

Computational Semantics and Pragmatics

Autumn 2013



Raquel Fernández
Institute for Logic, Language & Computation
University of Amsterdam

Outline

Last lecture:

- dialogue act and dialogue coherence
- turn taking

We'll discuss logistics of the course at the end.

From Speech Acts to Dialogue Acts

The concept of **dialogue act** (DA) extends the notion of speech act to incorporate ideas from conversation analysis and grounding models of dialogue.

It is the term favoured within computational linguistics to refer to the function or the role of an utterance within a dialogue.

- Taxonomies of DAs aim to cover a broader range of utterance functions than traditional speech act types
 - * importantly, they include grounding-related DAs (meta-communicative).
- They aim to be effective as tagsets for annotating dialogue corpora.

Dialogue Act Taxonomies: DAMSL

One of the most influential DA taxonomies is the DAMSL schema (Dialogue Act Markup in Several Layers) by Core & Allen (1997).

- Communicative Status
- Information Level
- Forward-looking Function
- Backward-looking Function

Explore the annotation manual:

<http://www.cs.rochester.edu/research/speech/damsl/RevisedManual/RevisedManual.html>

Utterances can perform several functions at once: possibly one tag per layer.

The taxonomy is meant to be general but not totally domain independent \rightsquigarrow it has been adapted to several types of dialogue.

DA Taxonomies: SWBD DAMSL

The SWBD DAMSL schema is a version of DAMSL created to annotated the Switchboard corpus. Here are the 18 most frequent DA in the corpus:

Tag	Example	Count	%
Statement	<i>Me, I'm in the legal department.</i>	72,824	36%
Continuer	<i>Uh-huh.</i>	37,096	19%
Opinion	<i>I think it's great</i>	25,197	13%
Agree/Accept	<i>That's exactly it.</i>	10,820	5%
Abandoned/Turn-Exit	<i>So, -/</i>	10,569	5%
Appreciation	<i>I can imagine.</i>	4,633	2%
Yes-No-Question	<i>Do you have to have any special training</i>	4,624	2%
Non-verbal	<i><Laughter>, <Throat_clearing></i>	3,548	2%
Yes answers	<i>Yes.</i>	2,934	1%
Conventional-closing	<i>Well, it's been nice talking to you.</i>	2,486	1%
Uninterpretable	<i>But, uh, yeah</i>	2,158	1%
Wh-Question	<i>Well, how old are you?</i>	1,911	1%
No answers	<i>No.</i>	1,340	1%
Response Ack	<i>Oh, okay.</i>	1,277	1%
Hedge	<i>I don't know if I'm making any sense</i>	1,182	1%
Declarative Question	<i>So you can afford to get a house?</i>	1,174	1%
Other	<i>Well give me a break, you know.</i>	1,074	1%
Backchannel-Question	<i>Is that right?</i>	1,019	1%

The average conversation consists of 144 turns, 271 utterances, and took 28 min. to annotate. The inter-annotator agreement was 84% ($\kappa=.80$).

<http://www.stanford.edu/~jurafsky/manual.august1.html>

DAs and Coherence

Why are dialogue acts so important? They play an important role in determining the coherence of a dialogue

In abstract terms a dialogue can be modelled as:

- A set S of **dialogue states**
- A set M of **dialogue acts** (“moves”)
- An **update function** $\delta : (S \times M) \rightarrow S$
- m is a **coherent next move** at a state s iff $\delta(s, m)$ is defined.

Several issues need to be worked out in detail, including:

- what information do dialogue states keep track of?
- what is the inventory of dialogue acts? DA taxonomies address this
 - * how do we recognise the DA of an utterance?
- what is the exact specification of the update function?
- what strategy can be used to choose a next dialogue act from a set of possible coherent next moves?

DA Interpretation: Cue-Based Model

How do we recognise the DA performed by an utterance?

The most common approach in computational linguistics is to use a probabilistic cue-based model:

- the listener uses cues in the input to infer a particular interpretation.
- use of several sources of knowledge: lexical, collocational, syntactic, prosodic, conversational-structure (the *micro-grammar* of each DA)
- **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. (\rightsquigarrow adjacency pairs)

Some References

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

Stolcke et al. (2000) Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech, *Computational Linguistics*, 26(3).

Keizer et al. (2002) Dialogue act recognition with Bayesian networks for Dutch dialogues. *Proc. SIGdial*

Klüwer et al. (2010) Using Syntactic and Semantic based Relations for Dialogue Act Recognition, *Proc. COLING*

Cuayáhuitl et al. (2013) Impact of ASR N-Best Information on Bayesian Dialogue Act Recognition. *Proc. SIGdial*

Dialogue Grammars

Let's assume we can recognise the DA performed by an utterance.

One possibility to account for coherence is to model the possible sequences of DAs by means a **dialogue grammar**.

- we may use a finite-state machine (regular grammar)
- we may use more powerful grammars
- we may use a probabilistic language model for sequences of DAs

These methods are used by simple commercial systems in limited domains. Overall they are too restrictive, they impose a structure.

Polanyi, Livia & Remco Scha (1984). A syntactic approach to discourse semantics. *Proc ACL*

Inferential Plan-Based Models

Another possibility is to model coherence using **logical inference** to reason about the intentions of the dialogue participants.

- based on epistemic logics (beliefs, desires, and intentions - BDI)
- The **BDI approach** is meant to be a general model of rational action that can be applied to conversation
- It proposes an axiomatization of BDI to account for
 - * what motivates our actions
 - * how to understand actions by others
- Therefore BDI approaches model in one single framework:
 - * DA recognition
 - * the states of the dialogue as the epistemic states of the participants
 - * the update function as logical inference

Inferential Plan-based Models

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
- a theorem prover

Plan-based approaches aim to explain **indirect speech acts**.

(1) Can you pass me the salt?

↪ Literal speech act: *yes-no question*

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

and also, for instance, answers that appear to be overinformative:

(2) Customer: When does the train to Montreal leave?

Clerk: At 3:15 at gate 7.

↪ the clerk recognises the plan of the customer and identifies possible obstacles and relevant information to solve them

Can you pass me the salt?

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.

BDI Approaches

For more details on the BDI axiomatization and the plan-inference rules see Jurafsky (2004) for a short summary and the original papers by Allen et al.

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

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.

Main influences of these approaches:

- Austin's and Searle's characterisation of speech acts in terms of felicity conditions that appeal to the mental attitudes of speakers
- Hintikka's logic of belief

BDI approaches have been used as the basis to implement conversational agents in the TRAINS/TRIPS projects.

- see the project's website for access to a dialogue corpus collected to develop the system, movies of the system in action, and links to publications. <http://www.cs.rochester.edu/research/trains/>

Allen et al. (2001) Towards Conversational Human-Computer Interaction, *AI Magazine*.

Allen et al. (2001) An architecture for more realistic conversational systems, in *Proc. of Intelligent User Interfaces*.

Information State Update Approaches

An alternative (less inference intensive) to BDI models is to focus on the public/conventional aspects of dialogue (*common ground*) rather than the private attitudes of the participants.

This is known as the **ISU approach**, which is influenced by the work of philosophers such as Lewis and Stalnaker:

- the dynamics of dialogue can be modelled using a game metaphor: participants (players) make moves that update an evolving *conversational scoreboard* that represents the information that has become common as a result of the dialogue.

Lewis. 1979. Score keeping in a language game. *Journal of Philosophical Logic*.

Stalnaker. 1979. Assertion. In *Syntax and Semantics IX*. Academic Press.

Carlson. 1983. Dialogue Games. *Synthese Language Library*. D. Reidel.

This is also the main idea of dynamic semantics:

- dialogue states represent configurations of the conversational scoreboard (the context)
- dialogue acts are context-change operators

Information States

The term **Information state** (IS) refers to the state of the dialogue: the dialogue context that gets updated with each dialogue move.

Different theories of the dynamics of dialogue will represent ISs differently. Some common IS components are:

- the commitments of the dialogue participants
- a stack of questions under discussion (QUD)
- the latest move made in the dialogue
- grounded and ungrounded information

ISs are typically represented as feature structures. For instance:

COM	<i>Set of Propositions</i>
QUD	<i>Stack of QUDs</i>
MOVES	<i>List of moves</i>
PENDING	<i>List of moves</i>

Ginzburg (2012) *The Interactive Stance*, OUP.

Update Rules

Dialogue acts trigger IS updates. Update rules are specified in terms of:

- **preconditions**: information that must hold in the IS for the rule to be applied
- **effects**: the resulting IS after application of the rule

Dialogue acts can be described according to their IS update potential. For example:

- Questions add an element to QUD
- Answers eliminate an element in QUD
- Acknowledgements move information from PENDING to MOVES and COMMITMENTS
- ...

Traum & Larsson (2000) The Information State Approach to Dialogue Management. In *Current and New Directions in Discourse and Dialogue*, pp. 325–353.

Interim Summary

- DA taxonomies: inventories of dialogue acts
- DA recognition is often modelled as cue-based probabilistic inference
- Approaches to dialogue coherence:
 - * **Dialogue grammars** describe and restrict possible sequences of DAs
 - * **BDI approaches** are general models of rational action; they model actions in conversations by reasoning with the mental attitudes of the participants
 - * **ISU approach** focuses on the public aspects of dialogue, on the common ground built by the participants during a conversation, and in how dialogue acts update the conversational scoreboard.

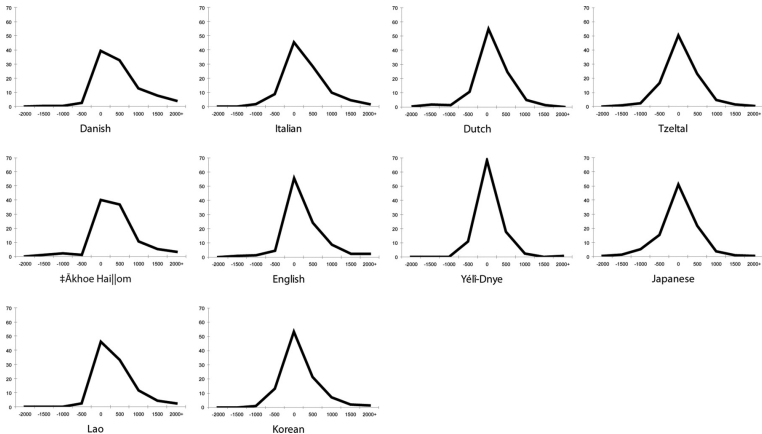
Note that “Information State Update” is in fact a general term that covers all these approaches, but is typically used to refer to the latter.

Turn Taking

Dialogue participants do not only need to make decisions about what to say, but also about *when to say it* \rightsquigarrow **timing**

- Turn-taking is one of the fundamental organisational principles of conversation.
- It is learned within the first 2 years of life
- There are some individual and cultural differences
 - * shy people pause longer and speak less and less often
- But there are **strong universal patterns**:
 - * tendency to minimize both overlap and gaps between turns

Distribution of Turn Transitions in 10 Languages



Stivers et al. (2009) Universals and cultural variation in turn-taking in conversation, *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*.

Turn Taking is Predictive

- Turn-taking happens very smoothly:
 - * Overlaps are rare: on average, less than 5% of speech.
 - * Inter-turn pauses are very short: $\sim 200\text{ms}$ (less than 500ms.)
 - ▶ even shorter than some intra-turn pauses
 - ▶ shorter than the motor-planning needed to produce the next utterance
- Turn-taking is not reactive but **predictive**.

Conversation Analysis Model

The seminal model of turn-taking was put forward by sociologists within the framework of Conversation Analysis (Sacks et al. 1974)

- According to this model, turns consist of **turn constructional unites** (TCUs) with **projectable** points that can be predicted beforehand.
- Such projectable points act as **transition relevance places** (TRPs) where turn transitions are relevant.
- Three rules govern the expected behaviour at TRPs:
 1. if devices to select a next speaker (e.g. questions) are used, the selected speaker takes the turn; else
 2. any other party may take the turn, or
 3. if no other participant takes the turn, then the current speaker may continue.

Predictions: no overlap and no gaps as the norm

Sacks, Schegloff, & Jefferson (1974) A simplest systematics for the organization of turn-taking in conversation. *Language*, 50:735–99.

Grounding Utterances and Turn-Taking

Backchannels ('*uhu*', '*mhm*') do not follow the CA model:

- frequently produced in overlap;
- not meant and not perceived as attempts to take the floor;

According to Clark (1996), the CA turn-taking rules do not apply to utterances at the **meta-linguistic level** of interaction:

- backchannels do not indicate floor competition
- their placement determines which part of the utterance they react to.

Clarification requests have slightly different constraints:

- they involve turn switching
- but the preceding turn can be resumed smoothly

- (3) A: They X-rayed me, and took a urine sample, took a blood sample.
Er, the doctor. . .
B: Chorlton?
A: Chorlton, mhm, he examined me. . .

Factors of Turn Delay

If gaps occur, it is typically for a reason \rightsquigarrow silence is informative.

- Non-answer replies are slower than answers to questions.
- Disconfirmation responses are slower than confirmations.
- Questions are responded to faster if the speaker looks at the addressee.

Stivers et al. (2009) Universals and cultural variation in turn-taking in conversation, *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*.

Turn-Taking Models

Subsequent research has focused on how to make more precise the notions of TCU and TRP.

- **How can TRPs be predicted?** humans are able to predict whether an utterance will continue and for how many words.
 - * syntactic closure plus acoustic information (rising/falling intonation; faster speaking rate);
 - * syntactic completion is context-dependent - pragmatic completion;
 - * certain prosodic patterns signal that the speaker plans to hold the turn beyond syntactic completion;
 - * word fragments and filled pauses are indicative of turn-hold.

Some recent references (not only on TRP prediction):

Gravano et al. (2009) Turn-yielding cues in task-oriented dialogue, *Proc. SIGdial*

K. Laskowski (2010) Modeling Norms of Turn-Taking in Multi-Party Conversation, *Proc. ACL*

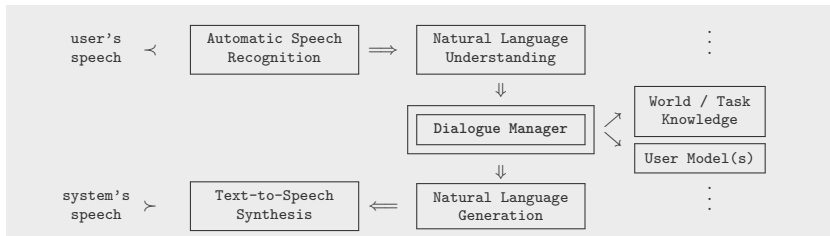
Selfridge and Heeman (2010): Importance-Driven Turn-Bidding for Spoken Dialogue Systems. *Proc. ACL*

Niebuhr et al. (2013) Speech Reduction, Intensity, and F0 Shape are Cues to Turn-Taking, *Proc. SIGdial*

Marisa Casillas (2013) Learning to take turns on time: perception and production processes involved in keeping inter-speaker gaps short. PhD thesis.

To Finish: Turn-taking Demo

Traditional architecture of a dialogue system:



Incremental architectures are currently being developed where modules receive input from other modules as available, and information flows in both directions, with “later” modules informing “previous” ones

- Demonstration video of the ‘*Numbers System*’, which implements incremental dialogue processing for smooth turn-taking:

<http://wwwhomes.uni-bielefeld.de/dschlangen/inpro/videos.html>

Skantze & Schlangen (2009) Incremental Dialogue Processing in a Micro-Domain, in *Proc. of SIGdial*.

Aist et al. (2006) Software architectures for incremental understanding of human speech, *Proc. Interspeech/ICSLP*.

Schlangen and Skantze (2009) A general, abstract model of incremental dialogue processing, in *Proc. of EACL*.

Final Projects

Any topic related to the themes covered in the course. A few ideas on possible types of projects (abstracting over particular topics):

- a quantitative corpus study of some interesting phenomenon
- a machine learning experiment using an existing corpus
- an analysis of data collected by yourself in an experiment
- an analysis and small extension of a paper from the literature
- an analysis of interesting connections between different approaches
- an extension of one of the homework exercises
- ...

Some options in this list may seem unfeasible to you, but they may be perfectly possible – don't abandon an interesting idea before discussing it with me!