The Visual-World Paradigm
Adult Language Processing

SS16 - (Embodied) Language Comprehension

Maria Staudte
(Embodied) Language Comprehension

- Language is “embodied” & “situated”:
  - **Embodied**: Language is inseparably bound to our body and our physical interaction with the world
  - **Situated**: Language is typically used and understood in context
Language is situated

- 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
Language is situated

• We see things when we listen and speak

• We often try to ignore the visual information — this requires cognitive control (cf. automatic processing)

• We often use the visual information — this requires cognitive control, capacity sharing and cross-modal integration
While on a *photographic/safari* in *Africa*, I managed to get a number of breathtaking shots of the wild *terrain*. These included pictures of rugged *mountains* and *forests*, as well as muddy *streams/winding their way through big game/country*. One of my best shots though was ruined by my scatter-brained *dog/Scotty*. Just as I had slowly *wormed* my way on my *stomach*...
Eyetracking

Use “visual attention” as index to underlying 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
The Measure

- Tracker samples at 60/250/500/1000 Hz
  - New sample every 16,6/4/2/every millisecond
- \(x,y\) - coordinates per sample
- Fixations & saccades
  - min duration of ~80ms = fixation
The Measure

- Fixation proportion vs probability
- Time to first fixation/saccade
- Fixation durations
- Inspection probability

The knight attacked the windmill on his donkey.
Lexical access over time

"Pick up the candle"

(Allopenna et al. 1998)
Attachment Ambiguity

(Tanenhaus et al. 2000)

Put the apple on the towel in the box.

location  direction

Put the apple on the towel in the box.

location

Referential contrast

No referential contrast
Incremental Semantic Interpretation

More visual referential ambiguity:

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

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

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

(Sedivy et al. 1999)
Predictive Listening

• Anticipation of objects (e.g. Altmann & Kamide 1999, 2007)

• Semantic requirements of a noun to be used as argument for a verb

• drink – object needs to be drinkable, liquid

“The man will drink the wine”
Referential & Anticipatory Looking

• Can information associated with a 3-place verb be used to anticipate the second post-verbal argument — during hearing the first?

• “The woman will spread/slide the butter…

  1. Inanimate goal (“on the bread”)

  2. Animate goal (“to the man”)

(Kamide, Altmann & Haywood, 2003)
Referential & Anticipatory Looking

1. The woman will spread the butter on the bread
Referential & Anticipatory Looking

2. The woman will slide the butter to the man
Referential & Anticipatory Looking

% trials with looks to target

- **appropriate**
- **inappropriate**
- **woman**

- the woman will
- spread/slide (Region 1)
- the butter (Region 2)
- on/to the
Asymetric distribution

- Looks to the **man**
  - Slide condition: 24.3%
  - Spread condition: 16.8%

- Looks to the **bread**
  - Slide condition: 24.3%
  - Spread condition: 27.4%

Bread could be a theme, too!
Composed predictions?

- Are the observed effects driven by lexical information associated with the verb?

- Are the effects due to the combination of distinct sources of information?

  - Can information about the agent be combined with the selectional restrictions of the verb to predict the theme?
Composed predictions?

1. The man will ride the motorbike
Composed predictions?

2. The girl will ride the carousel
Conditions

• Main conditions:
  • The man will ride the motorbike
  • The girl will ride the carousel

• To assess combinatory effects (vs agent effects):
  • The man will taste the beer
  • The girl will taste the sweets
Hypotheses

• Compare looks towards the motorbike after
  • The man will *taste*
  • The man will *ride*

• Only agent-based information -> No difference

• Compare looks towards the motorbike after
  • The *man* will ride
  • The *girl* will ride

• Only verb-based information  -> No difference
Hypotheses

• Compare looks towards the motorbike after
  • The **man** will taste
  • The **girl** will taste

• Agent-effect information -> man > girl

• Compare looks towards the motorbike after
  • The girl will **ride**
  • The girl will **taste**

• Verb-effect -> ride > taste
Composed predictions
Composed predictions

• Combinatory effects found

• (Almost) no simple agent or verb effects
  • More looks to motorbike after “girl will ride” than after “girl will taste”
Anticipatory Looks

- **Prediction:**
  - projection “forward in time”
  - Processor predicts a post-verbal argument / linguistic input
  - Projected structure evaluated against visual/mental context

- **Integration:**
  - current evidence
  - Thematic organisation of objects according to linguistic context
What’s the hen, egg...?

• “The man will ride ... “

➡ Does the linguistic structure trigger a predictive process?

➡ Does the visual context suggest a plausible argument?
“Coordinated Interplay Account”

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

• What takes **precedence**, world knowledge or scene information?

• What exactly is **predicted**?
In Experiment 2 we further investigate the verb-mediated influence of depicted events. Although allowing us to confirm findings on the tight temporal coordination (Experiment 1), the focus of Experiment 2 was on examining the relative importance of depicted events in comparison with stereotypical thematic role knowledge. Our first aim was to replicate that both stored world knowledge about likely thematic role fillers (Kamide, Scheepers, & Altmann, 2003) and depicted events (Experiment 1, and Knoeferle et al., 2005) can influence incremental thematic role assignment when only one of them is identified as relevant by the verb in the utterance. Second, within the same experiment we explored the relative importance of depicted events in comparison with world knowledge about likely role fillers for incremental thematic role assignment. To test this, utterances identified both a stereotypical agent and a different agent of a depicted action as relevant for comprehension. Crucially, because the utterance in this case did not uniquely identify only one scene entity as relevant for utterance interpretation, the comprehension system was forced to choose between the two available and relevant agents for the thematic interpretation of the utterance.

We exploited scenes such as the one presented in Fig. 11. A scene shows three characters (e.g., a pilot, a wizard, and a detective). The middle character is always a patient (i.e., not performing an action), and the other two characters (the wizard and the detective) always have an agent role. Each of these two characters (the wizard and the detective) can be qualified as agent in two respects: through stereotypical knowledge about what wizards and detectives typically do, and through the immediately depicted event that they are performing. The wizard is a stereotypical agent of a jinxing action, and depicted as performing a spying action. The detective is a stereotypical agent of a spying action, and is depicted as serving food to the pilot. Thus, stereotypical knowledge (e.g., wizard-jinxing) and depicted events (e.g., detective-serving-food) uniquely identify a (different) agent on an image as relevant for comprehension. In addition, one agent (e.g., the wizard) is depicted as performing an action (spying) that is a stereotypical action of the other agent (the detective; see Fig. 11).

Item sentences in Experiment 2 were unambiguous OVS sentences that related to patient–action–agent events (Fig. 11). Recall that German has a rich case-marking system where

![Fig. 11. Example image for Experiment 2.](image-url)

Theme = Pilot

Agent = Wizard/Detective

Stereotypical actions:
- Wizard & jinx
- Detective & spy

Depicted actions:
- Wizard & spy
- Detective & serve food

(Knoeferle & Crocker, 2006)
World knowledge vs scene info

In Experiment 2 we further investigate the verb-mediated influence of depicted events. Although allowing us to confirm findings on the tight temporal coordination (Experiment 1), the focus of Experiment 2 was on examining the relative importance of depicted events in comparison with stereotypical thematic role knowledge. Our first aim was to replicate that both stored world knowledge about likely thematic role fillers (Kamide, Scheepers, & Altmann, 2003) and depicted events (Experiment 1, and Knoeferle et al., 2005) can influence incremental thematic role assignment when only one of them is identified as relevant by the verb in the utterance. Second, within the same experiment we explored the relative importance of depicted events in comparison with world knowledge about likely role fillers for incremental thematic role assignment. To test this, utterances identified both a stereotypical agent and a different agent of a depicted action as relevant for comprehension. Crucially, because the utterance in this case did not uniquely identify only one scene entity as relevant for utterance interpretation, the comprehension system was forced to choose between the two available and relevant agents for the thematic interpretation of the utterance.

We exploited scenes such as the one presented in Fig. 11. A scene shows three characters (e.g., a pilot, a wizard, and a detective). The middle character is always a patient (i.e., not performing an action), and the other two characters (the wizard and the detective) always have an agent role. Each of these two characters (the wizard and the detective) can be qualified as agent in two respects: through stereotypical knowledge about what wizards and detectives typically do, and through the immediately depicted event that they are performing. The wizard is a stereotypical agent of a jinxing action, and depicted as performing a spying action. The detective is a stereotypical agent of a spying action, and is depicted as serving food to the pilot. Thus, stereotypical knowledge (e.g., wizard-jinxing) and depicted events (e.g., detective-serving-food) uniquely identify a (different) agent on an image as relevant for comprehension. In addition, one agent (e.g., the wizard) is depicted as performing an action (spying) that is a stereotypical action of the other agent (the detective; see Fig. 11).

Item sentences in Experiment 2 were unambiguous OVS sentences that related to patient–action–agent events (Fig. 11). Recall that German has a rich case-marking system where

“Den Piloten verzaubert / verköstigt / bespitzelt gleich…”
“The pilot is jinxed / served / spied on soon by…”

Theme = Pilot
Agent = Wizard/Detective
Stereotypical actions:
- Wizard & jinx
- Detective & spy
Depicted actions:
- Wizard & spy
- Detective & serve food
Precedence & Revision

“The pilot is spied on soon by the detective/wizard”
Concurrent & previous scene information

- Concurrent scene information extremely dominant!
- Eyes move to relevant objects, automatically? why?
- Relevant objects and/or relevant locations?
- Scene information changes, objects move/disappear
Blank-Screen Paradigm

1. scene shown
2. scene removed
3. “The woman will eat the cake”
Why such eye-movements?

• People look at relevant things in a scene
• People look at locations where relevant things have been
• Not for perceiving visual information
• For what?
Why such eye-movements?

• Hebbian learning / processing
  • Objects are located somewhere
  • Featural and spatial information is associated
  • One activates the other

• Looking at a location might activate what was at that location!
Why such eye-movements?

(Laeng et al., 2014)

a) Imagery
b) Free recall “What was the color of the beak of the toucan bird?”
c) Fixed view recall “…”
More changes...

- Scene info + utterance
  - Eye-movements to utterance-related objects in scene
- Scene changes/disappears
  - Eye-movements to where these objects were in the scene
- What if spoken utterance proposes changes to scene?
The “represented” scene

a. The woman will put the glass on the table
b. The woman is too lazy to put the glass on the table

She will pick up the bottle, and pour the wine carefully into the glass

(Altmann & Kamide, 2009)
The “represented” scene

Concurrent

Blank

(Embodied) Language Comprehension
Summary

• Eye-movements
  • are a result of (incremental) information integration
    • Lexical access
    • Syntactic ambiguity resolution
    • Interpretation of contrastive features in reference resolution
  • index predicted contents
  • help (?) memory retrieval
  • index online up-to-date representations of the context
Outlook

• Next time: “Visual World Paradigm II”
  • What takes precedence ✓
  • What is predicted?
References


