Gaze as function of instructions

-

and vice versa

Seminar
Winter term 2013/14

Attention & Language

- Listeners look at what they hear (Literature)
- Speakers monitor what listeners look at (Literature)
- How do speakers **use** what listeners look at? (Research Question)

How do we communicate?

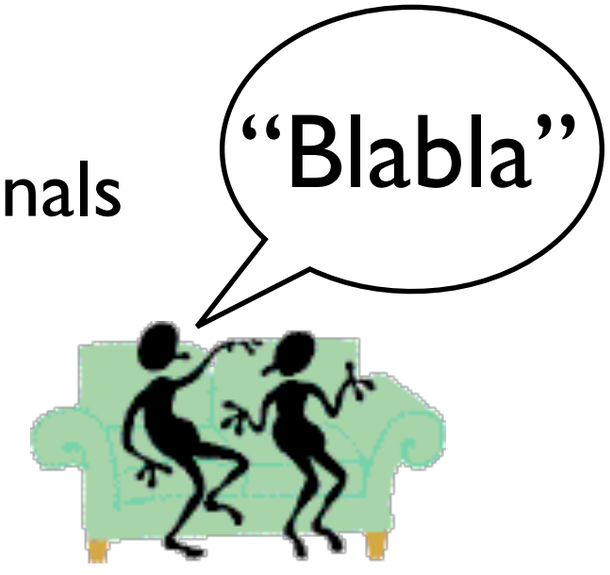
- Speech
- Non-Verbal Sounds
- Body Posture and facial expressions
- Gaze Direction
- Gesture

Speech

- Explicit: “Mary and John have an appointment at 2pm on 22nd of October 2013.”
- Vague: “This is too small.”
- General: “Mary likes cake.”
- Pragmatic: “I am cold” -> *“Please close the window.”*

Situated Speech

- Situated: (embodied) Speaker, Listener, environment, context
- Implicit speaker/listener's non-verbal signals
  - Unconscious?
- Always present
- Extremely rich (emotions, attitude, attention...)
- Situating and augmenting speech stream



Situated Speech

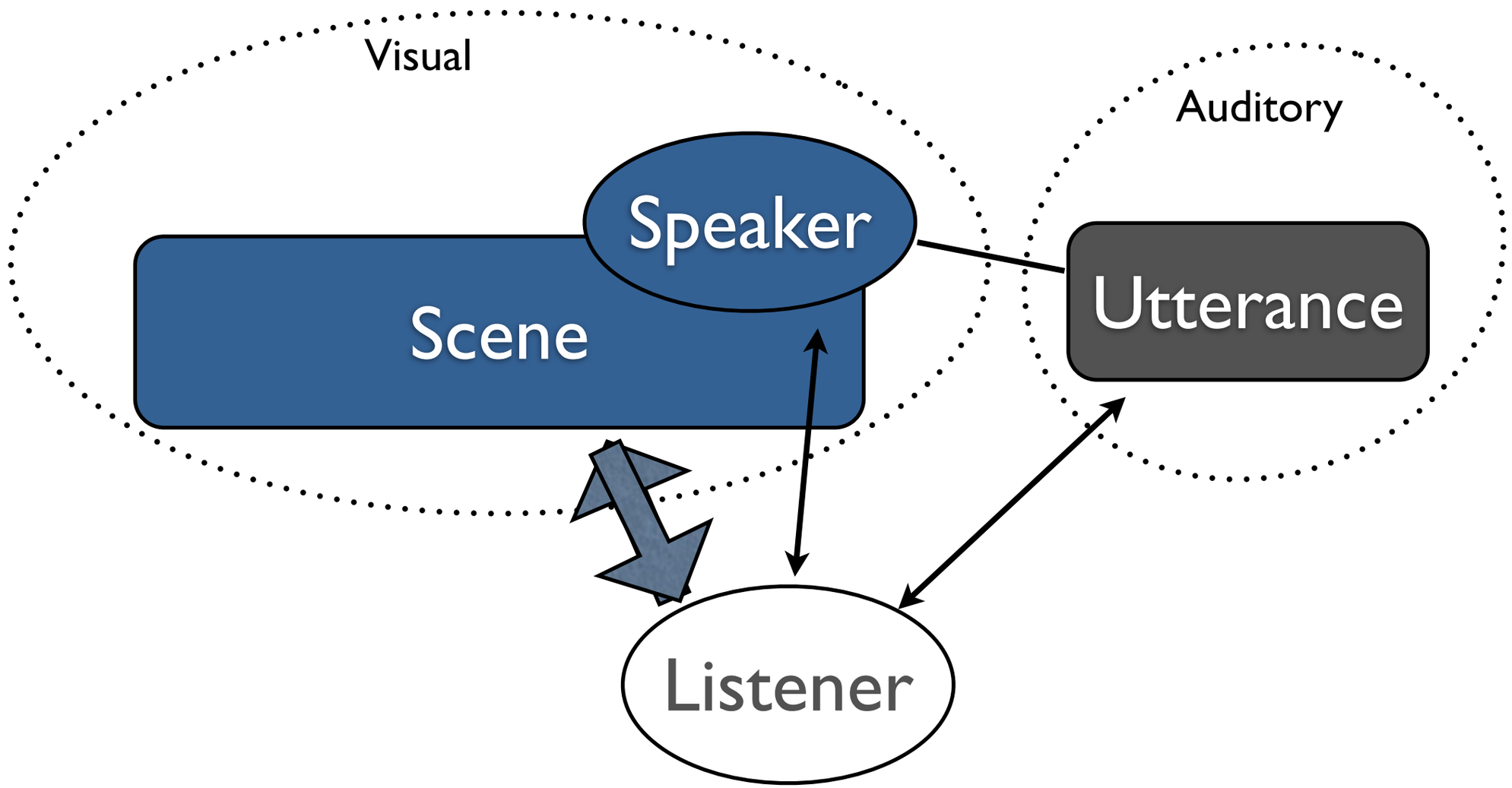
- Spoken language and environment provide huge amounts of information simultaneously
- Processing needs to be fast!
- Using one to facilitate processing the other:
  - Visual information (non-verbal cues)
  - Visual (scene) information
  - Linguistic information

(Visual) Attention

"Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence."

William James, 1890, "Principles of Psychology"

Situated Speech

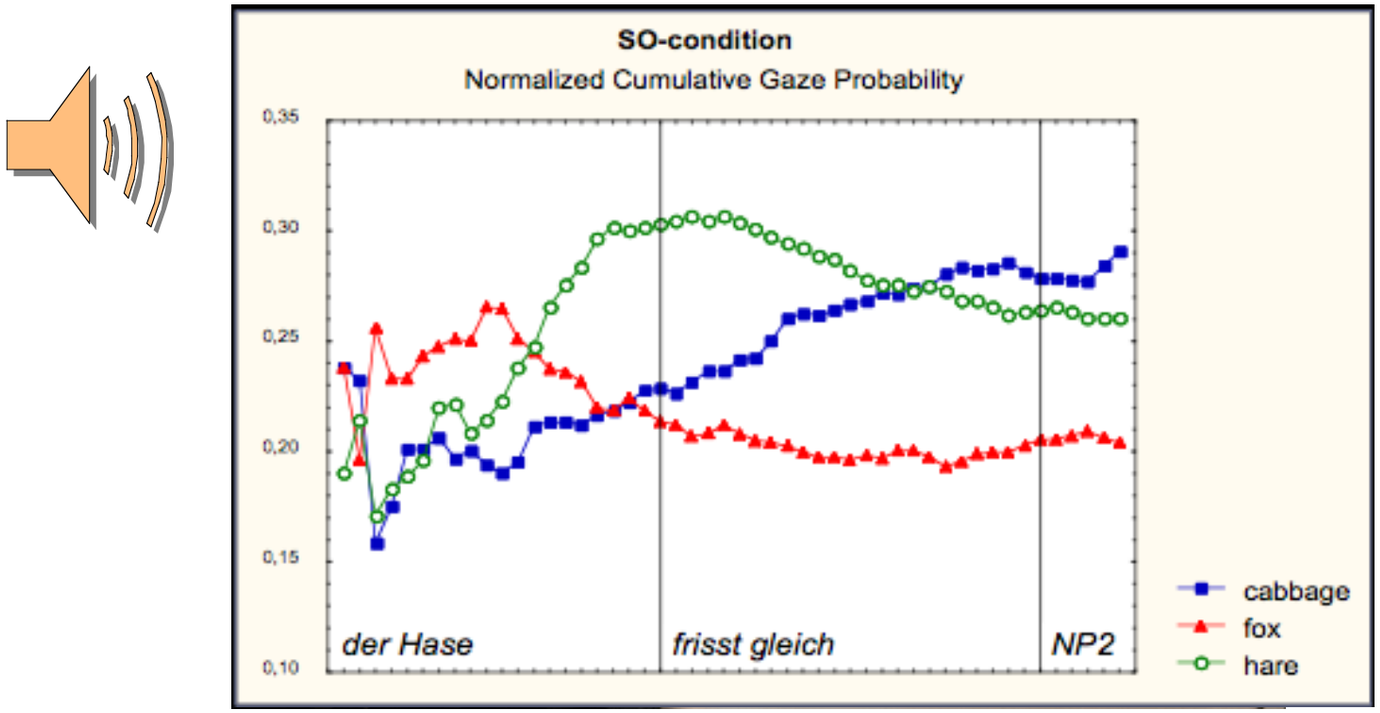


(cf. the Coordinated Interplay Account by Knoeferle & Crocker, 2006)

Attentional Cues

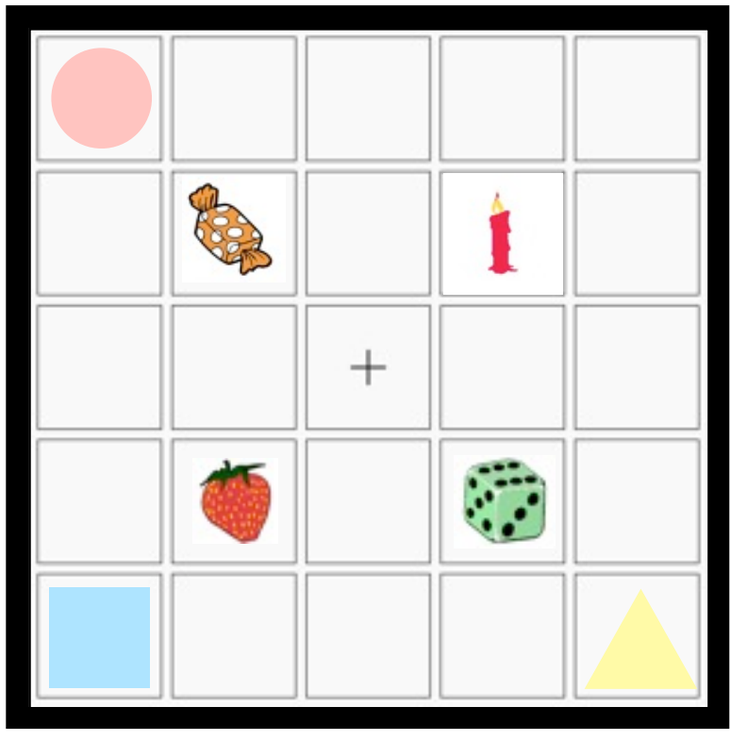
- Gaze is - by expressing the attentional focus - a reflection of what is currently processed.
- Visual World Paradigm:
    - Studies relying on and investigating eye movements during language comprehension/production as a cue to what is being processed and when
- Interaction Studies:
    - Study gaze as additional information channel reflecting and influencing communicative processes

Eye-tracking in scenes

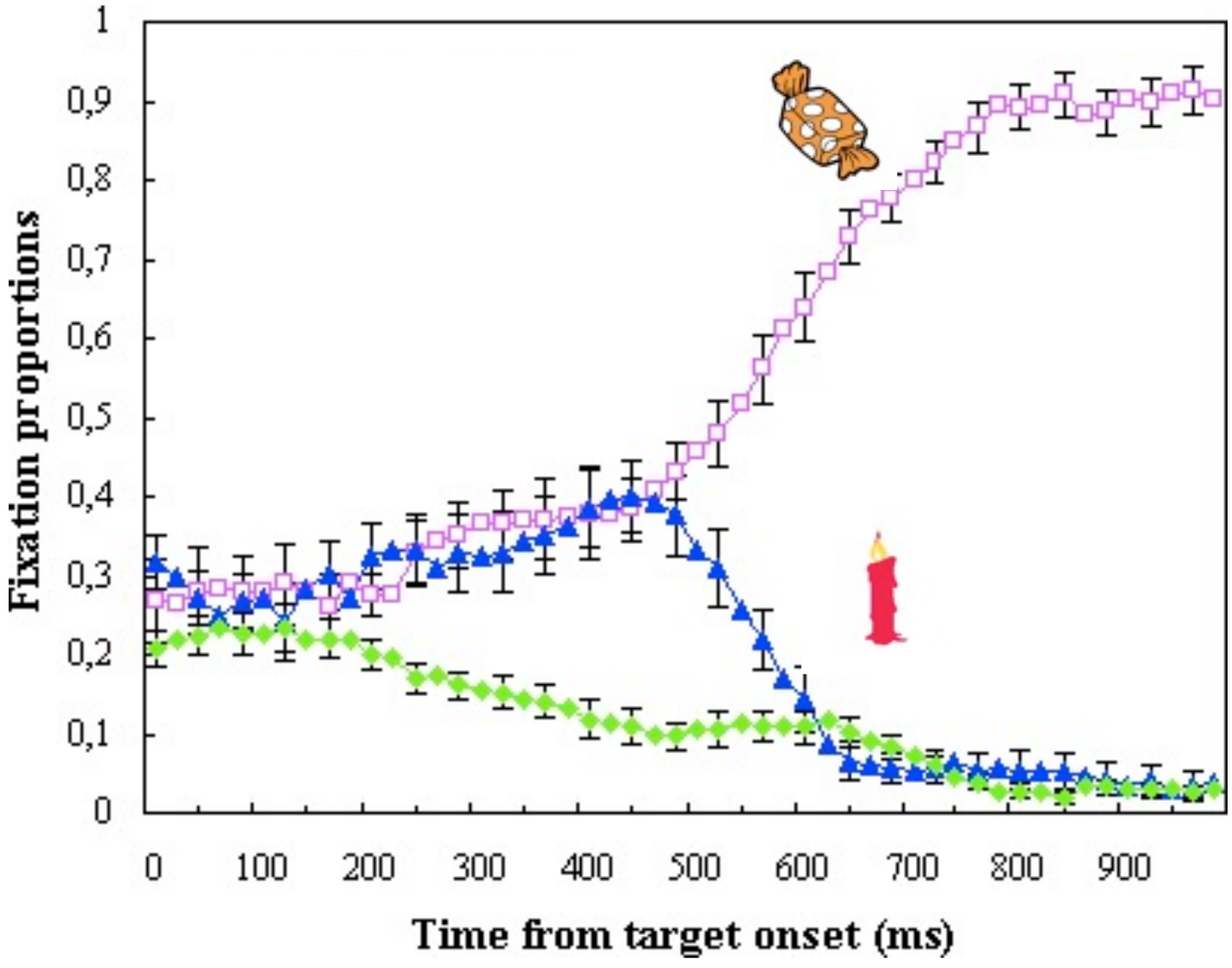


- Attention to objects in the scene is closely time-locked to comprehension
  - Makes it possible to use eye-tracking in scenes during utterance presentation to investigate spoken comprehension
  - Permits us to examine use of scene information for comprehension

Lexical access over time

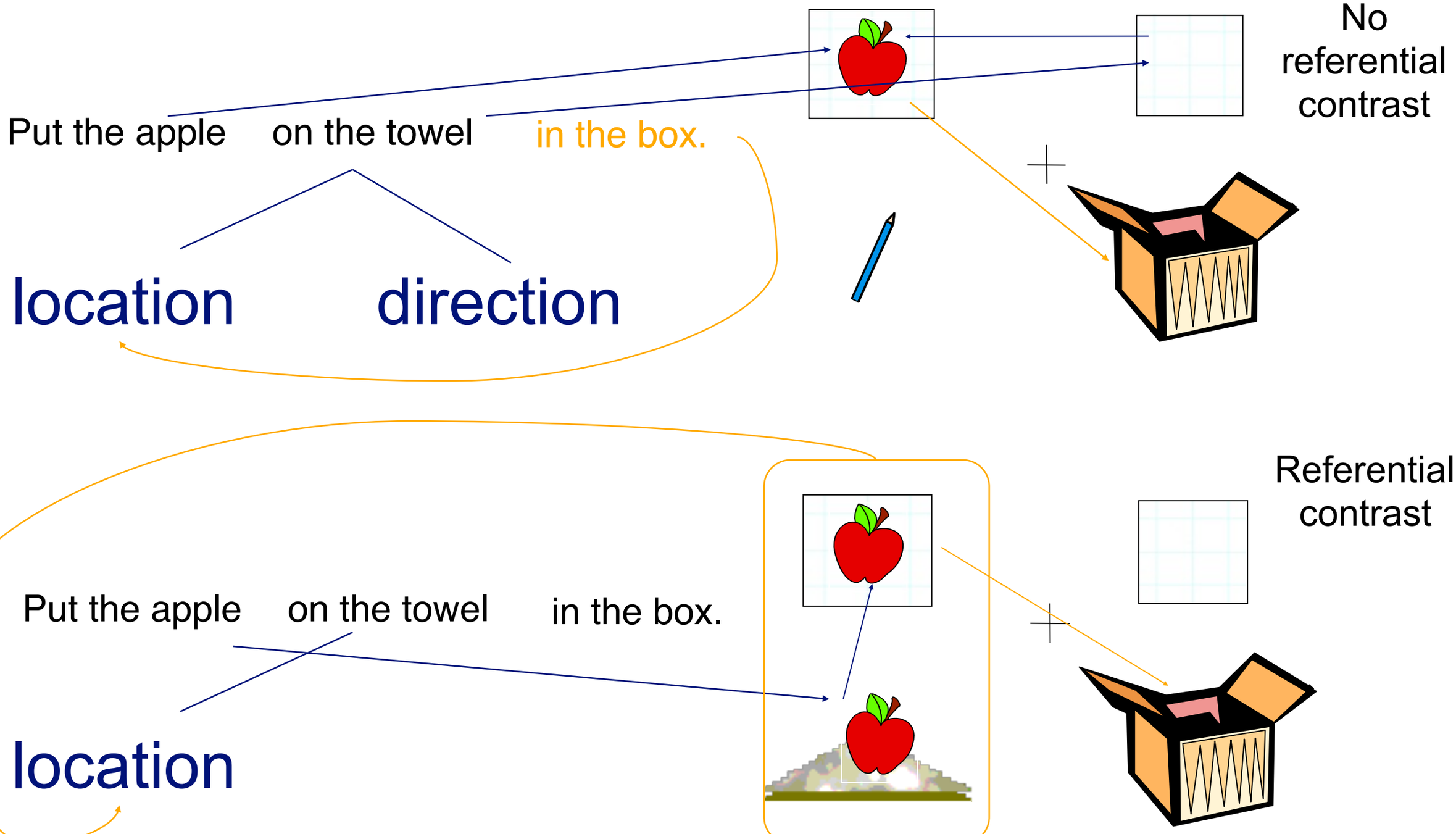


"Pick up the *candle*"



Visual World Studies

Tanenhaus et al. 2000



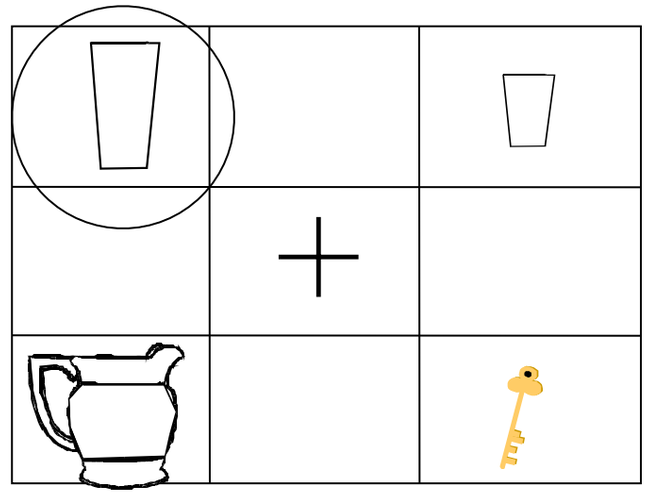
Incremental Semantic Interpretation

Sedivy et al. 1999

More visual referential ambiguity:

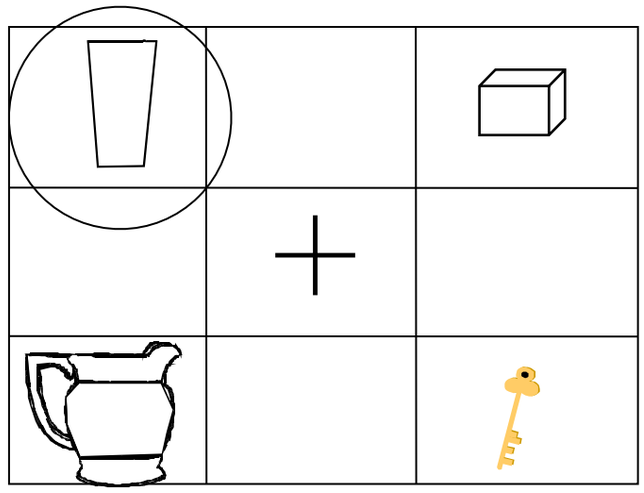
- Influence of **visual contexts** on
  - Mapping of reference to entities
  - Properties of objects (small, tall)

Two same-type objects
that differ in 1 property: size



Pick up the tall glass and put it below the pitcher.

No contrastive objects
of the same type



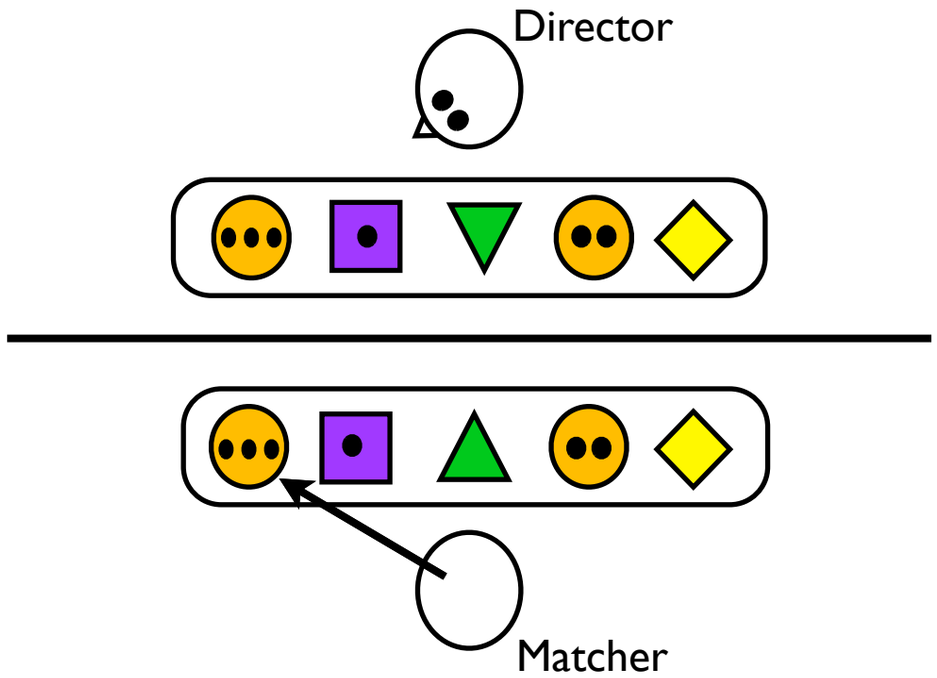
- More rapid looks to the tall glass before hearing "glass" in the contrastive than non-contrastive condition

Eye-Movements in Situated Speech

- Listener not only *hears* but also sees the speaker / is *seen by* the speaker:
  - Spoken utterances shape listener eye-movements
  - Speaker gaze shapes listener eye-movements
  - Listener eye-movements shape spoken utterances!?

Eye-Movements in Situated Speech

- Listener can see speaker's gaze
  - "Move the circle with ...
  - ".. three dots to location A."
- Speaker can see listener's gaze
  - "Move the circle with ...
  - ".. yeah, that, to location A."
- How useful is this?
- How does this affect language comprehension & production?



Hanna & Brennan 2007

Attention & Language

- Listeners look at what they hear
- Speakers monitor what listeners look at
- How do speakers **use** what listeners look at?

Listener Gaze in Interaction

- Backchannel, Turn-taking etc. (Sacks et al. 1997, Cassell et al. 1994...)
- Feedback for Speaker
  1. Allow speaker to encourage, repair, re-direct...
  2. Evaluate speaker's/system's efficiency

Feedback for Speaker

1. Allow speaker to encourage, repair, re-direct...

- Joint Lego building task (Clark & Krych 2004)
  - Face-to-face interaction is faster
  - Speaker needs less turns and less words, allows deictic expressions when joint workspace is visible
  - Speaker self-interrupts
  - Listener *uses* signals to elicit more info from speaker

Feedback for Speaker

2. Evaluate speaker's/system's efficiency

- How do listeners understand instructions? What is an efficient instruction?
  - Collect corpus of human generated instructions (Gargett et al. 2010, Lovelace et al. 1999)
    - Detect structure & features, e.g. landmark use
  - Generate instructions/REs and assess intelligibility (Garoufi & Koller 2010, Roger et al. 2011)
    - Measures of intelligibility: Task performance, satisfaction, ...
  - Navigation & Eye-tracking (Andersen et al. 2012, Kiefer et al. 2012)

Listener Gaze in Interaction

Attention & Language

- Listeners look at what they hear
- Speakers monitor what listeners look at
- How do speakers **use** what listeners look at?
  - Can we evaluate instructions (better) using eye-tracking?
  - Can we construct instructions (better) using eye-tracking?

Timing

"Push the **left** button to the..."

Timing

“...right of the flower. “

Timing

"...flower. - Yes, that one."

Timing

Organizational Things

- Topics:
  - One problem - different areas of research
  - Your interest? Select paper...
- More research & presentation:
  - Find one more paper related to your first
  - Make **two** short presentations (á 30-45min)

Organizational Things

- Learning goals:
  - Read & analyse a paper
  - Relate to another paper
  - Present (meet with me before & after)
  - Identify and collect relevant results and patterns
  - Sketch answer for research question(s)...

Organizational Things

- Website: http://www.coli.uni-saarland.de/~masta/WS13/gazefunction.html
- Meeting time: Mon 4-6 ?!
- Term schedule
- Seminar papers?

Questions?

- Contact me:
  - C7 4, room 3.19
  - masta@coli.uni-saarland.de