

Holsanova, J. (1997). Verbal or Visual Thinkers? Different ways of orienting in a complex picture. In: *Proceedings of the European Conference on Cognitive Science*. Manchester. (pp. 32-37).

Verbal or Visual Thinker? Different Ways of Orienting in a Complex Picture.

Jana Holsanova
Dept. of Cognitive Science
Lund University
Sweden
jana.holsanova@fil.lu.se

ABSTRACT

A complex picture was shown to 12 subjects, and each subject was asked to describe it from memory. Their spoken descriptions were recorded, transcribed and segmented into intonation units (Chafe 1994). The transcripts were then qualitatively analyzed with respect to how the picture had been conceptualized. As a result, two different styles of description could be identified. Perceiving space was dominant in the more static technical style, while perceiving time was dominant in the dynamic narrative style. A parallel can be drawn between the technical versus narrative style, and the visual versus verbal thinker (Grow 1996). In study 2 and 3, the relation of visual perception and spoken language description is further sketched.

Keywords

spoken descriptions, spatial relations, temporal relations, narrative schema, orientation, verbal thinker, visual thinker, visual and verbal focus of attention

STUDY 1: DESCRIBING A COMPLEX PICTURE

The global question in my ongoing study is: What can linguistic representations in spontaneous spoken descriptions of a complex picture reveal about our way to perceive and conceptualize the world?

In my first explorative study, A complex picture was shown to 12 Swedish subjects of different age (8-64 years), both sexes (6 women, 6 men) and different background (economics, linguistics, computer science, philosophy, psychology, medicine). After the picture had been removed from their sight, each subject was asked to describe what s/he had seen. The spoken descriptions were recorded, transcribed and then segmented into intonation units (Chafe 1994); each intonation unit being the linguistic expression of a

focus of attention. The transcripts were then qualitatively analyzed with respect to how the picture had been conceptualized. The more specific questions in this are: In what way do the twelve speakers present and structure information about the picture when translating between two modalities (i.e. between their visual perception and their spoken language description)? Can spoken descriptions reveal something about the subjects' way of thinking? Are there great differences across the twelve descriptions of the same picture?

As a result of the first study, two different styles of description could be identified. Perceiving space was predominant in the first style, while perceiving time was predominant in the other one.

The Technical Description Style

The first group exhibited a largely technical style in their descriptions. The perceptual primacy of spatial relations led to a more static description of the picture. A typical beginning of the technical style is shown in Example 1:

Example 1.

0601 well it's a **picture'**
0602 **rectangular'**
0603 and ... it was mainly **green'**
0604 with a blue **sky,**
0605 divided into foreground and background

The subjects with technical style used many nouns but almost no dynamic verbs. They preferred the passive voice and overused the existential expression "there is" as well as the verb "to be". They delivered very detailed descriptions of objects, object elements and colours, but all details seemed to be of equal importance. The picture was also decomposed into fields, which were then described systematically, using a variety of terms for spatial relations (*in the middle, on the left side, at the very bottom, high up at the top of the tree, in the background, in the horizontal position*). In this technical description style, localizations were very frequent (they appeared in 17-40% of all intonation units). Whatmore, the subjects who used the most spatial expressions, tended to express localisations in a more exact and more elaborate way (see Example 2):

Example 2.

---> 0834 in the middle of the picture' there's a **tree'**
0835 uh . and in it three **birds** are sitting
---> 0836 in ... to the left in the tree
---> 0837 at the very **bottom'**
---> 0838 or or the **lowest** bird on the left in the tree' she sits on her **eggs'**
---> 0839 uuh and above her' still on the left I think'
0840 there is a a **bigger** bird'

The spatial expressions used here involve three different orientation axes: up-down, left-right and back-front. Those orientations coincide with the internal spatial framework of an upright observer (Taylor/Tversky, 1992, Lang et al. 1991).

During the description process, the subjects establish an elaborate set of referential frameworks for the purpose of localisation: They referred a) to the picture as a whole (*central in the picture, in the background, in the left corner of the picture*), or b) to the picture elements already described, focusing either on its components (*in his left hand the man holds..., the twig on the left hand side of the tree ...*) or on the relations between these picture elements (*the left hand peasant*). Thereby, the speaker established a landmark that can be referred to in the following description (*above the twig, behind the two standing men*). Thus, when introducing a new object, the speaker does not need to localize it in the picture as a whole.

Also, spatial expressions were used for different functions. They conveyed information about orientation and position of objects and, indirectly, they marked the speakers' steps in the description process. Focusing and refocusing on picture elements and their relations was done with the help of spatial terms combined with phonetic stress, loudness and voice quality. Almost no discourse markers were used for this purpose.

The Narrative Description Style

The descriptions of the second group were determined by the perceptual primacy of dynamic events in the picture. These descriptions followed a narrative schema. The subjects started with an introduction of the main characters, their involvement in different activities, and a description of the scene. A typical beginning of the narrative style is shown in Example 3:

Example 3.

0702	it's early in the spring '
0703	and Pettson and his cat Findus are going to sow
0704	. uh . it's a spring landscape '
0705	there is . arable and pasture land and cows '
0706	a stone wall '

In some cases, subjects even added a specification of the season of the year and the general mood of the picture. The sequential description of events in phases was typical of this narrative style (see Example 4).

Example 4.

---> 0106 and it starts with Pettson digging up his little **garden'**
---> 0107 then he **rakes'**
---> 0108 and . then . he **sows'**
---> 0109 **beds out'**
---> 0110 uh yeah plants **potatoes** later on,
---> 0111 and when he's ready he starts sowing **lettuce'**

The dynamic quality in this narrative style was achieved mainly through the use of dynamic verbs in active voice (*watering, singing, flying, chasing, running, jumping, beating*), and by the use of temporal verbs (*to start, to finish*), temporal adverbs (*first, then, later on*) and verb coordinations expressing ingressive or durative action. The subjects specified activities on a time scale and reflected time differences in the two halves of the picture. However, they used very few spatial expressions (in 3-7% of their intonation units) and the localisations used were less exact (*at a distance, a little bit longer way off, in one corner, around, on the fringe of the picture, somewhere in the picture*).

Focusing and refocusing on the various picture elements and their relations was mainly done with the help of discourse markers. Discourse markers, sometimes also called discourse operators or attention markers, appear as conjunctions, adverbials, interjections, particles or lexical clauses (*well, but, cause, so now, anyway, and so, okay, and then, by the way, I mean*). "A discourse operator 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, cf. also Holmqvist & Holsanova 1997).

My results derive independent support from studies about writing problems. According to Grow (1996), visual thinkers have difficulties organizing expository prose because their preferred mode of thought is fundamentally different from the verbal thinkers. Whereas verbal thinkers constantly analyze, compare, relate, and evaluate, visual thinkers often list features without taking a position, imposing an order, or presenting an action. The descriptions of visual thinkers are static, they are not arranged in a dynamic sequence. Also, visual thinkers allow for many elements to appear at once, simultaneously, not marking the connections between them. The reader is expected to interpret and to make the connections. Furthermore, visual thinkers do not focus on words, but on such non-verbal values as line, color, texture, balance, and proportion.

When looking back at the results of my first study, these comparisons can be made: On the one hand, the dynamic, rhythmic and thus very lively narrative description style, where connections are explicitly done with the help of discourse markers, is close to the verbal thinkers' way of writing. On the other hand, the static character of the technical

description style, with the perceptual primacy of spatial relations, where the picture is decomposed into fields, many details are listed but no explicit connections are made, reminds us of the visual thinkers. Despite the fact that Grows analysis is based on written language, whereas mine is based on spoken language, a parallel can be drawn between the technical versus narrative style and the visual versus the verbal thinker.

STUDY 2: PICTURE DESCRIPTIONS IN A TASK ORIENTED COOPERATIVE SETTING

In another ongoing study, the picture descriptions will be produced in a task oriented cooperative setting. The following questions will be relevant: Will the nature of spatial expressions be different from that in my first study? How will the visual information be verbally transmitted in a context, where the speaker instructs a partner about the positions of picture elements with the aim to enable her/him to draw the picture? Will the observer perspective dominate or will the speaker switch between different perspectives?

My hypothesis is that the speakers a) will use numerous, exact, elaborate spatial expressions close to the technical style, b) they will describe both sizes, shapes, distances and angles in a more elaborate way, c) they may partly adopt the intrinsic, object-centered perspective. According to Tversky et al (1994:663) who examined subjects' descriptions of environments, "The spatial models formed from observing a scene seem to be the same as those formed from reading about a scene". My question is: Will the spatial models from perception be the same as those formed from listening to spoken descriptions?

STUDY 3: CORRELATIONS BETWEEN VISUAL AND VERBAL FOCUSES OF ATTENTION

In a third study, the same complex picture was shown to 4 subjects and their eye movements were registered during one minute. The eye tracker used was an SMI iView 50 Hz pupil and corneal reflex system. After the picture has been removed, each subject was asked to describe what s/he had seen. The subjects' eye movements were once again registered while the verbal descriptions were recorded. As in Study 1, the spoken descriptions were transcribed and segmented into intonation units. The question for the analysis was if there are any similarities between the visual focus patterns (both during perception and description phase) and the focus patterns in the spoken description.

Several studies have been made of our visual perception of pictures. Buswell's (1935) investigation of the way we look at pictures revealed some general patterns in eye-fixations. Buswell's (1935) investigation of the way we look at pictures revealed some

general patterns in eye fixations. Since Yarbus' classic observations (1967), eye movements have often been thought to reflect cognitive events. The question is only how much can be learned about cognitive operations by inspecting eye fixations patterns? One of the difficulties of relating fixations to cognitive processes is that fixation itself does not always indicate what properties are being acquired. The idea was to use the linguistic descriptions as a source of such indications. The main hypothesis in Study 3 is that we use similar mechanisms when we perceive information visually, when we recall it from memory, and when we verbalize it (Chafe 1994, Just and Carpenter, 1976).

CONCLUDING REMARKS

In study 1, a complex picture was shown to 12 Swedish subjects, and each subject was asked to describe it from memory. In spite of the fact that the subjects looked at the same picture during the same time with the same instructions, two different styles of description could be identified. In this way, the spoken descriptions revealed the subjects' way of thinking. Depending on whether the subjects primarily perceived space or time in the picture, they tended to use either the static technical description style, or the dynamic narrative style. They used different linguistic means when presenting and structuring information about the picture, according to how they perceived it visually.

When comparing linguistic and visual data, a number of questions emerge. Two of them are the following: How will the visual information be verbally transmitted in a context, where the speaker instructs a partner about the positions of picture elements with the aim to enable her/him to draw the picture? Can the eye movement patterns reveal verbal and visual thinkers? To answer these questions, more empirical work must be done, as indicated in Study 2 and 3.

REFERENCES

- Buswell G. T. (1935). How people look at pictures. A study of the psychology of perception in art. The University of Chicago Press, Chicago.
- Chafe, W. (1994). *Discourse, Consciousness, and Time*. Chicago University Press, Chicago.
- Chafe, W. (1996). How Consciousness Shapes Language. *Pragmatics and Cognition* 4, 1, 35-54.
- Finke, R. A. & Shepard R. N. (1986). Visual functions of mental imagery. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.). *Handbook of perception and human performance*. Wiley, New York.
- Grow, G. (1996). The Writing Problems of Visual Thinkers. <http://www-wane.scri.fsu.edu/ggrow/>. An expanded version of a print article that appeared in the refereed journal, *Visible Language*, 28.2, Spring 1994, 134-161.
- Herzog, G. (1995). From Visual Input to Verbal Output in the Visual Translator. Bericht Nr. 124. Sonderforschungsbereich 314. Universität des Saarlandes, Saarbrücken.

- Herzog, G. (1995). From visual input to verbal output in the visual translator. Bericht Nr. 124. Sonderforschungsbereich 314. Universität des Saarlandes, Saarbrücken.
- Holmqvist, K. & Holsanova, J. (1997). Reconstruction of Focus Movements in Spoken Discourse. In Liebert, W., Redeker, G., Waugh, L. *Discourse Construction*. Proceedings of ICLC 95. Benjamins, Amsterdam.
- 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, 81-83.
- Just, M.A., Carpenter, P.A. (1976). Eye fixations and cognitive processes, *Cognitive Psychology* 8, 441-480.
- Lang, E., Carstensen, K.-U., & Simmons, G. (1991). Modelling Spatial Knowledge on a Linguistic Basis. Theory - Prototype - Integration. (= *Lecture Notes on Artificial Intelligence* 481) Springer-Verlag, Berlin, Heidelberg, New York.
- Redeker, G. (1991). Linguistic markers of discourse structure. Review article. *Linguistics* 29 (1991), 1139-1172.
- Taylor, H. A., Tversky, B. (1992). Description and Depiction of environments. *Memory and Cognition*, 20, 483-496.
- Tversky, B, Franklin, N., Taylor, H. A., Bryant, D. J. (1994). Spatial Mental Models from Descriptions. *Journal of the American Society for Information Science* 45, 9, 656-668.
- Yarbus, A. L. (1967). *Eye movements and vision*. Plenum Press, New York.