Natural Language Generation
for
Embodied Conversational Agents

Kristina Striegnitz

ESSLLI 2008
Hamburg, Germany

Course Overview

Day 1: Introductions
   Embodied Conversational Agents, Role of non-verbal behavior in
   communication, Natural Language Generation

Day 2: Overview of Natural Language Generation
   Surface realization

Day 3: Generating referring expressions

Day 4: Discourse and dialogue

Day 5: Other issues: emotions and rapport
   Evaluation

Material

http://eagle.union.edu/~striegnk/courses/esslli2008/

slides, bibliography, links ...

Me

Kristina Striegnitz
striegnk@union.edu

Computer Science
Union College
Schenectady, NY, USA

before:
Northwestern University
(Postdoc)
Saarland University (PhD)
My research

natural language generation
- Embodied Conversational Agents
- referring expressions
- contextual reasoning
- games

You

- What's your name?
- Where are you from?
- What's your background?
- Why are you taking this class?

- Your email address. (for getting a notification when the slides are up on the web page)

Today

- Introductions: this course, me, you
- What are ECAs?
- Face-to-face conversation
  - non-verbal behaviors
  - their functions
  - Are they communicative?

Game
Today

- Introductions: this course, me, you
- What are ECAs?
  - Face-to-face conversation
    - non-verbal behaviors
    - their functions
    - Are they communicative?

What are Embodied Conversational Agents (ECAs)?

- usually *virtually* embodied
- participate in face-to-face dialogue
  - generate natural language and non-verbal output (making contributions to the dialogue as well as signaling dialogue state)
  - recognize and interpret verbal and non-verbal input
  - appropriately respond to and use turn-taking behaviors, feedback, clarification questions and other dialogue behaviors

Examples

Why ECAs?

Intuitively:

- face-to-face is the most natural way to communicate
- non-verbal behaviors play a communicative role
- seeing your partner makes communication more efficient

Two goals:

- Learn about human-human communication.
- Improve human-machine interaction.
**Today**

- Introductions: this course, me, you
- What are ECAs?
- **Face-to-face conversation**
  - non-verbal behaviors
  - their functions
  - Are they communicative?

---

**Properties of face-to-face conversation**

- Participants can see and hear each other.
- They share the same physical environment (at the same time).
- Sending and receiving is immediate and simultaneous.

  ➔ collaboration
  ➔ non-verbal behaviors

---

**Clark & Krych (2004)**

*Speaking while monitoring addressees for understanding, J. of Memory and Language*

---

**Four interactive conditions**

- workspace visible, faces visible
- workspace hidden, faces visible
- workspace visible, faces hidden
- workspace hidden, faces hidden
One non-interactive condition

Visible vs. non-visible workspace

Workspace visible
A: Take a short blue.
B: [Retrieves a short blue block.]
A: [Looks at B's block.] Put it at the end of the yellow close to the green.
B: [Places the blue block on the yellow block.]
A: [Looks at result.] Take a...

Workspace not visible
A: And then you're gonna take a blue block of four.
B: M-hm.
A: And you're gonna put it on top of the four blocks—four yellow blocks farthest away from you.
B: Which are the ones closest to the green.
A: Yeah
B: Okay. But the green's still not attached.
A: Yeah. And then...

Visually sharing a workspace is more efficient

Being able to interact helps

<table>
<thead>
<tr>
<th>interactive, workspace hidden</th>
<th>non-interactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>model errors</td>
<td>5%</td>
</tr>
<tr>
<td>block errors</td>
<td>0.8%</td>
</tr>
</tbody>
</table>
Interactions/Collaboration

Workspace visible
A: Take a short blue.
B: [Retrieves a short blue block.]
A: [Looks at B’s block.] Put it at the end of the yellow close to the green.
B: [Places the blue block on the yellow block.]
A: [Looks at result.] Take a...
Kinds of non-verbal behavior (by purpose)


- **emblems**: have a generally known definition (e.g., thumbs up in the US)
- **illustrators**: illustrate what is currently being said
- **affect display**
- **regulators**: maintain and regulate the back-and-forth nature of dialogue
- **adaptors**: fragments/reductions of behaviors to satisfy bodily needs (e.g., smoothing hair, pushing up glasses,...)

Mapping between specific behaviors and purpose

- no one-to-one mapping

Are non-verbal behaviors communicative?

Do speakers intend them to transport information?
Do listeners draw information from them?

- some obviously yes:
  - **emblems**:
  - **pointing**: “Take this and put it there.”
  - **emotional expressions in the face**: Ekman, Friesen, Ellsworth (1972, 1982)
- others controversial:
  - **gesture**
**Gesture categories (according to David McNeill)**

- (emblems)
- beat gestures
- iconic gestures
- metaphoric gestures
- deictic gestures

**Krauss et al.: Gestures are not communicative**

- Gestures do convey information related to the semantic content of the accompanying speech.
- BUT: relationship is imprecise and unreliable and information contributed by gesture doesn't seem to help

→ main purpose of gesture is to help the speaker (lexical access)

**Krauss et al.'s evidence**

- Assigning interpretations to gestures seen in isolation without accompanying speech:
  - Slightly more often than chance people give an interpretation that is more similar to the lexical affiliate (as chosen by a panel of judges) than to the lexical affiliate of a randomly chosen gesture.
  - Semantic category seems to play a role: actions easier than locations easier than objects and descriptions.
- Recognizing previously seen gestures/speech/speech and gesture
  - Greater than chance but lower than for speech or speech plus gesture.
  - No difference between speech and speech plus gestures.
- Identifying objects (shapes, sounds, tastes) based on speech only and speech plus gesture descriptions:
  - No difference between speech and speech plus gestures.

Beattie & Shovelton: Listeners do draw information from gestures

**Experiment:**
- Speakers are asked to retell a cartoon ⇒ 18 clips of iconic gestures in speech context.
- Listeners/viewers see 6 audio only, 6 video only, 6 audio and video clips and have to answer various questions about the objects in the clip (number, shape, size, movement, ...).

---

Beattie & Shovelton (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. Journal of Language and Social Psychology

---

Cassell, McNeill & McCoulough: Listeners do draw information from gestures

**Experiment:**
- Scripted video of a story. Gestures were designed
  - to sometimes be redundant to speech
  - to sometimes add information not contained in speech
  - to sometimes contradict information in speech
- Participants watched the video and then retold the story to somebody else.
- Result was analyzed for inaccuracies wrt. the text of the scripted video.

---

Experiment:
• Speakers described pictures which their interlocutors had to reproduce.
• Face-to-face setting.
• Analyze descriptions for cases where directional information is missing in the speech.

Results:
• Speakers who gestured omitted directional information more often than non-gesturers.

Conclusion:
• Gestures are intended to be communicative (and to provide the missing information).

Summary
• Face-to-face communication: most natural setting for human-human communication.
• Non-verbal behaviors seem to play important role for certain communicative tasks in human-human communication.
• For a full understanding of human-human communication, need to understand the role of non-verbal behavior.
• ECAs try to leverage non-verbal behavior for better human-machine communication.