

# Language and Vision from a Dynamic Perspective

Jana Holsánová

Dept. of Cognitive Science, Lund University  
Kungshuset, Lundagård, 222 22 Lund, Sweden  
jana.holsanova@lucs.lu.se  
phone +46 46 222 40 45  
fax: +46 46 222 97 58

Language and pictures are traditionally dealt with as two modalities with quite different properties, with different strengths and weaknesses. What I would like to find out is rather what language and vision have in common. In this study, participants were examining a complex picture and describing it simultaneously. Their eye movements were registered during inspection of the picture with a SMI iView eye tracker and their spoken language descriptions were recorded and analysed. The common denominator here is the way how we focus attention in language and vision.

In a dynamic view, both visual scanning of a picture and spoken language description of the picture can be conceived of with the help of a spotlight metaphor: The picture elements described lie in the focus of a spotlight and are embedded in a context (in the periphery). The spotlight then moves to the next picture element which pops up from the periphery and will lie in the focus for a while. Speakers zoom in and out when talking about different aspects and different portions of the picture. Eye movements are governed in a similar way. When scanning the picture visually, observers focus on one area of interest a time and the spotlight moves quickly to another region which is then scanned with the help of visual fixations. The picture becomes 'linearized' through the visual path. What has been formulated so far is, that language and vision proceed in brief spurts, have a focus and a periphery and 'linearize' the picture content in the process of observation perception, conceptualization and simultaneous description.

But apart from that, are there any comparable units in both modalities? What are they based on? Which size do they have? What common cognitive patterns underlie our competence in these disparate modes of thought? How is rhythm important for language and vision?

The aims of the study are twofold: a) to compare semantic and sequential patterns in the eye movement data and in the spoken language data and b) to find out if there are similar general strategies in the visual scanning and describing of the picture. In my presentation, I will give some preliminary answers to the questions above and verify the following hypotheses:

- A comparable unit are attentional superfoci rather than attentional foci.
- Hierarchy and rhythm are common for both modes.
- Zooming in and out in the picture is usual in both modes.
- There are structural similarities as well as similarities in the 'semantics'.

Summarizing superfoci are often followed by a list of details both in the visual and verbal data. A general survey is followed by a detailed examination.

Eye movements depict (mimic) functional relationships among picture elements, in repetitive 'rhythmic units, simultaneously to the verbal description.

- Vision and language have numerous procedures in common: e.g. comparing activities, reexaminations of areas or elements, recategorizing picture elements.
- Perceiving and describing 'visual units' presented in the picture, as well as creating 'mental units' (classes of objects with similar traits) during later parts of observation.
- Although both modes are 'linearized', there will be some procedural asynchronies:

Delay between the visual and verbal part.

Static visual focus in case of speech production problems.

Similarities between the patterns visually scanned during pauses and the following sequence of verbal description.

Focus is ahead of speech production. There is asynchrony of planning and production (object focused before described).

There will be an overlap between speech production and the next visual focus.

In my talk, I will focus on similarities between those two modes and exemplify comparable units and patterns.