Holsanova, J. (1996). Attention Movements in Language and Vision, In: *Representations and Processes between Vision and NL. Proceedings of the 12th European Conference of Artificial Intelligence.* Budapest, Hungary. (pp. 81-83).

Vision and Discourse

A PhD. project plan

Jana Holsanova
Lund University Cognitive Science
Kungshuset, Lundagård
222 22 Lund, Sweden
jana. holsanova@lucs.lu.se

1 Introduction

People are very good at transforming information from one code into another. Even a child can explain verbally the content of a photography s/he is looking at, or draw a picture which illustrates a story s/he is listening to. Most of us use drawings spontaneously when we speak. The problem is only that we know very little about how we handle pictures in our mind and how we translate between the different information codes. The aim of my project is to gain more insight into the process of mapping visual representations into linguistic representations. In an empirical investigation I will examine a) spontaneous descriptions of a picture, and b) spontaneous descriptions of visual events in film sequences. The key question will be the correlation between patterns in our visual perception and patterns in our spoken language production.

2 Stages of the analysis

The first stage of the analysis is a study of visual comprehension. The eye movements of about 20 subjects watching a film are to be registered with the help of an *eye-tracker* (iView, SMI). The eye fixations and focus movements will be recorded on videotape so that the focus patterns can be mapped out for each individual. A qualitative analysis of the eye fixations will then be made (the x and y co-ordinates on the video film will be correlated with the attended objects).

As a second stage, I plan to analyse the spoken retellings of the film. The 20 subjects will give spontaneous descriptions of visual events after having seen them. They are given the instruction to describe to another person in detail what they have seen in the film. Their retellings are then tape-recorded and transcribed on the computer. These transcripts will then be segmented into intonation units (Chafe 1994), which allows an analysis of attention focuses in the retellings.

As a third stage, I intend to compare the results of the visual experiment with the results of the discourse analysis. Of particular interest are the correlations between the focus movements when watching the film and the focus movements when retelling the film. If the focus of attention was directed towards an object during the phase of visual perception, that object should be focused on also in the verbal description. In the spoken presentation this object should be marked with stress, discourse markers, or accompanied by pauses, introductions etc. The longer the visual fixation, the clearer should the focus be in the subsequent discourse.

3 Attention Foci in Vision

This project is inspired by an experiment performed by Charlotte Baker (cf. Chafe 1980) which has yet never been published. She studied the connection between the eye movements when we look at a still picture and the corresponding description of the picture content from memory. Like Baker, I will start will a still picture.The eye-movements of the subjects will be measured with the instrumental help of an eye-tracker. Subjects will be shown a picture, and their eye movements will be monitored while they watch. Then the pictures will be removed, and the subjects will be asked to recall the picture from memory and describe its contents

verbally. The subject's verbal description of the picture will then be compared to the eye movement patterns.

My hypothesis is that the linguistic segments will correspond closely to the gaze movements. According to the studies of eye movements when looking at static pictures (Buswell 1935, Yarbus 1967), there will be two general patterns of perception: "One of these consists of a general survey in which the eye moves with a series of relatively short pauses over the main portions of the picture. A second type of pattern was observed in which series of fixations, usually longer in duration, are concentrated over small areas of the picture, evidencing detailed examination of those sections. ... the survey type of perception generally characterizes the early part of an examination of a picture, whereas the more detailed study, when it occurs, usually appears later." (Buswell 1935) .

In the other part of the investigation, I will use *dynamic pictures* ("Pear stories", 6 minutes long silent film, W. A. Palmer Films, Inc.) in my visual analysis. Still today, we do not know very much about the eye movement patterns when perceiving dynamic pictures.

Since we have to take the flow of time into account, focusing on still and dynamic pictures is by necessity different. When people look at still pictures, they can control their attention. They can choose what to look at, in what order and for how long. This is why it is interesting to see whether people talk about what they saw in more or less the same temporal order, and for more or less the same amount of time, as when they scan over the picture. My hypotheses are the following: When watching a film, the temporal foci of the viewers will more or less be guided by the progress of the film. The order of visual foci will thus be partly determined by the film. I am aware that the gaze direction may also be influenced by the memories, feelings, knowledge, and preferences of the viewers.

In addition to studying the correlation between visual and linguistic foci in individual subjects, I also hope to detect similarities in the visual and linguistic patterns of all subjects. It is not unlikely that all people concentrate on the same two or three objects in a picture or a film frame. It is however also possible that people focus their attention on very different objects, according to their individual associations, interests, and goals. In the latter case, the subjects' perceptions of the picture/film will differ between them. In effect, they will have seen different pictures and films, and the differences can be revealed by the eye-tracker measurements.

4 Attention Foci in Discourse

Without having access to the eye-movements, Chafe (1980) performed an experiment to see how people retell events in a silent movie. He found that subjects formulated in small units, progressed in brief spurts, jumped between different sequences, and added digressions and comments. This result was interpreted by Chafe in the following way: We get vast amounts of information from our perceptual system, our emotions and our memory. But according to the needs, interests and goals we currently have, we concentrate our attention to specific parts. The activated unit represents the amount of information that a person can devote his central attention to at a time. The linguistic expression corresponding to such a focus, Chafe terms *intonation unit* (Chafe 1994). Since the total amount of information that we perceive vastly surpasses the limited capacity of a focus, it is necessary for us to scan the information using several foci, one after the other. When we retell something that we recall from memory, our attention therefore moves sequentially from focus to focus as the retelling progresses.

There are different means for inducing focus movements during spoken discourse (cf. Holmqvist/Holsanova, in press). In spoken narrative retellings, *discourse markers* are the most frequent ones. Discourse markers reflect the planning process of the speaker and are often used to guide the attention of the hearer. Focused intonation units, as well as the particular relation between subsequent intonation units are marked by discourse markers. "(A discourse marker) is a word or phrase … that is uttered with the primary function of bringing to the listeners attention a particular kind of linkage of the upcoming utterance with the immediate discourse context." (Redeker 1991: 1168). The function of discourse markers can be illustrated by the

following piece of a narrative (from Redeker, 1993):

1 but we had a seamstress
2 and we were calling her Mietje.
3 -->But I think we were calling everyone Mietje back then
4 you know, I don't know why,
5 -->but anyway, so that was also a Mietje.
6 And uh - she was from Belgium.
7 And there were- she was a Belgian refugee,
8 -->'cause during during the war,
9 during the First World War
10 all those refugees were coming from Belgium,
11 and they were coming to Zealand
12 and they were looking for work there.
13-->And so SHE was our seamstress, ...

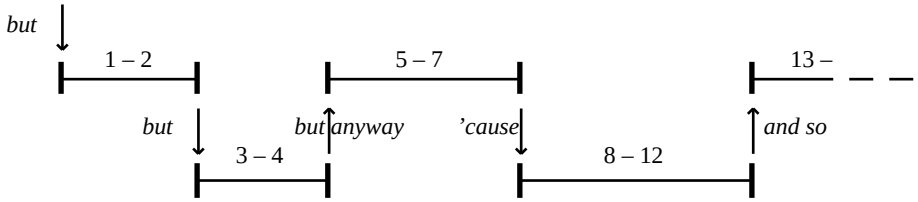

Fig 1: Focus structure of the narrative.

This narrative contains several side sequences, comments and digressions that are embedded in the main retelling track. The lines marked by arrows (3 to 5 and 8 to 13) are two examples of a side sequence. Lines 8 to 12 contain an explanation associated with the description of the main character, Mietje. The focus moves to a sidetrack signalled by the marker *'cause*. In line 13, the speaker closes this segment with the marker *and so* and refocuses back to the main track of the narrative. In fig 1, I suggest one way of illustrating the structure of this piece of the narrative.

5 Correlations between Vision and Discourse

My global hypothesis is that the transitions between intonation units and the more global units in discourse reflect the transitions of attention foci in vision. This hypothesis is based on the theory formulated by Chafe that similar mechanisms are used a) when we perceive information visually, b) when we recall it from memory, and c) when we verbalise it. "All these three processes may be guided by a single executive mechanism which determines what is focused on, for how long, and in what sequence." (Chafe 1980:16).

a) In my first investigation, based on a still picture, the linguistic segments will closely correspond to the gaze movements. Not only the parts of the picture, to which attention was paid, will be the same. Also, the order in which a subject focused on these parts will be identical in both cases. In other words, we will identify similar focus units in visual perception and in discourse production. The order of intonation units and the larger units in verbal decriptions will reflect the general order in which information was acquired visually.

b) In my second investigation, based on dynamic visual events in the Pear film, I also hope to detect similarities in the visual and discourse patterns. Some foci in vision will correspond to intonation units or other, more global units in discourse (such as series of punctual events followed by some kind of conclusion). However, the temporal order of the foci will differ. The subjects will add digressions and comments about the scenes, about the characters and about the film, according to their associations and preferences, they will jump between different aspects, and come back to the main track again. These transitions will be marked in their retellings through pauses, hesitations and discourse markers.

References:

Buswell, G. T. (1935): *How people look at pictures.* A study of the psychology of perception in art. Chicago: The University of Chicago Press, 1935.

Chafe, W. (1980): The Deployment of Consciousness in the Production of Narrative. In: W. Chafe (ed.) *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production.* Norwood, NJ: Ablex.

Chafe, W. (1994): *Discourse, Consciousness, and Time*. The University of Chicago Press: Chicago.

Holsanova, J. (1986): *Was ist Erzählen? Versuche zur Rekonstruktion des Alltagsbegriffs von Erzählen.* Unveröffentlichte Diplomarbeit. Humboldt Universität zu Berlin.

Holmqvist, K., Holsanova, J. (in press): Focus Movements and Internal Images of Spoken Discourse. In: *Discourse Constructions.* Proceedings of the ICLC '95.

Redeker, G. (1991): Linguistic markers of discourse structure. Review article. Linguistics 29 (1991), 1139-1172.

Redeker, G. (1993): The Attentional Structure of Spoken Discourse. Presentation at the Cognitive Linguistic Conference in Löwen, Juli 1993.

Schiffrin, D. (1987): *Discourse Markers*. Cambridge: Cambridge University Press.

Tomlin, R. S. (in press): Mapping Conceptual Representations into Linguistic Representations: The Role of Attention in Grammar. In: J. Nuyts and E. Pederson, *With Language in Mind*. Cambridge: CUP, pp. 1-21.

Tomlin, R. S. (in press): Focal Attention, voice, and word order. An experimental, cross-linguistic study. In: M. Noonan and P. Downing (Eds.), *Word Order in Discourse*. Amsterdam: John Benjamins, pp. 521-558.

Velichkovsky, B., M. Pomplun, and J. Rieser (1995): Attention and Communication: Eye-Movement-Based Research Paradigms. Forschungsberichte Band 11, Juni 1995. Dresden University of Technology.

Yarbus, A. L. (1967): *Eye movements and vision.* New York: Plenum Press.