

Visual Working Memory in Situated Language Comprehension

Emilia Ellsiepen, Matthew W. Crocker Saarland University, Saarbrücken, Germany

Motivation

- Visual context is taken into account when understanding language, but:
- How much of the visual context is maintained in memory?
 - How does visual context decay for purposes of language understanding?
 - What determines which parts of the visual context can be used?

If visual information is not immediately available, representations from visual working memory have to be used for situated language understanding

- How can we integrate a notion of visual working memory in a model of situated language processing?

Background

- Findings from Visual World Paradigm: Language guides gaze in a scene (Tanenhaus et al. '95, Altmann&Kamide '98..)
- Findings from Blank Screen Paradigm: Even when scene has vanished before the sentence starts, looks are directed to the regions, where objects used to be (Altmann '04)
 - this can be attributed to the use of spatial indexing of a scene (Richardson&Spivey '01)
 - has been shown for anticipation only in scenes with up to five scene entities
- Research on visual working memory suggests that only 5 objects can be accurately represented in working memory (Luck&Vogel '97)
- Research on spatial indexes: up to five can be kept, visually salient objects are more likely to receive one: e.g. sudden onset (Pylyshyn '01)

Experiment

Is anticipation of role fillers on a blank screen still possible, if number of scene entities exceeds visual working memory capacity?
Does the order with which objects are attended influence their language mediated access?

Materials:

30 items, each consisting of a display with 6 objects and 1 person and two sentences containing restrictive verbs (1 for counterbalancing), which picks out only one of the objects as a possible role filler

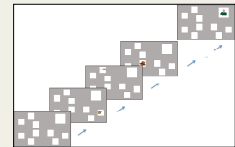
e.g. Der Vater raucht gleich die Pfeife.
The father smokes soon the pipe
The father will soon smoke the pipe



Counterbalancing: Der Vater schärft jetzt gleich das Messer.
The father sharpens now soon the knife
The father will soon sharpen the knife

Procedure:

- Objects and person appear one after the other in different locations on the screen
- After the last one disappeared, the sentence is played
- Eye-movements on the „blank“ screen are measured



Manipulation:

- Position of target object in the presentation sequence is manipulated: it can be 1st, 2nd, 3rd, 4th, 5th, or 6th object, the last is always the person

Verb Region

ANTICIPATORY EYE-MOVEMENTS:

People look at objects, they expect to be mentioned next based on the semantic restrictions of verbs, even before the object is mentioned (Altmann&Kamide '99)

Significant main effect of object
F1(1,30)=37.6 p<.001
people look on the target object more often than on a distractor

Significant main effect of position
F1(5,150)=2.51 p<.05
Marginal interaction
F1(5,150)=2.25 p=.073

Pairwise comparisons:

Significantly more looks to target only in position 1 and 2

PRIMACY EFFECT:

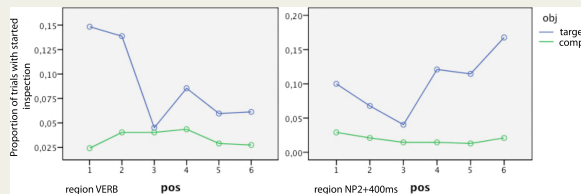
First items are remembered better, due to activated long-term memory representations (e.g. Atkinson&Shiffrin '68)

Results

NP2 Region

REFERENTIAL EYE-MOVEMENTS:

People look at objects that are mentioned (e.g. Tanenhaus et al. 95)



Significant main effect of object
F1(1,30)= 14,8 p<.005
people look on the target object more often than on a distractor

Significant main effect of condition
F1(5,150)=3.24 p<.01
Significant interaction
F1(5,150)=2.99 p<.05

Pairwise comparisons:

Significantly more looks to target only in position 4, 5 and 6

RECENCY EFFECT:

Last items are remembered better, because they are still in (perceptual) short term memory (e.g. Atkinson&Shiffrin '68)

Discussion

- We find anticipatory and referential eye-movements to former object locations:
 - Also in a context with 7 scene entities
 - But not for all serial positions
 - Primacy effect in anticipation
Verb-based anticipation relies on conceptual representation:
 - Primacy effect is usually due to activated long-term memory representations
 - Long-term memory representations are conceptual
 - Recency effect in referential eye-movements
Referential eye-movements can rely on perceptual representations, i.e. phonetical form of the word:
 - Recency effect is due to items being still in short term memory
 - Short term memory holds perceptual representations
- ⇒ There are memory capacity restrictions on the use of visual context in situated language processing
⇒ Anticipatory and referential eye-movements trigger access at different levels of representation: conceptual for anticipation and perceptual for referential eye-movements
⇒ This suggests that anticipation is not the prediction of phonological forms
⇒ Both can drive eye-movements on a blank screen, but show different serial effects

References

- G. T. Altmann (2004). Language mediated eye-movements in the absence of a visual world: the blank screen paradigm. *Cognition*, 93(3):79–87.
- G. T. Altmann and Y. Kamide (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73(3):247–264.
- R. C. Atkinson and R. M. Shiffrin (1968). Human Memory: A Proposed System and Its Control Processes. In: K.W. Spence & J.T. Spence (Hg.). *The Psychology of Learning and Motivation. Vol 2. New York: Acad. Press.*
- S. J. Luck and E. K. Vogel (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657):279–281. Zenon W. Pylyshyn (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80 (1-2):127–158.
- D.C. Richardson and M.J. Spivey (2000). Representation, space and hollywood squares. *Cognition*, 76:269–295.
- M. K. Tanenhaus, M. J. Spivey-Knowlton, K. M. Eberhard, and J. C. Sedivy (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268 (5217):1632–1634.