Fundamentals of Linguistic Interaction

Raquel Fernández

Institute for Logic, Language & Computation
University of Amsterdam

NASSLLI 2016
Topics on timing and turn taking

- Empirical facts
- Models: prediction vs. reaction, prediction + reaction
- Semiotics of timing (e.g., rhetoric and social significance)
- Development and turn-taking
Norm: little overlap, short gap.

- Lengthy silences carry semiotic significance (undesired or unexpected response; rhetorical effect)
- Overlaps (or interruptions) may be socially loaded (sign of dominance and authority).
Jim Lehrer: Do you believe you could do a better job than President Bush in preventing another 9/11-type terrorist attack on the United States?

John Kerry: [pause 0.278] Yes, I do. [pause 1.268] But before I answer further, let me thank you for moderating. [pause 0.588] I want to thank the University of Miami [pause 0.564] for hosting us.

Jim Lehrer: Mr. President, you have a ninety-second rebuttal.

George W. Bush: [pause 0.055] uh uh l- [pause 0.165] l, too, thank the University of Miami, and [pause 0.454] and uh [pause 2.116] and say our prayers are with [speeds up] the good people of this state, who’ve suffered a lot.
Timing may be socially loaded:


Common conjectures in sociolinguistic work:

1. *Interrupting and being interrupted*
   - (a) Men interrupt more often than women.
   - (b) Women are more often interrupted than men.

2. *Who interrupts who*
   - (a) Men interrupt more often women than other men.
   - (b) Women interrupt more often other women than other men.

3. *Function of interruptions*
   - (a) Women’s interruptions are more often collaborative than men’s.
   - (b) Men’s interruptions are more often competitive than women’s.
Social factors: gender

- 75 multiparty meetings with transcriptions, times, and dialogue acts
- 28 meetings with at least two male and two female participants.
- Focus on interruptions, which are different from overlaps.

<table>
<thead>
<tr>
<th></th>
<th>overlap</th>
<th>no overlap</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>interruption</td>
<td>1,233</td>
<td>2,037</td>
<td>3,270</td>
</tr>
<tr>
<td>no interruption</td>
<td>6,499</td>
<td>35,896</td>
<td>42,395</td>
</tr>
<tr>
<td>total</td>
<td>7,732</td>
<td>37,933</td>
<td>45,665</td>
</tr>
</tbody>
</table>
Examples

Participant ids:  \textit{me} = male;  \textit{fe} = female

\begin{itemize}
\item \textbf{interruption with overlap:}
  \begin{itemize}
  \item A: we should at least check that everybody here ==
    \begin{itemize}
    \item (Buw001 me013 3770.32-3772.5 s.%-)
    \end{itemize}
  \item B: i think everyone here is on the list .
    \begin{itemize}
    \item (Buw001 me011 3771.96-3773.64 s^na)
    \end{itemize}
  \end{itemize}
\end{itemize}

\begin{itemize}
\item \textbf{interruption without overlap:}
  \begin{itemize}
  \item A: let’s just == (Buw002 fe008 2636.37-2637.62 s^cs.%-)
  \item B: but they actually had a big list of like things that people had
    transcribed . (Buw002 me070 2637.73-2641.66 s)
  \end{itemize}
\end{itemize}

\begin{itemize}
\item \textbf{overlap without interruption:}
  \begin{itemize}
  \item A: and you can ask all the questions about how this all fits
    together . (Bed010 m010 386.25-389.72 s)
  \item B: that’s fine . (Bed010 m045 389.48-389.88 s^ba)
  \end{itemize}
\end{itemize}
Results


1. **Interrupting and being interrupted**
   - (a) Men interrupt more often than women. ×
   - (b) Women are more often interrupted than men. ✓

2. **Who interrupts who**
   - (a) Men interrupt more often women than other men. ✓
   - (b) Women interrupt more often other women than other men. ×
     
     but women overlap more with other women

3. **Function of interruptions**
   - (a) Women’s interruptions are more often collaborative. ×
   - (b) Men’s interruptions are more often competitive. ✓
     
     men interrupt more often by overlapping and with floor grabbers

Caution! other social variables are very likely to be confounds.
Longitudinal study of 12 mother-infant dyads in free-play interactions at six ages between 3 and 18 months.

- Children develop the temporal properties of turn-taking early in infancy (vocal exchanges).
- Overlap: first more than mothers; by 18 months similar to mothers.
- Gaps: significant increase at 9 months
- Overlaps and gaps of mothers remain stable over time.
• timing coordination – turn taking
• meaning coordination – dialogue acts and grounding
• style coordination - alignment and adaptation
• language acquisition in interaction
Two views of communication:

- Shannon (1948) - Information theory: information encoded by the sender, transmitted, and decoded by the recipient.
- Grice (1957) - Human communication is characterised by the process of *intention recognition*
  - Speech acts / dialogue acts / moves encapsulate intention
  - Intention is not fully determined by linguistic form
Goals and intentions beyond language

We have a strong tendency to ascribe *goals* and *intentions* to agents. Related to

- theory of mind: ability to model internal mental state of agents
- attribution of causation


Sensing actions by others triggers attribution of intentions, goals, causes. Speech act theory: conversations are made up of *linguistic actions*. 
Speech Act Theory

Initiated by Austin (*Who to do things with words*) and developed by Searle in the 60s-70s within philosophy of language.

Speech act theory grows out of the following observations:

- Typically, the meaning of a sentence is taken to be its truth value.
- There are utterances for which it doesn’t makes sense to say whether they are true or false, e.g., (2)-(5):

  (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?

- These (and generally all) utterances serve to *perform actions*.
- This is an aspect of meaning that cannot be captured in terms of truth-conditional semantics (⇝ *felicity conditions*).
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 ⇝ its *speech act*.

- **perlocutionary act**: the act(s) that derive from the locution and illocution of an utterance (effects produced on the audience)

---

Searle distinguished between five basic types of speech acts:

- **Representatives**: the speaker is committed to the truth of the expressed proposition (assert, inform)
- **Directives**: the speaker intends to elicit a particular action from the hearer (request, order, advice)
- **Commissives**: the speaker is committed to some future action (promise, oaths, vows)
- **Expressives**: the speaker expresses an attitude or emotion towards the proposition (congratulations, excuses, thanks)
- **Declarations**: the speaker changes the reality in accord with the proposition of the declaration (provided certain conventions hold), e.g. baptisms, pronouncing someone guilty.

Dialogue acts (term introduced by Bunt, 1994):

- **coherence and cohesion:**
  - inspired by dynamic semantics: moves as context-change actions (QUD, SDRT, ...several relevant courses at NASSLLI)
  - structure (e.g., *adjacency pairs*), forward-looking and backward-looking acts

  ```
  Waitress: What’ll ya have girls?
  Customer: What’s the soup of the day?
  Waitress: Clam chowder.
  Customer: I’ll have a bowl of clam chowder.
  ```

- **meta-communication:** acts for dialogue control
  [more on this tomorrow - grounding process]

- **multi-functionality:** utterances can perform more than one action at once.

DA taxonomies aim to be effective as tagsets for annotating dialogue corpora.

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

The taxonomy is meant to be general but not totally domain independent ~ it has been adapted to several types of dialogue.
DA Taxonomies: SWBD DAMSL

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

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

The average conversation consists of 144 turns, 271 utterances, and took 28 min. to annotate. The inter-annotator agreement was 84% ($\kappa=.80$).
On the Gricean view, it is possible for the same signal to correspond to different intentions:

The gun is loaded $\mapsto$ *threatening?* *warning?* *explaining?*

Conversely, the same intention can be realised by different signals:

*Requesting:*
  - A day return ticket to Utrecht, please.
  - Can you please give me a day return ticket to Utrecht?
  - I would like a day return ticket to Utrecht.

$\mapsto$ *How do we map from utterances to dialogue acts?*
Two computational models of the interpretation of dialogue acts:

- **Symbolic models**: based on epistemic logic (beliefs, desires, and intentions - BDI); use of logical inference to reason about the speaker’s intentions.

- **Probabilistic 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.

Symbolic Models

Classic symbolic models of dialogue acts aim to explain *indirect speech acts*

Can you pass me the salt?

→ Literal speech act [literal force hypothesis]: *yes-no question*
→ Indirect speech act after an inference chain: *request* (pass me the salt)

- S is cooperative, thus U has some aim
- S already knows the answer to the explicit question
- thus S must intend something other than asking
- ability to do something is a pre-condition for requesting
- therefore, given the context, S is probably *requesting* me to pass her the salt.

The *BDI approach* is meant to be a general model of rational action that can be applied to conversation:

- what motivates our actions
- how to understand actions by others
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/](http://www.cs.rochester.edu/research/trains/)

Allen et al. (2001) An architecture for more realistic conversational systems, in *Proc. of Intelligent User Interfaces.*
• the listener uses cues in the input to infer a particular interpretation.
• use of several sources of knowledge: lexical, collocational, syntactic, prosodic, conversational-structure

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.

$$d^* = \arg\max_d P(d|c) = \arg\max_d P(d)P(c|d)$$

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.
• **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)

More recent probabilistic models try to bypass feature engineering:


⇝ Not incremental and hence not compatible with fast, smooth turn-taking.

Proposal of an incremental framework:

Summary

**Timing:**
- social factors
- development

**Content:**
- conversation as intention recognition
- from speech acts to dialogue acts
- dialogue act taxonomies
- dialogue act recognition

**Resources:**
- see course website

**Tomorrow:**
- dialogue as joint action, the grounding process, audience design