View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Lund University Publications

Holsanova, J. (2020). Uncovering Scientific and Multimodal Literacy through Audio Description, in V. Eisenlauer & Karatza S. (eds.) 'Multimodal Literacies: Media affordances, semiotic resources and discourse communities'. Journal of Visual Literacy, Special Issue 39(3).

Uncovering scientific and multimodal literacy through audio description

Jana Holsanova, Cognitive Science Department, Lund university, Sweden

Abstract

Today's scientific texts are complex and multimodal. Due to new technology, the number of images is increasing, as is their diversity and complexity. Interaction with complex texts and visualisations becomes a challenge. How can we help readers and learners achieve multimodal literacy? We use data from the audio description of a popular scientific journal and think-aloud protocols to uncover knowledge and competences necessary for reading and understanding multimodal scientific texts. Four issues of the printed journal were analysed. The aural version of the journal was compared with the printed version to show how the semiotic interplay has been presented for the users. Additional meaning-making activities have been identified from the think-aloud protocol. As a result, we could reveal how the audio describer combined the contents of the available resources, made judgements about relevant information, determined ways of verbalising visual information, used conceptual knowledge, filled in the gaps missing in the interplay of the resources, and reordered information for optimal flow and understanding. We argue that the meaning-making activities identified through audio description and think-aloud protocols can be incorporated into instruction in educational contexts and can thereby improve readers' competencies for reading and understanding multimodal scientific texts.

Keywords

meaning-making processes; scientific and multimodal literacy; production and reception of multimodality; audio description; think-aloud protocol

1. Introduction

Today's printed and digital texts are not only linguistic but are also complex and multimodal. This means that, in addition to the linguistic text, they also contain images and graphic devices of various kinds: photos, drawings, diagrams, graphs, tables, maps, timelines, and flowcharts, as well as boxes, frames, and various layout elements.

However, knowledge about how we actually interact with these messages is limited. Although the composition of multimodal texts and their potential for meaning-making has been discussed in a social semiotic context (Kress & van Leeuwen, 1996/2006), and in a rhetorical context (Bateman 2008), there is still little known about how users actually are affected by and interact with text design, how they read complex texts, what attracts their attention and what does not, and how they integrate information from the language, images, and graphics (Holsanova 2014a).

Since the diversity and complexity of images in scientific texts is increasing, interaction

with complex texts becomes a cognitive challenge that requires several competences (Jewitt & Kress 2003, Kress & van Leeuwen 1996/2006, Unsworth & Cléirigh 2014, Holsanova 2019). It becomes more difficult to read and understand complex visualisations and to identify their relevance in specific contexts. In order to be able to use complex text consisting of language, images, and graphics in the classroom and in everyday life, it is necessary for users to possess new knowledge, skills, and competences (Behnke 2017). In particular, the use of images in science calls for specialised knowledge and competences. Learners need to be gradually and critically trained to use a visual scientific coding orientation (Unsworth 1997:35, Kress & van Leeuwen 1996/2006).

Researchers also formulate the need to redefine literacy in the current curriculum within the framework of multimodal literacy, something which is necessary for reading, viewing, responding to, and producing multimodal and digital texts (Walsh 2010:211). The first steps on this road have been taken. Unsworth & Macken-Horarik (2014) report that a new Australian curriculum for English emphasises the multimodal nature of literacy and requires students in primary and secondary schools to develop explicit knowledge about visual and verbal grammar as a resource for text interpretation and text creation. Lim (2018) applies an instructional approach to multimodal literacy, informed by Systemic Functional Theory, to teach multimodal texts and describes a trial of this in a secondary school in Singapore. In a Swedish context, Danielsson & Selander (2016) have developed a model for working with multimodal texts in education.

However, the concept of multimodal literacy still remains vague. Exactly which types of knowledge, skills, and competences does multimodal literacy include? How can we discover them? How can we help readers and learners achieve scientific and multimodal literacy? How can we enhance their ability to identify which meanings are created by the individual modes and which meanings are created by the interactions of these modes (Kress 2003)? In our study, we use a novel method – a combination of an audio description of a Swedish popular scientific journal and concurrent think-aloud protocols created during the audio description task – to uncover knowledge and competences necessary for reading, understanding, and creating multimodal scientific texts. The focus of the present article is on the users' activities during actual interaction with complex multimodal texts. The objective of the study is to reveal the dynamic interpretative processes of meaning-making.

After an introduction to the theoretical framework, methods, and material, the results from four steps of analysis are summarised. The resources of language, images, and graphics used in the journal are characterised to show how they are deployed in scientific explanation. The methods of concurrent think-aloud protocols and audio description are illustrated by the audio description of a diagram. The traces of meaning-making activities from the interaction with the multimodal journal – as revealed by the methods – are summarised. The results and methods are discussed. It is argued that the combination of audio description and think-aloud protocol can be successfully used as a novel method to reveal multimodal competences. Finally, conclusions are drawn concerning the possibility of incorporating the results of the study into tailored instruction for educational purposes, with the aim of improving novices' understanding of multimodal scientific texts.

2. Theoretical framework

In our empirical study, we apply a framework of social semiotic theories (Kress & van Leeuwen 1996/2006, Jewitt & Kress 2003, O'Halloran et al., 2012) – in particular concerning

both the visual construction of specialised knowledge (Unsworth 1997) and text-image relations (Martinec & Salway 2005, Unsworth & Cléirigh 2014) – in combination with cognitive theories on the reception of multimodality (Holsanova et al. 2006, Holsanova 2008, Holsanova & Nord 2010, Boeriis & Holsanova 2012, Holsanova 2014a,b, Holsanova 2016), and pragmatic theories of multimodal meaning-making (Bucher 2017). The focus is on the actual use of multimodal texts.

Based on the theory of metafunctions in verbal and visual communication (Kress and van Leeuwen (2006/1996), Unsworth (1997) examines meanings in images. He distinguishes three kinds of meaning: (a) ideational or representational, (b) interactive or interpersonal, and (c) textual or compositional. The ideational or representational meanings in images can be either *narrative* (involving action; mental or verbal processes) or *conceptual* (concerned with more abstract classification or decomposition of objects and processes). Interactive or interpersonal meanings in images are limited to the issue of modality (to what extent images are naturalistic representations of reality). Unsworth (1997) refers to the various representations of reality as coding orientations: *naturalistic* (e.g., colour photographs, movies, and video, which depict reality as it is seen 'naturally'), *realistic* (e.g., drawings and paintings that approximate the natural features of phenomena), or *scientific* (e.g., schematic and conventional line drawings in science). Finally, textual or compositional meanings of images concern the ways in which layout influences what kind of emphasis is given to images (e.g., relative salience and prominence of images on the page or screen).

Complex multimodal texts can be studied from the perspectives of both production and reception (Bucher 2007, Holsanova 2012b, 2014a,b). The production perspective focuses on the interplay between various resources, their contribution to the content of the message, and their orchestration in order to achieve a certain effect; this process is often referred to as *intersemiosis* (O'Halloran et al. 2012). The reception perspective is closely connected to recipients' ability to select, attend to, and process information, as well as to their ability to integrate information from various resources and to fill in the 'gaps'.

In previous research, eye-tracking methodology has been used in research on the *reception of multimodality*. Using eye movement measurements we are able to follow the reading and scanning processes in detail. We can trace exactly not only what is looked at, but also where, when, and how often (Holsanova 2014a). By complementing eye tracking with other measurements, and using a triangulation of methods (Holsanova 2012), researchers have been able to trace users' interactions in detail (Bucher 2017, Bucher & Niemann 2012, Holsanova 2001, 2008, Holsanova et al. 2006, 2009, 2012, Kaltenbacher & Kaltenbacher 2015, van Gogh & Scheiter 2009). However, eye tracking data does not tell us about recipients' understanding of the messages. We cannot conclude from the eye movement protocols alone what aspects and properties of an image element have been focused on, or at what level of abstraction. Visual fixation does not reveal which concept was associated with the element, or what the viewer had in mind. In order to trace the underlying thought processes, eye tracking has to be complemented with other measurements, and a triangulation of methods must be used (Holsanova 2012). Since the focus of our study is on meaning-making processes, the use of verbalisation in the form of audio description and think-aloud protocols is a good way to track these processes.

Audio description (AD) is primarily used to offer richer understanding and enjoyment for people with visual impairment and blindness. It is used to increase the accessibility of films, theatre performances, museum and art exhibitions – and complex printed materials. In

the specific context of the popular scientific journal, the task of the audio describer is to make the contents accessible by producing an aural version of the journal by transforming the contents of text, images, and graphics into speech. In order to do so, the audio describer must assess what to describe, determine how to describe it, and decide when to describe it (Holsanova 2016b). The audio describer is a mediator between the original producer and the end users. The goal of AD is 'to do justice to the communicative aim of the author(s) and the communicative needs of the recipients of the text' (Reviers 2017:34).

The *think-aloud method* is a qualitative research method where participants speak aloud while performing a task (Ericsson & Simon 1993, van Someren, Barnard & Sandberg 1994). In general, this method is used to provide insights about participants' thinking and decision-making processes, especially regarding language-based activities. By using AD and concurrent think-aloud protocols during the AD task, we are able to uncover multimodal reading in process.

Research on *reading in various media* shows that readers sometimes have difficulties navigating texts, finding relevant information based on text and image, reading and interpreting images and graphic representations, and making sense of the complete message of the text at hand (Holsanova 2010). Apart from that, the interpretation of tables, graphs, charts, and maps is based on rules, conventions, and prior knowledge. The reader must thus be able to combine perceptual and conceptual knowledge (what you see and what you know) to understand properly (Pettersson 2008, Unsworth & Cléirigh 2014). Some readers need instruction or guidance through a complex text with the help of various cues (Holsanova & Nord 2011, Holsanova 2014, Holsanova et al. 2009, Scheiter, Holsanova & Wiebe 2008). This support can be provided by a design that guides learners' attention towards its relevant aspects, by teachers' general instructions, or – as we suggest here – by tailored instruction informed by the meaning-making activities identified via AD and think-aloud protocols.

There is a need for empirical research on readers' interactions with multimodal messages in order to improve learners' competences in dealing with multimodal texts and complex visualisations. The specific research questions are: How do readers actually interact with multimodal documents? How do they understand the multimodal interplay? How do they navigate to find relevant information? How do they extract and integrate information from various resources?

3. Methods and material

In this study, we used data from AD of a Swedish popular scientific journal, *Forskning och Framsteg* [Research and Progress], in combination with think-aloud protocols recorded during the AD task to trace thought processes during meaning-making. The ultimate goal was to gain access to knowledge and competences necessary for reading, understanding, and creating multimodal scientific texts.

The reason why is it advantageous to use audio description to uncover meaning-making activities lies in the double role of the audio describer. AD is characterised as 'a complex cognitive-linguistic and intermodal mediation activity where creative meaning-making processes during production and reception coincide' (Braun 2007). On the one hand, the audio describer is a recipient of the text who must to read the text thoroughly; make sense of the language, images, and graphics and interpret their interplay; find semantic relationships among various resources; and identify the main message. On the other hand, the audio

describer is a producer who selects relevant pictorial information in the context of the message, verbalises it, integrates it with the content of the written text, and transforms all this into speech (Holsanova, forthc.). Thus, the audio described version of the journal demonstrates how the semiotic interplay of the message has been understood and presented for the end users while it also shows traces of integration and meaning-making. When we then compare the spoken version with the printed one, the results of the interpretative meaning-making activities are revealed. By using AD and concurrent think-aloud protocols during the AD task, we can trace thought processes and discover additional aspects of the dynamic meaning-making process.

The material consisted of three sources: (a) printed data from four issues of the Swedish multimodal popular science journal *Forskning och Framsteg* (2016–2018; 280 pages)[1], (b) spoken data from the audio version of these four issues (twelve hours of recordings of which ten hours were of popular scientific articles), and (c) spoken data from think-aloud protocols recorded during the AD task (two hours). The journal has close contacts with the Swedish research community and publishes articles on a wide range of topics: research and advances in astronomy and physics, the environment and ecology, energy and technology, chemistry, medicine and psychology, animals and nature, history and archaeology, philosophy and ethics, language, cognition, culture, social science, and economics[2].

The analysis of empirical data was conducted in four steps. First, the four issues of the printed journal were analysed concerning which semiotic resources are used in the journal and how text, images, and graphics are deployed in scientific explanation. Since the focus was on readers' activities and their use of the material, the approach was to characterise and summarise the types of text, images, and graphics found in the printed material. Second, the aural version of the journal was compared to the printed version to show how language, images, and graphics have been integrated and how the semiotic interplay has been presented for the end users. Third, the interpretative processes of meaning-making were revealed by think-aloud protocols during the AD task. Fourth, on the basis of the comparison between the printed and the aural versions, together with the think-aloud protocols, we gained insights into activities of meaning-making.

4. Analysis and results

In the following we present results of the four steps of analysis. Section 4.1 summarises the results of the multimodal analysis of the resources used in the journal, whereas Sections 4.2 and 4.3 illustrate traces of meaning-making processes extracted from the think-aloud protocols and from the AD, respectively. Finally, Section 4.4 summarises the main results of the study. It presents the meaning-making activities identified on the basis of all of the collected data. These activities show us the types of knowledge necessary for reading multimodal texts

---

[1] The journal has been published since 1966. Thirty thousand copies of each issue are published, of which ninety-eight percent go to subscribers. The printed edition of the journal is read by about 500,000 Swedes per year. On the Internet, *FoF* has more than 200,000 unique visitors per month. Half of the readers are men, half women. Retrieved 9 February 2020 from: https://fof.se/om-forskning-framsteg.

[2] The outline of the journal is as follows: Editorial, Readers' letters to the editor, Introduction, Advances in a number of specialised areas of research under the headings Humans, Environment, Technology, the Universe, Past & Present, Commentary, Questions & Answers, Everyday mystery (How things work), Books, Use your brain, Announcements, and Advertisements. Each printed issue is about seventy pages long, with the audio version being roughly three hours long.

that can also be applied for educational purposes. The following questions will be in focus: What activities is the audio describer involved in during reception of a multimodal popular scientific journal and during AD production? What meaning-making activities can be revealed through AD and think-aloud protocols? Which aspects of scientific and multimodal literacy does the audio describer exhibit in the process of AD production and in the final audio version of the journal?

4.1 Language, images, and graphics deployed in scientific explanation

This section presents the results of the multimodal analysis of the journal *Forskning och Framsteg* (*FoF*)[3]. The journal contains article texts, different types of images of varying complexity such as photos, graphs, tables, maps, timelines, information graphics and diagrams, as well as layout elements such as highlighted fact boxes and quotations. All of these elements contribute to the content of popular science explanations and invite the reader to construct meaning. In the following, the resources in *FoF* will be briefly characterised with a focus on the visual construction of specialised knowledge (Unsworth 1997, Unsworth & Cléirigh 2014).

In *FoF* the most frequent image type is a *photograph*. Photos can be characterised as narrative (involving action), with a naturalistic coding orientation (depicting reality as it is seen), containing portraits of people, animals, and environments. They also serve to identify the author or the researcher behind the article. Sometimes, *FoF* uses what are called genre photos with a generalised content (e.g., people with computers sitting around a table), vaguely associated with a topic or adding ambiance but not contributing to the main message of the article. Photos are rarely described by an audio describer. For the few that are described, this is done selectively, with a focus on those parts that are semantically relevant for the article they accompany.

FoF uses a large number of layout elements that can be referred to as *layout modules:* e.g., fact boxes, quotations, captions, annotations. These layout elements have compositional meaning by showing the salience of certain parts of the article text and by putting emphasis on certain aspects of the message. Their shape and layout vary. They are often visually delimited as a 'gestalt'; or highlighted by a colour background; or contain text accompanied by numbers, icons, or colour-coded keywords; and their connection to the article text is marked by graphical means (arrows and lines). They draw readers' attention to prominent parts of the article (Holsanova et al. 2006, 2009, Holsanova 2014a,b).

Further, *FoF* uses *diagrams, graphs, maps, tables,* and *timelines* to visualise the article content. These types of images can be characterised as conceptual (classify or decompose objects and processes), scientific (contain schematic and conventional representations), and multimodal (include captions, headlines, symbols, colours, annotations). They are only audio described when relevant and complementary to the main text (e.g., when the article text mentions a general tendency and the diagram contributes concrete details).

Information graphics (IG) is mainly used in *FoF* to explain complex processes and to show how things work in everyday life. IG can be characterised as both narrative and conceptual since it depicts action and the decomposition of objects. It has both a realistic and a scientific coding orientation since it contains images that approximate natural phenomena, as

[3] The author would like to thank Erika Sombeck who assisted in the analysis presented in 4.1.

well as schematic conventional representations. This type of visualisation is complex and multimodal. It includes other types of images and layout elements, such as maps, diagrams, timelines, text modules, headlines, numbers, drawings, arrows (indicating direction, connection, and movement), zoom-in boxes, speech bubbles, etc. With respect to navigation it sometimes offers multiple reading paths. A challenge for the audio describer is to find a logical, optimal way of presenting this and creating a narrative 'flow'.

4.2 Traces of meaning-making processes in think-aloud protocols. Case study diagram

By using concurrent think-aloud protocols during the AD task, we were able to reveal the steps in the dynamic meaning-making process during the interaction with the multimodal journal. The audio describer was asked to 'Comment on what you are thinking and doing in the process of audio description.' As a result, the audio describer reflected on his activities, mentioned problems, drew inferences, tested several versions, and verbalised a wide range of thoughts. The informant who produced the audio version of the journal, PL, is an authorised and experienced audio describer who has been in charge of AD for *FoF* since 2011. He has previous experience as an actor, vocal artist, and producer; has good verbalisation skills; and has no problems putting his thoughts into words.

To illustrate the methodology, we present one example in detail: the audio description of a *diagram* included in an article about the increasing use of the Internet and mobile phones (*FoF* 3/2018, p. 28, cf. Figure 1). The challenge for the audio describer has been to find the most exact, comprehensive, and recipient-friendly way of describing it. Example 1 shows a section of a think-aloud protocol recorded by the audio describer during the description of this diagram. He is commenting online while reading the diagram. Later on, he trys out a first version of AD on the fly, improves it, takes notes, and produces a final version of AD in the studio (cf Section 4.3).

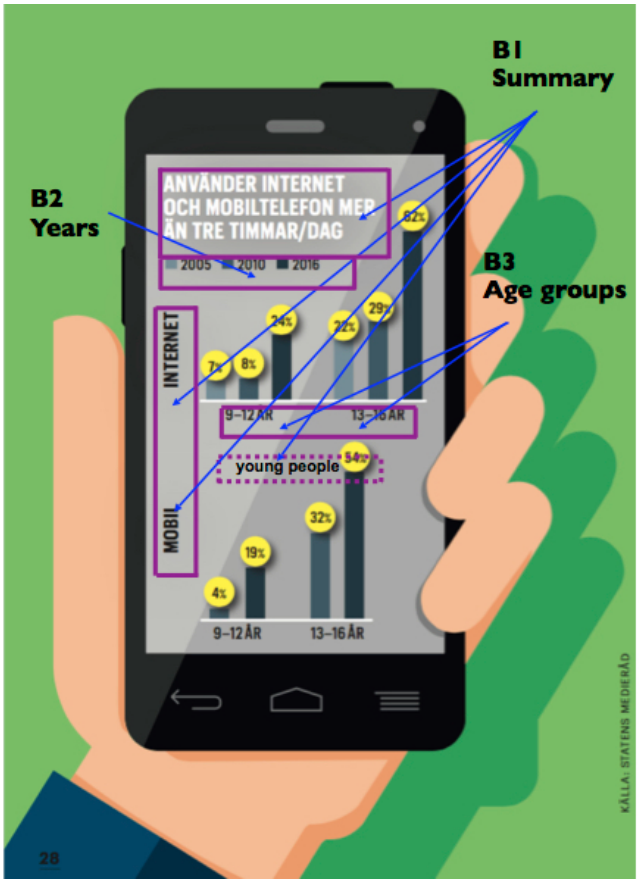
> Here is a bar chart, it is on page twenty-eight of *Forskning och Framsteg* 2018, issue number three, and it shows a mobile phone that a cartoon figure is holding in its hand. There are two charts and I see that the heading is 'Uses the Internet and mobile phone more than three hours a day'. … Does the heading belong to both charts? I think so. I also see that it concerns 2005, 2010, 2016 and that the first group refers to 9- to 12-year-olds with certain numbers and the next group to 13- to 16-year-olds. I think that I will need to do something repetitive here, alternatively that I say the numbers 2005, 2010, 2016 in succession and then I'll hope that the user will remember it. … So… there are two charts. I'll take them one by one. I really have to divide it. After the title and after the years and age groups. I will start with the Internet and then I think I need to reconnect to the heading 'Use of the Internet for more than three hours a day'. Then I'll go straight over to the years and percentages. I list them quickly, make a little rhythm of it all, and do the same for age group two. And through all that I have set up a system that makes it clear…. I see that the second chart with the mobile use has only two bars, not three. According to the colours, I can see that it only concerns the years 2010 and 2016.

Example 1: A section of the concurrent think-aloud protocol recorded during the AD task, transcribed and translated into English.

The journal article surrounding this diagram presents an overview and a general tendency stating that the use of the Internet and mobile phones is increasing. Since the diagram contains details about this increase that are not mentioned in the article text, the audio describer considers it to be relevant and decides to describe it. As we can observe in this extract, his thought processes concern identification of the type of images (figurative and diagrammatic) and the structure of the diagram and its integral parts, as well as the categories (age groups, years) and the relative quantities of different categories (percentages expressed by the bars). The audio describer is using proper terminology (bar chart); establishing semantic relations between the heading, the data visualisations, and the labels; reading the chart according to particular rules and conventions (interpreting bar colours); and reconnecting visual information to the overarching verbal heading. He is also expressing thoughts about how to group and divide the data presentation into coherent chunks as well as considering how to arrange it in order to create clear patterns so that the the most important facts will be easily understood and remember by the end users.

4.3 Traces of meaning-making processes in the audio description of a diagram

Next, we will have a look at the results of the audio description process, the final version of AD. The audio described version of the journal demonstrates how the semiotic interplay of the message has been understood and presented for the end users and in it can be seen traces of integration and meaning-making. Figures 1a,b illustrate the step-by-step process in the final AD version of the diagram (read from top to bottom). From left to right, these two figures show (a) areas of interest in the diagram the audio describer focused on during AD, (b) the final version of the AD formulated by the audio describer, and (c) a summary of the activities extracted during this process. The audio description also reflects prosodic features (emphasis, pauses, and chunking into units of speech). The first two columns illustrate the dynamic process of how the audio describer step-by-step focused visual and verbal attention on different aspects of the diagram, formulating one idea at a time (Holsanova 2001, 2008, 2011). For practical reasons, the AD is divided into two figures: an overview in Fig. 1a, followed by a detailed description in Fig. 1b. The areas of interest that we marked in the diagram have been established by the audio describer on the basis of existing units (e.g., heading) and groupings of existing units (ages, years), or created by the audio describer as a new unit (e.g., young people, years 2010 and 2016) – the last one marked by dotted lines.

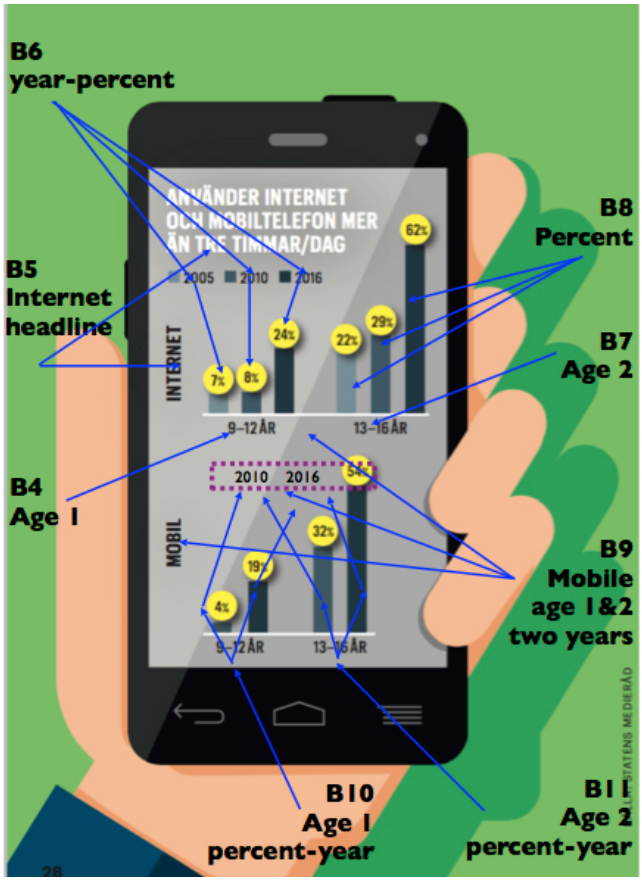


AUDIO DESCRIPTION	ACTIVITIES Overview
Here is a chart, showing young people's use of the Internet and mobile phones for more than three hours a day	B1 Summary • identifies type of image • summarises the age groups as 'young people' • names the categories: Internet and mobile phone • connects to the headline
in 2005, 2010, and 2016.	B2 • mentions years
There are two age groups: 9 to 12 year-olds and 13 to 16 year-olds.	B3 • mentions age groups
	DETAILS

Figure 1a: First part of the AD of a diagram (overview): (a) visualisation of the step-by-step focus on various areas of interest in the diagram; (b) final version of the audio description; and (c) extracted meaning-making activities.

Figure 1b: Second part of the AD of a diagram (details): (a) visualisation of the step-by-step focus on various areas of interest in the diagram; (b) final version of the audio description; and (c) extracted meaning-making activities.

We can observe that the audio describer reconnects several times to the heading, formulates sub-summaries, groups similar information into units in various ways, creates new units necessary for the description (e.g., based on inferences from colours in the diagram), and makes the description 'digestible' by segmenting and portioning information.



In the group of the 9 to 12 year-olds	B4 • mentions age group 1
who use the Internet for more than three hours a day	B5 • mentions category: Internet • reconnects to the headline
the percentages are 2005: 7%, 2010: 8%, and 2016: 24%.	B6 • clusters data: year -percent, year - percent, year - percent
For the 13 to 16 year-olds	B7 • mentions age group 2
the percentages are 22%, 29%, and 62%. respectively.	B8 • clusters data: percent, percent, percent
The use of mobile phones by the respective age groups is only shown for 2010 and 2016:	B9 • creates a sub-summary, mentions category mobile • refers to age groups • infers years from colors • establishes a new unit
9 to 12 year-olds, 4% in 2010, 19% in 2016;	B10 • mentions age group 1 • clusters data: percent - year, percent - year
13 to 16 year-olds, 32% in 2010 and 54% in 2016.	B11 A • mentions age group 2 • clusters data: percent - year, percent - year

4.4 Meaning-making activities and aspects of scientific and multimodal literacy

In this section, we summarise meaning-making activities identified on the basis of all the collected data. These activities and the associated questions below are the results of the application of this study to educating for multimodal literacy. They tell us the type of knowledge assumed by the scientific article in its different modalities.

Apart from the diagrams, we collected meaning-making activities from the AD of other types of images, layout elements, and visualisations found in *FoF*, such as tables, graphs, timelines, maps, and information graphics. The meaning-making processes are reflected in the aural version of the journal and also become traceable through the think-aloud protocols created during the AD. This method enables us to obtain direct access to cognitive and interpretative meaning-making processes. The various meaning-making activities of the audio describer were extracted on the basis of the comparison of the aural version of the popular scientific journal with the printed version and on the basis of think-aloud protocols during AD. The results illustrate how the audio describer – being both a recipient and a producer – extracted, combined, processed, and understood information from text, images, and graphics, and how he presented it for users.

The majority of the questions have been formulated by the audio describer himself. Some of the questions have been slightly reformulated by the present researcher, concerning, e.g., teminology (semantic links, segmentation). A small number of aspects have been partly expressed explicitly by the audio describer himself and partly inferred by the researcher based on a comparison of the printed original with the aural version (e.g., reorganising the con-

tent). The meaning-making activities have been organised by the researcher and are summarised in the following five groups:

(1) Assessment and decisions about relevant information. The first group of activities concerns the audio describer's judgements about the relevance of information expressed in the visualisations in the context of the whole article. Not all visualisations are necessary for understanding the overall content of the article (e.g., decorative and genre images). At the same time, there are time and space restrictions that constrain the production of AD. Therefore, an audio describer must assess what to describe and how to describe it, select relevant images and graphics (or their parts), and determine which of these are not relevant and can be eliminated. The following questions guide this process:
What parts does the article consist of? What is the overall idea of the article? What kinds of images are there? (bar chart, pie chart, timeline). What do they contain? What purpose do these have in the context of the message? Do these images contribute with relevant information? Or are they only decorative/genre images? Is the image content already expressed in the article text, in the caption, or elsewhere?

(2) Interpreting various types of images. The second group of activities concerns verbalisation of visual content by using prior conceptual knowledge. The audio describer needs to identify the type of image, use appropriate terminology, and decide how the image or its parts should be described verbally. In order to read and understand complex visualisations, the audio describer uses knowledge about the area of expertise, i.e., specialised knowledge, in order to interpret tables, graphs, charts, and maps, and conceptual knowledge for interpreting solid lines (connections, trends, directions), dotted lines (expected results), colours in maps or bar charts (various categories, years), etc. The following questions guide this process:
Which topic/area of expertise is this about? Which type of image is it? How should the image content be interpreted and verbalised? In how much detail should it be described?

(3) *Integrating various modes of representation during meaning-making.* The third group of activities concerns identification of semantic links between language, images, and graphics and integration of various modes of representation into a coherent whole. On the basis of prior knowledge and everyday experience, the audio describer combines the contents of the available resources, creating links between headline, image, caption, and annotations and filling in the gaps missing in the interplay of the resources. The following questions guide this process:
What is the relation between the content of the article text, the headline, images, graphics, annotations, etc.? Are the semantic links between related parts of the written text, the images, and the annotations marked (linguistically, graphically, or by spatial contiguity)? Or do these links need to be inferred and filled in?

(4) Reorganising the content. The fourth group of activities concerns ordering, segmenting, and semantic grouping of information. When verbalising information from complex visualisations such as information graphics, timelines, diagrams, and flow charts, the audio describer must determine where to start and which way to continue. Since these kinds of visualisations allow multiple ways of reading, it is necessary to choose entry points and reading paths and to find the logical sequential order in the complex message (information graphics in particular). Sometimes, it is necessary to go back and reconnect to the headline (graphs and diagrams). In other words, the audio describer reorders information for optimal flow and understanding. Apart from navigation in a complex image, the audio describer is also involved in segmenting information into meaningful chunks and grouping similar information into

larger units. He is, moreover, engaged in creating summaries and introductions to information modules (diagrams, timelines, information graphics). To support understanding, he delivers a global overview first, followed by a more detailed description of an image (cf. previous diagram example). The following questions guide this process:
In what order should information be presented? What is the logical sequence supporting the narrative? Which system should be established for segmenting information? Which types of information belong together and should be presented together? How should this unit be introduced and summarised? How can intonation, speech rate, emphasis, and pauses be used to group and highlight information?

(5) Facilitating understanding and cognitive processing. The fifth group of activities aims at providing optimal understanding and flow and at considering the mental capacity and working memory of the recipients. The audio describer tries to make the description short and comprehensive, conveys information in easily digestible portions, and repeats information for better understanding. When transforming written text, images, and graphics into speech, the audio describer is aware of the role of vocal delivery. He uses voice quality and prosody – intonation, speech rate, emphasis, and pauses of different length – to highlight important information and to group certain bits of information. The following questions (that the audio describer himself commented on) guide this process:
Is the image description understandable? Can I understand it myself if I close my eyes and listen? Will the user understand it and be able to remember the most important information? Can the user keep all this information in his head? Will the way I structure the content help the user to digest the information?

5. Discussion

Audio description has sometimes been defined as intersemiotic translation (following Jakobson's taxonomy of 1959) since it transforms images into words. But this definition does not give the whole picture. The audio describer was indeed partly concerned with the mediation of visually constructed specialised knowledge, but as the revealed interpreting processes suggest, his activities cannot be limited to replacing the visual part of the message with a verbal part. Rather, the audio describer was engaged in intermodal integration: he was making sense of verbal and visual information and interpreting images and graphics in connection with text. By taking a holistic grasp (Braun 2008), the audio describer supplemented what is lacking in the multimodal interplay to achieve a comparable understanding and experience for the audience (Holsanova 2020, Reviers 2017). The audio describer was thus involved in multimodal mediating activities. We would therefore argue that the activities revealed here constitute aspects of multimodal and scientific literacy necessary for reading, interpreting, understanding, and using complex multimodal texts (Jewitt & Kress 2003, Unsworth & Macken-Horarik 2014, Walsh 2010). Also, it shows that reception of multimodality cannot be restricted to decoding the semiotic resources. It is rather a complex process of deriving the content with the help of prior knowledge and on the basis of the context by making inferences (Bucher 2017, Holsanova 2014b, Wildfeuer 2012).

One of the limitations of this study is that the data collected stems from a single audio describer. There is, however, a practical reason for that. It is still not very common that popular scientific journals are made accessible by producing aural versions, including AD of images and graphics. To our knowledge, *FoF* is the only Swedish journal practising this and PL

is the only audio describer who has been hired for this task. However, future research could focus on the process of image descriptions when producing accessible teaching materials in various STEM subjects (Holsanova 2019). This would generate similar data and make it possible to collect meaning-making activities from several audio describers.

Although AD is an accessibility aid that is primarily used to offer a richer and more detailed understanding and enjoyment for end users with visual impairment and blindness (Holsanova 2016a,b, Holsanova et al. 2016), initial attempts have been made to use AD for other audiences, beyond those with visual impairment. For example, AD has been used for second language learners for improving lexical competences (Walczak 2016), for acquisition of reading and writing skills (Kleege & Wallin 2015), as an aural guide to children's visual attention (Krejtz et al. 2012a), and as a multimodal learning tool (Krejtz et al. 2012b). Researchers claim that audio description pushes students to practise close reading of visual material and deepens their analysis (Kleege & Wallin 2015). AD could also be useful for groups with print disability, e.g., it can support readers with dyslexia in interaction with complex multimodal texts by 'suggesting' where to look and which information to focus on and integrate. In addition, AD can be of benefit to users on the autistic spectrum and can provide guidance for readers with attentional disorders by directing readers' attention towards relevant information and by suggesting entry points and reading paths. Finally, modified AD can be used as a focalisation tool for cognitively challenged audiences and facilitate access to the emotional content in multimodal narrative texts (Starr 2017). In sum, it has been shown that AD in educational settings can support learners in extracting relevant information, increase their understanding, and facilitate comprehension.

6. Summary and conclusions

The focus of the current study was on the actual use of multimodal materials and the aim was to trace the interpretative processes of meaning-making in readers' interaction with complex multimodal texts. The study was conducted in the framework of social semiotic theories, in particular on the visual construction of knowledge and text-image relations. This was done in combination with cognitive theories on the reception of multimodality and pragmatic theories on multimodal meaning-making. A novel method based on recordings from the AD of a popular scientific journal and think-aloud protocols created during the AD was used to empirically study meaning-making processes. The audio described version of the popular scientific journal, containing spoken scientific explanations, reflected a number of of meaning-making processes and demonstrated how text, images, and graphics in complex multimodal texts have been integrated by the audio describer. In addition to that, the think-aloud protocols performed during the AD raised interesting questions and revealed a large number of important aspects and competences that are necessary for reading and understanding of multimodal scientific texts.

As a result, we were able to trace a variety of meaning-making activities that occurred during the recipient's dynamic interaction with complex multimodal texts. The audio describer – being first recipient and later producer – combined the contents of the available resources, made judgements about relevant information, determined ways of verbalising visual information, used conceptual knowledge, filled in the gaps missing in the interplay of resources, reordered information for optimal flow, and facilitated understanding and cognitive processing.

Based on our study, we suggest that the combination of AD and think-aloud protocols is a fruitful novel methodology for uncovering competencies necessary for processing and understanding multimodal scientific texts. We argue that the results of our study can be applied for educational purposes to promote multimodal and scientific literacy. In particular, we believe that the list of the meaning-making activities identified with the help of these methods can help students develop explicit knowledge about text-image integration and improve their competences in dealing with complex visualisations. By incorporating these aspects of multimodal meaning-making into tailored instruction for specially designed programmes on multimodal literacy, learners can be gradually and critically trained for reading and understanding scientific multimodal documents (Unsworth 1997, Kress & van Leeuwen 1996/2006).

7. References

Bateman, J. (2008). *Multimodality and genre: A foundation for the systematic analysis of multimodal documents.* London, UK: Palgrave Macmillan.

Behnke, Y. (2017). Der Pictural Turn – die Digital Natives und visuelle geographische Lernmedien. Herausforderungen beim Lehren und Lernen mit Bild und Text im Geographieuntrricht.[The pictural turn - Digital natives and visual geographic learning media. Challenges in teaching and learning with pictures and text in geography lessons], *Schulgeographie* [School Geography] #10-8-1. Heft 92 (Mai 2018), 59–66.

Boeriis, M. & Holsanova, J. (2012). Tracking visual segmentation. Connecting semiotic and cognitive perspectives. *Visual communication,* 11(3), 259–281.

Braun, S. (2007). Audio description from a discourse perspective: a socially relevant framework for research and training. *Linguistica Antverpiensia,* NS 6, 357–369.

Braun, S. (2008). Audiodescription research: state of the art and beyond. *Translation Studies in the New Millennium,* 6, 14-30.

Bucher, H.-J. (2007). Textdesign und Multimodalität: Zur Semantik und Pragmatik medialer Gestaltungsformen [Text design and multimodality: On the semantics and pragmatics of media design forms]. In K.S. Roth and J. Spitzmüller (Eds.) *Textdesign und Textwirkung in der massenmedialen Kommunikation* [Text design and text media effects in mass media communication] (pp. 49-76). Konstanz: UVK.

Bucher, H.-J. (2017). Understanding multimodal meaning-making: Theories of multimodality in the light of reception studies. In O. Seizov & J. Wildfeuer (Eds.) (2017, in press). *New studies in multimodality: Conceptual and methodological elaborations*. London/New York: Bloomsbury.

Bucher, H.-J. & Niemann, P. (2012). Visualising Science: The reception of powerpoint presentations. *Visual communication,* 11(3), 283–306.

Danielsson, K., & Selander, S. (2016). Reading multimodal texts for learning – a model for cultivating multimodal literacy. *Designs for Learning*, *8*(1), 25–36.

Ericsson, K. A. & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Holsanova, J. (forthc.). The cognitive perspective on audio description. In C. Taylor & E. Perego (Eds.), *Handbook of Audio Description*. Abingdon, UK: Taylor & Francis.

Holsanova, J., Johansson, R., & Lyberg-Åhlander, V. (2020, June). How the blind audiences receive and experience audio descriptions of visual events - a project presentation. *Pro-*

ceedings of the 3rd Swiss conference on barrier-free communication. Zürich, Switzerland.
Holsanova, J. (2020). Att beskriva det som syns men inte hörs. Om syntolkning [To describe what is visible but not audible. On audio description]. *HumaNetten,* 44, 125-146.
Holsanova, J. (2019). *Bildbeskrivning för tillgänglighet – riktlinjer, forskning och praktik* [Image description for accessibility – guidelines, research and practices]. Myndigheten för tillgängliga medier, rapport nr. 6. [Swedish agency for accessible media, Report #6].
Holsanova, J. (2016a). Cognitive approach to audio description. In A. Matamala & P. Orero (Eds.), *Researching audio description: New approaches*. (pp. 49–73). London, UK: Palgrave Macmillan.
Holsanova, J. (2016b). Kognitiva och kommunikativa aspekter av syntolkning [Cognitive and communicative aspects of audio description]. In J. Holsanova, M. Andrén, & C. Wadensjö (Eds.), *Syntolkning – forskning och praktik* [Audio description – research and practices]. Myndigheten för tillgängliga medier, rapport nr. 4 [Swedish agency for accessible media, Report #4], 17–27.
Holsanova, J., Wadensjö, C., & Andrén, M. (Eds.) (2016). *Syntolkning – forskning och praktik*. [Audio description – research and practices]. Myndigheten för tillgängliga medier, rapport nr. 4 [Swedish agency for accessible media, Report #4].
Holsanova, J. (2014a). Reception of multimodality: Applying eye tracking methodology in multimodal research. In C. Jewitt (Ed.), *Routledge handbook of multimodal analysis*. Second edition, (pp. 285–296). New York and Oxon, UK: Routledge.
Holsanova, J. (2014b). In the eye of the beholder: Visual communication from a recipient perspective. In David Machin (Ed.), *Visual communication*. *Handbooks of communication science*. (pp. 331–355). Berlin/Boston: De Gruyter.
Holsanova, J. (2012a). New methods for studying visual communication and multimodal integration. *Visual communication*, 11(3), 251–257.
Holsanova, J. (Ed.) (2012b). *Methodologies for multimodal research. Visual communication*, 11(3, special issue). London: Sage.
Holsanova, J., Holmberg, N., & Ek, J. (2012 May). Method for tracking reflected reading and multimodal learning of pupils with various abilities. *Proceedings of the designs for learning conference*, (pp. 92–94). Copenhagen, Denmark.
Holsanova, J. & Nord, A. (2011). Multimodal design: media structures media principles and users meaning-making in newspapers and net papers. In H.-J. Bucher, T. Gloning, & K. Lehnen (Eds.), *Neue Medie – neue Formate: Ausdifferenzierung und Konvergenz in der Medienkommunikation* [New Media – new formats: Differentiation and convergence in media communication]. (pp. 81-103). Frankfurt/ New York: Campus.
Holsanova, J. (2011). How we focus attention in picture viewing, picture description, and during mental imagery. In Sachs-Hombach, K. & Totzke, R. (Eds.), *Bilder, Sehen, Denken* [Images, seeing, thinking]. (pp. 291 – 313). Köln: Herbert von Halem Verlag.
Holsanova, J. (2010). *Myter och sanningar om läsning. Om samspelet mellan språk och bild i olika medier* [Myths and truths about reading. On the interaction between language and image in different media]. Stockholm, Sweden: Norstedts.
Holsanova, J., Holmberg, N. & Holmqvist, K. (2009). Reading information graphics: the role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology,* 23, 1215–1226.
Holsanova, J. (2008). *Discourse, vision, and cognition*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Holsanova, J., Rahm, H., & Holmqvist, K. (2006). Entry points and reading paths on the newspaper spread: Comparing semiotic analysis with eye-tracking measurements. *Visual communication,* 5(1), 65-93.

Holsanova, J. (2001). *Picture viewing and picture description. Two windows on the mind*. (Doctoral dissertation). Lund University Cognitive Studies 83.

Jakobson, R. (1959). On linguistic aspects of translation. In R. Brower (Ed.), *On translation.* (pp. 232–239). Cambridge, MA: Harvard University Press.

Jewitt, C. & Kress, G. (2003). *Multimodal literacy*. New York: Peter Lang.

Kaltenbacher, M. & Kaltenbacher, T. (2015). Seeing the unforeseen: Eye-tracking reading paths in multimodal webpages. In J. Wildfeuer (Ed.), *Building bridges for multimodal research: International perspectives on theories and practices of multimodal analysis.* (pp. 227-245). Frankfurt am Main: Peter Lang.

Kleege, G. & Wallin, S. (2015). Audio description as a pedagogical tool. *Disability Studies Quarterly,* 35(2).

Krejtz, I., Szarkowska, A., Krejtz, K., Walczak, A., & Duchowski, A. (2012). Audio description as an aural guide of children's visual attention: Evidence from an eye-tracking study. *Proceedings of the ACM symposium on eye tracking research and applications conference.* (pp. 99-106). New York: ACM.

Krejtz, K., Krejtz, I., Duchowski, A., Szarkowska, A., & Walczak, A. (2012). Multimodal learning with audio description: An eye tracking study of children's gaze during a visual recognition task. In *Proceedings of the ACM symposium on applied perception* (SAP'12). (pp. 83–90). New York: ACM.

Kress, G. & van Leeuwen, T. (2006 [1996]). *Reading images: The grammar of visual design.* London: Routledge.

Lim, F.V. (2018). Developing a systemic functional approach to teach multimodal literacy. *Functional Linguist,* 5(13).

Martinec, R. & Salway, A. (2005). A system for image-text relations in new (and old) media. *Visual communication,* 4(3), 337-371.

O'Halloran, K. L., A. Podlasov, A. Chua, & Marissa K. L. E (2012). Interactive Software for Multimodal Analysis. *Visual communication,* 11(3), 363–381.

Pettersson, R. (2008). *Bilder i läromedel*. Andra upplagan. [Images in teaching material. Second edition]. Tullinge, Sweden: Institutet för infologi.

Reviers, N. (2017). *Audio description in Dutch. A corpus-based study into the linguistic features of a new, multimodal text type*. (Doctoral dissertation). University of Antwerp.

Scheiter, K., Holsanova, J., & Wiebe, E. (2008). Theoretical and methodological aspects of learning with visualizations. In R. Zheng (ed.), *Cognitive effects of multimedia learning*. Hershey, PA: IGI Global. 67-88.

Starr, K. (2017). *Audio description and cognitive diversity: A bespoke approach to facilitating access to the emotional content in multimodal narrative texts for autistic audiences*. (Doctoral dissertation). University of Surrey, UK.

Unsworth, L. (1997). Scaffolding reading of science explanations: Accessing the grammatical and visual forms of specialized knowledge. *Reading*, 31(3), 30-42.

Unsworth, L. & Cleirigh, C. (2014). Multimodality and reading: The construction of meaning through image-text interaction. In C Jewitt (Ed.), *Routledge Handbook of Multimodal Analysis*. Second edition. (pp. 176–188). London, UK: Routledge.

Unsworth, L. & Macken-Horarik, M. (2014). Interpretive responses to images in picture books by primary and secondary school students: Exploring curriculum expectations of a 'visual grammatics'. *English in Education,* 49(1).
van Gogh, T. & Scheiter, K. (2009). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction,* 20, 95-99.
van Someren, M., Barnard, Y., & Sandberg, J. (1994). *The think aloud method: A practical guide to modelling cognitive processes.* London, UK: Academic Press.
Walczak, A. (2016). Foreign language class with audio description: A case study, In A. Matamala & P. Orero (Eds.). *Researching audio description: New approaches*. (pp. 187-204). London, UK: Palgrave Macmillan.
Walsh, M. (2010). Multimodal literacy: What does it mean for classroom practice? *Australian Journal of Language and Literacy*, 33(3), 211-239.
Wildfeuer, J. (2012). More than words. Semantic continuity in moving images. *Image & Narrative,* 13(4), 181–203.

8. Acknowledgements

I wish to thank the editors and two anonymous reviewers for helpful and valuable comments on an earlier version of this paper. I am grateful to the professional audio describer, PL, for volunteering for this research. Many thanks to *Forskning och Framsteg* for the permission to reproduce their material (*Forskning och Framsteg*/Istock 3/2018, p. 28). This work was supported by a grant from the Krapperup Foundation (Gyllenstiernska Krapperupsstiftelsen) KR 2019-0023.