1 Introduction

Hello and welcome to my dissertation. We are taking the philosophical speech act and defining it computationally. We will look at the various problems that speech acts hope to solve, and how lenses in one discipline fall short in others. To address this, we build a speech act set from scratch, using the linguistic cues of conversation. This data-first approach side-steps descriptional approaches that might not apply to other situations, and is unique to this research.

2 Defining Speech Acts

2.1 Philosophical Usage

Speech acts were first codified up by JL Austin and his student Searle. Their goal was to shore up a philosophy of language in which the spoken-word is truth-functional. Austin noticed that there are some speech acts which are true by virtue of being spoken (performatives), which means they are not simply descriptions of the world, but actions themselves — speech acts. Searle took this idea and ran with it, dictating that there are six speech acts (. . . list of acts. . .). Each act had its own conditions for being uttered felicitously and its own effects on the truth values of the world around.

Here we see that the truth-function of the words we say is paramount.

2.2 Linguistic Usage

The elephant in the room is Chomsky, but Chomsky himself claims to only work at the syntactic level, not semantic or pragmatic. Linguistics at large has still posited cognitive mechanisms in language that might be present to account for speech acts. Transformations, such as tag questions, change the structure of a sentence to create a differently inflected meaning than sentences not including these constructions. The difference between the two can be considered the action that they are performing. At its base, these linguistic analyses has given rise to the direct force hypothesis.

Talk about alternative approaches (e.g. Jackendoff's spatial language).

Here we see that the point of the speech act is to link the intention of speech to the words we say.

2.3 Psychological Usage

Psychologists have operationalized speech acts as pragmatic differences for the ways that words are interpreted in contexts of conversation. Conversation analysis does this by studying sequences within conversation to see how it is organized. For example, adjacency pairs are types of speech that are often found bordering each other (e.g. offer/accept, thank/welcome). Other, more experimental psychologists have looked at speech acts to change behavior. Lena's study shows that speaker switch changes expectations about incoming speech. Gisladottir, Bögels and Levinson used context utterances to show that otherwise identical utterances are processed differently in the brain if their speech act is different.

Here we see that the speech act of the words is contextually dependent.

2.4 Computational Usage

Computer scientists have been interested in speech acts for decades for use in conversational agents. Programming an agent to behave appropriately has turned out to be a tricky task. Developers realized that assuming that the literal words that people say is also what they mean leads to errors. Similarly, there are many ambiguities within language, and the speech act is one of these. A common approach to solving this problem is to create large corpuses of linguistic interactions tagged for speech acts. This leads to problems of which tag-set to use. DAMSL, ISO, $\text{SWBD}_{\text{DAMSL}}$, Matthias's Big 5, Rhetorical Structure Theory and others have proposed to be solutions to speech act problems in agents, but each has fallen short in its own way, showing a disconnect between the motivation and the implementation of these projects.

Here we see that speech acts are for refining the interpretation of language or constraining the responses we get to our own utterances.

2.5 Goals of this work

Each discipline has its own theory and use for speech acts. In this work, we want to draw on each of them in order to have a useful operationalization of speech acts. However, we take a data-first approach. We think that the psychological findings of Warnke & Levinson are promising, but lack theoretical grounding for differentiating speech acts online. DAMSL, ISO, RST, etc., by contrast, all fall short by imposing their own constraints onto the data. Linguistic structure has not proven robust to the open-set of conversational utterances, but their mechanisms of encoding and decoding intention, still seems to be present for people.

Therefore, in this work, we look at meta-linguistic parameters within conversation to detail a speech act set and built a model for recognizing this set. This allows us to create a tag-set, ensure that we are operating at the pragmatic (rather than lexical, syntactic, or even semantic levels), and build from the reasoning that people do every day, rather than assign values of the reasoning we believe them to have done.

3 Descriptive Work

3.1 Definitions of Terms

In this and following chapters, we will use several terms that are used in different ways in different parts of the literature. Therefore, we outline definitions here for:

- Speech Act
- Direct Speech Act
- Indirect Speech Act
- Pragmatic Completion
- Turn Construction Unit
- Turn
- Conversational Context

3.2 Indirect Speech Act Frequency

A re-telling of the CABNC work on indirect speech acts. A few hundred utterances were hand-tagged for sentence type (declarative, interrogative, imperative, none) and sentences were tagged for speech act based on a set motivated by sentence type (statement, question, command). The goal was to find how big of a problem indirect speech acts are in real-world conversation, even given a small set of speech acts. We found that indirect speech acts are somewhat common (˜8%) and that non-sentence utterances were very common (˜40%). We have an outline of where indirect speech acts are found and some ideas about how this study might be extended in future work.

3.3 Indirect Speech Acts and FTOs in Conversation

Paper published with Lena shows that FTOs do not differ in conjunction with being direct / indirect. The hypothesis is that if we did find a longer FTO for indirect speech acts, it could be evidence of extra cognitive processing. We did not find this, but we did find that the data was best explained by a model including only speech act. This suggests that the timing of conversation orients to the social work being done, not any supposed cognitive load. This is also evidence that the pragmatic level of conversation is separated from the syntactic level.

3.4 Turn Lengths

We look at 7 different formulation of Turn and TCU length and find that they are well described by exponential (or geometric) distributions. This means that, en masse, turns and TCUs have a constant hazard rate. We must therefore be projecting the end of turns by other means.

3.5 TRP Durations

TRP durations are sensitive to speaker switch. We found that in about 1/6 of cases, there was a 0ms TRP and the same speaker continued, 1/2 cases there was a speaker switch (TRP ~250ms) and 1/3 cases no speaker switch (TRP ~600ms). This shows that we are sensitive to what is being said when and by whom in conversation.

4 Choosing Data, Parameters, and Models

4.1 Data

The data comes from the Switchboard corpus. We took data from the Jurafsky project that annotated with $\mathrm{SWBD}_{\mathrm{DAMSL}}$ tags and the MSU transcription that had timestamps down to the millisecond. We added our own annotations for sentence type using a DistilBERT model trained on a few hundred hand-tagged examples.

4.2 Parameters

In order of access in a conversation:

4.2.1 Previous Speech Act

A fundamental part of this idea is that our options are limited in response to previous context, therefore previous speech is perhaps the most important parameter of the models.

4.2.2 Previous TRP, Previous Speaker Switch

Based on previous work by Warnke, we expect that speech acts depend upon which speaker takes their turn next.

Previous work by the author shows that speaker-switch is time-sensitive. The pregnant pause paper and preference organization also make the argument that we are more likely to do some speech acts (e.g. accept, agree) faster than others (e.g. reject, disagree). Therefore, we include timing information, too.

4.2.3 Sentence Type Probability

Previous work shows that while it is not perfect, sentence type does inform the speech act in many cases. Our work here seeks a more fine-grained schema than sentence type provides, but that may simply segment the sentence types more finely.

4.2.4 TCU Length

Our work shows that TCU length is exponential, but since turn end is projectable, we suspect that there is some interaction between pragmatic closer and speech act.

4.2.5 Next TRP, Next Speaker-Switch

While next TRP and Next Speaker-switch are not available during the turn itself, the reaction to an utterance may be an important part of its characterization. For example, a rhetorical question may be followed by speaker continuation while an inquiry would be followed by a speaker switch. We include TRP for similar reasons as above.

4.2.6 Next Speech Act

While our chief goal is to understand how responses are constrained by context, it is also important what constraints are created. This is characterized by following speech act.

4.3 Models

We looked into several models for clustering: Affinity Propagation, Mean Shift, DBSCAN, and OPTICS. Each has its own advantages and drawbacks, and since we weren't sure what the geometry of the problem is that we are addressing, we pursued each with the metrics listed below.

4.4 Metrics

Metrics had two types: known ground truth and unknown ground truth. The advantage of known truth was that they worked for all models. The disadvantage was that we know the ground truth had issues. The problem with the metrics that had an unknown ground truth is that they do not play well with the DBSCAN and OPTICS algorithms due to the geometry of those solutions. However, DBSCAN and OPTICS do come with their own metrics that determine whether points are core or non-core for their clusters, which gives a sense for how well-clustered the data is in the model, which is useful in this context.

5 Models and Iterations

5.1 Iterations

While there is considerable debate about the number of speech acts that is appropriate to a schema, our goal was about a dozen. Our thinking is informed by examining the list of acts in competing schemas and background research on preference organization. There were several different types of statement, question, command, and non-sentence utterances that we thought a dozen speech acts would cover without any individual being too rare.

Therefore, we iterated on our clustering algorithm X times, reducing the number of labels in each iteration. Since the accessible, ground-truth parameters were consistent throughout, they should allow us to maintain consistency across iterations. However, since there are more parameters (due to the one-hot encoding) pointing to the previous (and next) labels, this is still a major feature of the model

5.2 Interpretation

The clusters had parameters X, Y, and Z.

In practice, this meant that the algorithm had clustered around utterances like A B and C.

As some examples, we see that label N has words like W1, W2, W3, even though we did not train on words. Here at the end of the work, we can finally add post-hoc descriptive labels to the clusters. Our interpretation of these labels is as follows:

L1	Desc1
L2	Desc2

6 Conclusion and Future Work

This work attempts to re-situate speech acts in conversation into the data of conversation, rather than descriptive analysis. The work shows that we can create a model of conversation where each utterance constrains the next utterance in certain normative, predictable ways. Our model orients to these constraints and builds a set of speech acts based on conversational behavior, which can be used to improve the naturalness of agents, and also to explore how intersubjectivity is maintained in conversation despite ambiguity.