UiO **University of Oslo**

INF5820:
What is spoken dialogue?

Pierre Lison,
Language Technology Group (LTG)
Department of Informatics

Fall 2014

Outline

- What is spoken dialogue?

- A cognitive perspective

- Summary

Outline

- ## What is spoken dialogue?
  - ### Turn-taking
  - ### Dialogue acts
  - ### Grounding
  - ### Conversational implicatures
  - ### Deixis
- A cognitive perspective
- Summary

What is dialogue?

- Spoken ("verbal") + possibly non-verbal interaction between two or more participants

- Dialogue is a joint, social *activity*, serving one or several purposes for the participants

- What does it mean to view dialogue as a **joint activity**?

Turn-taking

- ## Dialogue participants take *turns*

  - ### Turn = continuous contribution from one speaker

  - ### Turn-taking is essentially a *resource allocation problem*

- ## Surprisingly fluid in normal conversations:

  - ### Minimise both gaps (no speaker) and overlaps (more than one speaker)

  - ### Interval between speakers is around 250 ms

[Duncan (1972): «Some Signals and Rules for Taking Speaking Turns in Conversations», in *Journal of Personality and Social Psychology*]

Turn-taking (2)

- ## How are turns taken or released?

- ## Wide variety of markers for turn boundaries:

  - ### Syntactic/semantic information (complete grammatical unit)

  - ### Dialogue structure (greetings followed by greetings, questions followed by answers, etc.)

  - ### Intonation (if falling, often signals that the speaker is finished)

  - ### Non-verbal cues such as eye gaze, gestures

  - ### Silence and hesitation markers (unfilled pauses ≠ filled pauses)

  - ### Social conventions (e.g. social status of speakers)

Example of turn-taking

Speaker 1:   han vil bo i skogen ?

Speaker 2:   # altså hvis jeg hadde kommet og sagt " skal vi flytte i skogen ? " så hadde han sagt ja

Speaker 1:   mm

Speaker 2:   men jeg vil ikke bo i skogen

Speaker 1:   nei det skjønner jeg

Speaker 2:   så vi må jo finne et sted som er mellomting og det jeg vil ikke bo utpå landet # i hvilken som helst (uforståelig) ...

**Speaker 1:**   * men det kommer jo an på hvor i skogen da

[«*Norske talespråkskorpus - Oslo delen*» (NoTa), collected and annotated by the Tekstlaboratoriet]
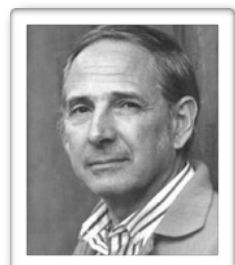

Dialogue acts

- ## Each utterance is an *action* performed by the speaker

  - The speaker has a specific **goal** (which might be only to establish or maintain *rapport* with the listeners)

  - The utterance produces specific **effects** upon the listeners, or the world

  - «*Language as action*» perspective

    [J. L. Austin (1955), *How to do things with words.*]
    [J. Searle (1969), *Speech Acts.*]
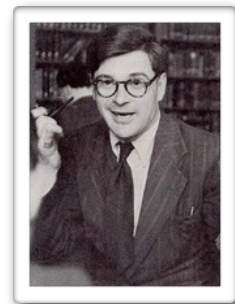
J.L. Austin (1911-1960)
philosopher of language

J. Searle (1932, -)
philosopher of language

Dialogue acts: example

- The mother reaction has a specific *purpose*

  - Communicating her suprise/anger, and stop Calvin

- Her question will trigger some *effects*:

  - A psychological reaction from Calvin (e.g. surprise)

  - Possibly a real-world effect as well (Calvin stopping his action)

Dialogue acts

- Searle's taxonomy:

  - **Assertives**: committing the speaker to the truth of a proposition. E.g.: *«The exam will take place on December 18th»*

  - **Directives**: attempts by the speaker to get the addressee to do something. E.g. : *«could you please clean up your room?»*

  - **Commissives**: committing the speaker to some future course of action. E.g.: *«I promise I'll clean up my room»*.

  - **Expressives**: expressing the psychological state of the speaker about a state of affairs. E.g.: *«thanks for cleaning up your room»*.

  - **Declaratives**: bringing about a different state of the world by the utterance. E.g.: *«You're fired»*.

Dialogue acts

- Current dialogue research is based on more elaborate taxonomies than Searle's *Speech Acts*

  - Various annotation frameworks

  - More focus on conversational phenomena

- Dialogue acts can be richly structured, with both:

  - an *internal* structure (arguments, adjuncts, etc.)

  - an *external* structure (rhetorical relations, references, etc.)

- We'll go into more details about this during the lecture on «spoken dialogue understanding»

Grounding

- Dialogue is a *joint, collaborative process* between the participants

  - Need to ensure *mutual understanding*

- Realised via the gradual expansion and refinement of their *common ground (CG)*

  - Common ground = shared knowledge

Speaker A's knowledge

Common ground

Speaker B's knowledge
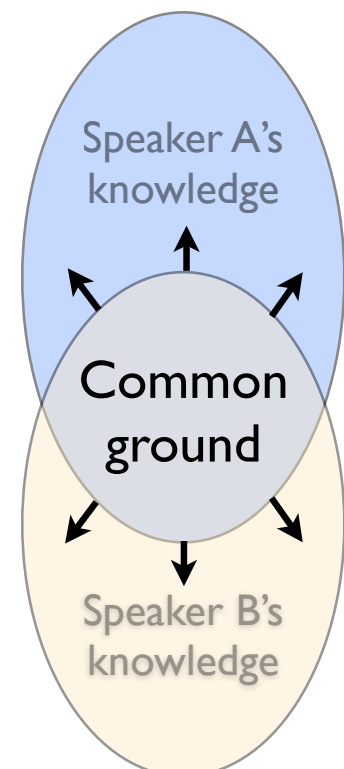
[H. H. Clark and E. F. Schaefer (1989),
«Contributing to discourse», in *Cognitive Science*]

Grounding

- *Grounding* is defined the process of gradually augmenting the common ground during the interaction

- Variety of signals and strategies

- Multiple levels:

  - *Contact* (attention to interlocutor)

  - *Perception* (detection of utterance)

  - *Understanding* (comprehension of utterance)

  - *Attitudinal* reactions

  [Jens Allwood (1992), «On discourse cohesion»,
  in *Gothenburg papers in Theoretical Linguistics*.]

Herbert H. Clark
psycholinguist

Jens Allwood (1947,-)
linguist

Grounding

- Grounding signals:

  - *Backchannels*: «uh-uh», «mm», «yeah»

  - *Explicit feedback*: «ja det skjønner jeg»

  - *Implicit feedback*: A: «I want to fly to Rome» → B: «there are two flights to Rome on Wednesday: ... »

- *Clarification* strategies:

  - «Did you mean to Rome or to Goa?», «could you confirm that ...»

- *Repair* strategies:

  - «OK, you're not going to Goa. Where do you want to go then?»

Examples of grounding

Speaker 1:	vi vasker den hver dag vi # vi har mopp
Speaker 2:	**mm** ## **ja** det er fort og faren til M27 legger nytt teppe han # det er gjort på to timer ## så det er fort gjort
Speaker 1:	**ja** ## da er ikke noe sak
Speaker 2:	vi har skifta teppe tre ganger allerede han gjør det gratis
Speaker 1:	**hæ ?**
Speaker 2:	vi har skifta teppe tre ganger og # han han ...
Speaker 1:	* **jeg skjønner ikke** hvorfor dere har teppe
Speaker 2:	jeg syns det var rart jeg **òg** # men e # (sibilant)

[«*Norske talespråkskorpus - Oslo delen*» (NoTa), collected and annotated by the Tekstlaboratoriet]

Examples of grounding

Speaker 1:	e # **nei** det er ikke mange
Speaker 2:	**ja * nei**
Speaker 1:	men heldigvis så var ikke Petter Rudi tatt ut denne gangen da
Speaker 2:	**ja** # jeg skjønner ikke hva han skal på landslaget å gjøre
Speaker 1:	* **nei han har ingen ting på landslaget**
Speaker 2:	**nei** # definitivt
Speaker 1:	å gjøre # han er ubrukelig
Speaker 2:	* moldensere
Speaker 1:	**hm?**
Speaker 2:	ja disse moldenserne
Speaker 1:	**en gang til?**
Speaker 2:	disse moldenserne
Speaker 1:	* **å ja** (fremre klikkelyd) # unnskyld # jeg hørte ikke hva du sa

implicit feedback
(repetition of *landslaget*)

clarification requests

[«*Norske talespråkskorpus - Oslo delen*» (NoTa), collected and annotated by the Tekstlaboratoriet]

Grounding

- Common ground is more than «knowledge than happen to be shared by all participants»

  - The participants must also know that it is shared (i.e. know that the others know it as well)

  - With two speakers A and B and common ground CG:

  $$\forall x, \ CG(x) \rightarrow \ knows(A, x)$$
  $$\wedge \ knows(B, x)$$
  $$\wedge \ knows(A, knows(B, x))$$
  $$\wedge \ knows(B, knows(A, x))$$
  $$\wedge \ knows(A, knows(B, knows(A, x)))$$
  $$\wedge \ ... \ \textit{ad infinitum}$$

Conversational implicatures

- Very often, part of the meaning of utterance is not explicitly stated, but only implied

  A: «Is William working today?»

  B: «He has a cold»

- How can we retrieve this «suggested» meaning, and go beyond literal interpretations?

  - Need to make some *assumptions* about the speaker's behaviour to help us infer the hidden part

Conversational implicatures

- Same idea again: dialogue as a *collaborative process*

- Grice's *Cooperative Principe*:

  - **Maxim of Quality**: «be truthful»

  - **Maxim of Quantity**: «be exactly as informative as required»

  - **Maxim of Relation**: «be relevant»

  - **Maxim of Manner**: «be clear»

Paul Grice (1913-1988)
philosopher of language

[Paul Grice (1975), *Logic and Conversation.*]

Conversational implicatures

- Based on the cooperative principle, one can draw *conversational implicatures*

  - All participants are assumed to adhere to the maxims

  - If an utterance initially seems to deliberately violate a maxim, the listener will then *infer* additional hypotheses required to make sense of the utterance

Conversational implicatures

A: «Is William working today?»

B: «He has a cold»

- At first glance, B seems to violate the maxim of relevance - he does not directly answer A's question

- But looking at the utterance more closely, we can read it as *implying* that (due to his cold) he is probably at home, and thus not working today

- This is because we assume that B is cooperative and wouldn't have uttered «he has a cold» if it didn't help answering A's question

---

Conversational implicatures

Hobbes' question is *suggesting* something about Calvin's need for schooling, without stating it explicitly

We can understand it because we assume that Hobbes' contribution is cooperative and thus relevant to the discussion
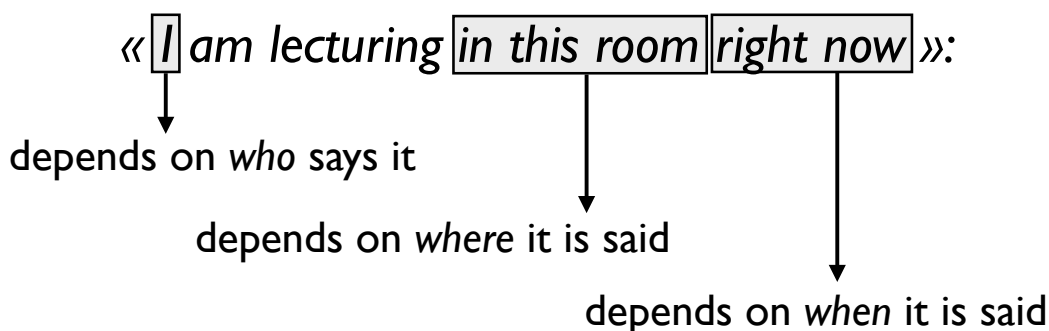
Conversational implicatures

When the cooperative maxims are violated, we can quickly notice it:

Which maxim is violated here?

Deixis

- Dialogue often *referential* to some spatio-temporal context

- Such references are called **deictics**

  - Related concepts: indexicals, anaphora

- The meaning of a deictic depends on the *context* in which it is uttered (including the speaker perspective)

« *I* am lecturing *in this room* *right now* »:

depends on *who* says it

depends on *where* it is said

depends on *when* it is said

Deixis

- ## Deictic markers:

  - *Pronouns*: «I», «you», «my», «yours»

  - *Adverbs of time and place*: «now», «yesterday», «here», «there»

  - *Demonstratives*: «this», «that»

  - *Tense markers*: «he just left»

  - *Others*: «the mug to your right», «go away!», «the other one»

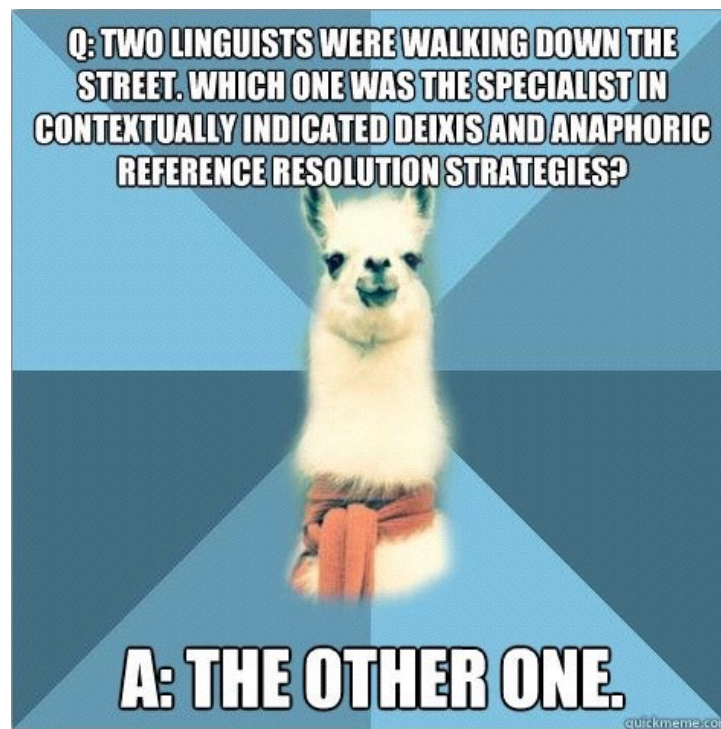 - *Non-verbal signs*, based on gestures, gaze, etc.

Deixis

- ## Deictics can refer to virtually anything:

  - Objects: «take that mug»

  - Events: «don't do that», «this car accident was awful»

  - Persons: «You're being an idiot»

  - Abstract entities: «This methodology is flawed»
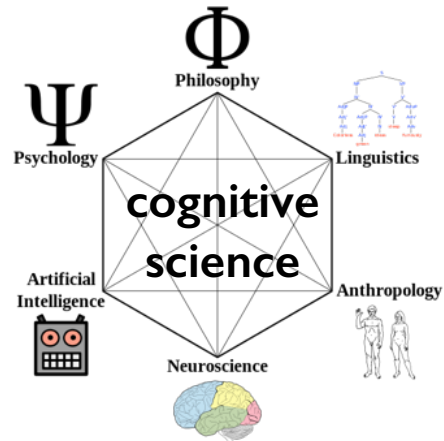
- ## Perspective is important:
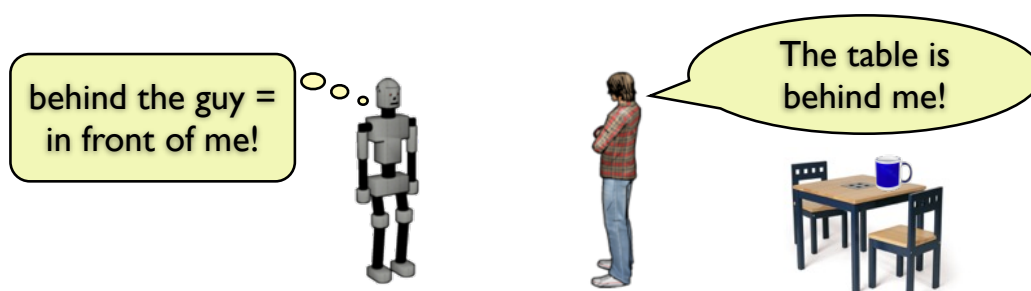
Deixis

Outline

- What is spoken dialogue?

- **A cognitive perspective**

- Summary

A cognitive perspective

- Cognitive science can offer us useful insights on the foundations of spoken dialogue

- Focus on two important ideas:

  - The human brain is *social*

  - The human brain is *predictive*

Social interactions

- Social interactions are a fundamental aspect of human cognition

  - Social interactions played a key factor in the evolution of the human brain (the *social brain hypothesis*)

  - We interact to share emotions, experience, and *participate in common activities*
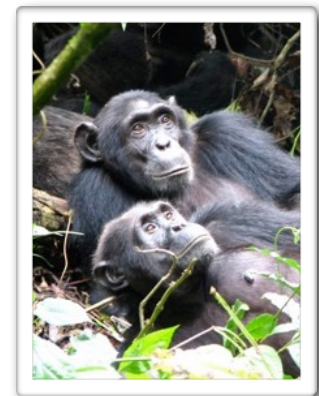
[Dunbar, R (1998), The social brain hypothesis. *Evolutionary Anthropology*]
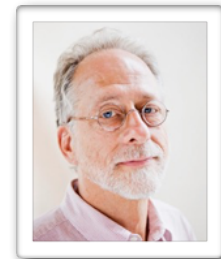
Social interactions

- Humans naturally view each other as goal-directed, *intentional agents*

  - Understand other agents in terms of belief, desires and intentions (*theory of mind*)

- But there's more: humans are able to *jointly attend* to external entities and establish *shared intentions*

[Dennett, D (1996), *The intentional stance.*]
[Tomasello, M (1999), *The cultural origins of human cognition.*]

Daniel Benett (1942, -)
philosopher of mind

Michael Tomasello (1950, -)
developmental psychologist

Alignment in social interactions

- Participants in a dialogue continuously **align** their mental representations

  - Notion of common ground discussed earlier

- But dialogue participants also align at a deeper level, by unconsciously *imitating* each other

- As the interaction unfolds, the participants automatically align their *wording*, *pronunciation*, *speech rate*, and *gestures*

[Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*]

Predictive mechanisms

- The brain does not receive information passively, it routinely projects *hypotheses* and interprets things in a particular way

- Top-down, **predictive** mechanisms crucial for human cognition

- People continuously predict what their interlocutor is going to say next, based on the current context

  - The prediction is performed *incrementally*

  - Exploitation of the broader context occurs *very rapidly* and guides all processing

[McRae, K., & Matsuki, K. (2009) in *Language & Linguistics Compass*]
[Van Berkum, J. J. A. (2010) in *Italian Journal of Linguistics*]

Outline

- What is spoken dialogue?

- A cognitive perspective

- **Summary**

Summary

- ## One key idea: dialogue is a **joint social activity**

  - The dialogue participants take *turns*

  - Each turn is composed of one or several *dialogue acts*

  - The participants cooperate to ensure mutual understanding (gradual expansion of *common ground*)

  - They interpret each other's utterances cooperatively (*conversational implicatures*)

  - This activity takes place in a given *context* which is crucial for making sense of the interaction (presence of *deictics*)

Summary

- ## We've also discussed some important ideas from cognitive science:

  - Social interactions and collaborative activities form a central part of human cognition

  - People naturally *align* their way of speaking (word choices, pronunciation, gestures etc.)

  - Human language processing is *proactive*: people are continuously predicting the next steps in the dialogue, and use these predictions to *guide* processing

Next week

- Next week, we'll talk about **speech recognition**

- We'll start by reviewing the core ideas of phonetics (the study of speech sounds)

- We'll then see how acoustic models and language models are employed to convert speech signals into text