm and Mihail Eric

Table of Contents

Motivational Example: Comey's Testimony[1]

RISCH: You put this in quotes—words matter. You wrote down the words so we can all have the words in front of us now. Theres twenty-eight words there that are in quotes, and it says, quote, "I hope'"—this is the president speaking—'I hope you can see your way clear to letting this go, to letting Flynn go. He is a good guy. I hope you can let this go. Now those are his exact words, is that correct?
COMEY: Correct.
RISCH: And you wrote them here, and you put them in quotes?
COMEY: Correct.
RISCH: Thank you for that. He did not direct you to let it go.
COMEY: Not in his words, no.
RISCH: He did not order you to let it go.
COMEY: Again, those words are not an order.

...

COMEY: ... the reason I keep saying his words is I took it as a direction... I mean, this is the president of the United States, with me alone, saying, I hope this. I took it as: this is what he wants me to do.

[1] Quotes replicated from [1]

Motivational Example cont'd

COMEY: ... the reason I keep saying his **words** is I took it as a direction... I mean, this is the president of the United States, with me alone, saying, I hope this. I took it as: this is what he wants me to do.

- I take Comey to be saying that, while Trump did not use the "*words*"—i.e. the *grammar*—of commanding, there were features of the context of utterance that led him to believe that Trump's "I hope..." utterance carried the force of a command.
- E.g. The speaker was the president of the United States, and they were speaking in confidence over a private dinner in the White House.

The Comeyan picture of directive force

- The **clause type** of an utterance determines its conventional effect.
  - E.g. A declarative assertion p commits the speaker to the truth of p.
- The **context** of the utterance helps to determine its additional effects.
  - E.g. When constructions like "I hope..." are interpreted as commands by virtue of their being uttered by an important person.

The Comeyan picture of directive force

- The **clause type** of an utterance determines its conventional effect.
  - E.g. A declarative assertion p commits the speaker to the truth of p.
- The **context** of the utterance helps to determine its effects.
  - E.g. When constructions like "I hope..." are interpreted as commands by virtue of their being uttered by an important person.
- The Comeyan picture shouldn't be surprising to anyone here.
- What I find disconcerting however, is there exists no data-driven account of the function which takes context and returns an illocutionary force.

Experimental Approach

- In general: Frame the prediction of a non-imperative utterance's directive force—i.e. **performative command** or **not**—as a machine learning task.

Experimental Approach

- In general: Frame the prediction of a non-imperative utterance's directive force—i.e. **performative command** or **not**—as a machine learning task.
- Use feature-based approach to representing facts about the context(s) of the utterances in our dataset. (e.g. "speaker is president of U.S.")

Experimental Approach

- In general: Frame the prediction of an utterance's directive force—i.e. **performative command** or **not**—as a machine learning task.
- Use feature-based approach to representing facts about the context(s) of the utterances in our dataset. (e.g. "speaker is president of U.S.")
- Learn classifiers (i.e. logistic regression) on these featural representations, and compare the performance of classifiers to see which contextual features are the best regressors of directive force/its absence.

High-level ML[2] overview

- Let $x^{(i)}$ denote an *input* variable. Here, A "featurized" representation—a vector—of an utterance and its context.
- Let $y^{(i)}$ denote its associated *output* variable. Here, whether or not the utterance i was interpreted as having directive force or not.
- Putting these together, let our *training set* be the collection of featurized utterances.

$$\{(x^i, y^i) : i = 1, \ldots, m\}$$

- Let $\mathcal{X}$ be the space in which our input vectors live: here, $\{0, 1\}^n$. And let $\mathcal{Y}$ be the space in which our output vectors live: here, $\{0, 1\}$.
- Then, provided such a training set, a *supervised learning* algorithm "learns" a function

$$h : \mathcal{X} \to \mathcal{Y}$$

such that $h(x)$ is a good predictor of its corresponding y.

[2] Notes summarized from [2]

Desiderata for a dataset...

1. Focus on a single utterance type, whose conventional effect is "far away from" the unconventional effect of directive force.

Desiderata for a dataset...

1. Focus on a single utterance type, whose conventional affect is "far away from" its conventional effect.
2. Simple consistent model world (to assure that one can define coherent, data-backed features)

Desiderata for a dataset...

1. Focus on a single utterance type, whose conventional affect is "far away from" its conventional effect.
2. Simple consistent model world (to assure that one can define coherent, data-backed features)
3. Avoid having to answer questions about the "intonational picture" :)

Desideratum 1: separability of conventional and non-conventional effects

- If possible, we want our dataset to consist of instances of a single utterance type, and we want that utterance type to respect the aforementioned separability.
- Informally, constructions like Trump's "I hope you do x" and "You should do x" are too close to imperatives to satisfy this criterion.

Desideratum 1: separability of conventional and non-conventional effects

- If possible, we want our dataset to consist of instances of a single utterance type, and we want that utterance type to respect the aforementioned separability.
- Informally, constructions like Trump's "I hope you do x" and "You should do x" are too close to imperatives to satisfy this criterion.
- **Our solution**: non-agentive declarative utterances about the locations of objects, which we call "locatives."

Locatives: an aside

CONTEXT: 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."

Locatives: an aside

CONTEXT: 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."

1. The addressee realizes he has the capacity to act on this information and goes to fetch the chair in question.

Locatives: an aside

CONTEXT: 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."

1. The addressee realizes he has the capacity to act on this information and goes to fetch the chair in question.
2. 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.

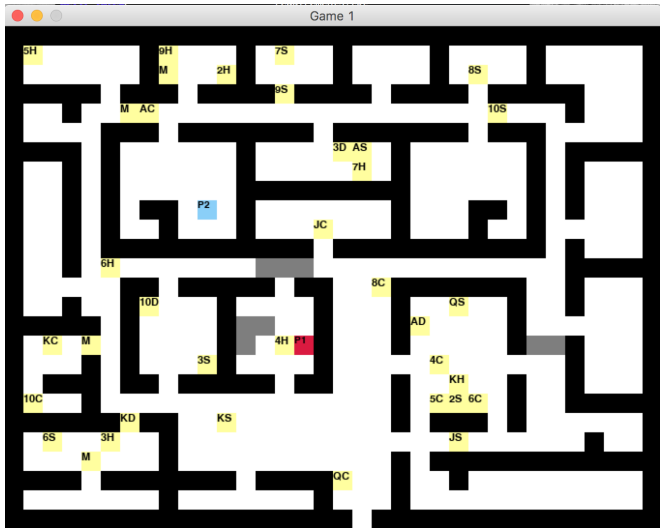
Locatives: an aside

CONTEXT: 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."

1. The addressee realizes he has the capacity to act on this information and goes to fetch the chair in question.
2. 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.

Desiderata 2+3: the Cards corpus[3]



- Two-player, web-based, collaborative game.
- Players are tasked with collecting six cards of the same suit, as they chat over a text interface.
- Constraints:
  - Players cannot see each other.
  - Players cannot see what they are each holding.
  - A single player can only hold three cards at a time.
- Critically, game transcripts record both the utterances made and the actions taken in the course of a game.

[3] For more details on the Cards corpus and its development see [3, 4, 5]

Locative performatives in the Cards corpus

⟨demo⟩

Constructing a training dataset from Cards corpus transcripts

1. Find instances of locatives in a random selection of transcripts that are ambiguous, out of context, as to who should act on them or whether anyone should act on them at all.
   - E.g. Ambiguous case:
   - E.g. Unambiguous case:
2. For each such instance, note whether:
   - **Utterance HAS directive force**: The agent acts on the card—either by moving to pick it up or by asking clarifying questions as to its whereabouts.
   - **Utterance DOES NOT have directive force**: The speaker acts on the card or no one acts on the card.

   The above cases are the *output variables* in our learning task.

Constructing a training dataset from Cards corpus transcripts

1. Find instances of locatives in a random selection of transcripts that are ambiguous, out of context, as to who should act on them or whether anyone should act on them at all.
2. For each such instance, note whether:
   - **Utterance HAS directive force**: The agent acts on the card—either by moving to pick it up or by asking clarifying questions as to its whereabouts.
   - **Utterance DOES NOT have directive force**: The speaker acts on the card or no one acts on the card.

   The above cases are the *output variables* in our learning task.
3. Annotate for the game state ($\sim$CG) as reflected by the utterances made by both players up to the time of the utterance.

Annotating for common-ground game state

A representation of the game state as inferable from the utterances committed to the chat dialogue.

- What cards are in player's hands
  - P1: "i have a 4 of hearts and a king of spades"
- Where are the players
  - P2: "where is 7h?"
  - P1: "it's in the middle room just in the tier under where you got the last card, if that makes sense?"
- What cards to the players say they need?
  - P2: "ok so we need to collect hearts then"
- Can a player act with respect to a card?
  - P1: "6S is located on the left of the screen about half way down. I can't pick it up - my hand is full"
- Which cards has a player mentioned that are not clearly in his or her hand?
  - P1: "I found KS"

Data size

- 94 locative utterances annotated with common ground and output labels.
- .8/.2 train/test split
- We're happy to pass on our annotations!

Features

CONTEXT: 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."

Features

CONTEXT: 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."

1. **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.

Features

CONTEXT: 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."

1. **Explicit Goal**:[4] 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.
2. **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.

[4] Public effective preference?

Baselines

- **Random**: predicts the addressee follow-up using a Bernoulli distribution weighted according to the class priors of the training data.
- **Bigram**: summarizes surface-level dialogue context via bigram features of all the utterances exchanged between players up to and including the locative utterance in question.

Results

F1 is the harmonic mean of precision and recall.
Precision: The number of true positives divided by the number of true positives and false positives.
Recall: The number of true positives divided by the number of true positives and false negatives.

Results

F1 is the harmonic mean of precision and recall.
Precision: The number of true positives divided by the number of true positives and false positives.
Recall: The number of true positives divided by the number of true positives and false negatives.

Model	F_1
Random	23.5
Bigram	58.9
Explicit Goal	76.2
Full Hand	**82.3**
Explicit Goal + Full Hand	77.7

Note: Bigram classifier uses 2,916 features.

Conclusions

- In collaborative tasks, directive force attached to locatives when they refer to objects relevant to an explicit goal, and particularly when the speaker cannot act.
- Single-feature classifiers present a potentially interesting method for testing linguistic hypotheses about context and illocutionary force.
- The role of the common ground is quite critical in computational models of dialogue.

Thanks!

Works cited I


F. Prose, "Words still matter," *The New York Review of Books*, 2017.


A. Ng, "Cs229 lecture notes," *CS229 Lecture notes*, vol. 1, no. 1, pp. 1–3, 2000.


A. Djalali, D. Clausen, S. Lauer, K. Schultz, and C. Potts, "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), Association for the Advancement of Artificial Intelligence, November 2011.


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

Works cited II

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