

# Figure Captions in Visual Interfaces

Bernhard Preim, Rainer Michel, Knut Hartmann, Thomas Strothotte

Otto-von-Guericke University of Magdeburg

Faculty of Computer Science

P.O.Box 4120

D-39016 Magdeburg, Germany

{*bernhard|rainer|hartmann|tstr*}@isg.cs.uni-magdeburg.de

## ABSTRACT

We present a general concept for the enhancement of visual interfaces with automatic figure captions describing a visualization. The incorporation of figure captions in interactive systems raises some fundamentally new questions as these images are exposed to changes. The dynamic aspects to be considered include the update and customization of figure captions.

We employ figure captions not only for the description of images but also for their modification leading to the introduction of *interactive figure captions*. A general architecture is developed and comprehensively described referring to two application domains: medical illustrations and geographic maps.

*Keywords:* Dynamic figure captions, interactive figure captions, medical illustrations, geographic maps

## 1 INTRODUCTION

Visualizations are produced to enable a viewer to extract information. For this purpose, visualizations are not merely a straightforward rendering of the data. Limited presentation space imposes restrictions to the visualization, resulting in omissions, exaggerations or displacements. Moreover, viewers wish to explore the data under a thematic focus. Thus, portions of the data may be presented with more detail or more comprehensively, while others may be simplified, shrunk or even left out.

As a result of sophisticated manipulations, complex images may arise which are difficult to interpret. In traditional print media, an image is therefore often accompanied by a figure caption which describes verbally how the image has arisen, what message it should convey and which details may be important, particularly when these are difficult to recognize.

Textual information to enhance the interpretation of images is often neglected because images are thought to be descriptive on their own. This is not entirely true, as the art historian GOMBRICH pointed out: “No picture tells its own story” [5]. WEIDENMANN, an educational psychologist examining

how to learn with images, builds on GOMBRICH’s work and argues that figure captions are crucial in a learning context [19]. They enhance the interpretation of images and make it easier to remember the contents of an image over a longer term [11].

In this paper, we show how the potential of figure captions can be employed for the enhancement of *visual interfaces*—interactive systems that produce images based on an underlying model. We describe the incorporation of figure captions with reference to anatomical illustrations and geographic maps. Furthermore, we discuss how to maintain consistency between visualizations and their captions. Besides the content selection (*what* to describe), the generation of figure captions requires the selection of phrases to verbalize this content. This incorporates a linguistic analysis of figure captions and other texts in the intended domain in order to generate appropriate phrases automatically.

This paper is organized as follows. First, we analyze figure captions in print media to derive the basic structure of descriptive figure captions (Section 2). A short review of related work is presented in Section 3. In Section 4, we discuss issues of figure captions specific in interactive systems. In Section 5 we show how to use figure captions to manipulate an image: we present *interactive figure captions*—captions with sensitive parts that the user can modify. A concept for the integration of figure captions in interactive systems is developed in Section 6. Based on this concept, applications of figure captions are presented in Section 7 for anatomy and in Section 8 for cartography. The final section summarizes the approach and provides an outlook on future work.

## 2 FIGURE CAPTIONS IN PRINT MEDIA

In traditional print media, diverse kinds of figure captions can be found. By no means all of them can be generated automatically or can be regarded as prototypes for interactive systems. To clarify to which figure captions we restrict ourselves, we propose the following classification. Following BERNARD [2], we refer to *descriptive figure captions* if the content of an image is described verbally.

We shall use this definition in a broader sense. Descriptive

figure captions describe a view on a model or a section of the real world. This definition also comprises phrases describing not only what is visible in the figure but also what is hidden, has been removed or what important objects are close to the depicted portion. Consequently, in our terms descriptive figure captions describe an image and its (spatial) context. Descriptive figure captions are employed e.g. to explain the construction of complex objects.

By contrast, *instructive figure captions* describe either how to interpret an image by directing attention of the viewer at important objects (typical phrases include “Look carefully at ...”) or what to do in the real world with the depicted objects. Instructive figure captions, for example, explain how to operate technical devices. These captions are often employed to describe different stages of a complex operation. How to use objects is described and it is assumed that the construction of these objects is known.

This classification covers a large portion of captions in reference materials, text books and repair manuals. Other figure captions, e.g. citations of people shown at a photograph in a journal, are beyond the scope of this classification.

In this paper, we deal with descriptive figure captions only. We show how the content of these captions can be derived automatically based on *structured models* and the user interaction with it. “Structured models” refers to models that consist of distinct objects together with some semantic information, at least the names of objects and the affiliation to categories.

By contrast, the generation of instructive figure captions requires not only semantic information about object names and categories, but also a considerable amount of information about the application domain, for example, about possible complications in the maintenance of objects and alternative ways to perform an operation under difficult circumstances.

In the following, we analyze descriptive figure captions in two application fields—anatomical atlases and maps. While figure captions in anatomy have developed over a long period of time, similar comprehensive descriptive texts are less wide-spread in cartography. Only in rare cases have legends been extended to inform about artifacts of the generation process [15].

## 2.1 Figure Captions in Anatomy

Anatomical atlases consist of large, often complex images which are not surrounded by textual information as in text-books (hence the name “atlas”). Figure captions are the only form of textual information available and do not interfere with other references to an image. We conducted a series of interviews with medical students; these revealed that figure captions are carefully studied to get an orientation in the study of complex images. This analysis is based mainly on the widely-known atlas of SOBOTTA [16].

Figure captions in anatomy follow a rather fixed structure.

The first items mentioned are generally the name of the depicted contents, the viewing direction and the important aspects (e.g. “muscles and sinews of a foot from lateral”). This information is essential if an unfamiliar picture of organs inside the human body is depicted. After the information about the image as a whole, important parameters of single objects are usually described. Among these parameters, manipulations that affect the visibility of objects are most important because they often influence the whole composition. Typical phrases include “[object<sub>1</sub>] has been removed to show [object<sub>2</sub>]” or “[object<sub>1</sub>] has been removed in the area of [object<sub>2</sub>]”. If small objects are important in a specific context, they must be enlarged to emphasize them. In this case, the context is preserved for better orientation, so that the surrounding objects cannot be enlarged. Such modifications are reflected by phrases like “[object] slightly enlarged”.

Besides geometric manipulations, *presentation variables* like colors and textures are often adapted to a specific context, e.g. to show spatial relations more clearly and to communicate which objects belong to the same category. In anatomy, objects of certain organ systems are colored uniformly according to accepted conventions<sup>1</sup>. In figure captions, the use of colors therefore needs to be described only if it differs significantly from the conventions or even conflicts with them.

An interesting facet is the generation of one figure caption for several images, for instance when images show the same model from different directions. Furthermore, visualizations may exploit the symmetry form of anatomic objects and show different aspects in both halves. In both cases, the similarities between the images are mentioned first, while the differences are described later. If horizontally arranged images are described by one caption, it is important that the left image be described before the right one, because there is a natural sequence of “reading images” from left to right.

Figure captions also depend on other textual components. In order to refer to objects via their name, they must either be labelled or circumscribed by characteristic features which are easy to recognize in the image.

## 2.2 Figure Captions and Legends for Maps

Maps are an abstract visualization of a part of the real world. Instead of presenting geographic objects as they appear on aerial photographs, symbols are used to present categories of objects. Depending on the map scale, legible symbols may become larger than the objects they represent. Resulting overlappings are removed using symbol displacement. With decreasing scale, the symbols are more and more simplified or even removed. These abstraction processes are referred to as *cartographic generalization* [8].

Color is used independently of the appearance of the objects in the real world. Instead, it is employed so that the discriminability of the symbols is ensured. In thematic maps usually

---

<sup>1</sup> Muscles are depicted red, nerves are yellow and bones are white

more than one thematic variable (e.g. population density and net income) needs to be presented, but size can only represent one. Therefore, color is often used as an additional degree of freedom to encode another thematic variable. Maps are, thus, constructed of symbols which do not resemble the objects they represent, neither in shape nor in size or color. Hence, maps are usually provided along with legends that explain the use of the symbols.

Besides this purpose, legends may serve more functions. SCHLICHTMANN identified major functions of map legends [15]. From his enumeration we derive the following:

**Provision of additional information about mapped objects.** A legend may contain information that is not shown in the map itself. For example in a map that shows the value of a thematic variable for each region, the legend may give an explanation of the symbols and in addition present them with their value for the whole area.

**Presentation of the underlying structure.** Sometimes the data inheres a structure that is not easily conveyed in the map. For example in maps for bedrock geology, the temporal sequence of the sediments is not encoded in the map and, thus, needs to be presented in the legend. This can be accomplished e.g. in such a way that besides the explanation of the textures or colors used for each sediment their age is described.

**Information about the underlying classification scheme.** If a variable may represent a large number of values or even continuous values it is not feasible to code all of them as different colors. Instead, categories are formed using domain knowledge (e.g. an income below \$200 is under the poverty line). Each category is assigned a color value and the legend can be used to reveal the underlying classification scheme.

**Provision of information about the visualization.** This is probably the most interesting aspect of the map legends' functionality, as it reveals what has happened in the abstraction process that transformed the real world data to the map. It covers in particular all aspects of presentation fidelity. For example, in a map about farm sizes in Canada you may find statements like "Values are rounded to the nearest five farms." indicating the error for the rendering of the number of farms.

Restrictions to the map fidelity are essential to the map reader. Map fidelity usually lessens with decreasing map scale. Small scale maps need a more comprehensive generalization to display all symbols with sufficient size and to resolve graphical conflicts between them [9]. Through these processes, objects of minor importance are removed and symbols are repositioned to prevent overlapping. Hence, the maps thus obtained no longer reflect the reality in full detail—they lie e.g. with respect to the completeness or distances between objects (cf. "How to lie with maps" by MONMONIER in [8]).

When maps are no longer manually created, the data about the fidelity can be gathered as a direct by-product of the visu-

alization process. Hence, it may be used to generate legends that reveal the inaccuracy that was introduced in the map and help the reader to determine which operations are still performable on the map.

Considering the various functions of map legends, it becomes apparent that they should be structured in different ways according to their purpose. For the most wide-spread use of map legends—symbol explanation—a schematic legend is used. This is a table-like legend where in the first column the symbol is displayed and in the second column its semantics is explained.

However, for the extended functionality this simple approach is not feasible. The information to be presented is more complex and not as easily structurable. For these cases, textual descriptions ought to be used. We refer to these natural language descriptions as figure captions, which is in accordance with the wide-spread use of the term outside the area of cartography. We will focus on the generation of figure captions for maps describing the visualization process, i.e. the operations that have been performed on the data to obtain the map and their impact on the reliability of the map.

### 2.3 Generalized Structure of Figure Captions

In order to derive a common approach to the generation of descriptive figure captions, a generalized structure is presented in the following.

Figure captions describe the depicted content and the view on an underlying model. This model may be three-dimensional, as in anatomy, or two-dimensional, as in cartography. The view on the model is characterized by a viewing direction (in 3D) or by the visible portion (in 2D). Besides this geometric aspect a view is determined by thematic aspects on which the visualization focuses (e.g. water ways in maps or the human skeleton in anatomy).

Since visualization is a sophisticated process, manipulations that restrict the image fidelity ought to be described in the caption. This covers the mentioning of single objects whose visibility or position was modified during the visualization process.

Often, there is no uniform scale in the whole visualization. Instead, the scale is adapted to the density of symbols and objects or their size. Since these modifications usually remain unnoticed to the untrained eye, their existence needs to be described in a figure caption.

Conventions, e.g. for the use of colors, reduce the need for comments in figure captions. The variety of the formulations in figure captions is high, but the basic structure is fixed and some typical phrases dominate in each domain.

## 3 RELATED WORK

The incorporation of figure captions in interactive systems is a new approach to enhance the usability of visual interfaces.

Figure captions in a narrow sense, that is captions that describe images previously generated, were first developed by MITTAH et al. [7]. This work has been carried out in the context of the SAGE-project, where complex diagrams are generated. MITTAH et al. argue that users can deal with more complex diagrams if they are complemented by explanatory captions. To put it in other words: complex relations that must be depicted in several images (without captions) can be integrated in one diagram when explained appropriately.

Some systems produce what we introduced as instructive figure captions (recall Section 2). In particular, in the WIP project (Knowledge Based Information Presentation, see WAHLSTER et al. [18]) and in the COMET project (Coordinated Multimedia Explanation Testbed, see FEINER and MCKEOWN [4]) technical illustrations and textual descriptions are generated to explain the repair and maintenance of technical devices. The textual components are generated based on large knowledge bases.

The selected content is presented within different media (graphics, animation and text) which are used either complementary, for example a verbal description conveys information which is hard to visualize, or redundantly. The redundant information presentation aims at reinforcing the message, e.g. by appropriate cross-references. In this *media selection* process the suitability of a medium to convey information is considered. Moreover, sophisticated *media coordination* facilities are employed.

This situation is considerably different from the problem we describe in this paper. In WIP and COMET, graphics and text generation mutually influence each other: the textual expressions generated affect the graphics generation and do not only comment them. The content of the figure captions often goes over and above the content of the figure.

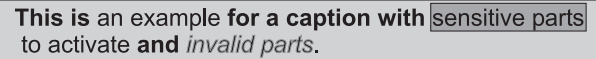
## 4 DYNAMIC FIGURE CAPTIONS

For enhancing interactive systems by using figure captions, ideas can be borrowed from print media. This orientation, however, is limited to static aspects. The incorporation of figure captions in interactive systems raises some fundamentally new questions. Figure captions describe images exposed to changes and therefore must be updated. Furthermore, incremental changes of figure captions are useful for easy identification of changing parts. It must be decided whether a figure caption should be legible in its entirety or, for example, be embedded in a dialogue with scrollbars. This section discusses possible answers to these and other related questions.

### 4.1 Layout Considerations

It is important to present the contents in a pleasing and legible way. Fonts and their sizes are important parameters to achieve these goals. Figure captions refer to the image as a whole and, therefore, have a higher priority than other textual components, e.g. labels, which refer only to parts of an

image—therefore, they should be displayed larger. In particular, a figure caption should be arranged so that it is easily perceived. While these observations are also valid for static figure captions, additional issues arise in interactive systems. Figure captions may contain sensitive parts for selection and even invalid parts—these features must be communicated with an appropriate layout. The sketch in Figure 1 presents a suggestion for such a visualization.



This is an example for a caption with sensitive parts to activate and *invalid parts*.

**Figure 1: A caption has sensitive parts (with an underlying rectangle) and invalid parts (light grey and italic). Bold parts represent fixed phrases which are based on the templates (cf. Section 6.1).**

Moreover, figure captions in interactive systems may change in length over the time. It is desirable that the caption be changed incrementally so that those parts of a caption which do not change remain at their positions. This strategy may, however, conflict with a well-balanced layout with captions centered under an image.

Figure captions in interactive systems are thus exposed to various changes. Their position, their presentation and even their content may change. We therefore refer to them as *dynamic figure captions*.

### 4.2 Adaptable Figure Captions

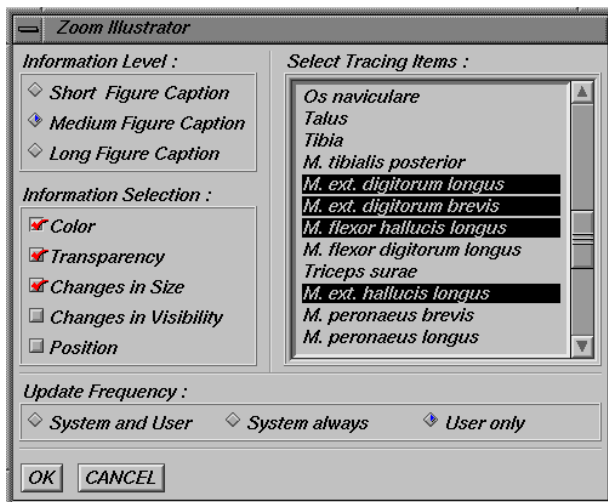
While figure captions in print media are necessarily static, dynamic figure captions can be tailored to the needs of the current user. Such an adaptation is useful because the overall amount of information to mention in a figure caption can become quite large. Adaptation facilities should enable users to control the content selection. Several parameters are particularly important for adaptation:

**Content Selection.** The caption generation process can be adapted to what kind of information is to be presented and to which level of detail. The information selection option (see the top left part of Figure 2) allows a user to request a notification when a certain property of the image has changed, i.e. when a presentation variable, like visibility or color has been altered, or when geometrical aspects of the image (like the size or the shape of a single object or its relative position) have been manipulated (see the bottom left part of Figure 2).

In addition, a user may control the amount of information by specifying the level of detail (see the upper left part of Figure 2). Selecting short figure captions means that only the most important changes in each selected category are presented whereas long figure captions inform the user about every modification to the selected property.

**Object Monitoring.** Users should be able to express their interest in specific objects or regions (parts of the under-

lying model). Based on such a specification, the user will be informed about all changes that affect the objects or regions traced and of their current state (see the list in the right part of Figure 2). For instance, if an object becomes hidden, the system comments on the visibility of such objects (e.g. “the traced [object<sub>1</sub>] is currently hidden by [object<sub>2</sub>] and [object<sub>3</sub>]”). This is similar to debugging tools which allow the user to monitor certain variables.



**Figure 2: Adaptation of figure captions. In the left part, the content selection is customized. In the right part, the user selects objects for monitoring.**

### 4.3 Updating Figure Captions

Nearly all interactions on images lead to incomplete or even invalid figure captions. This raises the question of when the caption is to be updated and who initiates the update process. In any case, the system should detect whether changes in the image affect the figure caption. In this case, the system or the user can initiate an update. Between these two variants, hybrid approaches are possible where the system initiates an update only after major changes and the user can always initiate an update.

**Update on explicit request only.** The caption is updated only upon the user’s request. The system indicates invalid parts (recall Figure 1).

**Automatic update.** Compared to the other variants, automatic updates by the system require fast generation of captions. This may confuse users because of the high frequency of changes in the corresponding area of the screen. However, the advantage of this approach is that captions remain consistent with the image at all times.

**Hybrid variant.** The caption is updated automatically only if radical changes in the image occur. Examples are considerable changes in the viewing specification (e.g. via a rotation) leading to a change in the visibility of a large number of objects or the removal of a class of objects due to filter-

ing operations. Another example is the incorporation of an additional view which requires the figure caption to describe both views.

As there is no optimal variant that is suitable for all application domains; leaving the decision on the initiation of an update up to the user seems to be the preferable option (see the bottom line in Figure 2).

## 5 INTERACTIVE FIGURE CAPTIONS

Figure captions, as we know from a variety of printed material, comment an image. They provide background information about the creation of an image and its context.

So far, we have described the incorporation of *dynamic* figure captions in visual interfaces. Although specific issues, like the update and adaptation emerged, the basic function of figure captions remains the same—to guide the interpretation of images.

As we show here, figure captions can be used for a very different purpose—they can be manipulated and cause an update of the image. With this approach figure captions serve as input for the graphics generation process. Therefore, we refer to these captions as *interactive* figure captions.

Usually, an image can be manipulated with very different handles. There are dialogues or handles to invoke transformations, material editors to change presentation variables and other dialogues to initiate filtering operations. The interactive manipulation of the corresponding parts of a figure caption, now, unifies these interaction facilities.

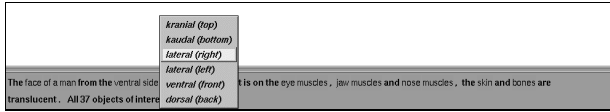
Interactive figure captions are also attractive in another respect—a figure caption does not only describe an image but also makes it possible to modify it. The image is no longer superior to the caption in that the caption is dependent on the image, but both stand on the same level and offer interaction facilities.

In this section, we demonstrate first results of this attempt. Note that this approach differs from the graphics and text generation in WIP and COMET (recall Section 3). In our concept, either the caption describes an image *or* the image is guided by the caption. The generation processes are still autonomous.

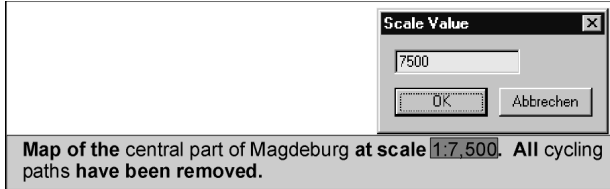
It is necessary to guide the user through this interaction. An unrestricted editing of the caption may lead to the specification of infeasible or ambiguous commands. Moreover, the user may not be aware of what interaction options are available. Guiding the user also avoids the enormous problems coming along with natural language understanding.

Therefore, sensitive regions are provided (recall Figure 1) which display simple dialogues with alternative choices for this item when selected. These can be pop-up menus, sliders, text edit fields for the specification of numbers (see Figure 3 and 4). Pop-up menus, which are used for the selection out of a small number of items, can be derived from the specifi-

cation of the lexical mappings of the corresponding template variable. Using this approach, a modification of the external mapping specification is always kept consistent with the pop-up menu.



**Figure 3: Interactive figure captions.** The selection of “lateral” causes a pop-up menu to appear containing the six main viewing directions.



**Figure 4: Interactive Figure Captions.** Using a text edit window to modify the scale value.

However, using these interaction facilities is not always convenient: three-dimensional widgets allow more direct control on 3D transformations, while WYSIWYG-color editors may be more intuitive to select colors than a list of color names. On the other hand it is important that the selection of a sensitive part in a caption leads to a uniform reaction. As a trade-off between dedicated interaction facilities and consistent feedback, the pop-up menus which are displayed after the selection can contain an item that invokes a dedicated editor, tailored to the interaction task.

The application of interactive figure captions reveals that it is not desirable to modify all variables of a figure caption. In particular, it is questionable whether modifications should be possible which would invalidate large portions of the caption. If the user, for instance, can alter the subject of the visualization the rest of the caption becomes useless. As a rule of thumb, parts of the caption that describe *how* something is depicted should be interactive while parts that describe *what* is depicted should not.

The interactive usage of captions is also limited by the fact that figure captions must correspond to generatable images. The system can only offer attributes for manipulation if it can actually handle an alternative specification for this attribute. If the system describes that “[object<sub>1</sub>]” is hidden by “[object<sub>2</sub>]”, the user cannot be allowed to delete “[object<sub>2</sub>]” or to replace it by another object, because it may be impossible to generate such a visualization.

## 6 INTEGRATION OF FIGURE CAPTIONS IN INTERACTIVE SYSTEMS

In this section we develop a concept for the integration of figure captions in interactive systems with respect to the is-

ssues discussed above. In particular, the generation of figure captions and the update process is considered.

### 6.1 Template-Based Generation

In the field of text-generation there are basically two approaches which differ considerably in the implementation effort, on the one hand, and in the quality of the textual descriptions generated, on the other hand: *template-based generation* and *natural language generation*

The template-based approach is a low-cost and low-quality approach which is based on prepared phrases, called *templates*. A popular example of template-based text generation is the APPLE MACINTOSH BALLOON Help System [13]. Templates contain variables as place-holders which are substituted by the current values (e.g. colors, viewing directions). This replacement requires that the numerical values representing the state of an interactive system are converted into verbal expressions. In its simplest form there are straightforward mappings of numerical values from an interval to a word or phrase.

Natural language generation (NLG) does not rely on prepared phrases but generates the whole expression from a semantic representation. Obviously, NLG is much more flexible, as it can consider many parameters. Thus, documents can be generated in different languages and different styles (concise or verbose, informal or polite) based on the same semantic representation<sup>2</sup>. The generation of sentences takes into account what has already been generated. Furthermore, similar expressions are merged. However, this flexibility is gained at the expense of a higher effort, which pays off especially in larger texts.

Templates, on the other hand, can be generated with less computational effort. They are appropriate for shorter texts that follow a rather fixed structure. Figure captions are an application for which acceptable results can be expected with templates.

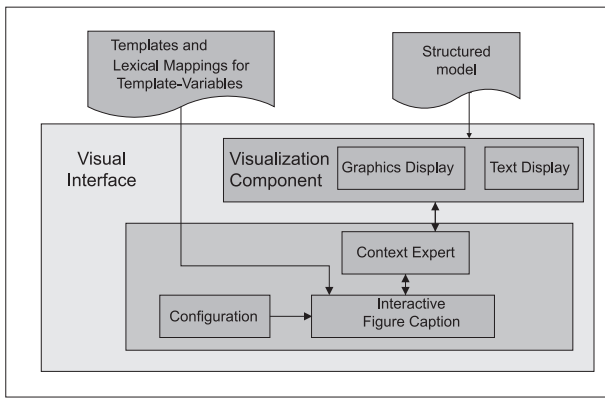
### 6.2 An Architecture for Figure Captions in Visual Interfaces

The generation of figure captions that remain consistent with the image requires that all changes to a visualization (e.g. an illustration, a map or a diagram) are represented explicitly in data structures. As we pointed out in Section 2, figure captions also depend on textual components, e.g. on the availability of labels. Therefore, the *visualization component* consists of a text and graphics display (cf. the architecture in Figure 5).

In order to describe the modifications that have been applied to the visualization, it is necessary to represent not only the cumulative state of an image and the corresponding textual elements, but also the interaction history. For this purpose,

<sup>2</sup>For a comprehensive comparison of both approaches refer to [13] and [14].





**Figure 5: Architecture of a visual interface with dynamic figure captions.**

an agent is required which is informed whenever the visualization changes. This agent manages the context of the interface and is therefore called the *context expert*<sup>3</sup>.

The Context Expert communicates with the *graphics display* and the *text display* and with the *interactive figure caption*. It analyzes the changes and initiates the text generation based on the user's specification (recall Section 4.2). The generation is also supplied with a separate description of sentences (templates) and the lexical mappings for template variables (often numerical values).

If figure captions are manipulated to control the visualization, a message is sent to the Context Expert (representing the requested changes) and the visualization.

### 6.3 Basic Scheme

The incorporation of figure captions in interactive systems requires, in general, a number of steps to be carried out. Based on our experience, we consider this to be:

1. a task analysis,
2. an analysis of the visualization algorithms,
3. an event analysis,
4. the content selection and
5. the linguistic realization

**(1) Task Analysis.** This step entails the identification of typical visualizations in a specific domain including groups of visualizations that are often viewed at in unison with one another. The analysis is performed with respect to the operations that viewers usually perform on these images in order to extract the requested information. As a result, the operations are identified that should be supported by a class of

<sup>3</sup>This terminology is related to the reference model for Intelligent Multimedia Presentation Systems, see [3]. This reference model has been established as a joint effort of researchers. The core of the reference model is an architectural scheme of the key components of multimedia presentation systems.

visualizations. This yields directly the properties of the image that must be preserved, in order to achieve a presentation that fulfills a viewer's demands.

**(2) Analysis of the Visualization Algorithms.** Based on the properties to be preserved, visualization algorithms are examined as to whether they may conflict with the above mentioned information extraction. Such conflicts may emerge from several constraints that affect the visualization process, e.g. legibility, visual discriminability, screen resolution or limited presentation areas. As a result of the application of these algorithms, important properties may be violated which should be signalled to the viewer. Hence, the application of such algorithms yields candidates for the content selection for a verbal description of the visualization.

**(3) Event Analysis.** In the third step, an *event protocol* is established comprising the events that may effect the verbal description. In general, these events reflect the operations identified in Step 2 that may restrict the extractable information. The events are sent to the *context expert* (recall Figure 5) and analyzed with respect to the actual results of the operations, i.e. whether any preservable properties of the visualizations have been violated globally or locally. Note that user-initiated operations, which usually need no further comment of the system, may have side effects that do require a notification.

**(4) Content Selection.** In this step, the events that have actually lead to modifications in the image are prioritized, sorted and filtered according to their impact and the user's specification (recall Section 4.2). Related events may be combined, events of minor importance may be removed, user-initiated events may be ignored if they have no severe side effects. As a result, a subset of the original events together with information about causal and temporal order is used for the content selection.

In order to ensure the consistency between the figure caption and the image, the statements already generated must be checked as to their validity. This applies also to the validity of lexical mappings for template variables. This process involves an analysis of the effects of events generated, e.g. about the visibility of objects after a rotation of a 3D model. As a result, it may be sufficient to indicate what is no longer valid (recall Section 4.2).

**(5) Linguistic Realization.** Based on a linguistic analysis to identify typical phrases in the domain, the text generation is prepared. This analysis includes figure captions and other kinds of textual information related to the description of images. This covers three steps:

**(a) Text Structure Analysis.** The sequence of text blocks is analyzed resulting in a description of the contents of each block, e.g. the existence of a title at the beginning, giving a concise description of the depicted contents and a description of the single modifications in the following text blocks.

**(b) Sentence Structure Analysis.** If the template-based approach is taken, the result of this analysis is a set of templates

(natural language expressions) together with conditions as to how templates should be selected. The templates should comprise typical phrases and be as concise as possible. They are grouped into categories from which no more than one is selected in the generation process. It is surprising that even similar parameters, like viewing directions or positions, may be expressed differently in different domains.

**(c) Lexical Realization of Template Variables.** The final step is the mapping of numerical values to template variables. In the case of the template-based approach for each attribute an interval of the numerical value is directly mapped to a sequence of words. With this approach, a value is always described in the same way. Colors are an example where this is useful. For this attribute, a widely accepted color naming system has been developed by BERK et al. [1] which can be used to describe colors objectively.

In some cases, absolute mappings using fixed intervals are not suitable. Values are considered differently depending on the range of the values for this quantity in a given visualization; what is large in one context may be very small in another. These issues are discussed in more detail in [17].

## 6.4 Representation of Events

All operations that affect the textual or graphical display emit an event notifying the Context Expert. Operations can affect the whole model, single objects, objects of a region or a certain category. Therefore the range of an operation is an important parameter.

As described earlier, modifications may be caused by the system, e.g. to adapt the visualization to the context. For the viewer's comprehension it is crucial whether something changes on his or her request or by an automatic process. Therefore, all events include a parameter indicating who initiated the change: the *user* or the *system*. If the system initiated a manipulation, its reason is also recorded. An event, therefore, has a name and several parameters:

- the range of the event,
- the initiator of the event and
- the reason for the event and additional parameters enabling the evaluation of the degree of change.

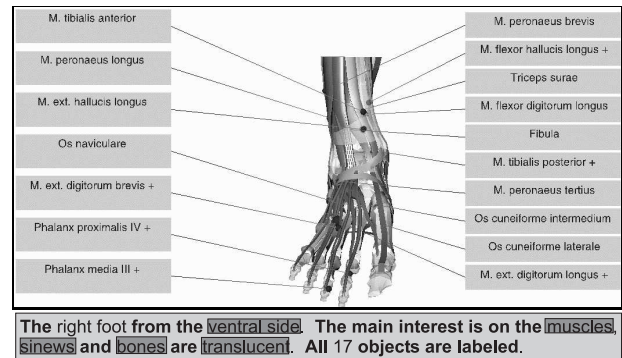
## 7 DYNAMIC CAPTIONS FOR ANATOMIC ILLUSTRATIONS

The application of dynamic figure captions in anatomy is directly related to the analysis of anatomical atlases as discussed in Section 2.1. We have incorporated figure captions in the ZOOM ILLUSTRATOR, an interactive illustration system [12].

### 7.1 Parameters of Visualizations

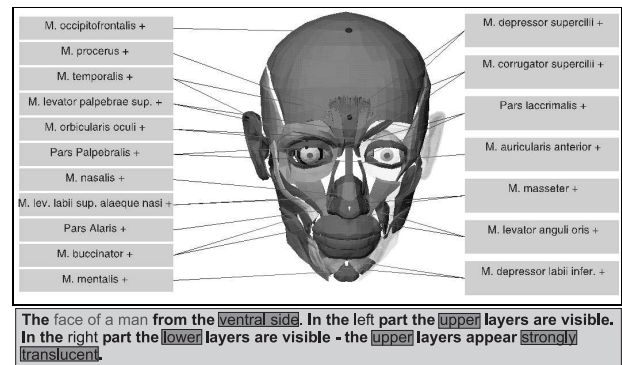
**(1) Task Analysis.** When anatomical illustrations are studied, it is important to be able to integrate the depicted con-

tent into a larger context. The user should be able to interpret what is shown from which direction. Furthermore, the location and topological relations of objects are of interest. In anatomy, visualizations are labelled. It is not clear, however, whether all objects of a category are labelled or whether some of them are not, e.g. due to space restrictions (recall Section 2.1).



**Figure 6: The ZOOM ILLUSTRATOR has produced an initial view which is described by an appropriate figure caption.**

From this analysis we derive the important parameters of a visualization. These include parameters of the textual and the graphics display. The state of the textual information is represented by the set of objects which are labelled or textually explained. The state of the graphical display includes the number of instances (usually one or two placed next to each other) and the viewing directions for each instance. Other parameters describe the state of individual objects. Objects are often rendered in a semi-translucent manner, e.g. to expose an object otherwise occluded. Transparency is the most important presentation variable because it does not only affect the object modified but also objects behind it.

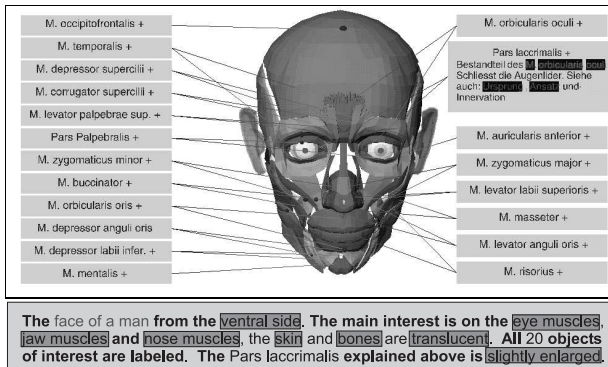


**Figure 7: The differences between the two halves are described in the figure caption.**

**(2) Analysis of the Visualization Algorithms.** For interactive 3D illustrations the most important analysis concerns the visibility of objects after changing the viewing direction.



The 3D model has to be analyzed concerning what objects are hidden by what objects and to what extent as well as to which objects are now visible. Because this analysis is computationally expensive, a pre-processing is employed. Another aspect of this step is the analysis of color contrasts (users may directly modify colors which has side effects for contrasts).

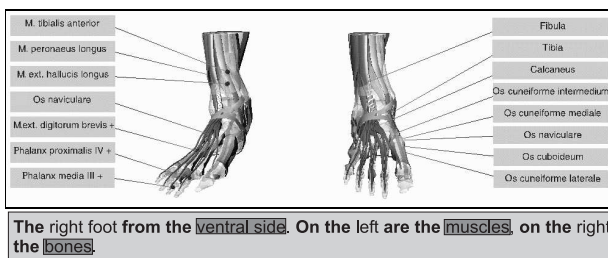


**Figure 8:** The figure caption includes comments as to how the sizes have changed to explain an object.

**(3) Event analysis.** All events are represented as messages in the Context Expert (recall Figure 5). An example for an event is

*rotated (model [left], user, [40,30,40]),*

which records that the user has rotated the left instance of the model with the last parameter specifying the amount of the rotation. Thus all important information about the event, its range and initiation is preserved. The effect of the event, such as that objects become visible and other disappear are not recorded. To describe the visibility of objects, a visibility analysis is employed (rays from the camera into the model are traced and as a result the sequence of objects hit is returned).



**Figure 9:** The caption for two instances of a 3D model consists of the similarities of both instances followed by the peculiarities of the left and right image.

**4 Content Selection.** In the content selection process, events are filtered as to whether they affect the global state or only single objects. Furthermore, it is analyzed whether visibility or “only” presentation variables change. The filtering pro-

cess considers both the a priori importance of events and the user specification.

## 5 Linguistic realization.

**(a) Text Structure Analysis.** The text structure analysis resulted in 7 categories for which templates are required. The caption should start with a description of the model view which includes the name of the model, the viewing direction and the aspects selected. Different templates are required, depending on, for example, whether one model view is presented or two are presented simultaneously. The second category of templates describes the filtering process of labels (whether all relevant labels could be displayed or not). Other categories include system decisions of how to emphasize objects, the description of rendering styles and attributes. As a result, these categories are arranged in a sequence which represents the order in which they are realized.

**(b) Sentence Structure Analysis.** The definition of templates yielded with 27 templates which represent a small yet sufficient set of phrases for the composition of figure captions. The first category includes all templates which describe the overall view. Let us give an example of the templates for the first category, where [Model] represents the name of a model to illustrate, [Direction] a viewing specification and [AspectList] a set of important categories:

1. The [Model] from [Direction].
2. The [Model] from [Direction<sub>1</sub>] and [Direction<sub>2</sub>].
3. The [Model] from [Direction<sub>1</sub>] and [Direction<sub>2</sub>]—on the left the [AspectList<sub>1</sub>], on the right [AspectList<sub>2</sub>].
4. The [Model] from [Direction]—on the left the [AspectList<sub>1</sub>], on the right [AspectList<sub>2</sub>].

With this collection, we have four templates of which one is selected by the system as the first sentence of the figure caption. If there is only one instance of the 3D model depicted, the first template is chosen. If there is a left and a right instance of the 3D model, one of the templates 2, 3 or 4 is chosen.

**(c) Lexical Realization of Template Variables.** We need phrases to name colors, transparency values and viewing directions. The naming of viewing directions considers conventions of the medical domain. In medical images, the frontal view, for example, is referred to as *ventral* which is more exact because ventral means “from the stomach”. Thus a reference point of the human body is used to name a viewing direction. In analogy, other viewing directions are named accordingly with reference to the human body.

## 7.2 Examples

We start with an initial view of the system immediately after the user has specified which model should be depicted and which aspects are of interest. The system produces an initial view of this model. From this specification, the system decides from which direction the model should be presented

and which objects to label. The figure caption (Figure 6) describes this and informs the user that all objects relevant to the specification of interesting objects are labelled<sup>4</sup>.

Let us look at another example: The model of a face has been selected for illustration. The user has manipulated the presentation so that the left and the right parts of the model look different. The figure caption indicates that the objects near the surface have been rendered semi-translucent in the right part of the model (see Figure 7).

The next image shows the model of a face after the user has requested an explanation for a small muscle. The system has automatically enlarged this muscle to emphasize it. This side effect is reflected in the figure caption (see Figure 8).

Finally, let us have a look at two instances of a 3D model and its description in Figure 9.

## 8 CAPTIONS IN CARTOGRAPHY

Dynamic figure captions have proven useful to describe the generalization operations applied to maps. This section discusses their incorporation in a system to create tactile town maps for blind people by sighted users—the MAP WIZARD [6]. Since tactile symbols are considerably larger than those in maps for sighted persons, the problems related to the map creation process are basically the same but occur to a larger extent. This system supports automatic and interactive cartographic generalization where the operations are applied iteratively until all conflicts could be removed.

**(1) Task Analysis.** Although there is a variety of cartographic visualizations some standard products exist, like topographic maps, town maps or nautical charts. The latter serve for vessel navigation, i.e. they allow the determination of the course and inform about shallow places. Town maps (to which we restrict ourselves in the following), by contrast, present the street network and important public buildings. They are consulted by users to search a path from point A to point B in a town.

Hence, it is essential for a reader to know if the entire street network is presented and if all positions are depicted correctly. Modifications that concern the features along the route are especially important with this respect. Should any such manipulation have occurred in a map, it may be useful for the reader to find them documented in a figure caption for this map.

**(2) Analysis of Visualization Algorithms.** As described in Section 2.2, restrictions to the size of the output area may entail reductions in the map fidelity. This comprises the completeness of the presentation as well as the positional accuracy. This process of cartographic generalization consists of several operations: simplification or smoothing, for example, modifies the outline of buildings or the curvature of streets.

<sup>4</sup>The captions are slightly enlarged. The bold parts result from the templates, while the words in plain text result from the substitution of variables. The sensitive parts are surrounded by a rectangle (recall Figure 1).

Displacement leads to an enlargement of the area between symbols and, thus, to a repositioning of these symbols and often of the symbols in the local and farther environment, too. In the course of selection, streets or buildings, for example, of low priority are removed.

**(3) Event Analysis.** Whenever such an operation is performed on the data within the context of map generalization, an event is generated and sent to the Context Expert (cf. the system architecture in Figure 5), e.g.:

- *smoothing (street name, system, tolerance value  $\epsilon$ , number of removed nodes)*

refers to an event where a number of nodes have been removed in a street using a tolerance value of  $\epsilon$  to simplify the street's curvature.

- *displacement (cause, system, list of streets used as origin, amount of displacement)*

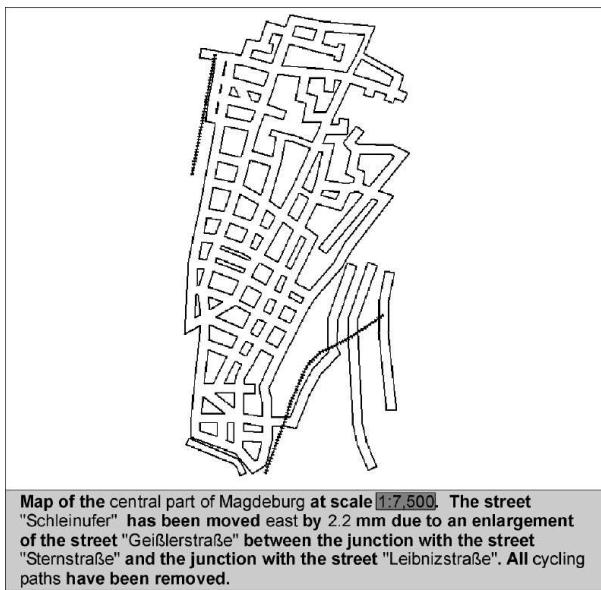
refers to an event where a displacement has been initiated originating from a street or several streets with a certain amount.

The concept of event lists is particularly suitable for figure captions for maps since the generalization process is implemented using the object-oriented design. The symbols that detect a conflict with a surrounding symbol analyze the type of conflict and may decide autonomously which generalization operation should be initiated. In addition, other conflicts are detected on a global scale and may be solved globally or again locally having a symbol performing the appropriate operation.

However, this leads to a distributed processing where the information about what is performed and why is spread out over the structure. Generating events for each operation and collecting them centrally is hence especially suitable. The Context Expert manages these events. It sorts them by priority and analyzes them to detect logical relations.

A particular analysis is necessary for objects that are traced because the user expressed a special interest in them. For this purpose, the symbols of those objects are set into a state where they detect any modification they are undergoing and generate a message accordingly. This message is then sent to the Context Expert for analysis. Thus, a street, for example, may register a repositioning and signal this. The Context Expert, however, receives in addition to this message information from another street stating that in order to solve a local graphical conflict with a river nearby the street has initiated a displacement. The Context Expert is now able to relate both events to one another and to generate a comment informing the map reader that the street he showed particular interest in has been moved due to a displacement that has been propagated throughout the map.

**(4) Content Selection.** The Context Expert sorts all events by priority and according to the user's specification (recall



**Figure 10: Tactile map with figure caption produced by the MAP WIZARD.**

Section 4.2) mentions only the most important ones in the figure caption.

**(5) Linguistic Realization.** As already mentioned in Section 2.2, the majority of map legends are schematic explanations of the coding. There are only a few maps that have long textual legends which may serve as a sample. We have, thus, decided to use phrases like:

- The street [street name<sub>1</sub>] has been moved [compass direction] by [amount] mm due to an enlargement of the street [street name<sub>2</sub>].
- The overlap of [amount] mm between [street name<sub>1</sub>] and [street name<sub>2</sub>] has been removed.

Note that this is only a subset of the templates for the description of the selection and displacement of symbols. Other templates consider rivers and railway lines in addition to streets and deal with the problem of overlaps among more than two linear features.

In general, it is not always straightforward to describe a position in the map, which is internally represented as a node (latitude and longitude) or a street segment (edge between two nodes). So far, we have chosen to mention the linear feature to which it belongs, e.g. the junction if applicable or the part of a street between two junctions.

For the verbalization of distances, rounded values in metric units are appropriate. Directional information is given using compass directions (north, north west, ...). However, sometimes describing a displacement vector as "[amount] mm in [compass] direction" is less appropriate than "[amount] mm perpendicular to [street name]" or a similar formulation.

## 9 CONCLUDING REMARKS

Descriptive figure captions contribute to a correct interpretation of complex images which arise in interactive visualization systems. This has been discussed for applications as different as medical illustrations and maps. Descriptive figure captions make users aware of important manipulations and, thus, make complex image generation processes transparent. Figure captions explain system-initiated manipulations and their reasons as well as complex user manipulations. Dynamic figure captions have some peculiarities compared to their counterparts in traditional media. Dynamic figure captions

- should be consistent with the image they describe. Therefore, the system must analyze whether captions become invalid or incomplete,
- can be tailored to the user's needs with appropriate configuration facilities and
- can be used for the parameterization of the images they describe (interactive captions).

Figure captions enrich an illustration with information which cannot be conveyed by schematic legends and labels alone.

The incorporation of figure captions requires an internal representation of the state of the interface. This representation, however, is not an extra effort exclusively for the generation of figure captions but can be used in a number of ways, e.g. to undo operations, to allow to reset to an earlier stage in an interactive session or to adapt the system's behavior to the discourse of human computer interaction.

As a result of the development of interactive figure captions, not only new interaction facilities emerged but interactive access to graphics has also been extended to attributes inaccessible so far. Interaction facilities with captions have been restricted to the replacement of single values until now—additional interaction facilities, like the insertion and removal of items in figure captions are worth to be studied.

Although figure captions are just emerging, they serve to accomplish some widely-recognized usability goals (see for example NIELSEN [10]): *Make the system's state visible* and *Speak the user's language*. Figure captions present important information about the state of a system. While this effect could partly be achieved with a formal output in the form of spreadsheets (e.g. viewing direction: [Direction], displacements: [List of Displacements]) figure captions are stronger related to the terminology users are familiar with. The terminology employed is carefully adapted to the domain e.g. to students of medicine or cartographers. Figure captions in visual interfaces are new and therefore require careful validation.

The linguistic approaches employed so far are straightforward. An important area for future work is to refine these approaches, including natural generation methods. It remains an open question under what circumstances the additional

possibilities of natural language generation justify the extra effort compared to the template-based approach.

A possible application of figure captions left for future study is their use as bookmarks. Bookmarks are used to set landmarks in large information spaces. It would be very useful to automatically generate meaningful names for the organization of bookmarks in visualization systems.

Short figure captions can serve for this task. They are consistent with the state of the visual interface and provide the necessary information to recall the generated images. Finally, if an interactive system makes it possible to produce screen shots, a corresponding figure caption can be saved as well to enhance retrieval tasks. For this purpose, detailed figure captions are required because screen shots are often looked at later in the absence of the interactive situation in which they have been generated.

## ACKNOWLEDGMENTS

The authors want to thank Torsten Sommerfeld who implemented figure captions for anatomic illustrations. Our thanks go also to Sylvia Zabel for carefully proofreading the paper.

## REFERENCES

- [1] BERK, T., BROWNSTONE, L., AND KAUFMANN, A. A new color-naming system for graphics languages. *IEEE Graphics and Applications* (May 1982), 37–44.
- [2] BERNARD, R. M. Using extended captions to improve learning from instructional illustrations. *British Journal of Educational Technology* 21, 3 (1990), 215–225.
- [3] BORDEGONI, M., FACONTI, G., MAYBURY, M. T., RIST, T., RUGGIERI, S., TRAHANIAS, P., AND WILSON, M. A standard reference model for intelligent multimedia presentation systems. *Journal for the Development and Application of Standards for Computers, Data Communication and Interfaces* (1997).
- [4] FEINER, S. K., AND MCKEOWN, K. R. *Intelligent Multimedia Interfaces*. AAAI Press, 1993, ch. Automating the Generation of Coordinated Multimedia Explanations, pp. 117–138.
- [5] GOMBRICH, E. H. *The Sense of Order - A Study in the Psychology of decorative art*, 2nd ed. Phaidon Press, London, 1984.
- [6] MICHEL, R. Computer-supported symbol displacement. In *Proceedings of 18th International Cartographic Conference* (1997), L. Ottoson, Ed., pp. 1795–1803.
- [7] MITTAH, V., ROTH, S., MOORE, J. D., MATTIS, J., AND CARENINI, G. Generating explanatory captions for information graphics. In *Proceedings of IJCAI* (Montreal, December 1995), pp. 1276–1283.
- [8] MONMONIER, M. *How to lie with maps*. University of Chicago Press, 1996.
- [9] MÜLLER, J. C., LAGRANGE, J. P., AND WEIBEL, R. *GIS and Generalization: Methodology and Practice*. Taylor and Francis, London, 1995.
- [10] NIELSEN, J. Enhancing the explanatory power of heuristics. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'94)* (April 1994), Human Factors in Computing Systems, ACM, pp. 152–158.
- [11] NUGENT, G. C. Deaf students learning from captioned instructions: The relationship between the visual and the caption display. *Journal of Special Education* 17, 2 (1983), 227–234.
- [12] PREIM, B., RAAB, A., AND STROTHOTTE, T. Coherent zoom of illustrations with 3D-graphics and text. In *Proceedings of Graphics Interface '97* (Kelowna, May 1997), pp. 105–113.
- [13] REITER, E. Natural language generation versus templates. In *Proceedings of the 5th European Workshop on Natural Language Generation* (Leiden, Niederlande, 1995).
- [14] REITER, E., AND MELLISH, C. Optimizing the costs and benefits of natural language generation. In *Proceedings of IJCAI (Chamberg, Frankreich, August 28–September 3, 1993)*, Morgan Kaufmann, pp. 1164–1169.
- [15] SCHLICHTMANN, H. Functions of the map legend. In *Proc. of 18th International Cartographic Conference* (Stockholm, July 1997), p. 430.
- [16] SOBOTTA, J. *Atlas der Anatomie des Menschen*, 19th ed., vol. 2. Urban & Schwarzenberg, Munich, Vienna, Baltimore, 1988.
- [17] STAAB, S., AND HAHN, U. "tall", "good", "high"—compared to what? In *Proceedings of IJCAI (Nagoya, Japan, August)*.
- [18] WAHLSTER, W., ANDRÉ, E., W. FINKLER, H.-J. P., AND RIST, T. Plan-based integration of natural language and graphics generation. *AI-Journal* (1993), 387–427.
- [19] WEIDENMANN, B. Informative Bilder (Was sie können, wie man sie didaktisch nutzt und wie man sie nicht verwenden sollte). *Pädagogik* (September 1989), 30–34.