

Computational Semantics and Pragmatics

Autumn 2011

Raquel Fernández

Institute for Logic, Language & Computation
University of Amsterdam



Plan for Today

Introduction to Dialogue Modelling:

- Speech Act theory
- Interpretation of Speech Acts / Dialogue Acts:
 - * Plan-based inference models
 - * Cue-based probabilistic models

Dialogue Modelling

- Dialogue is the most basic setting for **language use**.
- Dialogue Modelling is concerned with designing formal systems that model aspects of conversation.
- Dialogue is a form of **interaction**: it involves multiple participants that need to coordinate.
 - * **content coordination**: utterances in a dialogue are connected to form a coherent discourse; speakers need to avoid misunderstanding.
 - * **interaction coordination**: turn-taking (who speaks when) and integration of language with other modalities (gestures, gaze, ...)
- Dialogue Modelling has strong connections with empirical (corpus-based) and computational (dialogue systems) research.

Dialogue Systems

Dialogue Systems are artificial agents that can communicate with humans using natural language.

- Two different lines of research in the field of dialogue systems:
 - * an **applied**, engineering oriented line, that sees DSs as interfaces that are useful to get some tasks done
 - * a **theoretical**, foundational line, that sees DSs as computational models of language-capable agents and hence as tools for understanding human communication (cognitive science)
- DSs are **end-to-end agents**: they need to say something about all levels of language processing, from perception to understanding and production.
- Implementing DSs forces us to make decisions on how to deal in an operational way with aspects related to language use
↪ *computational pragmatics*

Dialogue Corpora

A corpus is a machine-readable collection of text or speech that can be used for qualitative *and quantitative* analyses of particular phenomena of interest.

Types of information that can be present in dialogue corpora:

- Transcriptions of speech (orthographic or phonetic)
- Audio: speech waveforms
- Video: images of the participants (gestures, gaze...) or the setup

Annotations can be present at each of these three layers

Dialogue Corpora

Two main types of dialogue corpora:

- **Free conversation**
- **Task-oriented dialogue**: more pragmatically constrained setting, allows analysis of dialogue strategies with respect to task success.

An orthogonal dimension of classification: number of participants

- **Two-person dialogue**
- **Multi-party dialogue**: non-trivial turn-taking and addressing, dialogue structure, participant roles, leadership, ...

Corpora of human-computer interaction:

- **Real human-computer dialogue** (for system evaluation)
- **Simulated human-computer dialogue**: Wizard of Oz setup, where the human believes she's interacting with a machine but the machine's operations are simulated in whole or in part by a human (the wizard).

A Dialogue Transcript

From Levinson (1983) on Conversation Analysis (Schegloff 1972).

B: I ordered some paint from you uh a couple of weeks ago some vermilion
A: Yuh
B: And I wanted to order some more the name is Boyd
A: Yes // how many tubes would you like sir
B: U:hm (.) What's the price now eh with V.A.T. do you know eh
A: Er I'll just work that out for you =
B: = Thanks
(10.0)
A: Three pounds nineteen a tube sir
B: Three nineteen is it =
A: = Yeah
B: E::h (1.0) That's for the large tube isn't it
A: Well yeah it's the thirty-seven c.c.s.
B: Er, I'll tell you what I'll just eh eh ring you back I have to work
out how many I'll need. Sorry I did- wasn't sure of the price you see
A: Okay.

Levinson (1983) *Pragmatics*, Cambridge University Press.

Schegloff (1972) Sequencing in Conversational Openings. In *Directions in Sociolinguistics*, pp. 346–380.

Some Key Units of Analysis

Telephone conversation between two participants, Switchboard Corpus:

A.1: Okay, {F um. } / How has it been this week for you? /
B.2: Weather-wise, or otherwise? /
A.3: Weather-wise. /
B.4: Weather-wise. / Damp, cold, warm <laughter>. /
A.5: <laughter> {F Oh, } no, / damp. /
B.6: [We have, + we have] gone through, what might be called the four seasons, {F uh, } in the last week. /
A.7: Uh-huh. /
B.8: We have had highs of seventy-two, lows in the twenties. /

- **Turns**: stretches of speech by one speaker bounded by that speaker's silence – that is, bounded either by a pause in the dialogue or by speech by someone else.
- **Utterances**: units of speech delimited by prosodic boundaries (such as boundary tones or pauses) that form *intentional units* – that is, that can be analysed as an action performed with the intention of achieving something (→ *dialogue acts/speech acts*).

Speech Act Theory

Intuitively, conversations are made up of sequences of **actions** (*questioning, acknowledging,...*). This common-sense view of dialogue is at the root of **speech act theory** – initiated by Austin and developed by Searle in the 60s-70s.

Speech act theory grows out of the following observations:

- There are utterances for which it doesn't make sense to say whether they are true or false.
 - (1) The director bought a new car this year.
 - (2) I apologize for being late.
 - (3) I promise to come to your talk tomorrow afternoon.
 - (4) Put the car in the garage, please.
 - (5) Is she a vegetarian?
- Utterances serve to perform actions.
- The **speech act** of an utterance is an aspect of meaning that cannot be captured in terms of truth-conditional semantics

Felicity Conditions

Speech acts are characterised in terms of **felicity conditions**: conditions under which utterances can be used to properly perform actions (specifications of appropriate use).

Searle identifies four types of felicity conditions:

<i>Conditions</i>	REQUESTING	PROMISING
propositional content:	Future act A of H	Future act A of S
preparatory	a) S believes H can do A b) It is not obvious that S do A without being asked	a) S believes H wants S doing A b) It is not obvious that S would do A in the normal course of events
sincerity	S intends to do A	S wants H to do A
essential	The utterance counts as an attempt to get H to do A	The utterance counts as an undertaking to do A

These conditions can be seen as dimensions on which a speech act can go wrong, but also as constitutive of particular speech acts.

Types of Acts

What are exactly the actions that are performed by utterances?
Austin identifies three types of acts that are performed simultaneously:

- **locutionary act**: basic act of speaking, of uttering a linguistic expression with a particular phonetics/phonology, morphology, syntax, and semantics.
- **illocutionary act**: the kind of action the speaker intends to accomplish, e.g. *blaming, asking, thanking, joking,...*
 - * these functions are commonly referred to as the illocutionary force of an utterance.
 - * the term speech act is commonly used to refer to the illocutionary act/force of an utterance
- **perlocutionary act**: the act by which the locution and illocution of an utterance produce a certain effect on the addressee.

Relations between Acts

Locutionary vs. illocutionary acts:

- The same locutionary act can have different illocutionary forces in different contexts:

The gun is loaded \rightsquigarrow *threatening?* *warning?* *explaining?*

- Conversely, the same illocutionary act can be realised by different locutionary acts:

Three different ways of carrying out the speech act of requesting:

- (6) A day return ticket to Utrecht, please.
- (7) Can I have a day return ticket to Utrecht, please?
- (8) I'd like a day return ticket to Utrecht.

Illocutionary vs. Perlocutionary acts:

- Illocutionary acts are intended by the speaker and are under the speaker's full control.
- Perlocutionary acts are not always intended and are not under the speaker's control.

Speech Act Interpretation

There isn't a one-to-one relation between a locutionary act and its illocutionary force. How can we derive the speech act performed by an utterance?

Two computational models of the interpretation of speech acts:

- **Inferential plan-based models:** based on epistemic logic (beliefs, desires, and intentions - BDI); use of logical inference to reason about the speaker's intentions.
- **Probabilistic cue-based models:** the surface form of the sentence is seen as a set of cues to the speaker's intentions; use of probabilistic machine learning models.

Both models use a kind of inference: the hearer infers something that was not contained directly in the semantics of the utterance.

Daniel Jurafsky (2004) Pragmatics and Computational Linguistics. *Handbook of Pragmatics*. Oxford: Blackwell.

Plan-based Inference Models

Plan-based approaches are motivated by **indirect speech acts**.

According to the so-called **literal force hypothesis** each type of surface form is conventionally associated with a particular illocutionary force. Indirect speech acts are derived by inference.

(9) Can you pass me the salt?

Literal speech act: *question*

Indirect speech act after an inference chain: *directive* (pass me the salt)

The BDI model is based on three components:

- an axiomatization of belief / desire / intention, and of action and planning inspired originally by the work of Hintikka (1969)
- a set of plan inference rules, which codify the heuristics of the system
- a theorem prover

Allen & Perrault (1980) Analyzing Intention in Utterances, *Artificial Intelligence* 15(3).

Perrault & Allen (1980) A Plan-based Analysis of Indirect Speech Acts, *Computational Linguistics* 6(3):167-182.

Plan-based Inference Models

Given these three components and an input sentence, a plan-inference system can interpret the correct speech act by simulating an inference chain along the following lines, as suggested by Searle:

1. X has asked me a question about whether I have the ability to pass her the salt.
2. I assume that X is being cooperative in the conversation (in the Gricean sense) and that her utterance therefore has some aim.
3. X knows I have the ability to pass her the salt, and there is no alternative reason why X should have a purely theoretical interest in my ability.
4. Therefore X's utterance probably has some ulterior illocutionary point. What can it be?
5. A preparatory condition for a directive is that the hearer have the ability to perform the directed action.
6. Therefore X has asked me a question about my preparedness for the action of passing X the salt.
7. Furthermore, X and I are in a conversational situation in which passing the salt is a common and expected activity.
8. Therefore, in the absence of any other plausible illocutionary act, X is probably requesting me to pass her the salt.

Drawbacks of the BDI Approach

The model requires that the utterance has a single literal meaning on which the inference rules operate to produce a non-literal interpretation.

- there is no clear evidence (psycholinguistic or otherwise) for the temporal primacy of literal interpretation.
- for many speech act types beyond question, statement, and request, it is not clear what the literal force would be. (e.g. *'yeah'*)

The Cue-based Model

The cue-based model downplays the role of literal meaning and thinks of the listener as using different cues in the input to help decide how to build an interpretation.

- use of several sources of knowledge (cues): lexical, collocational, syntactic, prosodic, conversational-structure.

The cue-based model represents a more empirical tradition which goes beyond the classic notion of speech acts to model more kinds of conversational functions → **dialogue acts**

- The concept of **dialogue acts** extends the notion of speech act by paying attention to the functions that utterances play in a broader sense. It is inspired by different research areas:
 - * work on grounding by Clark and colleagues (more on this next week)
 - * work on dialogue structure (connection to previous/coming utterances)

Cues for the DA Identification

The cue-based approach relies on identifying features that are probabilistically associated with some dialogue act.

The characteristic lexical, grammatical, prosodic, and conversational properties of DAs are sometimes called their *micro-grammar*

- **Lexical and Syntactic Cues:** words/phrases that occur more often in particular DAs. presence of particular words, such as 'please' (requests), word order (questions), tag particle 'right?' in final position (declarative questions or checks)
- **Prosodic Cues:** final pitch rise (polar questions and declarative questions); loudness or stress can help distinguish 'yeah' agreement from backchannel.
- **Conversational Structure Cues:** 'No it isn't' is an agreement after 'It isn't raining' and a disagreement after 'It is raining'. 'yeah' is more likely to be an agreement after a proposal. (↗ adjacency pairs)

Cue-based Algorithms

Cue-based models are supervised machine learning algorithms trained on a dialogue corpus hand-labeled with DAs.

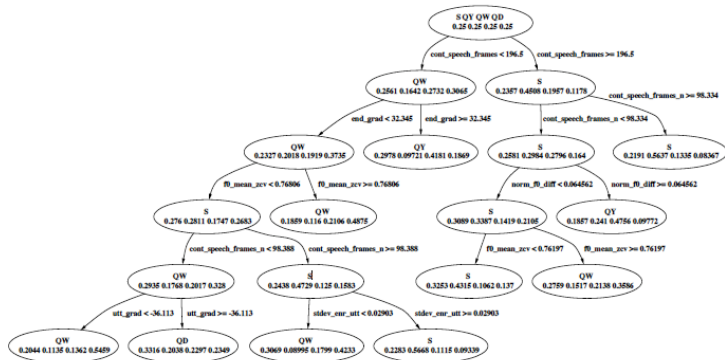
Typically, a statistical classifier is trained for each particular type of DA. The classifier learns to recognise the combination of features that suggests the presence of a question, an assessment, an inform...

Given the observed cues c , the goal is to find the DA d^* that has the maximum posterior probability $P(d|c)$ given those cues.

$$\begin{aligned}d^* &= \operatorname{argmax}_d P(d|c) \\ &= \operatorname{argmax}_d P(d)P(c|d)\end{aligned}$$

We need to choose the DA that maximises the product of two probabilities: the prior probability of a DA $P(d)$ and the likelihood $P(c|d)$ of observing a particular combination of features when a particular DA is present.

An Example Decision Tree



Decision tree for the classification of statements (S), yes-no questions (QY), wh-questions (QW) and declarative questions (QD), using acoustic features (slope of F0, average energy, duration measures, etc.)

Shriberg et al. (1998) Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech? *Language and Speech*, 41:439-487.

Summing Up

Speech act interpretation:

- plan-based inference models: use of reasoning that seems critical for cooperative conversation, but limited coverage and expensive computational methods
- cue-based models: features of the surface form as the basis for inference, wider coverage and more efficient computational methods (although need for annotated corpora)

Next week:

- taxonomies of dialogue acts
- grounding: the process by which participant build common ground