MULTIMODAL COMPREHENSION OF GRAPH-TEXT CONSTELLATIONS: AN INFORMATION PROCESSING PERSPECTIVE

Dissertation

zur Erlangung des akademischen Grades
Dr. rer. nat.
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
der Universität Hamburg

eingereicht beim Department Informatik von

CENGIZ ACARTÜRK

aus IZMIR

NOVEMBER 2009

Genehmigt von der
Fakultät fur Mathematik, Informatik und Naturwissenschaften
Department Informatik der Universität Hamburg

auf Antrag von
Prof. Dr. Christopher Habel (Erstgutachter/Doktorvater)
Prof. Dr. Wolfgang Menzel (Zweitgutachter)
Prof. Dr. Barbara Kaup (Drittergutachter)

Hamburg, den 19.02.10 (Tag der Disputation)

Abstract

Documents containing information in different representational modalities (e.g., text, pictorial and diagrammatic illustrations, graphs) are widely used in personal and professional settings of daily life, both in printed media and in human-computer interaction environments, such as WWW. In particular, graph-text constellations are one of those cognitive means by which we accomplish high-level problem solving and communication tasks. This dissertation performs a systematic, scientific investigation of multimodal comprehension of graph-text constellations by humans—as natural cognitive systems—for providing empirical basis and scientific models for the design and development of artificial cognitive systems that have the capacity to support users during their design and production of graph-text constellations. First, an overview of the existing approaches in graph comprehension and language comprehension is presented. Second, a theoretical framework is developed by extending the reach of the existing approaches to embody comprehension of graphical and linguistic entities under the framework of the common conceptual basis for multimodal comprehension. Third, various aspects of the interaction between language comprehension and graph comprehension are empirically investigated by conducting experiments with human participants. The empirical investigations reveal how users integrate the information contributed by the representations in different modalities, how the changes in one modality lead to substantial influences in comprehension of the entities in the other modality, and how inappropriate design of graphs and graph-text combinations give rise to difficulties during the course of multimodal comprehension. Based on the analysis of those difficulties in comprehension from the perspective of HCI, this dissertation lays the foundations for appropriate design of graph-text constellations and for the development of software tools that would provide users with support for appropriate design during their course of designing and producing graph-text constellations.

Zusammenfassung

Dokumente, die Informationen in verschiedenen Repräsentationsmodalitäten (z.B. Text, bildhafte und diagrammatische Darstellungen, Graphiken) enthalten, sind im privaten und beruflichen Alltag sowohl in den Printmedien als auch in Anwendungsszenarien der Mensch-Computer-Interaktion (MCI), wie z.B. im WWW, weit verbreitet. Insbesondere sind Graph-Text-Konstellationen eines jener kognitiven Mittel, mit denen wir anspruchsvolle Probleme und Kommunikationsaufgaben bewältigen. Diese Dissertation enthält eine systematische und wissenschaftliche Untersuchung zum Verstehen von Graph-Text-Konstellationen durch Menschen—als natürlichen kognitiven Systemen—und stellt damit ein empirisches und wissenschaftliches Modell für das Design und die Entwicklung von künstlichen kognitiven Agenten bereit, die in der Lage sind, Nutzer beim Design und bei der Entwicklung von Graph-Text-Konstellationen zu unterstützen. Als erstes wird ein Überblick zu den existierenden Ansätzen zum Verstehen von Graphen und zum Sprachverstehen gegeben. Als zweites wird ein theoretisches Rahmenwerk entwickelt, das die existierenden Ansätze derart erweitert, dass das Verstehen von graphischen und linguistischen Entitäten im Rahmen einer gemeinsamen begrifflichen Basis für multimodales Verstehen behandelt wird. Als drittes werden diverse Aspekte der Interaktion von Sprachverstehen und Graphverstehen durch Experimente mit menschlichen Teilnehmern empirisch untersucht. Die empirischen Untersuchungen zeigen, dass und in welcher Weise Menschen Informationen, die in verschiedenen Modalitäten repräsentiert sind, integrieren, dass und in welcher Weise Änderungen in einer Modalität wesentliche Einflüsse auf das Verstehen von Entitäten in anderen Modalitäten haben, und dass und in welcher Weise ein ungeeignetes Design von Graphiken und Graph-Text-Kombinationen zu Schwierigkeiten beim multimodalen Verstehen führen. Ausgehend von der Analyse dieser Schwierigkeiten beim menschlichen Verstehen stellt diese Dissertation aus der Perspektive der MCI Grundlagen für das Design von Graphik-Text-Konstellationen bereit und ebensos Prinzipien für die Entwicklung von Softwaretools, die Nutzer beim geeigneten Design während der Konstruktion und Produktion von Graph-Text-Konstellationen unterstützen.

Acknowledgements

First and foremost, I would like to thank my *Doktorvater* Christopher Habel, who has been my intellectual best friend as well as being an excellent academic supervisor. Countless number of long and inspiring discussions have provided me with the opportunity to learn conducting both collaborative and independent research during whole time span of the study.

I am grateful to Wolfgang Menzel from the University of Hamburg and Barbara Kaup from Tübingen University for accepting to review the dissertation and for their very valuable and constructive feedback. Special thanks go to Carola Eschenbach, who provided detailed comments and suggestions to many aspects of the study, including written material as well as oral presentations.

This dissertation project was supported by DFG (German Science Foundation) in International Research Training Group IRTG 1247 ‘Cross-modal Interaction in Natural and Artificial Cognitive Systems’ (CINACS). I am grateful to the CINACS project organizers, whose financial and academic support made this dissertation possible. Special thanks go to Jianwei Zhang who is the spokesperson of CINACS, to Maosong Sun, and to other CINACS professors both at the University of Hamburg and at Tsinghua University for their useful comments and suggestions in our colloquia and internal meetings.

I would like to thank Kürşat Çağıltay from the Middle East Technical University (METU) for his collaboration and support for providing me with experimental facilities. Özge Alaçam helped me a lot during the time that I conducted empirical work at METU HCI Research and Application Laboratory. I give special thanks to Deniz Zeyrek, Cem Bozşahin, Ceyhan Temürcü, and Annette Hohenberger from METU Cognitive Science Program for their intellectual support and friendship at different stages of my academic career.

I am grateful to my flatmate and colleague Sascha Jockel, and to my close daily partners at the University of Hamburg: Tian Gan, Christian Graf, Mario Maiworm, Patrick McCrae, Martin Weser, and Mohammed Elmogy. I express my gratitude to my friends Hildegard Westermann, Özgür Özçep, and Emel Özçep for their close friendship and support. It has been a real pleasure to work with Tatjana Tetsis and other CINACS administrative staff, who spent much effort to make the progress of the CINACS project possible. I also appreciate my student assistants Jan Schlüter, Isabelle Streicher, Huixin Zheng, Elnara Babayeva, and Abdulilah Dozen, who provided technical support at different stages of the study.

Special thanks go to Oya Sezginer and Hatice Bay for their generous support during the final stages of publication. Many friends encouraged me to focus on my research and to end up with research progress, including Nazan Ağırmatlıoğlu, Selin Poçan Barutçuoğlu and Engin Barutçuoğlu, İbrahim Çalışır, Efkan Çatıker, Çağrı Çöltekin, İbrahim Elibal, Gökhan Eryol and Feyza Taşkazan Eryol, Barış Mücen, Emre Sezginer, Ulaş Sunata, Gökçe Türkmendağ, and Murat Ulubay. Şahabettin Demirel deserves special thanks, who has been one of my best intellectual friends.

Last but not the least, my family and my parents-in-law have showed great patience and support during my education, including the doctoral studies. This dissertation would not be possible without their huge effort to encourage me to conduct research without interruption.

I dedicate this dissertation to my wife, Banu, who showed the greatest patience and love.

TABLE OF CONTENTS

Section 1

Graph-Text Constellations: Fundamental Cognitive Aspects of Multimodal Information Processing

1 Graph-Text Constellations: A Cognitive Science Perspective

As humans, we use a variety of cognitive means to accomplish high-level problem solving tasks in daily life; in particular

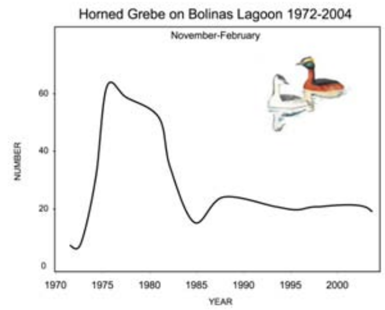
- we use number systems and arithmetics to solve numerical problems,
- we plan travel routes by using geographic maps, or topological maps such as the transit maps of public transport systems,
- we design organizational charts to show the structure of organizations and the relationships among its parts. We produce information graphics, such as statistical bar graphs and line graphs, to analyze behavior of physical, economic, or social systems.

Additionally, we use those means of problem solving to communicate with other people; in particular

- we produce and interpret written or spoken utterances to express and understand ideas or thoughts,
- in collaborative problem solving, we use numbers and equations to discuss solutions to numerical problems,
- we draw sketch maps or mark routes on the transit maps to show the travel routes to other people,
- we use organizational charts and information graphics to convey hierarchical and quantitative information in presentation settings.

The subject of our activity in accomplishing these tasks comprises *cognitive entities* such as our knowledge, ideas, plans and *entities in the environment* such as language in written or spoken form, maps, diagrams, and information graphics.

Consider the use of information graphics and language in a graph-text constellation, as exemplified in Figure 1.1. That is an excerpt from a waterbird consensus project report, which started in the 1970s. The graph-text constellation is composed of a population trend graph that depicts the number of a waterbird species, namely horned grebe, between 1972 and 2004, and a text that describes the population trend.



Bolinas Lagoon Population Trends

From a peak of about 60 wintering birds in 1976, numbers have declined to about 20 birds currently.

Figure 1.1. A sample graph-text constellation from "Waterbird Census at Bolinas Lagoon, Marin County, CA" by Wetlands Ecology Division, Point Reyes Bird Observatory (PRBO) Conservation Science (redrawn based on the original). The graph and the paragraph in the figure are placed side by side for the purpose of demonstration; see the census report for the originals: http://www.prbo.org/cms/366, retrieved on September 25, 2009. The constellation includes a population trend graph depicting the number of wintering birds in the lagoon and verbal statements about the population trend.

The graph *depicts* and the text *describes* the knowledge of the producer of the graph-text constellation about the population trend of the horned grebe on the lagoon. In other words, the entities in the environment—in this case, the graph-text constellation—externalize cognitive entities, i.e. the knowledge of the producer. In part 1.1 below, cognitive entities are called *internal*, *mental*, or *conceptual representations* and entities in the environment are called *external representations*.

1.1 INTERNAL AND EXTERNAL REPRESENTATIONS

The study of internal representations has been the main domain of research in cognitive science and cognitive psychology since the 1960s. The information processing approach to human mind, namely the *computational theory of mind* (CTM) considered human mind as an information processing system and as a form of computation (Putnam, 1961; Fodor, 1975; see Horst, 2005, for a review). The central hypothesis of CTM has been that thinking can be investigated in terms of representations and computational procedures that operate on those. Internal representations have been widely accepted as essential for cognition, and the CTM approach has dominated research in different domains including the relation between cognition, language, and vision. This view is connected to Artificial Intelligence (AI) in that it assumes the existence of internal representations that are analogous to computer data structures, and the computational procedures that are similar to computational algorithms.

Historically, the CTM was connected to several empirical research domains that have been influential in cognitive science. Computational accounts have been offered for various types of mental phenomena including problem solving and reasoning (as a special case of problem solving, Newell & Simon, 1972, 1976; Fodor, 1975), Human

Computer Interaction (Card, Moran, & Newell, 1983), language learning and use (Chomsky, 1959; 1965; Fodor, 1975; Pinker, 1989), and vision (Marr, 1982; Osherson, Kosslyn, & Hollerbach, 1990).

Newell and Simon (1972) proposed a computational account for human problem solving. Their approach was to recognize problem solving as a search in the problem space of candidate states and the use of heuristics such as means-ends analysis for the search (viz., the problem space hypothesis). Specifically, according to the problem space hypothesis "the rational activity in which people engage to solve a problem can be described in terms of (1) a set of states of knowledge, (2) operators for changing one state into another, (3) constraints on applying operators and (4) control knowledge for deciding which operator to apply next" (cf. the problem space principle by Card et al., 1983, p. 87). Newell and Simon applied the problem space hypothesis to a set of classical problems such as the missionaries and cannibals problem and the Tower of Hanoi problem.

Newell and Simon's (1972) approach paved the way for the development of human information processing models for Human Computer Interaction (HCI) research such as the GOMS model (Card et al., 1983), and the development of computational architectures for the study of human cognition such as EPIC (Meyer & Kieras, 1997a, b; Kieras & Meyer, 1997), Soar (Laird, Rosenbloom, & Newell, 1987; Newell, 1990) and ACT-R (Anderson, 1993; Anderson & Lebiere, 1998). In particular, the GOMS (Goals, Operators, Methods, and Selection rules) model investigates users' interaction with computers in terms of component tasks and a set of elementary actions such as perceptual, motor, and cognitive processes (John & Kieras, 1996a, b; Kieras, 1997; see Chapter 9 for a discussion of the GOMS approach).

The other two influential empirical research domains that CTM was connected to were Chomskian linguistics and Marr's theory of vision. Chomsky challenged the behaviorist approach to language acquisition (Skinner, 1957) by proposing the existence of a universally shared, innate set of linguistic principles (viz., universal grammar). Fodor (1975) argued that the Chomskian approach required an inner language of thought, which in turn conceived by computationalists as the potential for investigating the principles of universal grammar in computational terms.

Marr (1982) and Marr and Poggio (1977) proposed a computational approach to vision. According to this view, three distinct levels of analysis are necessary to understand vision as an information processing system: The *computational level* is the abstract computational theory of the system. The *algorithmic level* is concerned with the implementation of the computational theory, i.e. the choice of representation for the input and output and the choice of the algorithms that transform one representation into another. The third level, namely the *implementation level*, is concerned with the physical realization of the system, e.g. the detailed computer architecture or in biological vision, the neural system. Marr's theory of vision is connected to CTM due to the algorithmic level where internal representations and algorithmic processes are involved (Horst, 2005).

Although the CTM approach has been the dominant view in cognitive science, there has been much disagreement in the nature of cognitive structures and the computational procedures that operate on those. A common view in CTM has been that reasoning is a process that operates on abstract sentential (also called *propositional* or *symbolic*) representations (Newell & Simon, 1972, 1976; Fodor, 1975, 1983; Pylyshyn, 1984). In this view, human cognitive architecture is assumed to be composed of informationally encapsulated modules that set architectural restrictions on the interaction between language and sensory-motor systems (e.g., vision, audition). The representations in sensory-motor systems are transduced into symbolic representations with a combinatorial syntax and semantics, which provides further support to higher cognitive functions such as language, memory, and thought processes. This view has been called *amodal symbol systems* approach (Barsalou, 1999, 2008). Historically, this view led into the separation of reasoning and perceptual processes, which resulted in independent development of the study of perception and the study of reasoning strategies (Chandrasekaran & Narayanan, 1992).

More recently, Hauser, Chomsky and Fitch (2002) ascribed a purely syntactic character to the computational core of the "faculty of language". Pinker and Jackendoff (2005) and Jackendoff and Pinker (2005) argued that the computational core should not be restricted to syntactic aspects; similar combinatorial systems, for example for the conceptual/semantic level, seem to be necessary for interfaces between the linguistic system and the sensory systems (Landau & Jackendoff, 1993; Barsalou, 1999; Jackendoff, 1996, 2002). In particular, Jackendoff (1996, 2002) proposed conceptual and spatial representations as essential parts of the interface between the linguistic system and the sensory systems (cf. the representational modularity hypothesis; see Chapter 2).

Moreover, the situated theories of meaning (cf. the theories of grounded cognition, see Barsalou, 2008, for a review) have presented advantages over the purely symbol-based—ungrounded—theories of meaning (Glenberg & Robertson, 2000; Zwaan & Madden, 2005). For instance, Glenberg and Robertson (2000) presented empirical evidence that embodied theories of meaning could capture the meaning of novel situations whereas ungrounded theories of meaning cannot. Mental imagery—the quasi-perceptual experience that resembles perceptual experience, but occurs in the absence of the appropriate external stimuli for the relevant perception (Thomas, 2009)—has been proposed as evidence for the existence of analog mental representations (Kosslyn, 1980, 1994a; Kosslyn, Ganis, & Thompson, 2003, also see Pylyshyn, 1973, 2003, for the debate on mental imagery). Nevertheless, despite its theoretical significance, the research on *grounding* has not been in a mature state in cognitive science. Integration of grounding with the classical paradigms is needed for it to be fully accepted (Barsalou, 2008).

In applied computer science, for instance in robotics, the specification of the processes that maintain the correspondence between perceptual data and abstract representations that refer to objects is called *anchoring* (Coradeschi & Saffiotti, 2003). Anchoring was offered as a solution to the so-called symbol grounding problem (Harnad, 1990).[1] However, the particular nature of the anchoring specifications has been under investigation.

In summary, the study of the nature of internal representations and the computational procedures that operate on those has been a continuing debate in cognitive science, producing partial solutions to problems in different domains of research.

In the past several decades, the theory of distributed cognition has brought a complementary perspective to cognition by extending its reach beyond the individual to incorporate the interaction between humans and the entities in the environment, i.e. *external representations*. (Zhang & Norman, 1994; Hutchins, 1995; Hollan, Hutchins, & Kirsh, 2000; Zhang & Wang, 2009).

Humans perform well in complex cognitive tasks when they interact with external representations. The study of reasoning strategies has shown how information processing tasks can be distributed across internal mind and external environment (Reisberg, 1987; Zhang & Norman, 1994; Cox, 1999). For instance, one can multiply two-digit numbers mentally, whereas the multiplication of four-digit numbers requires constructing the external representations of interim results, e.g. the interim values as the numbers on paper. Those uses of external representations exemplify their use as *memory aids*. The trivial reason for the role of external representations as memory aids is the limited storage and processing capacity of human working memory. On the other hand, there are further functional differences between internal and external representations, which underlie the role of constructing external representations in enhancing cognitive performance other than their role as memory aids. Reisberg (1987) states that by means of externalizing a representation, we create the relevant input for the perceptual processes which give access to knowledge and skills that are otherwise unavailable to us. In this way, external representations enhance cognitive performance by facilitating examination and development of one's own ideas in addition to being used as memory aid.

---

[1] Harnad (1990) describes the symbol grounding problem by the following questions: "How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads? How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?" (p. 335).

Further investigations on the role of external representations in cognitive performance has shown the advantages of using *perceptual assistance* for improving humans' performance in a variety of tasks such as mental imagery tasks, second language reading and reading of comic stories. In particular, Anderson and Helstrup (1993) showed that the provision of drawing support increases human performance in creating recognizable patterns from simple shapes. Miyake (1997) presented empirical evidence for higher performance of readers in constructive reading of comic stories (i.e., a better understanding of the relationships between the characters in the story) when they were asked to take simple notes often. In other problem solving domains such as creative design the facilitating role of constructing external representations on constructing design ideas and concepts has been revealed (Goldschmidt, 1991; Purcell & Gero, 1998; Do, Gross, Neiman, & Zimring, 2000; Suwa & Tversky, 1997, 2002).

In part 1.2 below, the role of external representations in problem solving and communication is discussed in more detail.

1.2 EXTERNAL REPRESENTATIONS IN PROBLEM SOLVING AND COMMUNICATION

Language, in the form of written text or spoken utterances, is the most frequently used type of external representation in problem solving tasks of daily life and in communication. Arabic and Roman numerals, and mathematical equations exemplify other types of *symbolic* (i.e., *propositional* or *sentential*) external representations than language. Additionally, in specific domains such as scientific activity, science education and economics, experts use diagrams and information graphics (henceforth, graphs) in communication as well as in problem solving tasks.[2] These representations exemplify *depictive* (also called *pictorial* or *imagistic*) external representations (see 1.3 for a discussion of the types of external representations).

Making certain information explicit whereas making other implicit, particular external representations, either symbolic or depictive, reveal how computational procedures of human mind operate. Revisiting the multiplication example presented above, consider using Roman numerals instead of Arabic numerals to multiply two four-digit numbers. Although the Arabic number system and the Roman number system are the two alternative external representations of the same abstract number system, they employ different operations of computational procedures. The difficulty of multiplication with Roman numerals compared to multiplication with Arabic

[2] The term *graph* is used for graphical representation of quantitative information in Cartesian coordinates (i.e., statistical graphs such as line graphs and bar graphs). Other terms are used interchangeably for the same meaning in the following chapters: *statistical graphics*, *information graphics*, *graphics*, *graph,* and *Cartesian graph*.

numerals is neither due to the recent conventions (which are based on the knowledge of Arabic numerals) nor due to the lack of our deep knowledge of Roman numerals. The difficulty is due to the incompatibility between our computational capabilities and the Roman numerals as a representational number system. This also explains why arithmetics were not developed before the discovery of the Arabic number system (Marr, 1982; Zhang & Norman, 1995).

Consider another example from the domain of depictive representations. The two graphs in Figure 1.2 show the number of annual sunspots between 1749 and 1924; the data were provided by Yule (1927) and graphed by Cleveland (1993).

(a)

(b)

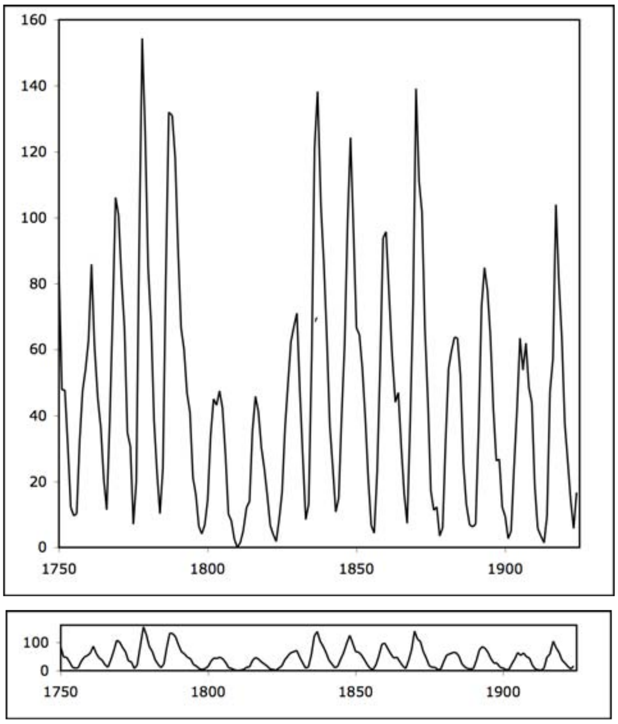


Figure 1.2. Yearly sunspot numbers between 1749 and 1924 with two different aspect ratios (redrawn based on the data provided by Yule (1927), and graphed by Cleveland (1993)).

Although the two graphs are the representations of the same data points, it is easier to interpret that the cycles rise more rapidly than they fall by investigating the graph in Figure 1.2b in comparison to the graph in Figure 1.2a (Cleveland, 1993).

In summary, in addition to enhancing cognitive performance by providing perceptual input, particular external representations can influence our reasoning strategies in different ways. This property of external representations is relevant not only to problem solving and reasoning but also to communication, as discussed below.

A closer look into problem solving strategies by expert and non-expert people reveals differences in constructing and interpreting external representations. Experts have more abstract knowledge of domain-specific problems than novices, whereas novices often rely on problems' literal features (Chi, Feltovich, & Glaser, 1981). Accordingly, experts rely on a richer variety of external representations in different formats compared to non-experts (Tabachneck-Schijf, Leonardo, & Simon, 1997).

A chemist, for instance, conducts an experiment and draws a line graph to interpret the relationship between the measured variables. Consider the two graphs in Figure 1.3. The graphs represent the relation between two variables, temperature and pressure in such an experiment conducted on a chemical system.

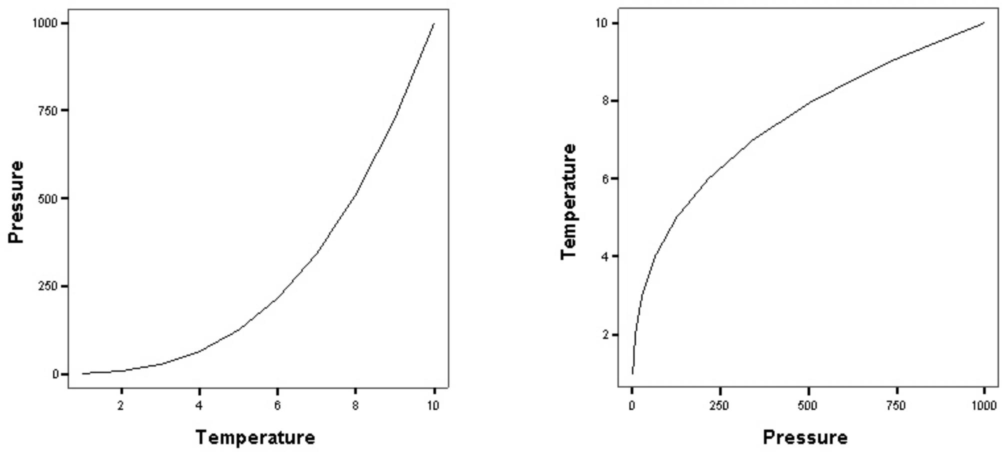


Figure 1.3. The presentation of the same relation with two perceptually different graphs.

The two graphs in Figure 1.3 represent the same relationship between temperature and pressure. For an expert chemist, it is a simple task to interpret that the graphs represent the same relation. However, for a non-expert person, this interpretation is difficult because the two graphs induce perceptually different messages.[3]

Consequently, when used for communicative purposes, graphs that have different perceptual properties may provide non-expert people with perceptual inputs that may lead to different conceptualizations and misconceptualizations of the presented information about the domain entities. The findings in graph comprehension studies with novice students provide support to this claim. For instance, Elby (2000) states that in a velocity-time graph "the flat horizontal line can be taken to activate *stillness*, an element of cognitive structure associated with lack of motion" (p. 485, emphasis in

[3] The term *induce* is used to mean humans' interpretation of the information provided by a particular external representation. The formal sense of the term *induce* in mathematics and logic is not intended here.

original). As a consequence, the velocity-time graph of a car with a flat horizontal line leads to misconceptualizations in students in that the students are likely to conclude that the car is motionless (also see Leinhardt, Zaslavsky, & Stein, 1990 for a review of students' common misconceptions and difficulties in interpretation of mathematical functions and graphs).

In summary, reasoning by non-expert people is more open to the influence of perceptual properties of external representations compared to reasoning by expert people. This difference underlies the previous findings about non-experts' difficulties in constructing and interpreting graphs (Elby, 2000) and representations that involve graphs and text (Tabachneck-Schijf et al., 1997; Paoletti, 2004). On the other hand, depictive external representations, such as graphs and diagrams are widely used to augment language in communication settings. In magazine articles or textbooks, either in print or on screen, graphs and diagrams augment written text; in presentation settings, such as lectures or seminars, they augment spoken language.

Human Computer Interaction (HCI), being concerned with the "design, evaluation and implementation of interactive computing systems for human use" (Hewett, 1992, p. 5)[4] is connected to the study of external representations from the perspective of cognitive science. The advantages of using depictive representations for some types of problem solving have been a topic of experimental research in the areas of cognitive science (e.g., Monaghan & Stenning, 1998) and of HCI (e.g., Scaife & Rogers, 1996) during the past decade, leading to principles and guidelines for designing graphical interfaces. The importance of design and development of human-computer interfaces that are able to support users to produce external representations has readily been recognized in HCI. These interfaces include software packages for expert use, as well as word processors, spreadsheet software, and presentation software tools that are used by non-expert people.

Nevertheless, technology has not sufficiently advanced to provide intelligent decisions about the correspondences between the information provided by representations in different formats (Cox, 1999). In particular, the available software tools offer limited capability to augment language by graphs. The distributed cognition approach, bringing the interaction between the internal mind and the external environment into the foreground, provides a cognitive framework for understanding interactions between people and technologies (Hollan et al., 2000). On the other hand, the distributed cognition approach has not yet focused on the distribution of information processing between the internal mind and external

[4] The quotation is the part of the definition of HCI given by ACM SIGCHI/ACM Special Interest Group on Computer-Human Interaction, Curricula for Human Computer Interaction: "Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them" (Hewett, 1992, p. 5).

representations of different types in the environment. Recent approaches that propose a cognitively oriented viewpoint consider design as *construction of representations* (Visser, 2006). Therefore, the investigation of specific types of external representations and the interaction between specific types of representational modalities is important for the analysis of the interaction between humans and external representations from the perspective of HCI and cognitive science. In part 1.3 below, external representations are discussed with respect to their types and modalities.

1.3 TYPES OF EXTERNAL REPRESENTATIONS AND REPRESENTATIONAL MODALITIES

External representations are not unitary; there are many different types of external representation. They can be classified according to (orthogonal) properties such as analogue and non-analogue representations, arbitrary and non-arbitrary representations, static and dynamic representations, and linguistic and non-linguistic representations according to the media of representation and the modality of representation (Palmer, 1978; Bernsen, 1994).

In particular, two types of external representations are in the focus of current cognitive science research. The first type is often characterized as linguistic or language-oriented representations that contain natural and artificial languages as well as numeric symbols. They are also called *symbolic*, *sentential*, or *propositional* representations. The second type of external representation is often characterized as being pictorial, diagrammatical or image-like representations such as pictorial illustrations, maps and iconic diagrams.[5] This type of representations is called *analog, diagrammatic, spatial* or *depictive* representations (Kosslyn, 1980). In this dissertation, the terms *symbolic representation* and *depictive representation* are used for these types of representations.

Symbolic representations and depictive representations exemplify two types of *representational modality*. The difference between the two lies in the relationship between the representing entity and the represented entity. The relation between a symbolic representation and the represented entity is *arbitrary*, i.e. no structural correspondence is required between a word and the entity represented by the word. On the other hand, depictive representations have structural correspondence between the representing entity and the represented entity. In other words—whether they are internal or external—depictive representations share relevant inherent constraints

[5] The term *iconic* is used to mean the structural similarity between the diagram and the represented entities. For instance, schematic illustrations of mechanical devices are iconic diagrams. Maps can also be classified as iconic diagrams, though with a different level of abstraction than schematic illustrations of mechanical devices (Hegarty, Carpenter, & Just, 1991).

with the domain they represent, whereas symbolic representations must code such constraints that are implicitly embodied in depictive representations (Palmer, 1978).

Charts, abstract diagrams and graphs (e.g., flowcharts, organizational charts, line graphs, bar graphs etc.) are depictive representations that share the relevant inherent constraints with the domain they represent at an abstract level; they lie between symbolic representations and pictorial illustrations. As stated by Winn (1987), "from words, they inherit the attribute of abstraction; but like pictures, they exploit spatial layout in a meaningful way. Their abstract nature makes them well suited to explaining how processes work where realistic pictures would fail" (p. 152).

Moreover, graphs are recognized as a distinct class of visual representation being different from charts and abstract diagrams as well as pictorial illustrations, maps, and iconic diagrams. Like abstract diagrams and charts, graphs have a higher level of abstraction compared to pictorial illustrations, maps and iconic diagrams. The spatial information derived from statistical graphs represents abstract relations between entities in the domain of discourse, whereas the spatial information derived from pictorial illustrations, maps and iconic diagrams represents spatial configurations between entities mostly in the physical world.[6] Furthermore, graphs convey quantitative information; the spatial configuration of graphical entities corresponds to the relationship between the domain entities with respect to a set of rules and conventions.

The subtypes of graphs are also abundant. Line graphs are generally used to convey trend information whereas bar graphs better convey information for making comparisons between specific measurements (Kosslyn, 1994b, 2006; Zacks & Tversky, 1999). Scatter plots are often used to show distribution of data points in the value space; and pie charts are used to show part-whole relationships. The survey studies on graphs in scientific publications and in news magazines show that line graphs that plot change of a quantitative measurement over time are the most frequently used graph type. Line graph is also the earliest type of graph with Cartesian coordinates. Line graphs are followed by bar graphs, scatter plots, and pie charts (Cleveland & McGill, 1985; Butler, 1993; Peden & Hausmann, 2000; Spence, 2005).

In summary, depictive representations do not comprise a homogenous class of representation. In particular, graphs, being constructed and interpreted according to a

[6] The term *spatial information* is used for the unifying information type that underlies different types of information. It can be derived from information sources in various representational or sensory modalities. Verbal information, either written or spoken, can be a source of spatial information. Information in different sensory modalities, such as haptic and auditory, can also be sources of spatial information. The derivation of spatial information from different modalities requires translation or unification of information types between the source modalities (Landau & Jackendoff, 1993).

set of rules and conventions, constitute a distinct class of depictive representation compared to iconic diagrams, maps, and pictorial illustrations. In contrast to pictorial illustrations, the knowledge of *graph schema* plays an important role in interpretation of graphs.[7]

The systematic investigation of the underlying mechanisms of interpretation of graphs by humans is relatively new. The focus of graph research in the last several decades has mostly been *visual decoding* or *pre-attentive vision* processes, few studies having investigated *conceptual processes* in interaction with *perceptual processes* (e.g., Carpenter & Shah, 1998, see Chapter 2 for a review of graph research).[8] Moreover, the relationship between graphical means of communication and language has been discussed from a general psychological perspective, e.g., Kosslyn's (1989) analysis of charts and graphs as means for communicating quantitative information, and Pinker's (1990) graph comprehension architecture. However, systematic investigation leading to a common base for characterizing the similarities and differences between modality-specific means of communication, and thus between representational modalities, is rare. Consequently, relevant prerequisites for the specification and realization of intelligent artificial systems that are able to construct representations in different modalities for multimodal communication and problem solving are missing. The part 1.4 below discusses multimodal communication and problem solving in humans as natural cognitive systems.

1.4 MULTIMODAL COMMUNICATION AND PROBLEM SOLVING

Multimodal constellations are widely used in communication and problem solving settings of daily life, e.g. in newspaper articles and in scientific and educational settings such as in scientific articles, textbooks, and classroom presentations. There are many cases where one representational modality has an advantage over the other in terms of computational efficiency in processing by humans (Larkin & Simon, 1987). However, depictive representations are not generally per se explanatorily adequate when presented in isolation. Therefore, depictive representations are often accompanied by language. As a consequence, *multimodal* constellations are used abundantly in communication and problem solving settings.

[7] The *graph schema* concept is discussed in Chapter 4. Briefly, the term *schemata* is used for the *long term memory* structures that include the necessary information for interpretation of specific external representations such as diagrams and maps. Schema construction is a common aspect that underlies comprehension of abstract visual representations.

[8] Cleveland and McGill (1985) give the following definition for *visual decoding*: "the instantaneous perception of the visual field that comes without apparent mental effort" (p. 828). The term *perceptual processes* (of graph comprehension) is also used for the same meaning. The term *conceptual process* is often used to describe higher-level processes such as interpretation of graphical entities in problem solving rather than perceptual processes.

The term *multimodal* is used in different senses by different disciplines. In cognitive neuroscience the term is often used to mean sensory channels (e.g., visual, haptic, olfactory etc.) rather than representational modalities (e.g., depictive, symbolic etc.). In educational psychology and instructional design, the term *multimedia* and *multiple representations* are often used to mean presenting information in multiple representational modalities.[9] In this dissertation, the term *multimodal* is used as shorthand for the presentation of information in multiple representational modalities, as exemplified by Figure 1.1 at the beginning of this chapter. Accordingly, multimodal communication and problem solving refer to using multiple representational modalities in one action of communication or one process of problem solving. The two relevant modalities are text (i.e., language in written form) and graphs, unless stated otherwise.

For humans—as natural cognitive systems—a successful use of external representations in communication and problem solving requires the integration of information induced by the representations in different modalities. Although humans exhibit a tendency for the integration—guided by almost automatically performed processes without conscious effort—the nature of this integration is not straightforward. Foremost, humans do not receive visual information in a way that a camera recorder does. Kosslyn's (1994b) maxim *the mind is not a camera* reflects this characteristic of human visual system. Humans organize visual information by their eyes and brain. The interaction between perceptually induced (bottom-up or data-driven) information and conceptually induced (top-down) information characterizes human visual information processing. To exemplify this, consider the frequently cited example, the duck-rabbit in Figure 1.4 (originally noted by Jastrow, 1890).

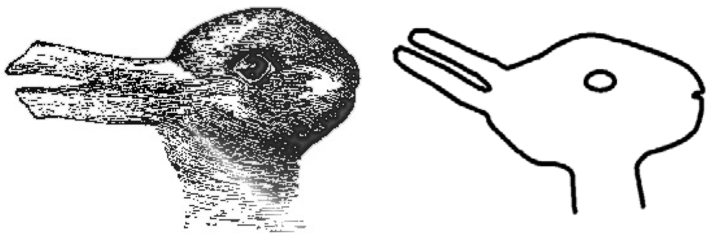


Figure 1.4. Two versions of the duck-rabbit.

The figure can be seen in two distinct ways: as a duck facing left or a rabbit facing right. Depending on the explicit instruction or the expectations, the reader may see a duck or a rabbit. The duck-rabbit example is often used to show that human

[9] On the other hand, Mayer (2009) states that in the context of learning and instructional science the term *multimedia learning* is more accurately called "dual-mode, dual-format, dual-code, or dual-channel learning" (p. 5) for emphasizing the distinction between *verbal* and *pictorial* presentation formats.

perception is not solely specified by the stimulus; rather, it is influenced by conceptual information.[10]

Moreover, humans have the ability to interpret depictive representations as belonging to different types. In other words, humans are able to make multiple levels of abstraction from depictive representations. For example, consider the depictive representation in Figure 1.5.

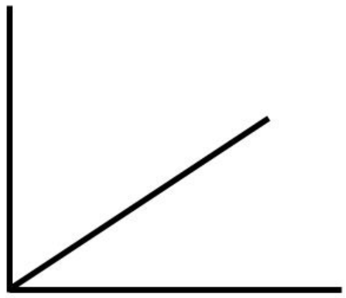


Figure 1.5. The same representation as a sketch map or a graph.

Depending on verbal instruction, humans are able to interpret the representation in the figure as a sketch map or as a line graph (Schiano & Tversky, 1992). Again, that exemplifies the influence of conceptual—in this case verbal—information on perceptual information.

Another peculiar characteristic of human perception and cognition is that humans' mental representation of space and the processing of spatial information have qualitative characteristics (cf. qualitative spatial reasoning), mostly restricted to processing by topological terms (Habel & Eschenbach, 1997). In general, the abstract topological and geometrical structures are relevant building blocks of conceptual representations. They are needed not only for communication about physical space, but also for types of using what often are called 'figurative language' (Habel, 1990; Forbus, 1995; Habel & Eschenbach, 1997; Eschenbach, Habel, Kulik, & Leßmöllmann, 1998; Eschenbach, Tschander, Habel, & Kulik, 2000; Tschander, Schmidtke, Habel, Eschenbach, & Kulik, 2003).[11]

[10] It should also be noted that the ambiguity exemplified by the duck-rabbit is not available to internal representations (Chambers & Reisberg, 1985; Reisberg, 1987). Chambers and Reisberg (1985) showed that the participants, after a short presentation of the ambiguous figures, were not able to reconstruct their mental images for an alternative construal of the form. However, when they drew a picture from their mental image, they were able to reinterpret their own drawings. This finding also shows how information processing is distributed across the internal mind and external representations (cf. the distributed cognition theory).

[11] Neurological evidence for the interaction between language and space is controversial at the current state of the art. At the first glance, neural organization of language and space seems separated. Nevertheless, as Chatterjee (2001) states, "language and spatial representations probably converge at

Accordingly, although graphs are typically depicted in Cartesian coordinate system, their interpretation by humans is based neither on Cartesian coordinates nor on Euclidean space. Moreover, graphs exhibit the properties of both depictive representations and symbolic representations. Graphs exhibit the properties of depictive representations by using the space in a non-arbitrary manner. In particular, the spatial configurations of graphical entities in the graph space correspond to the relationships between the represented domain entities. Graphs exhibit the properties of symbolic representations as well. The distinctive property of symbolic representations compared to depictive representations is that the former have clear syntactic structures (i.e., rules of formation, Kosslyn, 1980). Graphs are cognitive human artifacts that have a clear syntactic structure whereas other types of depictive representations such as pictorial illustrations do not. The syntactic analysis of a graph can be used for further analyses at semantic and pragmatic level (Kosslyn, 1989; Pinker, 1990; Tversky, 2004). As implied by being constructed and interpreted with respect to a set of rules and conventions, graphs have a communicative nature. The common aspect of graphs (in contrast to pictorial illustrations, iconic diagrams, maps etc.) and language is that both are means of communication with a grammatical structure.

As stated in 1.3, the relation between graphs and language—as means of communication with grammar—has been investigated by Kosslyn (1989) and Pinker (1990) from a general psychological perspective. Kosslyn (1989) offered an analytic scheme that systematically isolated constituents of graphs. He investigated the structures of the constituents and the interrelations between the structures at syntactic, semantic and pragmatic level. In Kosslyn's framework, the analysis at the syntactic level investigates graph components as syntactic constituents, without taking into account what they represent. The semantic analysis focuses on configurations of graph components and their relations, as well as what they represent. The analysis at the pragmatic level is concerned with the message that is conveyed by graph components. Kosslyn (1989) offered acceptability principles and guidelines for appropriate design of graphical representations. The analytical scheme offered by Kosslyn provided detailed information and sample analyses for application of the acceptability principles at the syntactic and semantic levels.

Pinker (1990) described graph comprehension in terms of the interaction of information induced by perceptual information sources and conceptual information sources. The perceptually induced information, such as shape and position of graph line segments, are transformed into *visual array* (cf. the primal sketch and 2½ dimensional sketch of Marr, 1982). The visual array is transformed into *visual*

an abstract level of concepts and simple spatial schemas … spatial context of communication can influence the production and comprehension of language." (p. 55). This approach is also compatible to Jackendoff's (1996, 2002) representational modularity hypothesis (see Chapter 2).

description by employing visual encoding processes (cf. visual routines, Ullman, 1984). The visual description represents information about relative spatial positions of graphical and textual entities.[12] The visual description is then transformed into internal conceptual and spatial representations via instantiation of the *general graph schema*. Whereas visual encoding corresponds to the phonological, morphological and syntactic stages of language comprehension, graph schema instantiation corresponds to the semantic and pragmatic stages (see Chapter 2 for an overview of Pinker's graph comprehension architecture and Chapter 4 for a cognitive architecture for multimodal comprehension of graph-text constellations).

The underlying idea in both studies is to transfer the widely used analysis of language in terms of syntactic, semantic and pragmatic level to the analysis of graphs. Nevertheless, neither Kosslyn (1989) nor Pinker (1990) extended their approach for the analysis of the relationship between overt linguistic entities and graphical entities as external representations. Kosslyn (1989) classified alphanumerical constituents of graphs as "labels"; however, he did not investigate the relationship between linguistic entities and the corresponding graphical entities. This dissertation extends Kosslyn's (1989) and Pinker's (1990) approach for the investigation of the relationship between graphical entities and linguistic entities in graph-text constellations.

Although the previous research in cognitive and educational psychology has obtained many fundamental insights in multimodal comprehension and learning (Glenberg & Langston, 1992; Glenberg & McDaniel, 1992; Hegarty & Just, 1993; Mayer, 2001/2009, 2005a; Schnotz, 2005), the specific domain of research has been constellations of verbal entities and pictorial illustrations rather than graph-text constellations with a few exceptions (e.g., Mautone & Mayer, 2007). Accordingly, those models and experiments have seldom focused on the analytical level of the semantics and pragmatics of referring expressions.

The relationship between linguistic entities and graphical entities in graph-text constellations exhibit different characteristics than the relationship between linguistic entities and pictorial entities in picture-text constellations. Both comprehension of graph-text constellations and comprehension of picture-text constellations are perceptually grounded. However, pictorial illustrations are external representations that directly represent the external world whereas graphs are external representations that represent the external world by means of mediating mental conceptualizations. In other words, graphs are externalizations of our mental conceptualization of the external world.

[12] In this dissertation, the term *textual entity* is used for *linguistic entities* in written language. Since the focus of the dissertation is written language rather than spoken language, the term linguistic entity is used to mean textual entities. The term *spoken language* is explicitly used where needed.

This difference between pictorial illustrations and graphs leads to significant differences between their relationships to accompanying linguistic entities such as verbal descriptions, where a verbal description can be seen as a specific instance of multimodal communication or problem solving. For instance, in graph-text constellations, a linguistic entity such as 'the peak' can be interpreted as a reference to a peak in the graph or as a reference to a peak in the domain of the graph (e.g., a peak in temperature) via the mediating link between the linguistic entity 'the peak' and the peak in the graph. In other words, humans construct different types of inter-modal (or cross-modal) relations (e.g., coreference, coherence etc.) in multimodal communication and problem solving by means of graph-text constellations compared to the ones in multimodal communication and problem solving by means of picture-text constellations. In particular, comprehension and generation of graph-text constellations require the construction of referential links between the layer of graphical entities, the layer of linguistic entities and the layer of domain entities by means of the graph schema knowledge, whereas the relationship between the layer of pictorial entities and the layer of domain entities is directly perceived—rather than being constructed by cognitive effort—in picture-text constellations.

In summary, human multimodal communication and problem solving exhibit peculiar characteristics that should be considered in the design of intelligent systems that have the capability for producing and interpreting graph-text constellations and the capability for supporting end-users during their course of designing and producing graph-text constellations. Firstly, human visual system is characterized by the interaction of bottom-up, perceptual processes and top-down, conceptual processes. Secondly, human cognition of space has qualitative properties rather than the metric properties obtained by the Euclidian or Cartesian analysis of the space. Finally, there are modality-specific differences between different types of depictive external representations. In particular, the differences between graphs and pictorial illustrations result in different construction of referential links between the layers of linguistic, depictive and domain entities in graph-text constellations compared to picture-text constellations, thus leading to different *layers of multimodality*. The layers of multimodality in graph-text constellations are discussed in more detail below.

1.5 LAYERS OF MULTIMODALITY IN GRAPH-TEXT CONSTELLATIONS

One of the most frequently cited arguments for combining representational modalities is the *argument of division of labor between representational modalities* (Habel & Acartürk, 2007). It can be exemplified by a characterization given in the *Publication Manual* of the American Psychological Association (APA):

> Tables are often preferred for the presentation of quantitative data in archival journals because they provide exact information; figures typically require the reader to estimate

> values. On the other hand, figures convey at a quick glance an overall pattern of results. They are especially useful in describing an interaction—or lack thereof—and nonlinear relations. A well-prepared figure can also convey structural or pictorial concepts more efficiently than can text. (5th ed., 2001, p. 176)

Although the main focus of interest in this characterization is scientific texts in the domain of psychological research, the characterization is also applicable to multimodal constellations in non-scientific journals, newspapers, so forth. The construction of referential links between the layers of a multimodal constellation is central to analyzing the division of labor between representational modalities. Figure 1.6 shows a schematic representation of multimodal constellations, and the structure of different types of referential links and their composition in a multimodal constellation.

(a)

(b)

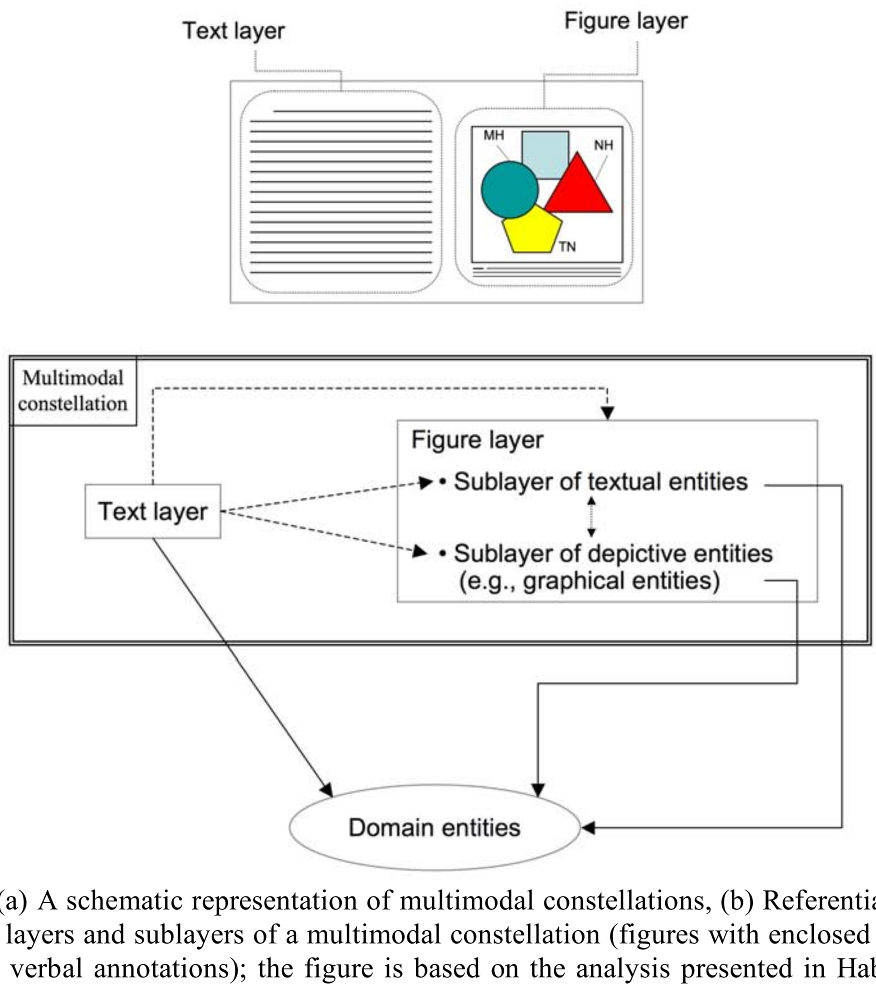


Figure 1.6. (a) A schematic representation of multimodal constellations, (b) Referential links between the layers and sublayers of a multimodal constellation (figures with enclosed textual entities, e.g. verbal annotations); the figure is based on the analysis presented in Habel and Acartürk (2007).

The schematic representation shown in Figure 1.6a is the typical layout of a multimodal constellation. The term *text layer* is used for textual entities in the accompanying text, such as the textual entities in the paragraphs of a multimodal constellation. The term *figure layer* is used for the figure, which is usually separated by a frame from the text layer. In multimodal constellations, figures are usually presented with enclosed textual entities (e.g., as in verbally annotated graphs). For instance, the figure layer in a graph-text constellation includes the graph (proper) and graph-related verbal information such as the graph title, axis labels and verbal annotations. In terms of the terminology of external representations, the term *sublayer of depictive entities* is used for the depictive representations in the figure layer; and the term *sublayer of textual entities* is used for the symbolic representations in the figure layer. The arrows in Figure 1.6b show *referential relations* (alternatively, *referential links*) between the layers of a multimodal constellation.

A systematic investigation of the construction of referential links by humans during the course of multimodal comprehension is fundamental for the study of communication and problem solving by means of multimodal constellations. From a linguistic point of view, the process of *referring* is the core of comprehension.[13] It is constituted by a *referential expression* that refers to an *entity* in the domain of discourse. Based on this, a reader has to establish *coreference*, the backbone of text coherence, by employing internal representations that mediate between language and the domain of discourse. In Figure 1.6b, the arrow from the text layer to domain entities shows this referential relation.

An integrated interpretation of a multimodal constellation requires the construction of additional types of referential relations by the reader. Foremost, linguistic expressions such as ‘as shown in Figure 1’ refer to the figure layer as a representational whole. The arrow from the text layer to the figure layer shows this referential relation. Moreover, there exist referential links between depictive entities in the figure layer and the layer of domain entities. For instance, a peak in the graph corresponds to a highest value of the measured domain entity. The arrow from the sublayer of depictive entities to domain entities shows this referential link.

A particular word or phrase in the text layer, such as ‘the peak’ can be interpreted as a peak in the graph or as a peak in the domain, thus resulting in an explicit language-to-graph reference (to the sublayer of depictive entities in the figure layer) or in an implicit language-to-domain reference via a mediating language-to-graph link. Previous studies on verbal description of sketch maps, which is a specific instance of multimodal constellation, show that the figure layer and the layer of domain entities

[13] Comprehension and generation of multimodal constellations are two facets of multimodal communication and problem solving. The focus in this dissertation is the comprehension perspective; however, the approach is also applicable to generation of multimodal constellations by humans as well as by intelligent artificial systems (Habel & Acartürk, 2007).

can be simultaneously described by humans, depending on the experimental settings of the verbalization task (Tappe & Habel, 1998). In graph-text constellations, these two alternative reference relations can be characterized as 'finding a peak in the accompanying graph' vs. 'constructing a peak in the course of domain entities'. Accordingly, different types of knowledge interact for the construction of referential links in multimodal comprehension of graph-text constellations: general knowledge of graphs (viz., graph schemata), which is used to detect gestalt atoms in graphs (e.g., the representation of increase and decrease processes etc.), and world knowledge of domain entities. The construction of referential links in graph-text constellations is the main topic of investigation in the following chapters.

The terms *multimodal constellation* and *graph-text constellation* are used in this dissertation instead of the terms *multimodal document* and *graph-text document* since the term *constellation* covers a broader range of graph-text combinations compared to the term *document*. For instance, web pages, posters, and maps with verbal annotations are multimodal *constellations* rather than multimodal *documents*. In particular, verbally annotated graphs exemplify a frequently used type of graph-text constellation that is not a graph-text document. Rather, verbally annotated graphs (as a specific subcase of combination of graphs and language) are used in graph-text constellations with a separate text-layer–figure-layer layout. This suggests that beyond the—traditionally discussed—type of *text-layer–figure-layer multimodality* there is a second kind, namely the *figure-internal multimodality* of depictive representations and figure-internal text, e.g. labeling of axes, verbal annotations to parts of graph lines or bars, and so on. Graph-internal multimodality and diagram-internal multimodality are specific cases of figure-internal multimodality (Acartürk, Habel, & Cagiltay, 2008). In graphs, axis labels present information about the domain of the graph; verbal annotations present information about values, as well as events and processes represented by the graphical entities. If the figure layer involves such enclosed textual entities the reader has to construct the referential links between the sublayer of textual entities and the sublayer of depictive entities within the figure layer. Such verbal information within the figure layer mediates the construction of referential links between the text layer and the sublayer of depictive entities. Verbal annotations are the frequently used entities that exemplify one case of graph-internal multimodality. Figure 1.6b shows that both types of multimodality can be involved in multimodal comprehension (see Chapter 7 for empirical investigations on verbally annotated graphs).

In summary, the analysis of the construction of referential links between the layers of a multimodal constellation (namely, the text layer, the figure layer, and the layer of domain entities) is fundamental to a systematic analysis of multimodal communication and problem solving. Moreover, in addition to the text-layer–figure-layer multimodality, figure-internal multimodality of depictive entities and textual entities within the figure layer is necessary for the systematic analysis. This dissertation investigates humans' construction of referential links between the layers of graph-text constellations with respect to both types of multimodality.

1.6 SUMMARY

We use symbolic and depictive external representations and their constellations, namely multimodal constellations, as cognitive means to accomplish high-level problem solving tasks and communication in daily life. In the past several decades of research in cognitive science, the study of external representations has brought a complementary perspective to the study of internal representations by showing that the information processing tasks are distributed across internal and external representations (cf. the distributed cognition approach, Zhang & Norman, 1994; Hutchins, 1995, among many others). The studies have shown that beyond their function as memory aids, external representations facilitate one's own ideas by providing perceptual input, which gives access to knowledge and skills that are otherwise unavailable to us (Reisberg, 1987; Cox, 1999).

It has already been recognized that particular types of external representations have advantages over the others in terms of computational efficiency when processed by humans (Larkin & Simon, 1987). Nevertheless, modality-specific differences between different types of external representation have not been well investigated in the current state of the art.

The focus of research in this dissertation is statistical graphs, in particular line graphs that represent change of a quantitative measurement over time. Graphs comprise a distinct class of depictive representations compared to pictorial illustrations, visual scenes and diagrams. Graphs are externalizations of our mental conceptualization of the world rather than being direct representations of the external world. They are human artifacts that are generated and interpreted according to a set of rules and conventions. This implies that graphs are, like language, grammatical means of problem solving and communication. Accordingly, they can be analyzed at syntactic, semantic and pragmatic level. This characteristic of graphs has been investigated from a general psychological perspective (Kosslyn, 1989; Pinker, 1990). However, a systematic investigation of the relationship between graphical entities and linguistic entities has not been performed.

Graphs, like other depictive representations, have certain limitations caused by their implicit nature. In particular, non-expert people experience perceptual difficulties during early stages of information extraction from line graphs. Therefore, graphs are often used to augment language in graph-text constellations rather than being used in isolated forms. On the other hand, despite their widespread use in daily-life settings, available computer technologies offer limited capacity to augment graphs by language.

The design and development of intelligent software tools that support the design and production of multimodal constellations by end-users should be grounded in the investigation of those computational procedures in humans as natural cognitive systems and the capabilities and limitations of artificial cognitive systems. Human multimodal communication and problem solving have peculiar computational characteristics that are recently beyond the capabilities of artificial cognitive systems. Firstly, human visual information processing is characterized by the interaction of bottom-up, perceptual processes and top-down, conceptual processes. Secondly, spatial cognition by humans has qualitative properties that are difficult to analyze in terms of the metrical analysis of space, such as the Euclidian or Cartesian analysis of the space. Finally, modality-specific differences between depictive representations lead to the construction of different types of referential links between linguistic entities, depictive entities, and domain entities in multimodal constellations (cf. the layers of multimodality). HCI, within the framework of design as construction of representations (Visser, 2006), is the domain of research where artificial and natural cognitive systems interact. This dissertation presents a systematic investigation of graph-text constellations at the interface between natural and artificial cognitive systems. It proposes a theoretical framework for the integration of information induced by different representational modalities, presents experimental studies that investigated how humans perform the integration, and based on the empirical findings, it suggests design guidelines for the development of intelligent software tools that have the capability to support end-users during their course of designing and producing graph-text constellations.

The dissertation is composed of four sections. Section 1 includes two chapters: In Chapter 2, graphs and language are discussed within the context of the previous studies on graph comprehension, language comprehension, and multimodal constellations in relevant disciplines. In Chapter 3, eye tracking paradigm, i.e. the analysis of eye movement recordings of humans, is presented as a research paradigm for a systematic investigation of graph-text constellations; in particular, for the investigation of the construction of referential links between linguistic entities and graphical entities. In Chapter 4, the theoretical framework of *the layers of multimodality* is extended for developing a cognitive architecture for multimodal comprehension of graph-text constellations. Section 2 includes four chapters: Chapter 5 discusses the relevant factors in the empirical investigations that are presented in subsequent chapters of Section 2. Chapter 6, Chapter 7, and Chapter 8 present empirical investigations on graph-text constellations. In particular, Chapter 6 presents empirical investigations on the construction of coreference relations by means of the graph schema and the role of syntactic complexity in graph-text constellations with a separate text-layer–figure-layer layout. Chapter 7 presents empirical investigations on the construction of coherence relations by means of causal attribution and temporal aspect in verbally annotated graphs, i.e. graph-internal multimodality. Chapter 8 presents empirical investigations on the construction of coherence relations by means of scalar congruency between graphical and textual entities in graph-text constellations. Section 3 is composed of two chapters: Chapter 9 presents a GOMS-style cognitive processing model that investigates a specific case of multimodal comprehension of graph-text constellations—multimodal comprehension of spatial prepositional phrases—within the framework of the architecture presented in Chapter 4. Chapter 10 presents guidelines and principles for appropriate design of graph-text constellations and for the design of intelligent software systems that have the capability to support non-expert end-users during their course of designing and producing graph-text constellations. Section 4 concludes the dissertation with Chapter 11.

2 Graphs and Language as Means of Reasoning and Communication

This chapter presents the investigation of graph-text constellations as means for reasoning and communication within the context of the previous studies conducted in relevant disciplines. In 2.1, graphs are discussed as a distinct class of depictive representation. An overview of the previous studies in graph comprehension in 2.2 is followed by a reinterpretation of graphs as quasi-symbolic depictive representations with grammatical structure and a reinterpretation of graph comprehension as interaction between bottom-up, perceptually driven and top-down processing of information in 2.3. Interaction between language and visual percepts is discussed in 2.4 within the framework of the representational modularity hypothesis (Jackendoff, 1996, 2002). In 2.5, interaction between graphical entities and textual entities in graph-text constellations is investigated in terms of the interaction between entities at a conceptual level, mediated by the knowledge of graph schema. In 2.6, the construction of coreference links between graphical and textual entities is further extended to involve coherence relations; in particular, the construction of coherence relations by means of causal attribution and the construction of coherence relations by means of scalar congruency between graphical and textual entities are discussed. The experimental investigations presented in the following chapters are introduced shortly with respect to their relationship to the presented approach.

2.1 GRAPHS AS A DISTINCT CLASS OF DEPICTIVE REPRESENTATION

Graphs are cognitive artifacts. Their development goes back to the eighteenth century. Line graph that plots the change of a quantitative measurement over time (viz., time-series graph) is the earliest type of graph with rectangular coordinates. Time-series graph is also the most common type of graph today (Cleveland & McGill, 1985; Butler, 1993; Peden & Hausmann, 2000; Spence, 2005). According to Tufte (1983), the use of time-series graphs in scientific writings started about in the 1700s. Time-series graphs in rectangular Cartesian coordinates appeared the first time in the works of W. Playfair (1786) as “lineal arithmetic”, in J. H. Lambert’s (1779) writings as “Figuren”, and in J. Watt’s works as “diagrams” (see Hankins, 1999, for a historical review).[1] Sample graphs from those early studies can be seen in Appendix A.1.

Despite the introduction of graphs in the early eighteenth century, researchers did not welcome graphs until the twentieth century. Tufte (1983) states that statistical graphs were seen as tools for lying and cheating people until Tukey's works were published in the 1960s. Funkhouser (1937) presents the following review of criticisms by the scientists of the nineteenth century:

> However, a number of … statisticians looked with disfavor upon this method of handling statistical data. Jacques Peuchet wrote in 1805, "No one will ever believe that such methods [graphic] can serve any useful purpose in the study of statistics. They are but plays of the imagination as foreign to this science as the details of natural history or of topography with which some writers unfortunately wish to embellish it". P.A. Dufau also offered an adverse criticism, "We have renounced the use of graphic procedures employed by many statisticians because they appear to us not to have sufficient rigor". … Carl Knies expressed his disapproval of "die bildliche Methode" by saying, "Outside of its use as a pedagogic means, it is only a plaything without importance." (p. 295) [2]

The skepticism of the early scientists about graphs is partially justified by the implicit nature of depictive representations and perceptual difficulties in graph comprehension, as discussed in Chapter 1. However, a competent use of graphs and diagrams facilitates scientific reasoning (Larkin & Simon, 1987; Cheng & Simon, 1995; Catley, 2006). As external representations, graphs provide the relevant input for the perceptual processes that give access to our spatial knowledge and skills. Visually representing abstract relations between concepts, graphs give insight to the reader that cannot be easily captured by alternative forms of representations. In particular, tables are generally used to represent precise information, whereas graphs are more effective than tables for representing relations between variables since graphs have a better potential to exploit visual information processing in humans (cf. Gestalt principles, Wertheimer 1923/1938; Köhler, 1923). As long as there is no perceptual cue, such as a change in the number of digits, it is difficult to interpret trends from tables. Moreover, the presentation of data by means of graphs can reveal

[1] Earlier researchers, such as C. Huygens (1629-1695) and E. Halley (1646-1742) used non-time-series line graphs that depicted the relationship between two measured quantities. The development of time-series graphs, however, is often attributed to William Playfair; see Spence (2000, 2005), Friendly and Denis (2005), Wainer and Velleman (2001), and Beniger and Robyn (1978) for a review of the history of statistical graphics; Fienberg (1979) for a history of graphical methods in statistics; and Costigan-Eaves and MacDonald-Ross (1990) for a review William Playfair's works.

[2] Joseph Peuchet, *Statistique élémentaire de la France* (Paris, 1805), p. 33; P. A. Dufau, *Traité de statistique; ou, Théorie de l'étude des lois* (Paris, 1840), p. 141; and Carl Knies, *Die Statistik als selbständige Wissenschaft* (Kassel, 1850), p. 114: all quoted from H. Gray Funkhouser, Historical Development of the Graphical Representation of Statistical Data, *Osiris*, 1st Ser., 1937, 3:269-404, on p. 295.

underlying trends and anomalies that are not always easily captured by statistical analysis. Levitin (2002) suggests that a qualitative analysis by graphing data should precede statistical analysis in experimental research:

> The first thing one should do with experimental data is to graph them in a way that clarifies the relation between the data and the hypothesis. Forget about statistical significance testing—what does the pattern of data suggest? Graph everything you can think of—individual subject data, subject averages, averages across conditions—and see what patterns emerge. Roger Shepard has pointed out that the human brain is not very adept at scanning a table of numbers and picking out patterns, but is much better at picking out patterns in a visual display. … Once you have made all your graphs, look them over for interesting patterns and effects. Try to get a feel for what you have found, and understand how the data relate to your hypotheses and your experimental design. A well-formed graph can make a finding easy to understand and evaluate far better than a dry recitation of numbers and statistical tests can do. (p. 128-129)

Loftus (1993) goes so far as to claim that "hypothesis testing is overrated, overused, and practically useless as a means of illuminating what the data in some experiment are trying to tell us. … graphical presentation methods are a much better way to provide such illumination, particularly given the ease with which present computer technology allows such methods to be implemented" (p. 250).

Despite their widespread use, there are limited survey data about the use of graphs in printed and online material. The available data for graph use in the domain of psychology show that the use of data graphs in psychology journals has been increasing since 1940s. In particular, Butler (1993) reported a survey of "pictures" (e.g., pictures of people, animals, equipment etc.), "data graphs" (e.g., line charts, bar charts, scatter plots, maps etc.), and "conceptual diagrams" (flowcharts, network diagrams, Venn diagrams etc.) in introductory psychology textbooks (a total of 7879 pages) and journals (a total of 6534 pages) that were published between 1939 and the 1990s. Butler's survey shows that in the introductory psychology textbooks, pictures were the most frequently used type of depictive representation, which was followed by data graphs and conceptual diagrams. Additionally, the number of pictures in the books increased continuously between 1939 and the 1990s, whereas data graphs were more frequently used in the in the 1960s than all other years (partly due to the prevalence of chart recordings of animals between the 50s and 70s); there was no major change in the use of conceptual graphs. The results of the survey for the psychology journal articles reflected a different pattern. Data graphs were more frequently used than pictures and conceptual diagrams in the journals. The use of data graphs increased continuously between 1939 and the 1990s. The use of pictures in the journal articles increased after the 1960s. The use of conceptual diagrams did not show a major change between 1939 and the 1990s.[3] Recently, accompanied by the development of software tools for designing and producing graphs, the use of data graphs has been increasing in non-scientific settings such as newspaper articles and user blogs on WWW, as well as in scientific publications. However, research on depictive representations—in particular, research on specific types of depictive representations such as graphs—is not in a mature state (Scaife & Rogers, 1996). The class of depictive representations is investigated with respect to its subclasses below.

The classification of depictive representations into pictures, data graphs, and conceptual graphs in Butler's (1993) survey has been used independently by other researchers. Although there is no universally accepted taxonomy of depictive representations (Ainsworth, 2006), graphs are often classified as a distinct type of depictive representation. Myers (1988) offered a set of categories ordered with their abstract properties: *photographs*, *drawings*, *maps (of places or of bodies)*, and *graphs*, *models* and *tables*. Hegarty et al. (1991) offered a taxonomy of "diagrams", i.e. the graphical representations in scientific and technical material by their definition. The taxonomy included three classes: *iconic diagrams*, *schematic diagrams*, and *charts and graphs*. In iconic diagrams—for instance, in photographs and line drawings of objects—the spatial relations between the depiction and the represented entities are isomorphic. In schematic diagrams—for instance, in flow charts, organizational charts and linguistic tree diagrams—the spatial relations between graphical depictions and represented entities are not isomorphic. Instead, functional features are at the foreground in schematic diagrams. On the other hand, charts and graphs are different from iconic diagrams and schematic diagrams in the sense that charts and graphs represent quantitative information. In charts and graphs, the mapping between the representation and the represented entities is between the quantifiable attributes of graphs and the quantifiable attributes of the referents.

In Bernsen's (1994) taxonomy of generic unimodal modalities, "static graphs" are classified as one type of modality among 28 types. Static graphs, exemplified by bar charts, pie charts and dot charts, are classified as a distinct type of modality compared to diagrammatic pictures (e.g., maps, cartoons) and non-diagrammatic pictures (e.g., real world representations or pictures).

Graphs are classified as a distinct type of depictive representation not only by researchers but also by readers. Lohse, Rueter, Biolsi and Walker (1990) and Lohse,

[3] It should also be noted that scientific disciplines do not use graphs equally. In an earlier study, Cleveland (1984) conducted a survey of fifty articles from a sample set of journals in seven different disciplines, which were published in 1980-1981. He found that researchers in natural science used graphs more frequently than researchers in mathematical and social sciences. This is partly due to more reliance on inferential statistics in social sciences compared to natural sciences (Best, Smith & Stubbs, 2001).

Biolsi, Walker and Rueter (1994) conducted exploratory analyses for classification of "visual representations". In both studies, they asked the participants to name, rate and sort a set of depictive representations according to a set of dimensions based on perceptual similarity such as concrete-abstract, temporal-nontemporal, numeric-nonnumeric and so on. As a result, Lohse et al. (1990) obtained five major clusters: *graphs and tables*, *maps*, *diagrams*, *networks*, and *icons*. Lohse et al. (1994) obtained a more detailed classification that included eleven major clusters: *graphs*, *tables*, *graphical tables*, *time charts*, *networks*, *structure diagrams*, *process diagrams*, *maps*, *cartograms*, *icons*, and *pictures*. Despite their limitations, such as the limited number of stimuli and perceptual similarity as having been the basis of classification, as acknowledged by the authors of the studies, these two studies show that graphs are recognized as a distinct class of depictive representations by readers as well as by researchers.

These findings in survey studies present support for the claim in Chapter 1, which stated that graphs exhibit different properties with respect to the relationship between representing entities and represented entities compared to pictorial illustrations.

In summary, graphs, being cognitive artifacts developed by humans, comprise a distinct class of depictive representation. Although graphs were introduced in the eighteenth century, the advantages of a competent use of graphs in scientific reasoning have been recognized by scientific disciplines in the past century. Recently, the importance of graph literacy in the modern society has been acknowledged by the organizational institutions of education. The U.S. National Council of Teachers of Mathematics (NCTM, 2000) states that students should learn to use such visual tools as graphs, diagrams, and maps during middle school; and during high school they should develop facility in using them (Novick, 2006). On the other hand, despite their recognized importance, modality-specific research on graphs and graph comprehension has been limited. In particular, empirical research on subtypes of graphs and graph-text constellations has been scarce. In 2.2, previous research on interpretation of graphs by humans is reviewed in connection with the research presented in this dissertation.

2.2 AN OVERVIEW OF PREVIOUS STUDIES ON INTERPRETATION OF GRAPHS BY HUMANS

The initial studies on interpretation of graphs by humans were conducted in the 1920s and 1930s in the domains of psychology and statistics. These studies were, for the most part, exploratory studies on different graph types and their role in learning. In particular, Eells (1926) reported an experimental study for comparing "circle diagrams to show component parts" (i.e., pie charts) and bar diagrams. The purpose of the study was to defend the use of circle diagrams to show component parts against the criticisms of the statisticians. Washburne (1927) conducted an experimental investigation of graphic, tabular and textual representation of quantitative material with school children. Washburne presented the material in tables, bar graphs, line graphs and text, and asked the participants a set of posttest questions. He found that different representations of the data fostered learning of different types of fact. Based on the findings, Washburne proposed a set of "general rules for the appropriate use of the various graphic, tabular, and textual forms" (p. 375), in other words *design principles and guidelines* in the recent terminology of HCI.[4] Graham (1937) investigated the factors that affect "observation" of bar graphs. He investigated the influence of a set of spatial properties such as nearness or remoteness from the scale axes, horizontal or vertical presentation, the spacing between the bars, and the width and length of the bars. He found illusory tendencies in participants' observation of bar graphs such as overestimating short bars that were far from the scale axes, narrow bars, and coarse scale units. Funkhouser (1937) presented an early and comprehensive review of the historical development of graphical representation of statistical data.

The study of the role of graphs in learning, both in isolated form and within linguistic context, has continued to date: Vernon (1953) reported that presenting graphs with text is advantageous as long as the readers are educated graph readers. Feliciano, Powers, and Kearl (1963) found that the method of presentation of statistical information (i.e., short tables, long tables, bar graphs) was a significant factor in learning. They reported a set of presentation methods ordered by their effectiveness in learning (from the most effective to the least effective): horizontal grouped bar graph reinforced with text, short and simple table reinforced with text, graph without textual reinforcement, short table without textual reinforcement. Culbertson and Powers (1959) found that using labels in graphs was more effective than using labels in a separately positioned legend in learning with graphs. In a later study, Milroy and Poulton (1978) compared the three methods of labeling graphs: direct labeling on the graph line, inserting a key below the graph line, and inserting a key in the position of graph caption. They reported that direct labeling resulted in improved reading speed compared to other two methods. More recently, Tabachneck-Schijf et al. (1997) emphasized non-experts' difficulties with multiple representations. Paoletti (2004) reported that the integration of visual information (in particular, diagrams and graphs) and verbal information was hardly achieved by students; they observed that students often neglected illustrations, relying on information provided by the text. However, Eilam and Poyas (2008) reported that learning with statistical graphs and text resulted in better retention and transfer than learning with text-only homework material.[5]

[4] See Chapter 10 of the dissertation for design principles and guidelines for graph-text constellations.

[5] Retention and transfer are often used to measure learning outcomes. Retention measures remembering; transfer measures understanding (Mayer 2001/2009).

The studies on learning with graphs and graph-text constellations, and at a broader level the studies on depiction-text constellations have resulted in divergent findings. As a reaction to the divergent findings, several theories of multimodal learning have been developed since the last decade for reaching a unified theoretical framework for the analysis of multimodal constellations (e.g., the cognitive theory of multimedia learning by Mayer, 2001/2009, 2005b, the cognitive load theory by Sweller, van Merriënboer, & Paas, 1998, and the integrated model of text and picture comprehension by Schnotz, 2005). However, due to the abundance of different types of illustrations, the illustration-type-specific differences have been seldom considered in detail in educational psychology research. The theoretical approaches have focused on picture-text constellations or diagram-text constellations rather than graph-text constellations; modality-specific research on graphs and graph-text constellations is lacking.[6] Therefore, implications of the available theories of multimodal learning with picture-text constellations on comprehension and learning with graph-text constellations are not clear at this stage.

Initial studies of graph interpretation by humans within the framework of the information processing approach to human mind appeared in the 1980s (cf. the computational theory of mind, see Chapter 1). Most of the initial research performed in the 1980s investigated *perceptual processes* that are carried out during the initial stages of interpretation of graphs by humans. Cleveland and McGill (1984, 1985) defined graph perception as "the visual decoding of the quantitative and qualitative information encoded on graphs" (1985, p. 829). Their definition of the term *visual decoding* is "the instantaneous perception of the visual field that comes without apparent mental effort" (p. 828). They identified a set of elementary perceptual tasks for information extraction from graphs. From the most accurate to the least accurate, these tasks were perceiving position along a common scale, position on identical but nonaligned scales, length, angle and slope, area, volume, density and color saturation, and color hue. They suggested a guideline for graph construction, which states that "graphs should employ elementary tasks as high in the ordering as possible" (1984; p. 531). In a later study, Cleveland (1993) offered a framework for a three-stage visual decoding in graphs: (1) quantitative/categorical and physical/scale classification of information on display, (2) visual processes of graphical displays, namely *pattern perception and projection* and *table-lookup*, (3) visual operations that carry out visual processing of graphical displays, namely *detection*, *assembly*, and *estimation* for pattern perception and projection, and *interpolation* and *matching* for table-lookup. Cleveland proposed that the study of speed and the accuracy of these visual operations are crucial for the investigation of the effectiveness of information display method.

[6] This is astonishing, since thorough discussions about types of illustrations started in the 1980s (Alesandrini, 1984; Peeck, 1987; Winn, 1987).

Later studies investigated further processes of graph interpretation, often called the processes of graph comprehension, though the border between perception and comprehension was set differently by these studies. In particular, Simkin and Hastie (1987) investigated how the interaction between graph type and judgment (task) type affected quantitative information extraction from bar graphs and pie charts. They used a set of operators—elementary information processes—for the analysis of the interaction: *anchoring*, *scanning*, *projection*, *superimposition*, and *detection*. The elementary information processes were based on Pinker's (1981) graph perception processes, Kosslyn's (1980) mental image processing, and Ullman's (1984) visual routines approach to visual scene perception. The difference between Simkin and Hastie's (1987) approach and the other approaches of the time is that the former went beyond perceptual processes by proposing a cognitive framework for comprehension of bar graphs and pie charts. Simkin and Hastie stated that they investigated the mental processes that intervene between early vision and mental representation.

Lohse (1993) developed a software tool—called *UCIE*—for simulation of graph perception. Lohse stated that UCIE was a simulation model for underlying perceptual and cognitive processes during information decoding from graphs. His model was based on assumptions about the logical sequence of eye fixations, short term memory capacity, duration limitations, and difficulty of information extraction in each glance. Based on the assumptions, his model predicted reaction times. Lohse's approach is an early cognitive modeling approach for graph comprehension. More recently, simulation models of graph comprehension in different domains have been developed (e.g., Peebles & Cheng, 2001, 2002, 2003, for a cognitive model of problem solving with Cartesian graphs in ACT-R/PM). Shah and Carpenter (1995) and Carpenter and Shah (1998) identified three processes for the interpretation of line graphs: (1) encoding of the major visual patterns, (2) translation of visual features to conceptual relations, (3) determining the referents of the concepts being quantified and associating those referents to the encoded functions (p. 45). They proposed that an integrated sequence of pattern-recognition, interpretive processes, and integrative processes are carried out during comprehension of line graphs. Lohse (1993) and Carpenter and Shah (1998) are also the two initial studies that used eye tracking as the method of investigation of graph comprehension by humans (see Shah & Hoeffner, 2002, and Shah, Freedman, & Vekiri, 2005, for a review of research on comprehension of quantitative information in visual displays).

In summary, the initial studies from the perspective of the information processing approach were conducted in the 1980s with a specific focus on perceptual processes of graph comprehension. In the past two decades, further processes of graph comprehension have been investigated. Both the studies on perceptual processes of graph comprehension and the studies on further processes of graph comprehension have had a *process-oriented* perspective rather than a *structure-oriented* perspective. In other words, these studies have investigated the processes of graph perception and comprehension rather than structural aspects of graphical entities. One exception is the seminal study of Kosslyn (1989), which emphasized the clear grammatical structure of graphs and presented a structural analysis of charts and graphs from a general psychological perspective. In particular, Kosslyn stated that the processes of graph comprehension could be investigated at syntactic, semantic, and pragmatic level: At the syntactic level, graph components are investigated as syntactic constituents, devoid of what they stand for, whereas the semantic level is concerned with what is represented by a graphical entity and what are the relations between graph components. The pragmatic level is concerned with the message conveyed by the graphical entities.[7] Kosslyn's (1989) analytical approach has not been further advanced by empirical research.

The closest view that considers graphs as representations with grammatical structure was proposed by Pinker (1990). In the graph comprehension architecture presented by Pinker, internal mental representations mediate between graphical entities and domain entities in a similar way to language comprehension where internal representations mediate between linguistic entities and domain entities. Pinker's graph comprehension architecture is discussed below.

2.3 GRAPHS AS QUASI-SYMBOLIC DEPICTIVE REPRESENTATIONS WITH GRAMMATICAL STRUCTURE

As stated in Chapter 1, human visual information processing is characterized by the interaction between perceptual processing of bottom-up information and top-down information processes such as the knowledge of graph schema, world knowledge of domain entities, and the requirement of the task. The most explicit investigation of graph comprehension as the interaction between bottom-up perceptual processes and top-down processes of comprehension was proposed by Pinker (1990).

Pinker (1990) proposed a theory of graph comprehension and a cognitive architecture for the structures and processes that compute the structures during graph comprehension. The information flow diagram for the proposed cognitive architecture is presented in Figure 2.1.

[7] Futrelle (1990) presents a kindred approach for syntactic and semantic analyses of graphs. He analyzes graphs in three stages: Vectorization and OCR for producing geometric and text objects; syntactic analysis (identification of scale lines, labels, numerical value, data points, and their relation to the scale lines); semantic analysis for capturing concepts and other statistical information. The semantic analysis ultimately merges this information with the accompanying text. Futrelle (1999) divides semantic analysis into further substages. Also see Tversky (2004) and Engelhardt (2006, 2007) for kindred but broader-level grammatical approaches to visual representations.

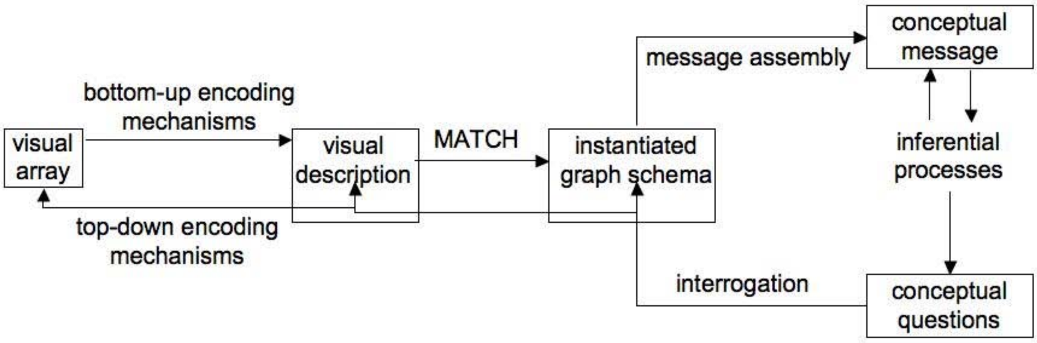


Figure 2.1. Pinker's graph comprehension architecture (redrawn based on Pinker, 1990, p. 104)

In Pinker's theory of graph comprehension, the *visual array* roughly corresponds to primal sketch and 2½ dimensional sketch in Marr's (1982) computational theory of vision. The *visual description* is the structural, symbolic representation of the graph. The visual description specifies the identity of graphical entities and relations among them in terms of propositional (i.e., symbolic) representations. Pinker uses a node-link graphic notation for visual descriptions. In visual description, the variables stand for perceived entities, and the predicates stand for relations among the perceived entities.[8] The processes that create visual description from visual array are called *visual encoding processes*. Ullman (1984) proposes mechanisms (viz., visual routines) that detect the relationships between the perceived entities (i.e., the predicates) in the visual array. To sum up, the visual description describes graph primitives such as lines in line graphs and rectangles in bar graphs, their properties such as color and texture, shape predicates such as orientation, curvature, smoothness (i.e., the physical dimensions of graphical entities), and the spatial relations among the graph primitives.

The visual description representation is transformed into a more abstract representation by the instantiation of the *graph schema*. The graph schema specifies the mapping between physical dimension of graphical entities and the appropriate mathematical scales. The transformation between visual description and instantiated graph schema is governed by the MATCH process that recognizes individual graphs as belonging to a particular type. This representation—called instantiated graph schema in Pinker's architecture—is then transformed into the *conceptual message* by the *message assembly* process that creates a conceptual message out of the instantiated graph schema.

---

[8] Also see Miller and Johnson-Laird (1976) for kindred representation of depictive entities in terms of propositional representations. See Chapter 4 of the dissertation for a discussion of such a language for graph-text constellations, which is based on a language for verbally-instructed route navigation (CRIL, Conceptual Route Instruction Language) proposed by Tschander et al. (2003).

Information flow in the reverse direction takes place if the graph comprehender starts with a *conceptual question*: As a first step, the conceptual question is transformed into the instantiated graph schema by the *interrogation process* that retrieves or encodes new information on the basis of conceptual questions. The conceptual message and the conceptual question are the information that are extracted or assumed to be extracted from the graph. According to Pinker (1990), a set of *inferential processes* applies mathematical and logical inference rules between the conceptual message and the conceptual questions.

In summary, graph comprehension includes a set of processes that transform graphically induced information—i.e., information induced by external graphical representations such as axes of a graph frame, tick marks on the axes, and graph lines—into internal (i.e., conceptual and spatial) representations. The key concept in Pinker's (1990) architecture is the graph schema. The graph schema is the interface that transforms information between the visual description and the conceptual message, and between conceptual questions and the visual description. Without the knowledge of the graph schema, the external visual representations on a page or on screen can be interpreted as lines or geometric figures but not as part of a statistical information graphics. In multimodal comprehension of graph-text constellations, the graph schema provides the basis for coreference construction between graphical and linguistic entities. Before going into the details of coreference construction in graph-text constellations, the interaction between language and visual percepts is discussed below.

2.4 THE INTERACTION BETWEEN LANGUAGE AND VISUAL PERCEPTS

Language comprehension, in its most basic form, includes a set of processes that transform linguistically induced information, i.e. information induced by external linguistic representations such as words, phrases, and sentences, into internal conceptual and spatial representations that mediate between linguistic entities and domain entities (Jackendoff, 2002). These processes include linguistic processes at phonological, syntactic and semantic level. This broad view of language comprehension has been widely accepted by language researchers, though with disagreements on the nature of internal representations and the processes that manipulate those representations.

On the other hand, there is no widely accepted approach to the underlying nature of humans' cross-modal abilities, such as the ability to talk about what we see. The interaction between language and visual percepts has been a matter of debate since the 1980s. As introduced in Chapter 1, language and sensory-motor systems (e.g., vision, audition) were conceived as domain-specific and informationally encapsulated modules by Fodor's (1983) *modularity of mind hypothesis*. In this view, the representations of sensory-motor systems are transduced into symbolic representations, which in turn support higher cognitive functions such as language. Since language and vision are two informationally encapsulated modules, operation of one module does not interfere with the other. In other words, the presence of visual context does not influence the operations of the faculty of language. However, in the last two decades, psycholinguistics has intensively investigated the interaction of—in particular, spoken—language comprehension and visual perception (Ferreira & Tanenhaus, 2007; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995, cf. the visual world paradigm).[9] According to this interactionist view, in contrast to Fodor's (1983) modularity of mind hypothesis, perceptual processes mediate online language comprehension processes. Jackendoff (1996, 2002) proposed the *representational modularity hypothesis*, which states that conceptual and spatial representations as essential parts of the interface between the linguistic system and the sensory systems. The representational modularity hypothesis is discussed in more detail below.

Jackendoff (1987, 1996, 2002) proposes that the mind/brain encodes information in different formats, and there is a module that is responsible for encoding information in each format. This is called the representational modularity hypothesis (see Figure 2.2). The modules in the representational modularity hypothesis are different than Fodor's (1983) input-output modules for cognitive faculties in that in the representational modularity hypothesis the modules specify the individual characteristics of processes. Moreover, there are interface modules between the representational modules. The interface modules are shown by arrows in Figure 2.2.

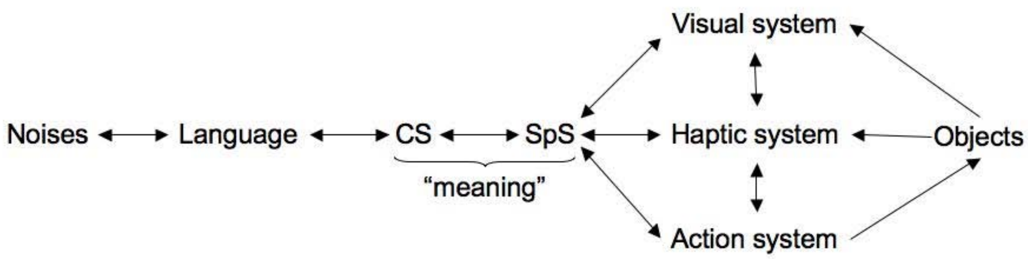


Figure 2.2. A schematic representation of the component modules and the interface modules in the representational modularity hypothesis, redrawn based on Jackendoff (2002, p. 348).

Jackendoff (1983) states that the conceptual structure is the level of mental representation at which linguistic, sensory, and motor information are compatible (cf. the conceptual structure hypothesis). According to Jackendoff,

[9] The term *visual world* is used to mean gazes on a visual display that depict a natural scene, directed by spoken language. The focus of the visual world paradigm is then on utterance mediated eye movements (Knoeferle, Crocker, Scheepers, & Pickering, 2005). Also see Pickering, McElree and Garrod (2004) for a criticism on its potential to provide a transparent window on language processing.

> There must be levels of mental representation at which information conveyed by language is compatible with information from other peripheral systems such as vision, nonverbal audition, smell, kinesthesia, and so forth. If there were no such levels, it would be impossible to use language to report sensory input. We couldn't talk about what we see and hear. (p. 16)

Afterwards, Jackendoff (1987, 1996, 2002) and Landau and Jackendoff (1993) proposed a further structural division of meaning into *conceptual structure* (CS) and *spatial structure* (SpS).[10] The CS encodes aspects of understanding such as category membership and predicate-argument structure, whereas the SpS encodes spatial properties of the physical world such as object shape, and visual object identification and categorization. The CS is algebraic and propositional but not strictly linguistic, whereas the SpS is geometric and quasi-topological but not strictly visual (Jackendoff, 2002, p. 348).[11] The CS is not strictly linguistic for several reasons. First, the CS has a hierarchical structure rather than a linear order, in contrast to syntax. Second, the CS includes LCS (Lexical Conceptual Structure), which distinguishes syntactically similar concepts such as CAT and DOG.[12] Finally, the verbs of syntax are functions and the objects/subjects of syntax are arguments in the CS. Concerning the SpS, it is not strictly visual because it is able to encode spatial layout in different modalities. The SpS representations are image schema representations or mental model representations, rather than image-like representations. In other words, SpS representations are more abstract representations than skeletal images (cf. Marr, 1982; Biederman, 1987).

There is an overlap between CS and SpS in that "the notions of physical object, part-whole relationships, locations, force, and causation have reflexes in both systems. It is these shared components that enable the two systems to communicate with each other through an interface of the usual sort" (Jackendoff, 2002, p. 347). In addition, concerning the spatial terms in language, path (or trajectory) and place (or location) have reflexes both in CS and SpS. The linguistic terms such as *on*, *in* (location) and *through* (path) have conceptual counterparts in the CS and geometric (or quasi-topological) counterparts in the SpS.

---

[10] The early versions of the representational modularity hypothesis (Jackendoff, 1987, 1996) use the term *spatial representation (SR)*, whereas Jackendoff (2002) uses the term *spatial structure (SpS)* instead of SR. Jackendoff (2002) states that the reason for using the term *structure* instead of *representation* is the emphasis of *intentionality* by the term *representation*, which is not intended in the representational modularity hypothesis. The relationship between representation, intentionality, and structure is beyond the scope of this dissertation; the terms representation and structure are used interchangeably.

[11] It should be noted that the term *proposition* is not used for *proposition* in truth-conditional logic. The expressions of propositions in Jackendoff's approach refer to the world as we conceptualize it, rather than to real or possible worlds, in contrast to truth-conditional logic (Jackendoff, 1996, also see Palmer, 1978, on language as the paradigm case of propositional representations).

[12] In this dissertation, concepts are shown by capitalized small letters.

To sum up, the representational modularity hypothesis investigates the cross-modal interaction (i.e., inter-modal interaction) between language and visual percepts in terms of the relationship between conceptual structure (CS) and spatial structure (SpS). In this dissertation, the representational modularity framework is taken as the basis for multimodal comprehension of graph-text constellations where graphs are visual percepts with grammatical structure.

Multimodal comprehension of a graph-text constellation is successful as long as the internal representations of the information induced by different modalities result in an integrated and, in principle, a coherent interpretation. Therefore, a systematic investigation of comprehension of graph-text constellations requires the specification of correspondences between the information induced by different representational modalities and internal mental representations, and the specification of coreference relations among internal mental representations. In particular, the relationship between graphical entities and linguistic entities is revealed by the concepts that are accessible both by textual entities (e.g., the term 'peak' in text) and graphical entities (e.g., the peak as a graphical entity) via the graph schema. The graph schema provides a syntactic and combinatory system at the interface between CS and SpS. A detailed investigation of computational aspects of the interaction between graphical and linguistic entities is presented in Chapter 4, where a cognitive architecture for multimodal comprehension of graph-text constellations is presented and in Chapter 10, where a GOMS-style cognitive processing model for multimodal graph-text constellations is presented. In 2.5 below, the interaction between bottom-up and top-down processes in comprehension of graph-text constellations is discussed.

2.5 LINGUISTICALLY GUIDED COMPREHENSION OF GRAPHS: CONSTRUCTION OF COREFERENCE RELATIONS BY MEANS OF THE GRAPH SCHEMA

Linguistic entities in graph-text constellations, in particular paragraphs that accompany graphs, are usually incomplete verbal statements about the domain entities that are represented by the graphical entities. The *completeness* of the verbal statements in graph-text constellations is an outcome of the argument of division of labor between representational modalities (see Chapter 1). Bernsen (1994) states that analog (i.e., depictive) representations "illustrate" linguistic discourse, whereas language "annotates" analog representations.[13] Therefore, linguistic entities bring certain aspects of graphical entities to the foreground in graph-text constellations, whereas leaving the others at the background. In other words, comprehension of

[13] Bernsen (1994) classifies statistical graphs as *analog* (i.e., depictive) representations because graphs represent data according to selected dimensions of interest. In other words, according to Bernsen, graphs represent dimensionality, despite the lack of pictorial similarity to the entities in the represented world.

graphs in graph-text constellations is guided by the linguistic entities that provide top-down task information. The graph schema instantiation plays the role of a knowledge source that provides the basis for the comprehension processes. The information provided by the linguistic entities corresponds to the conceptual questions in Pinker's (1990) graph comprehension architecture.

The processing of perceptually induced information, on the other hand, is also influenced by visual salience of the graphical entities. The research on human perception suggests that humans decompose shapes into parts by using part boundaries, which are often specified by visually salient regions (Hoffman & Singh, 1997; Cohen & Singh, 2007). Visual salience is often determined by a set of *critical points* in the terminology of geometry. These critical points in a closed contour are *maxima*, *minima*, and *points of inflection*. Attneave (1954) stated that the points of maximal curvature are the points where information is concentrated most, along occluding contours (see Feldman & Singh, 2005, for a formal derivation). Expanding the concept of critical points, Freeman (1978) appended *discontinuities* in curvature, *endpoints*, *intersections*, and *points of tangency* to the list. De Winter and Wagemans (2004, 2006) presented a database of observer responses to curvature extrema in part segmentation studies. Their studies showed that observers gave priority to select negative minima for cutting curves. Positive maxima and inflection points followed negative minima. The formal analysis of visually/informationally salient regions also provides a tool for predicting behavioral measures, such as attention and eye movements (Feldman & Singh, 2005). The prediction is that attention and eye movements would be directed to the informative regions of a curve.

These findings have implications for perceptually driven—i.e., not-linguistically-guided—inspection of line graphs. An initial prediction is that in a line graph, the points of maximal curvature would be the points where information is concentrated most, as Attneave (1954) stated for occluded object boundaries. The results of a small-scale study of visually salient regions in line graphs are presented in Appendix A.2. The results of the participants' eye movement recordings suggest that humans employ the geometrically critical points for visual decomposition of graphical entities in line graphs. Moreover, humans inspect the maxima regions more than they inspect the minima regions. A more detailed analysis of visual decomposition of line segments is beyond the scope of this dissertation. The focus in this dissertation is the investigation of the guidance by the linguistic entities by means of the graph schema instantiation in graph-text constellations, rather than a detailed investigation of the role of visual/informational salience of graphical entities. Experimentally, the differences between the inspections of the graphs with different accompanying linguistic phrases would reveal the influence of linguistic guidance on graph comprehension. Eye movement recordings of readers, in principle, should reflect this difference. In Chapter 3, eye movement recording is proposed as a research paradigm for the analysis of graph-text constellations. Chapter 6, Chapter 7, and Chapter 8 present experimental investigations, which show how accompanying linguistic phrases in graph-text constellations influence inspection of graphical entities by human participants. Chapter 9 employs a task analysis method for the development of a GOMS-style model for prediction of readers' eye movements on graphical entities after their reading of linguistic entities. In the following part of this chapter, the investigation of the construction of inter-modal coreference relations between graphical and textual entities in graph-text constellations is extended to the analysis of conceptual relations that are beyond the construction of coreference relations by means of the graph schema.

2.6 CONSTRUCTION OF COREFERENCE RELATIONS IN GRAPH-TEXT CONSTELLATIONS: FURTHER CONCEPTUAL RELATIONS

In the 1970s, text comprehension research focused on the surface structure of the text and semantic meaning structure that is constructed out of the surface properties. In those studies, the surface structure is conceived as the verbatim representation of the text; the surface structure preserves wording and syntactic structure. The semantic meaning structure (viz., the textbase) is conceived as a propositional network where the propositions that share arguments are connected via the mechanism of argument overlap (Kintsch & van Dijk, 1978; Kintsch, 1988). In the 1980s, it was proposed that readers construct a model of the situation presented by the text, i.e. a *coherent* representation of what is described in the text. Further experimental studies have shown that aspects of the situation could be observed in comprehension processes of the readers (Zwaan, Langston, & Graesser; 1995; Johnson-Laird, 1983, van Dijk & Kintsch, 1983; see Zwaan & Radvansky, 1998, for a review). A set of dimensions was proposed as coherence relations constructed during text comprehension. In particular, according to Zwaan et al. (1995) readers monitor and update their current situation model on a number of event indices: temporality, spatiality, protagonist, causality, and intentionality (cf. the *event indexing model*; also see Graesser, Millis, & Zwaan, 1997).

The previous research on text comprehension implies that a systematic analysis of comprehension of graph-text constellations should involve the investigation of the aspects of inter-modal coreference relations that are beyond the construction of coreference relations via the graph schema, namely coherence relations. This investigation corresponds to a level of analysis that is beyond the syntactic and semantic analyses of graphical entities and their relationship with textual entities. This is also the level of analysis where graph-text constellations are used for communicative purposes. In this dissertation, two dimensions of coherence relations are investigated: causal attribution and scalar congruency between graphical and textual entities, as discussed below.

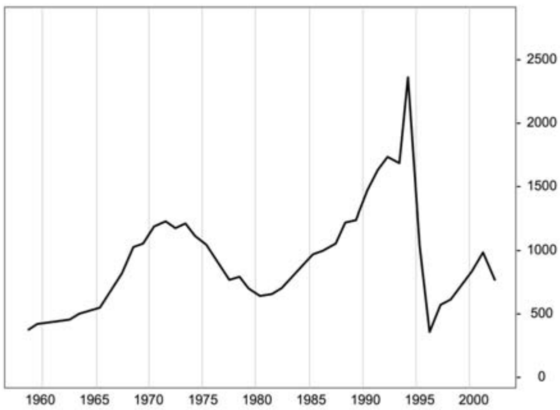
2.6.1 Causal Attribution

In language comprehension, causal attribution (or causal inference) is a relatively well-investigated phenomenon. In the event indexing model for text comprehension (Zwaan et al., 1995), causality is one of the five conceptual dimensions that are monitored by the readers. A break in continuity of these dimensions influences the comprehension of text.

From the perspective of psychology, most of the traditional approaches on causal attribution focus on the role of covariation information, i.e., statistical relations between events. These studies assume that humans make causal attributions by interpreting data points. However, covariation information is not the sole determinant of causal inference. Humans use other cues, such as intervention—i.e., the influence of manipulating certain variables on the effects—and temporal order; causes precede their effects in making causal attributions to events (Lagnado & Sloman, 2006). Moreover, causal attribution can depend on different types of information (Ahn, Kalish, Medin, & Gelman, 1995). The majority of research on causal reasoning and causal models, nevertheless, deals with inferring causality from information provided by experiments or observations of directly perceived (or indirectly mediated) events, and the research on the role of the modality of representation on causal attribution is scant except for a few studies (e.g., Majid, Sanford, & Pickering, 2006).

The conversational character of language and graphs in instructional and scientific documents can influence causal attribution by explicit argumentation or implicit argumentative hints (see Hilton, 1990, for text, and Oestermeier & Hesse, 2000, for language and diagrams). On the other hand, despite its widespread use, the conceptual dimension of causality is not a well-investigated phenomenon in graph-text constellations except for a few studies (e.g., Oestermeier & Hesse, 2000; Green, 2005).

Line graphs present a special case of covariation information and temporal continuity. They are particularly suitable for the investigation of causal attribution since the graphical entities represent processes and events. This is exemplified by a graph-text constellation shown in Figure 2.3, excerpted from the annual report of an ecological study project about moose and wolf population in the Isle Royale.



Moose have been slowly increasing since 1996, when most of the population died of starvation. In the past year, this sustained increase was halted as an outbreak of winter ticks led to higher moose mortality last spring, followed by reduced recruitment in the 2002 cohort. The tick outbreak coincided with a continent-wide moose mortality event that extended from New Hampshire to Alberta, perhaps resulting from an unusually warm autumn in 2001. Another contributing factor was a late and cold spring in 2002. During winter 2002-2003, there were record-low snow depths, air temperature was unusually low, and wind speed was above average.

Figure 2.3. A sample graph-text constellation from "Ecological Studies of Wolves on Isle Royale. Annual Report 2002-2003" by Rolf O. Peterson and John A. Vucetich. School of Forest Resources and Environmental Science Michigan Technological University Houghton, Michigan USA 49931-1295. 31 March 2003. The constellation includes a population trend graph depicting the number of moose in Isle Royale and verbal statements about the population trend. The graph and the paragraph in the figure are placed side by side for the purpose of demonstration. The graph is redrawn based on the original provided in the report (p. 3); the wolf population graph line was omitted for clarity.
http://www.isleroyalewolf.org/wolfhome/home.html, retrieved on October 11, 2009

The graph-text constellation shown in the Figure 2.3 is composed of a population trend graph that depicts the number of moose between 1959 and 2003 in the Isle Royale, and a single paragraph that is excerpted from the report. The text describes a set of domain events such as increases and decreases and their causes such as 'starvation', 'an outbreak of winter ticks', and so on. The events and processes (i.e., increases, decreases and the changes from an increase to a decrease or from a decrease to an increase) that are caused by those events are also referred to by graphical entities. Consequently, multimodal comprehension of the graphical entities and the textual entities in the graph-text constellation result in a coherent interpretation during comprehension.

Moreover, verbal annotations in line graphs—in particular, event annotations that represent information about external events and their relation to the events and processes represented by the graph lines—can facilitate the construction of coreference relations between graphical entities and linguistic entities within the graph-layer of graph-text constellations (cf. the graph-internal multimodality, see Chapter 1). This is exemplified in Figure 2.4, which is an excerpt from another study on the Isle Royale moose-wolf progression by N. C. Heywood.

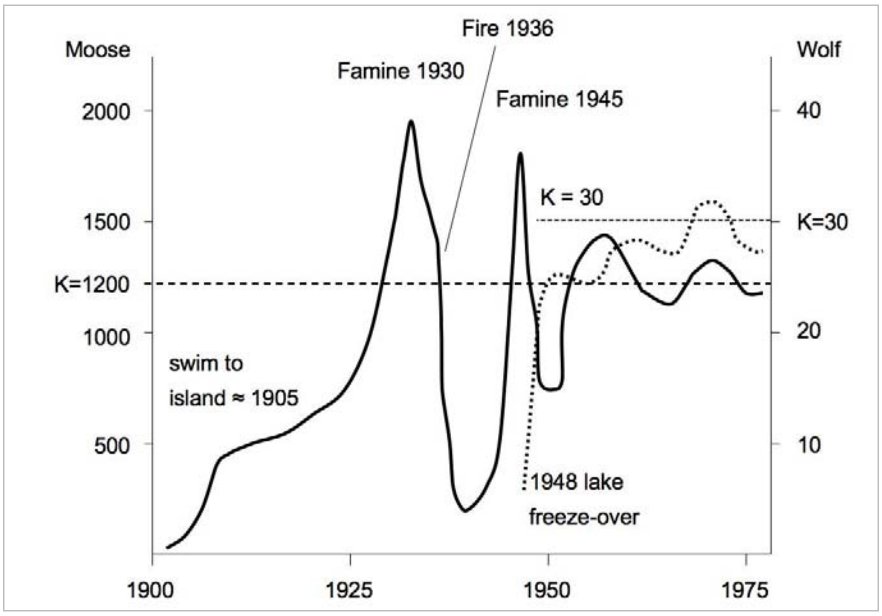


Figure 2.4. A sample verbally annotated line graph: Isle Royale Moose/Wolf Progression. The figure is redrawn based on a line graph used in N. C. Heywood's course material in Biogeography.
http://www.uwsp.edu/geo/faculty/heywood/Geog358/Population/Populate2.htm, retrieved on October 11, 2009). The figure goes back to Harris, A. and Tuttle, E. Geology of National Parks. ISBN 0-8403-2810-9.

In verbally annotated line graphs, the construction of the correspondence between verbal annotations and graphical entities can be induced by spatial contiguity between verbal annotations and the graph line or prominent parts of the graph line (e.g., 'Famine 1930' or '1948 lake freeze-over' in Figure 2.4) or by explicit 'pointers', called 'annotation icons' in this dissertation, (e.g., the straight line between 'Fire 1936' and the graph line). The causal attributions concerning the relationship between the events and processes referred to by the verbal annotations (e.g., 'Famine 1930' and 'Fire 1936') and the events and processes referred to by the graphical entities (e.g., the increases and decreases represented by the graph lines) results in a coherent interpretation during multimodal comprehension of textual and graphical entities. Chapter 7 presents experimental investigations of causal attribution in verbally annotated graphs. In particular, the influence of the position of verbal annotations on participants' judgments concerning the causal attributions between the events and processes represented by the verbal annotation and the events and processes represented by the graphical entities, and the shape of the graph lines on participants' judgments concerning the temporal aspect of the events and processes representation by verbal annotations are investigated.

From the perspective of HCI, design decisions such as the positioning of verbal annotations, and more generally the positioning of linguistic entities in the graph layer present a challenge, in particular for automatic generation by artificial systems, in pattern recognition and in diagram parsing (e.g., Futrelle, 1999). Experimental investigation of causal attribution in comprehension of verbally annotated graphs by humans constitutes a necessary step that provides the background for implementation in software programs. In the current state of the art, in popular software programs such as spreadsheet tools as well as the scientific software tools that are used to draw graphs, the facilities for the design of verbal annotations have a relatively limited capability. A set of problematic cases and possible technical solutions are exemplified in Chapter 10, where principles and guidelines are presented for designing software tools that have the capacity to support end-users during their course of designing and producing verbally annotated graphs.

2.6.2 Scalar Concepts

The second dimension of coherence relation between graphical and textual entities is based on the concept of *scale*. A basic component of human cognition is the ability to order objects with respect to their gradable properties, and make comparison between them. The findings on this aspect of human cognition in linguistics, psychology, and philosophy show that humans establish relations between abstract representations of measurement rather than direct relations between objects (Bierwisch, 1989; Kennedy, 1999, and references therein). These abstract representations of measurement are called *scale*.

The linguistic expressions of scale and the closely connected concepts of *vagueness* and *gradability* in syntactic categories have been well investigated in natural language semantics (Sapir, 1944; Cresswell, 1977; Klein, 1980; Ludlow, 1989; Eschenbach, 1995; Kennedy, 1999; Rotstein & Winter, 2004; Kennedy & McNeally, 2005, among many others). Scalar adjectives and scalar adverbs have been the two frequently investigated syntactic categories. On the other hand, in psychology and behavioral sciences, the concept of scale has been investigated as a means for scientific measurement rather than a basic component of human cognition. Accordingly, the term *scaling* is used to mean measuring with respect to quantitative attributes. In particular, in his *theory of scale types*, the psychologist S. S. Stevens proposed that four different types of scales are used in all scientific measurement (Stevens, 1946): nominal, ordinal, interval, and ratio. Briefly, in the nominal scale, labels are assigned to measured objects or events. In the ordinal scale, the numbers assigned to objects or events represent the rank order of the entities; however, the differences among the assigned numbers are not comparable. In the interval scale, the differences but not the ratios among the assigned numbers are comparable. In the ratio scale, the ratios among the assigned numbers, as well as the differences, are comparable (also see the objections to Stevens' theory of scale types, Michell, 1986; Narens & Luce, 1986; Velleman & Wilkinson, 1993, among others). The classification of scale types proposed by Stevens is widely used in introductory chapters of research method books in psychology and behavioral sciences as a basis for measurement.

In addition to the introduction of scale types for measurement, a set of design guidelines is often presented, which states that the graph type should be selected according to the type of the scale of the domain entities. In particular, a line graph should be used to represent a domain value that has an interval scale or a ratio scale, whereas a bar graph should be used to represent a domain value that has an ordinal scale. Kosslyn (1989), in his discussion of the semantics of quantitative relational information, suggested that the information content of a graph can be interpreted by investigating the relation between the scale types used in the individual axes and by the way the points are paired by the specifiers (e.g., graph lines, bars). Kosslyn considered the possible combinations of nominal, ordinal, interval, and ratio scales (also see Zhang, 1996, for a taxonomy of relational information displays using Stevens' classification). To sum up, inconsistency between the represented scale type and the representing scale type leads to misconceptualizations in graph comprehension.

The focus of the *scale* concept in this dissertation goes beyond selecting an appropriate graph type to represent a domain value with a certain scale type. In multimodal comprehension of graph-text constellations, the variations in the axes of a line graph may result in a different conceptualization of the processes and events represented by the graphical entities compared to the conceptualization of the processes and events represented by the linguistic entities. As a result, these differences may lead to processing difficulties in the construction of coherence relations between graphical and textual entities.

Although several studies in multimodal diagram-text constellations and picture-text constellations emphasized the capability of language to permit expression of ambiguity and the lack of linguistic expressiveness in diagrams (Stenning & Oberlander, 1995; Schnotz, 2002; Ainsworth, 2006), the investigations in graph comprehension did not go beyond the unimodal representation of graphs. Chapter 8 presents empirical investigations of the interaction between the representation of scale by graphs and the representation of scales by linguistic expressions. In particular, gradable (i.e., scalar) adjectives and gradable adverbs in graph-text constellations that have a separate graph-layer–text-layer layout are investigated.

From the perspective of HCI and graph design, adjusting the range of axis values of the graph, the designer has the opportunity to decide a combination of axis ranges for the best representation of the intended message. In other words, the designer can adjust the values on the axes such that the visual impression reflects the appropriate information (Kosslyn, 1994b; 2006). The concept of scale, as a coherence dimension, is important for the design of graph-text constellations since the vagueness of language gives opportunity to the designer of a graph-text constellation to select appropriate linguistic expressions that accompany the graph, in addition to giving the opportunity to adjust the ranges of the axes. Therefore, empirical investigations of the interaction between spatial configurations of graphical entities in the graph space and the gradable adjectives and adverbs in the text layer of a graph text constellation provide the background for the design of artificial cognitive systems that have the capability to inspect the relationship between graphical and textual entities in graph-text constellations. The implication of the empirical studies that are reported in Chapter 8 are discussed in Chapter 10, where design principles and guidelines are devised for graph-text constellations.

2.7 SUMMARY

Understanding the nature of the interaction between parts of a document (or a hypertext interface) with different representational modalities is a complex task with respect to internal characteristics of the information content and to the type of representation used other than text (i.e., pictorial illustrations, diagrams, graphs, and so on). Graphs and pictorial illustrations have different characteristics that are relevant to multimodal comprehension. Pictorial illustrations—also called "representational pictures" by Alesandrini (1984)—can informally be characterized by their visual resemblance to the objects they stand for. Pictorial illustrations and their referents have spatially similar layouts (i.e., iconic similarity), which, for instance, guarantees an optically veridical mapping from the visual-world object to the external representation in the case of photographs.[14] Even line drawings must have a high degree of systematic resemblance to the depicted entity to be able to function successfully in a multimodal constellation. Moreover, pictorial illustrations do not possess an internal syntax in the sense of representational formats as discussed by Kosslyn (1980, p. 31). On the other hand, graphs are representational artifacts that possess internal syntactic structures. The syntactic analysis of a graph is fundamental for succeeding processes of semantic and pragmatic analyses in graph comprehension (Kosslyn, 1989; Pinker, 1990). From the perspective of the development of cognitive architectures and models for multimodal comprehension of graph-text constellations, the formal characteristics of graphs present advantages over pictorial illustrations.

This chapter presented the investigation of graph-text constellations as grammatical means for problem solving and communication in connection with the previous work conducted in relevant disciplines. An overview of the previous studies on graphs shows that, despite their recognized importance and widespread use today, modality-specific research on graph comprehension and graph-text constellations is lacking. The research on graphs was mainly based on the role of graphs as learning material

[14] The veridicality commitments of drawings are less strict, but—as Chabris and Kosslyn (2005) argue for the subcase of caricatures—this property of drawings can facilitate processing.

until the 1980s. The initial studies on graph comprehension from the perspective of the information processing approach have been conducted since the last several decades. These studies focused on processes of graph perception and comprehension rather than the structural properties of graphs.

On the other hand, graphs, having a clear grammatical structure, can be conceived as quasi-symbolic depictive representations that allow the analysis at syntactic and semantic levels, in a similar way to the analysis of language. The graph schema is introduced as the key concept for the basis of construction of coreference relations between graphical and textual entities in graph-text constellations. The construction of coreference links between graphical and textual entities is further extended to involve coherence relations, in particular causal attribution and the correspondence between linguistic expressions of scale and graphical representation of scalar information in graph-text constellations.

From the perspective of HCI, the analysis of the construction of inter-modal coreference relations by humans is a necessary step for devising principles and guidelines for designing graph-text constellations. The design principles and guidelines, which are distributively presented in the chapters where empirical investigations are reported (Chapter 6, Chapter 7, and Chapter 8), are summarized, and the implications for the development of software tools that support end-users during their course of designing and producing graph-text constellations are discussed in Chapter 10. In Chapter 3, eye tracking is introduced as the research paradigm for a systematic empirical investigation of graph-text constellations.

3 Eye Tracking as a Research Paradigm for a Systematic Investigation of Graph-Text Constellations

This chapter introduces the eye tracking methodology and proposes it as an empirical research paradigm for a systematic investigation of graph-text constellations. In 3.1, the available research methodologies in HCI and cognitive science are reviewed and their use in the dissertation for analyzing comprehension of graph-text constellations is introduced. In 3.2, the eye tracking paradigm is introduced as a research method in the study of human comprehension. In particular, the relationship between visual attention and eye movements is discussed; dependent variables and techniques for analysis of eye movement data are introduced; and the models of eye movement control in reading, the state-of-the-art research on eye movements in visual scene viewing, and the influence of textual entities on eye movements in picture-text constellations are presented. In the last part of the chapter, in 3.3, the promise of the eye tracking paradigm for analyzing comprehension of multimodal graph-text constellations is discussed.

3.1 RESEARCH METHODS IN HCI AND COGNITIVE SCIENCE

HCI and cognitive science are interdisciplinary fields of research, which use research methods that cut across different research domains, such as cognitive psychology, educational psychology, psychophysics, and linguistics. These methods involve controlled experiments; questionnaires, in-depth interviews, focus groups, and usability inspection methods such as heuristic evaluation and cognitive walkthrough; formal and representational analyses; and cognitive architectures and cognitive modeling (see Cairns & Cox, 2008, for a review of research methods in HCI).

Controlled experiments measure different aspects of human behavior, such as response time, psychophysical reactions, subjective judgments, and eye movement behavior. The method of conducting controlled experiments has been adopted from research methods in psychology and it has been used widely to study human cognition and to evaluate human-computer interfaces for usability (Blandford, Cox, & Cairns, 2008). In a controlled experiment, causal relationships between two or more variables are investigated in terms of statistically obtained significant effects. Usually, the results obtained from an experimental group of participants are compared to the results obtained from a control group (viz., between-subject design); or the results obtained from an experimental design condition are compared to the results obtained from a control design condition (viz., within-subject design). The difference between the experimental group/condition and the control group/condition is the investigated aspect of the stimuli (viz., the dependent variable). The comparison between the experimental group/condition and the control group/condition is performed in terms of the measured variable (viz., the independent variable). For instance, in a between-subject design a difference between two interfaces, such as two different menu structures on a web page, can be investigated in terms of the resulting differences between total stimuli inspection time (i.e., gaze time) of the participants in the experimental group and the inspection time of the participants in the control group, whereas in a within-subject design the investigation is performed in terms of the resulting differences between total stimuli inspection time of the same participants under the two experimental conditions (i.e., the two interfaces). In both cases, the dependent variable is the difference between the two interfaces and the independent variable is the gaze time of the participants. Moreover, experimental investigations in educational psychology and instructional design focus on the role of introspective judgments measured by *self-reported difficulty*, *retention*, and *transfer*. By definition, retention measures *remembering* whereas transfer measures *understanding* (Mayer, 2001).

Questionnaires, in-depth interviews, focus groups, and usability inspection methods such as heuristic evaluation (Nielsen, 1994) and cognitive walkthrough (Lewis, Poison, Wharton, & Rieman, 1990; Wharton, Rieman, Lewis, & Poison, 1994) are HCI methods, which emerged as a response to the need for timely interface evaluation systems after the fast changes in the development of interface technologies in the past several decades. Questionnaires consist of a set of questions that participants are asked to complete. In-depth interviews are one-to-one interviews between an investigator and a participant, whereas focus groups involve one investigator and a group of participants in a single session (see Adams & Cox, 2008 for a review). A kindred approach, the *thinking aloud* method (Ericsson & Simon, 1980, 1993) was developed as a formative and empirical research method for gathering data in interface evaluation. The method has been used in the evaluation of strategic thought and in specifying structures and rules underlying verbal production. The thinking aloud method has been used by researchers in cognitive science and artificial intelligence (e.g., Tabachneck-Schijf et al., 1997), as well as by researchers in subdisciplines of HCI such as usability analysis (also see Russo, Johnson, & Stephens, 1989, and Guan, Lee, Cuddihy, & Ramey, 2006, for the objections concerning the ecological validity of the thinking aloud method, and alternative proposals). Briefly, the heuristic evaluation method is based on usability specialists' judgments about the components of the interface with respect to their compliance with the established usability principles. The cognitive walkthrough method simulates a user's problem solving steps by checking whether the memory content and the goal of the user lead to the next correct action (Nielsen & Mack, 1994).

Formal and representational analyses of interactive systems facilitate automation of interface evaluation. Formal techniques are used to construct a model that captures interactive key features of a design, which, in turn, provide the capability of determining if a property is valid of the constructed model (viz., model checking; see Harrison, Campos, & Loer, 2008, for a review). Representational analysis investigates the relationship between represented entities and representing entities in terms of the structural properties of representations and their relation to the structural properties of relevant tasks. For instance, Zhang (1996) presented an analysis of representational properties of graphic and tabular displays (viz., relational information displays, RIDs) within the framework of the distributed cognition approach (Zhang & Norman, 1994). Zhang (1996) analyzed the relationship between RIDs and structures of tasks in terms of the mapping between the information perceivable from a RID and the information required for the task (also see Zhang, 1997, for a representational analysis of cockpit information displays).

Cognitive architectures specify the underlying infrastructure for intelligent systems, i.e. the fundamental aspects of a cognitive agent that stays constant over time and across different application domains, such as short-term and long-term memory systems, the representations in these memory systems, and the functional processes that operate on the representations (Langley, Laird, & Rogers, 2009). Recently, the three major cognitive architectures used in HCI and cognitive science are Soar (Laird et al., 1987; Newell, 1990), ACT-R (Anderson, 1993; Anderson & Lebiere, 1998), and EPIC (Meyer & Kieras, 1997a, b; Kieras & Meyer, 1997). Cognitive modeling, on the other hand, specifies how humans perform specific information processing tasks during their interaction with complex information displays. Cognitive models range from descriptions of general steps to complete a task to simulation models that predict user behavior during task performance with an interface (Cox & Peebles, 2008). An early cognitive modeling approach, the GOMS family of analysis techniques, was presented by Card et al. (1983). The GOMS (Goals, Operators, Methods, and Selection rules) modeling techniques investigate users' interaction with computers in terms of component tasks and the elementary actions such as perceptual, motor, and cognitive processes (John & Kieras, 1996a,b; Kieras, 1997; see Chapter 9 for a discussion of the GOMS approach; see Gray, Young, & Kirschenbaum, 1997; Cox & Peebles, 2008; Langley et al., 2009; and Sun, 2009, for reviews of computational modeling and computational architectures in HCI).

The research methods in HCI and cognitive science that are presented above are interdependent techniques used for systematic analyses of human-computer interfaces. The outcomes of the presented methodologies can be used to devise principles and guidelines for designing interactive systems. This dissertation presents a theoretical framework for the development of an architecture for multimodal comprehension of graph-text constellations (Chapter 4), a set of controlled experiments that employed eye tracking and complementary techniques (e.g., recall and judgment reports) in Chapter 6, 7, and 8, and the application of the GOMS approach for modeling a set of specific information processing tasks in sentence-graph verification tasks (Chapter 9). The specification of principles and guidelines for designing graph-text constellations and software tools that have the capacity to support end-users during their course of designing and producing graph-text constellations are presented as the complementary step of the systematic investigation in Chapter 10. In the following part of the chapter, the eye tracking methodology is introduced as a research paradigm in the study of human comprehension.

3.2 EYE TRACKING AS A RESEARCH PARADIGM IN THE STUDY OF COMPREHENSION

Research on eye movements goes back to the nineteenth century. Rayner (1998) defines three eras in eye movement research. In the first era—between 1879 and 1920—researchers investigated eye movement characteristics such as *eye movement suppression* (i.e., the lack of perceiving during saccades—the ballistic jumps of the eye on the stimuli) and *saccade latency* (i.e., the time to initiate an eye movement). The second era was concurrent with the behaviorist research trends in psychology. In the second era, eye movement research did not attract the interest of researchers compared to the first and the third era. In the third era—after the 1970s—eye movement recording systems and data analysis methods were developed. Recently, the terms *eye movement recording* and *eye tracking* are used for recording eye movement data by eye tracking devices and analyzing the recorded data by integrated or third party software programs.

Eye tracking devices can be classified into four types: (a) *electrooculogram (EOG)* eye trackers use electrodes to record the time when the eyes move but they do not provide other information such as saccade length, (b) *scleral search coil* eye trackers use cables that are physically connected to the eye; they are not used in studies that involve human participants, (c) *infrared* eye trackers are the most frequently used eye trackers recently; they have advantages over the other types such as allowing head movements, and (d) *video-based* eye trackers, also called *dual-purkinje* eye trackers, have limitations such as the requirement of keeping the head still during recording, and sensitivity of the recordings to pupil size; however, they have higher resolution compared to the other types (see Richardson & Spivey, 2004a, b; Duchowski, 2007, for a review of technological development of eye trackers).

There are different types of eye movement. In humans, *pursuit* eye movements are used to track moving objects; *vergence* for looking at objects while they are getting closer or farther; and *vestibular eye movements* for compensation during head or body movements (see Richardson, Dale, & Spivey, 2007, for a review). These three types of eye movement are generally investigated in research on the oculomotor system (the system that controls eye movements).

Within the framework of the information processing approach (cf. the computational theory of mind, see Chapter 1), *saccades* and *fixations* are the two most frequently investigated eye movement types, often measured as a reflection of cognitive processes. Saccades and fixations are also the most frequently used eye movements in humans. The terms *saccade* is used for the fast movement of the eye with a speed of approximately 500 degrees per second. The term *fixation* is used to mean the fixated position of the eye between two saccades. Saccades generally occur about four or five times per second. Fixations generally last between 40 to1,000 ms with an average of 250 ms (Haberlandt, 1994).

Rayner and Sereno (1994) describe three regions with respect to the visual acuity of the eye during a fixation: (a) *foveal region*, which corresponds to the region of 2° visual angle around the point of fixation,[1] (b) *parafoveal region*, which corresponds to the region of 5° visual angle on each side of the point of fixation; (c) *peripheral region*, which is the region outside the foveal and parafoveal regions. From the foveal region to the parafoveal region and the peripheral region, visual acuity drastically reduces. Although the visual acuity is reduced, readers can process information in this parafoveal region and peripheral region.

The motivation for studying eye movements for analyzing comprehension is the relationship between eye movements and visual attention. This relationship is discussed below.

3.2.1 The Relationship Between Eye Movements and Visual Attention

Attention is a fundamentally inner and introspective cognitive process. The meaning of visual attention can be illustrated by means of the visual display presented in Figure 3.1.

[1] The 1° visual angle approximately corresponds to 3-4 letters or digits or a size of thumb at arm distance; a 1024x768 pixels image corresponds to 22.5° x 16.2° at the viewing distance of 900 mm.

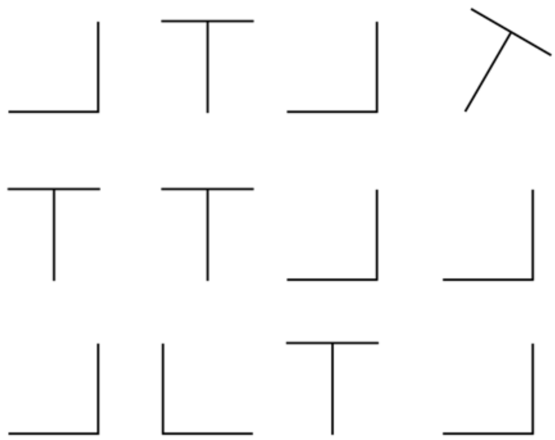


Figure 3.1. A visual display that can be used to illustrate the influence of perceptual properties of the stimuli and the influence of observer's task on visual attention (redrawn based on the figure presented by Posner & Fernandez-Duque, 1999).

A short inspection of Figure 3.1 quickly captures the tilted T. In other words, the *attention* of the observer is drawn to the tilted T, because it is different from the background formed by the other Ts and Ls. This behavior exemplifies the influence of perceptual properties of stimuli on attention, which is a component of human vision (Duchowski, 2007). On the other hand, if the observer is given a task, for instance, if the task of the observer is to find a target L, the observer guides his/her visual search to find the target; in other words, the observer diverts his/her *attention*. This behavior exemplifies voluntary attention (Posner & Fernandez-Duque, 1999).

The relationship between stimuli and attention mechanisms has attracted researchers' interest since the foundation of psychology (James, 1890; Von Helmholtz, 1925). After the 1950s, with the development of information processing approach to human cognition, attentional phenomena have been a subject of interdisciplinary research involving psychophysics, cognitive neuroscience, computer science and educational science, as well as psychology (Duchowski, 2007).

The exemplified interaction between perceptual properties of the stimuli and the task of the observer has implications on the study of attention from the perspective of the information processing approach. First, although purely stimuli-based models of attention provide a potential basis for computational modeling of human vision, they are incomplete *per se* since attentional phenomena are characterized by the interaction between perceptually driven data and top-down cognitive factors, e.g. the goal of the observer. Second, the study of the relationship between cognitive factors and the control of attention is necessary for understanding the underlying mechanisms of human visual information processing.

The study of attentional phenomena is important for cognitive science, HCI, and education research, not only for understanding the underlying mechanisms for the purpose of computational modeling, but also for studying its implications in comprehension, learning, and usability of human-computer interfaces. For instance, when observers are required to split their attention between sources of information in different representational modalities, such as a diagram-text constellation that has a separate layout of the diagram layer and the text layer, the integration of information induced by depictive and textual entities results in high *cognitive load*, compared to a multimodal constellation where textual information and depictive entities are presented with an integrated, rather than separate layout (cf. the split attention effect introduced by Chandler & Sweller, 1992; Sweller et al., 1998, within the framework of the cognitive load theory; the spatial contiguity principle introduced by Mayer, 2001/2009; see Chapter 7 for a more detailed discussion).

Like other cognitive processes, attention has been investigated by using the measures of behavioral performance that is correlated with attention (Motter, 1999). For the analysis of visual attention, eye movements have been often used as a measure of behavioral performance. The early eye movement research in the 1960s and 1970s was mainly based on selective attention on pictures (e.g., Yarbus, 1967). One of the research goals was to understand the factors that influence what can be seen in a single fixation. In addition, the factors that influence fixation locations on pictures were listed as a set of stimulus variables (e.g., informative or interesting regions in scenes) and top-down factors such as viewer expectations and goals, viewer characteristics such as age, culture and domain expertise (see Levie, 1987, for an early review of literature on research on pictures). More recently, Henderson (2003) emphasizes the interaction between bottom-up stimuli-based information that is generated from the image and top-down memory based knowledge that is generated from internal visual and cognitive systems for eye movement control in scene viewing. On the other hand, in other tasks, such as daily tasks like food preparation "[eye movement] control is seen as primarily 'top-down', and influenced very little by the 'intrinsic salience' of objects" (Land & Hayhoe 2001, p. 3359). Consequently, the relationship between stimuli-based factors and top-down factors is under debate.

According to Kowler (1999), "there are two natural links between eye movements and visual attention. One is the role played by attention in oculomotor control. The other is the way in which eye movements provide overt indicators of the locus of attention during the performance of complex cognitive tasks, such as reading or visual search" (p. 307). The studies in the 1980s showed that attention is diverted to a location before the eyes are fixated on that location. According to Pashler's (1998) *mandatory shift hypothesis*, "a shift in attention is mandatory before execution of a saccade" (p.81). These findings indicate a correlation between executed saccades and attention shifts.

On the other hand, attention can be diverted without executing a saccade. The mandatory shift hypothesis does not deny that visual attention does not have to be coincident with fixated location (Posner, Snyder, & Davidson, 1980). Humans can voluntarily divert their attention to another location while keeping their fixations at a certain location, i.e. without executing a saccade. Therefore, although in natural and unconstrained eye movements eye position and visual attention are highly likely to be coupled (Findlay & Gilchrist, 1998),[2] it should be acknowledged that consecutive eye movements of observers on stimuli reveal a partial path of locus of attention. This is a well-known problem and a basic assumption in eye tracking research. It has at least two implications for the study of cognitive processes, such as attentional phenomena, by eye tracking methodology. First, experimental investigations often should employ complementary methods such as measuring recall, retention, and transfer or subjective judgment reports where applicable. Second, cognitive models that predict or simulate eye movements on complex visual displays are not always realistically deterministic.

Although eye movement analysis as a measure of cognitive processes has potential caveats, compared to alternative research methodologies, eye tracking provides researchers with robust experimental data, in particular, data about online comprehension processes. Therefore, it has been widely used for the analysis of cognitive tasks in reading and scene viewing, as well as for the analysis of human-computer interfaces in usability testing. Most of the research findings on eye movements are from the domain of eye movement control in reading, which is presented below.

3.2.2 Models of Eye Movement Control in Reading

According to Rayner (1998), “eye movement measures can be used to infer moment-to-moment cognitive processes in reading … and that the variability in the measures reflects on-line processing” (p. 376). In reading research, the eye tracking method—like other methods such as the key-press, moving-window or stationary-window methods—is one of the self-paced timing techniques. In these techniques, the underlying assumption is that the readers read a passage “at a pace that matches the internal comprehension processes” (Haberlandt, 1994, p. 8). In other words, the data obtained by these techniques are assumed to reflect internal comprehension processes. Moreover, it is also assumed that the readers comprehend words as soon as they see them (viz., the immediacy hypothesis) and the readers process the word which is fixated (viz., the eye-mind hypothesis; Just & Carpenter, 1980). In reading research, the analysis of eye movement characteristics can provide data for the investigation of cognitive processes of pronoun resolution and coreference, lexical

[2] See Richardson et al. (2007) for the review of behavioral and neurological support.

and syntactic ambiguity, and the influence of semantic and discourse context on these processes (Richardson et al., 2007).

During reading, humans fixate for an average of 200–250 ms and then make a saccade movement for a space of approximately seven to nine characters (see Rayner, 1998; Duchowski, 2002, 2007; Richardson et al., 2007; Webb & Renshaw, 2008, for reviews). The variability is high: 100–500 ms fixation duration and 1–15 character space depending on textual and typographical factors such as line length.[3] Humans generally make longer fixations at the end of sentences and phrase boundaries compared to the fixations within a phrase (Just & Carpenter, 1980). Approximately 10-15% of the fixations are regressions (i.e., the re-reading of the text). A text that is difficult to understand results in long fixation durations, short saccades and high percentage of regression (Rayner & Sereno, 1994). It is generally accepted that there is no perceiving during saccades. Saccade durations, which have an approximate value of 5–8 ms in reading, are generally neglected in the calculation of gaze times. In most cases, gaze time is taken to be the sum of fixation durations (Just & Carpenter, 1980).

Several computational models of eye movement control in reading have been proposed over the last decades (see Reichle, 2006, for a review). According to Reichle (2006), the initial models of eye movement control in reading can be divided into two classes: (a) *cognitive-control models*, which assume that the completion of a cognitive event such as identification of a word cause the eyes to move; (b) *oculomotor-control models*, which assume that the oculomotor system maintains steady progress during reading by moving the eyes. However, as stated by Reichle, more recent theories make more fine-grained distinctions.

The two frequently cited cognitive-control models of eye movement control in reading are the E-Z Reader model (Reichle, Rayner, & Pollatsek, 2003; Reichle, Pollatsek, & Rayner, 2006), and the SWIFT model (Richter, Engbert, & Kliegl, 2006). The differences between the two models are based on the underlying assumptions about cognitive processes during reading. The E-Z Reader is serial (i.e., one word at a time), whereas the SWIFT is parallel (i.e., more than one word can be processed in the perceptual span region). The E-Z Reader has a cognitive engine for

[3] These data reflect eye movement characteristics under unconstrained reading conditions, i.e., without time limitations. Recent studies show that humans are also able to read text normally when they are allowed to fixate on the text only for a short period of 50–60 ms. This shows that the minimum required time to process information in reading is 50 to 60 ms under constrained conditions (Liversedge, Rayner, White, Vergilino-Perez, Findlay, & Kentridge, 2004). These results are obtained by using eye-contingent display eye trackers that are able to follow the fixation locations and display a visual mask on a fixated location after a short period of time. However, the implications of the visual mask paradigm on reading without time limitations are not clear at this stage. This dissertation focuses on the inspection of stimuli without time limitations.

driving saccades, whereas in the SWIFT model saccades are automatically driven by cognitive influence.

The E-Z Reader model and the SWIFT model are the two most comprehensive models of eye movement control in reading. In the current state of the art, there is no eye movement control model that investigates the interaction between linguistic entities and entities in other representational modalities. Recently, integration of visual scene information and spoken language during language processing is investigated from a psycholinguistics perspective under the framework of so-called visual world paradigm (Ferreira & Tanenhaus, 2007). Eye movement in visual scene viewing and the visual world paradigm are presented below.

3.2.3 Eye Movements in Visual Scene Viewing

It has been known that eye movement characteristics in visual scene viewing differ from eye movement characteristics in reading; for instance, there are no conventionally specified directional characteristics in scene viewing (Duchowski, 2002). However, compared to eye movement control in reading, less is known about eye movement control in visual scene viewing (Rayner, 1998; Henderson, 2003; Rayner, Smith, Malcolm, & Henderson, 2009).[4]

In early research on eye movements in visual scenes, the average fixation duration on pictures was assumed to be approximately 300 ms (Rayner, 1998). It was also assumed that the gist of an entire picture could be understood in 300 ms. For instance, Hegarty et al. (1991) state that in complex pictures such as photographs humans can abstract the general topic in 300 ms but more fixations are needed for further information extraction (also see Rayner & Pollatsek, 1992). Later studies showed that the gist information from a photograph could be extracted in as short as 30–50 ms (Henderson & Ferreira, 2004) and sufficient information can be encoded in 50–75 ms in object identification (Rayner, 1998). However, according to Rayner et al. (2009) humans need at least 150 ms in each fixation to be able to process information in visual scenes.

In the current state of the art, the research on eye movements in visual scene viewing is not at a mature state. It is difficult to generalize the recent findings on visual scene viewing due to the abundant types of visual representations and due to the lack of large-scale studies except for a few (e.g., Wooding, 2002; Wooding, Muggelstone, Purdy, & Gale, 2002). However, research in psycholinguistics since the last two decades has shown that language comprehension—particularly spoken language—interacts with visual perception (Ferreira & Tanenhaus, 2007). This was particularly

[4] Researchers use the terms *visual scene viewing* and *visual scene perception* interchangeably. In this dissertation, the former is used due to the unclear borders among perception, cognition, and comprehension.

evident for cases such as syntactic ambiguity resolution. These findings are largely based on the saccadic eye movements to objects in the visual world immediately after hearing the relevant words (Tanenhaus et al., 1995). The term *visual world* is then used to mean eye movements on a scene that are guided by spoken language. In other words, the focus of the visual world paradigm is on utterance mediated eye movements (Knoeferle et al., 2005).

The findings of the empirical investigations that employed the visual world paradigm have introduced a critical approach to the traditional view in psycholinguistics which states that language comprehension is informationally encapsulated from other perceptual and cognitive systems (Fodor, 1983, see Chapter 2). The visual world paradigm is applicable to visual displays that depict natural scenes, such as customer passing money to a salesclerk, but not to visual displays that depict objects, such as a set of random objects arranged on a table (Henderson & Ferreira, 2004).[5] Although the findings within the framework of the visual paradigm have provided clear evidence for the interaction between spoken language and vision in visual scenes, the implications for multimodal comprehension of graph-text constellations are not clear at this stage.

Since the past two decades, a number of studies have employed eye tracking methodology for the analysis of the influence of written language (i.e., textual entities) on eye movements in picture-text constellations. A short overview of these studies is presented below.

3.2.4 The Influence of Textual Entities on Eye Movements in Picture-Text Constellations

As stated by Rayner, Rotello, Stewart, Keir, and Duffy (2001), the eye movement characteristics when text and pictures have to be integrated in the comprehension process have been rarely investigated. In domains such as newspaper reading almost all research has been performed as case studies (Holsanova, Holmqvist, & Rahm, 2006) except for a few (e.g., Garcia & Stark, 1991). Given their widespread use, the lack of research on eye movements in multimodal constellations is surprising (Rayner et al., 2001; Duffy, 1992).

Several studies reported participants' tendency to look at the textual entities first, when the picture and the textual entities are presented together; in other words, inspections of pictorial illustrations are largely text directed (Underwood, Jebbett, & Roberts, 2004; Rayner et al., 2001; Hegarty et al., 1991; Ratwani, Trafton, & Boehm-Davis, 2008).

[5] Henderson and Ferreira (2004) emphasize that there are differences between the display of natural scenes and the display of objects with respect to the interaction between vision and language, in particular, when they are used as stimuli in the visual world paradigm.

Underwood et al. (2004) found that in their sentence-picture verification experiments the fixation durations on the pictures were characteristically longer than the fixation durations on the sentences. Carroll, Young, and Guertin (1992) investigated eye movements of the readers of cartoons. They found that mean fixation durations on the cartoons' picture (270 ms) and on the cartoon's caption (236 ms) were shorter than the ones in scene viewing. They stated that people briefly look at the cartoon, with as few as three fixations; then they read the caption carefully. After reading the caption, they fixate on the objects and the characters mentioned in the caption. Similar findings were reported for subtitles in television programs (d'Ydewalle & Gielen, 1992).

It is not possible to make generalizations from the limited number of studies that investigated the influence of textual entities on eye movements on pictorial illustrations. There are myriad types of depictive representations and modality-specific research on depictive representations and their interaction with language is lacking. Moreover, as discussed in Chapter 1 and Chapter 2, graphs comprise a distinct type of depictive representation, different from pictorial illustrations and visual scenes. In the following part of the chapter, the promise of eye tracking paradigm for the analysis of comprehension of multimodal graph-text constellations is presented.

3.3 THE PROMISE OF EYE TRACKING PARADIGM FOR ANALYZING MULTIMODAL GRAPH-TEXT CONSTELLATIONS

Eye tracking has been used for the investigation of comprehension processes in more abstract depictive representations such as cartographic representations and graphs, compared to pictorial illustrations and visual scenes (e.g., see Montello, 2002, for a review of cognitive map design, a subdiscipline of cartography). However, in the current state of the art, eye tracking is not a widely used research method in graph comprehension research despite its potential to provide robust data for analysis. As stated by Koerner (2004), "eye tracking data can provide insight into the moment-by-moment processing during graph comprehension by providing an indication of the allocation of attention to different elements of a graph during the course of comprehension or problem solving" (p. 470).[6]

Few researchers investigated eye movements on graphs for specific purposes: Lohse (1993), based on the GOMS approach, developed an early simulation model of graph perception. As presented in Chapter 2, Peebles and Cheng (2001, 2002, 2003)

[6] Koerner (2004) uses the term *graph comprehension* in a broad sense to mean comprehension of graphical entities in abstract hierarchical graphs that represent the relationships between non-numerical entities or concepts, such as tree diagrams and family trees. However, the quoted statement is applicable to statistical graphs as well as hierarchical graphs.

developed a simulation model of problem solving with two specific types of Cartesian graph (namely, function and parametric graph) using the framework of the ACT-R/PM cognitive architecture (Anderson, 1993; Anderson & Lebiere, 1998; Byrne & Anderson, 1998). Carpenter and Shah (1998) investigated participants' eye movements on three-variable line graphs during their verbalization of graphs to provide empirical evidence for their model of perceptual and conceptual processes in graph comprehension. In addition, Renshaw, Finlay, Tyfa, and Ward (2003) performed a comparison of 2-D and 3-D line graphs in a usability study; they showed that certain design features in graphs influence eye movements in a predictable way.

Although those previous studies made different assumptions and used different techniques for the analysis of eye movement data, they paved the way for the use of eye tracking method in graph comprehension research. They used language for specification of the tasks either on the same or in a separate screen. However, a systematic investigation of the relationship between graphical entities and textual entities was beyond the scope of those studies. In the succeeding part of the chapter, eye movements are introduced as input increments in comprehension of graph-text constellations.

3.3.1 Eye Movements as Input Increments in Comprehension of Graph-Text Constellations

The studies in language comprehension and production widely accept incremental architectures, where incremental models are often depicted as cascade of processes (Kempen & Hoenkamp, 1987; Levelt 1989). In language production, "the stream of information … is not continuous but consists of pieces of knowledge, called increments. In other words, the process of conceptualization presupposes conceptual building blocks, which are manipulated and combined to new and more complex conceptual structures" (Guhe, Habel, & Chandler, 2004, p. 8). Eye movements, as they have been studied so far in spoken language research, reveal the incremental and interactive nature of language. Richardson et al. (2007) state that humans are "gradually influenced by the incremental delivery of linguistic information, and eye movements exhibit the continuous, partially active representations that arise during processing" (p. 334).

During the course of comprehension of a graph-text constellation, a gaze shift between the text layer and the graph layer shows that input increments are generated from the two representational modalities. Following the mandatory shift hypothesis (Pashler, 1998), the gaze shift also indicates a shift in attention.[7] Consecutive gaze

[7] The research on eye movements indicates that detailed information processing with highest visual acuity is serial (i.e., serial foveal vision) whereas a parallel parafoveal or peripheral information processing dictates the next focus of attention (Duchowski, 2007, cf. "what" and "where" description of the selective nature of visual attention). However, the nature of the relationship between input

shifts between textual and graphical entities, in turn, indicate that input increments that are generated from textual and graphical entities are integrated under the assumption that the reader has the goal of constructing a coherent interpretation of the stimuli. Therefore, the nature of the integration of the information generated from input increments can be empirically investigated by testing the influence of changes in one representational modality on the input increments generated from the other modality. The empirical investigations reported in Chapter 6, Chapter 7, and Chapter 8 present various aspects of such an influence. In particular, Chapter 6 investigates the influence of syntactic complexity of textual entities on input increments generated from graphical entities by analyzing gaze shifts between graph layer and text layer of graph-text constellations. Chapter 7 presents empirical investigations for the bridging role of verbal annotations in graph-text constellations, the role of verbal annotations on causal attribution in verbally annotated graphs, and the influence of graphically induced information on temporally ambiguous events and processes represented by verbal annotations. Chapter 8 shows how changes in spatial configurations of graphical entities influence comprehension of scalar adjectives and adverbs in the text layer of graph-text constellations by investigating gaze times of the readers on sentences. See Chapter 5 for the specification of independent variables of eye movement that are used in the empirical investigations.

3.3.2 The Analysis of the Interaction Between Stimuli-Driven Factors and Top-Down Factors

In graph comprehension, compared to scene viewing, an additional knowledge source, i.e. the knowledge of graph schema should be specified as top-down information, besides the task information. During the course of comprehension of a graph-text constellation, a gaze shift from a textual entity to a graphical entity often indicates *graph schema instantiation*. As stated in Chapter 1, during the course of comprehension of a graph-text constellation, a linguistic entity such as ‘the peak’ can be interpreted as a reference to a peak in the graph or as a reference to a peak in the domain, often via the mediating link between ‘the peak’ and the peak in the graph. In other words, the input entity that is generated from ‘the peak’ may lead to ‘finding a peak in the graph’ or ‘constructing peak in the course of domain entities’ via the mediating link. Consecutive gaze shifts between graphical entities and textual entities, i.e. the consecutive input increments generated by different modalities, reveal readers’ effort to construct those inter-modal coreference relations.

Moreover, during the course of multimodal comprehension of a graph-text constellation, comprehension of linguistic entities often specifies the task for the comprehension of graphical entities, by providing a top-down factor in comprehension of the graph. Empirically, the interaction between stimuli-driven

increments and the construction of conceptual increments from input increments is beyond the scope of this dissertation.

factors (e.g., visual salience of graphical entities) and task information—as specified by comprehension of textual entities—can be investigated by comparing participants' fixation locations in graphs after their reading of linguistic entities and by comparing participants' gaze patterns on graphs. Chapter 6 and Chapter 7 present the empirical investigations of such comparisons.

3.3.3 Implications for Computational Architectures and Modeling

Eye movements constitute an interface between cognitive structures/processes and entities in the environment, i.e. external representations. They provide valuable data for understanding how computational procedures of human mind operate. The locations and in some cases the logical sequence of locations that should be fixated for extraction of information required by a specific task can be predicted by computational models. Eye movements in experimental setting can then be used to evaluate the efficiency of the model.

However, there are several drawbacks for the use of eye movements as a basis for computational modeling of multimodal comprehension of graph-text constellations. First, an input increment generated from a textual entity and the input increment generated from the corresponding graphical entity does not have to be strictly sequential. Eye movements are metabolically cheap with a low trigger threshold compared to other motor movements (Richardson et al., 2007). According to Koerner & Gilchrist (2004), the "lavish" (p. 493) characteristic of eye movements reflects the payoff between executing eye movements for extracting visual information from the display and performing reasoning tasks with the stored information. Hayhoe, Bensinger, and Ballard (1998) propose that the frequent use of eye movements in spatial reasoning is a processing strategy that allows reduced processing costs. In other words, information is obtained from the scene as it is needed, rather than from a previously built-up mental model of the surroundings (Land & Hayhoe, 2001). Although it is predicted that visual memory processes play a crucial role in integration, the nature of the processes of integration and the role of working memory components in multimodal comprehension have not been known, except for the consensus on the processing and capacity limitations of the memory systems (Hayhoe et al., 1998; Baddeley, 2000; Gyselinck, Jamet, & Dubois, 2008; Cowan, Saults, & Morey, 2006; Seufert, Schütze, & Brünken, 2009).

Moreover, during multimodal comprehension of a graph-text constellation, the reader has freedom for choosing one representational modality over the other for information extraction, as well as the freedom for shifting or not shifting the gaze to the graph. Previous studies show that students experience difficulties in constructing and comprehension multimodal representations (e.g., Tabachneck-Schijf et al., 1997; Paoletti, 2004). Accordingly, the analysis of eye movements reveals non-deterministic behavior as well as deterministic behavior of readers.

In conclusion, as in other fields of research on computational architectures and modeling, such as language comprehension, there is a trade-off between developing coarse-grained cognitive architectures and modeling for specific information processing tasks. Chapter 4 presents a coarse-grained cognitive architecture for multimodal comprehension of graph-text constellations. As a specific instance of the multimodal comprehension architecture, Chapter 9 presents a GOMS-style cognitive processing model for multimodal comprehension of spatial prepositions, which is based on the decomposition of the tasks specified by the linguistic entities.

3.3.4 Implications for Design Guidelines and Principles

Eye tracking methodology has been used in HCI research since the 1990s (e.g., Goldberg & Kotval, 1999; Cowen, Ball, & Delin, 2002; Poole & Ball, 2005, among many others, see Webb & Renshaw, 2008, for a review of eye tracking in HCI). Eye movements, being indicators of cognitive processes, are used for the analysis of processing difficulties in visual displays, which in turn lead to guidelines and principles for design. This dissertation uses the findings of empirical investigations for devising a set of principles and guidelines for designing graph-text constellations and designing software tools that have the capability to support end-users during their course of designing and producing graph-text constellations. The guidelines are presented distributively in the chapters where empirical investigations are reported, and they are summarized and extended in Chapter 10.

3.4 SUMMARY

The eye tracking technique is a frequently used method for the study of human comprehension. Within the framework of information processing approach to human cognition, the basic motivation for studying eye movements is the relationship between eye movements and cognitive processes; in particular, the relationship between eye movements and visual attention. Although the research on eye movement control in reading is relatively well developed, the research on eye movements in visual scenes and the research on eye movements in picture-text constellations are novel domains of research. In the recent state of the art, the eye tracking technique has found limited use in graph comprehension. This chapter introduced eye tracking methodology as an empirical research paradigm for a systematic investigation of multimodal comprehension of graph-text constellations. In particular, the promise of eye tracking for its potential to reveal incremental characteristics of multimodal comprehension processes and to reveal the interaction between stimuli-driven and top-down factors during comprehension were discussed, and the implications for computational architectures, computational modeling, and design guidelines and principles were presented. In Chapter 4, a cognitive architecture for multimodal comprehension of graph-text constellations is proposed.

4 A Cognitive Architecture for Multimodal Comprehension of Graph-Text Constellations[1]

The purpose of this chapter is to introduce a cognitive architecture for multimodal comprehension of graph-text constellations. In 4.1, the theoretical stance of the dissertation is presented with respect to Jackendoff's (1983) modes of theoretical description. In 4.2, the processing of spatial concepts in linguistic entities and the processing of graphical entities in graph-text constellations are discussed. In 4.3, the architecture is introduced.

4.1 THE THEORETICAL APPROACH: MODES OF DESCRIPTION

The theoretical investigation of multimodal comprehension of graph-text constellations is connected to the answer to a more general question: How do we talk about what we see? To answer this question, Jackendoff (1983) proposed an approach, which investigates complementary aspects of cognitive psychology and the study of language comprehension.

Jackendoff (1983) defined five modes of description available to a theoretical approach in cognitive psychology: phenomenology, physiology, functional architecture, process, and structure (p. 4). At the two ends, there are *phenomenological* description and *physiological* description. The former deals with the character of experience (e.g., participants' introspective judgments), whereas the latter deals with brain structure and functions. The third mode of description, namely the description of the *functional architecture*—in particular, memory components for explaining how we talk about what we see—specifies the capacities for processing information. The other two levels are the *structure* mode of description and the *process* mode of description.[2] Concerning the question how we talk about what we see, the structure mode of description (henceforth, the structural mode of description) specifies the formal characteristics of linguistic and visual information and the relationship between the two, whereas the process mode of description specifies, in

---

[1] This chapter is based on the two published articles on multimodal comprehension of graph-text constellations: Habel and Acartürk (2007); Habel and Acartürk (2009).

[2] The structure mode of description and the process mode of description correspond to Marr's (1982) computational level and algorithmic level.

addition, the characteristics of the algorithms for computing linguistic and visual information in real time (p. 5).

Comprehension of graph-text constellations by humans, from the perspective of phenomenological description, can be seen as composed of almost *automatically* performed cognitive processes that underlie the integration of information induced by the different modalities.[3] The phenomenological mode of description, on its own, is not appropriate for giving an explanatory account of the underlying nature of multimodal comprehension of graph-text constellations. The physiological mode of description, in its current state of the art, is able to investigate the perceptual processes that precede comprehension. However, the scope of the physiological mode of description is constrained by the limitations of its empirical methods of investigation, which has not been developed recently to provide an explanatory account of comprehension processes. On the level of the functional architecture the role of memory components in multimodal comprehension is a novel field of research. As stated in Chapter 3, theoretical implications of the previous memory research on multimodal comprehension are not always clear; and experimental evidence is scant (Baddeley, 2000; Gyselinck et al., 2008; Cowan et al., 2006; Seufert et al., 2009). Consequently, the theoretical level of investigation in this dissertation is comprised of the structural mode of description and the process mode of description.

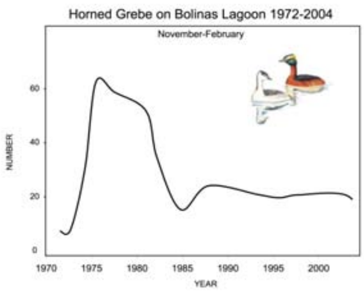
The structural mode of description precedes the process mode of description. As stated by Jackendoff (1983), "without a theory of structure, we cannot know what a theory of process has to explain" (p. 6). Graph-text constellations are particularly advantageous over picture-text constellations for an analysis of the structural aspects, since graphs have inherent grammatical structure, which allows syntactic and semantic levels of analysis. In parts 4.2 and 4.3 below, processing of spatial concepts in linguistic entities and processing of graphical entities in graph-text constellations are presented respectively, in connection with the structural characteristics of the investigated entities.

[3] As stated in Chapter 1, the term *induce* is used to mean humans' interpretation of the information provided by a particular external representation. The formal sense of the term *induce* in mathematics and logic is not intended here.

4.2 PROCESSING OF SPATIAL CONCEPTS IN GRAPH-TEXT CONSTELLATIONS

Investigation of comprehension of linguistic and graphical entities in graph-text constellations is exemplified by two excerpts from the waterbird census report that has been excerpted in Chapter 1. The graph-text constellations are shown in Figure 4.1.

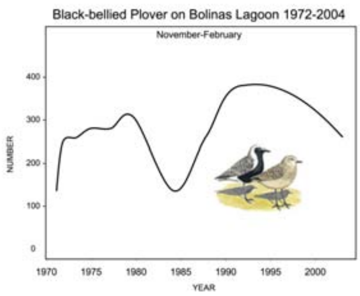
(a)



Bolinas Lagoon Population Trends

From a peak of about 60 wintering birds in 1976, numbers have declined to about 20 birds currently.

(b)



Bolinas Lagoon Population Trends

The Black-bellied Plover population has peaked twice in the past 35 years on Bolinas Lagoon. In 1980, the population peaked at 300 birds. After a drop to 120 in 1985, the population rebounded to 400 in 1995. Over the past ten years, the population has been decreasing. In 2005, there were about 260 Black-bellied Plovers on the lagoon.

Figure 4.1. Two sample graph-text constellations from “Waterbird Census at Bolinas Lagoon, Marin County, CA” by Wetlands Ecology Division, Point Reyes Bird Observatory (PRBO) Conservation Science (redrawn based on the originals). The graph and the paragraph in the figure were placed side by side for the purpose of demonstration. See the census report for the originals: http://www.prbo.org/cms/366, retrieved on September 25, 2009. Each constellation includes a population trend graph depicting the number of wintering birds in the lagoon and verbal statements about the population trend.

The graph-text constellations in the report present information about waterbirds in a lagoon, such as their appearance, migration, wintering and diet characteristics, and population trends. The population trend graphs represent the number of waterbirds between 1972 and 2004. The excerpts in Figure 4.1 exemplify verbal information about the population trends of two of the waterbird species in the lagoon, namely horned grebe and black-bellied plover. In 4.2.1, the processing of spatial concepts in linguistic entities is discussed. The processing of graphical entities is discussed in 4.2.2.

4.2.1 Processing of Spatial Concepts in Linguistic Entities

From the linguistic point of view, the process of *referring*, the core of comprehension is constituted by the referential expressions ‘peak of about 60’ and ‘declined to about 20’ in the graph-text constellation presented in Figure 4.1a. These expressions correspond to entities in the domain of discourse. Based on this, the reader has to establish coreference, the backbone of text coherence, by employing internal spatial and conceptual representations that mediate between language and the domain of discourse. The graphical entities, such as the maximum of the line graph in Figure 4.1a, correspond to the domain entities that are referred to by the linguistic expressions, such as ‘about 60’.[4]

Therefore, particularly relevant to the multimodal comprehension of graph-text constellations is the analysis of spatial terms in language. From the perspective of conceptual semantics—in particular, from the perspective of the investigation of interaction between spatial language and spatial cognition—the starting point of the analysis is the *lexemes* provided by *lexicon*. Within the framework of Jackendoff’s (1996, 2002) representational modularity hypothesis, lexicon maps words to their meaning, i.e. the constellations of conceptual structure (CS) and spatial structure (SpS). The lexical entry of a spatial term corresponds to a spatial concept, which has reflexes in CS and SpS, and in a spatial configuration in the external world (see Chapter 2 for the representational modularity hypothesis). The mathematical description of spatial concepts is given by their geometric characterization. The entities in the analysis of spatial terms in natural language are schematized by Eschenbach et al. (1998, p. 179), as shown below.

In language, the highly relevant types of expression that have reflexes both in conceptual structure (CS) and spatial structure (SpS) are *shape nouns*, *verbs of change in space* (i.e., spatial verbs), *spatial prepositional phrases*, and *spatial adverbial modifiers* (Eschenbach et al., 1998; Habel, 2005).

In Figure 4.1a, the term ‘peak’ corresponds to a domain entity, which is also referred to by the peak in the graph. The concept of PEAK is a shape concept, like other shape

[4] As stated in Chapter 2, linguistic entities in graph-text constellations are usually incomplete verbal statements about the domain entities that are represented by the graphical entities; linguistic entities bring certain aspects of the graphical entities to the foreground whereas leaving the others in the background. The *completeness* of the verbal statements in graph-text constellations is an outcome of the argument of *division of labor* between representational modalities (see Chapter 1 for the argument of division of labor).

concepts such as CORNER and KINK.[5] A peak in a line graph is a specific case of the more general concept of PEAK in the sense that graph peaks are directed geometric entities. Therefore, a peak in a line graph is part of the PATH, where the term *path* is used for directed, linear, and bounded entities (Eschenbach et al., 2000). The PATH concept is discussed in more detail in the context of spatial verbs below.

The common underlying conceptualization of spatial verbs is the presence of a BEARER OF MOTION traversing a PATH OF MOTION.[6] For instance, consider the frequently used verb ‘decline’ in multimodal graph-text constellations. The lexical entity ‘decline’ and similar entities such as ‘decrease’, ‘fall’, and ‘lose weight’ involve the process concept DECREASE_OF_VALUE*(_TEMP, _VALUE, ...)* provided by the lexicon in their conceptual structure. The DECREASE_OF_VALUE is the exemplification for the PATH OF MOTION where *TEMP* and *VALUE* are path arguments. The temporal argument *TEMP* of the process concept DECREASE_OF_VALUE stands for the temporal interval during which the whole process occurs. The process concept DECREASE_OF_VALUE then stands for a specification of a mapping from the temporal domain to the value domain, or—using the terminology of topology—for a PATH in the value space (Habel, 1990; Eschenbach et al., 2000).[7] Accordingly, in Figure 4.1a, the beginning of the temporal interval is explicitly specified in the sentence “From a peak of about 60 wintering birds in 1976, numbers have declined to about 20 birds currently”.

Concerning the spatial properties of the process concept DECREASE_OF_VALUE, the DECREASE-process holds the necessary condition *VALUE(BEGIN(DECREASE))* > *VALUE(END(DECREASE))*, which is independent of the value space. If the *SOURCE* and *GOAL* of the value path are specified, the value of *SOURCE* determines *VALUE(BEGIN(DECREASE))* and the value of *GOAL* determines *VALUE(END(DECREASE))*. These conditions are part of the geometric specification of the DECREASE concept. They can be used to reason about values, even in cases in which *SOURCE* or *GOAL* specifications are missing (cf. Geometric Concept Specification, GCS; Eschenbach et al., 2000). In Figure 4.1a, the lexical information of ‘decline’ provides *SOURCE* and *GOAL* arguments to be filled optionally. The sentence specifies information for both

[5] See Eschenbach et al. (1998) for a detailed investigation of geometric and ontological aspects of shape concepts underlying the semantics of nouns; in particular, they present the analysis of the German shape nouns ‘Ecke’ and ‘Knick’ (in English, ‘corner’ and ‘kink’).

[6] According to Habel (2005), CHANGE OF LOCATION and CHANGE OF ORIENTATION are the two aspects of conceptualization that underlie the verbs of spatial change. In this dissertation, the former is focused by the analysis of the verbs such as *increase* and *decrease* whereas the latter is left for further research. In multimodal comprehension of graph-text constellations, the CHANGE OF ORIENTATION aspect can be linguistically induced by a verb such as *turn* and graphically induced by trend reversals.

[7] See Eschenbach et al. (2000) for the geometry of paths as bounded linear oriented structures and the schemata for the corresponding lexical entries.

arguments, namely ‘peak of about 60’—via a *from*-prepositional phrase (PP)—and ‘about 20’—via a *to*-PP.

To sum up, GCS is a knowledge source that contributes to processing of spatial properties of linguistic entities by specifying geometric aspects of the spatial properties of the concepts that are referred to by linguistic entities. As stated in Chapter 1, such abstract topological and geometrical structures are relevant building blocks of conceptual representations in general, not only needed for communication about physical space, but also for a type of usage what is often called ‘figurative language’ (Habel & Eschenbach, 1997).

A second knowledge source that contributes to processing of spatial properties of linguistic entities is the linguistically induced Spatial Reference System (SRS_L). The SRS_L mediates the mapping of spatial expressions in language to spatial configurations of objects in the world. In the literature on spatial cognition, two major types of reference frames have been proposed, namely, the intrinsic or object-centered reference frame, and the deictic or observer-centered frame reference frame (e.g., Levinson, 1996; Carlson, 1999). A detailed formal specification of SRS_L is beyond the scope of this dissertation (see Jackendoff, 1996, for a more detailed specification of reference frame types, and Eschenbach, 1999, for the formal specifications of reference systems in spatial cognition).

The second frequently used type of expression that has reflexes both in conceptual structure (CS) and spatial structure (SpS), spatial prepositional phrases and spatial adverbial modifiers, are discussed below.

Spatial prepositional phrases and spatial adverbials introduce modification on the concepts of spatial change. PPs can be divided into two types, namely directional and non-directional (i.e., local). Directional PPs can further be divided into source, goal, course, and shape PPs (Eschenbach et al., 2000). Modification by source PPs and goal PPs are exemplified in (1).

(1) a. Since 1972, when there were only about 5 Gadwalls on Bolinas Lagoon, the population has increased.[8]
b. Since 1975 when the species only occurred irregularly in winter on Bolinas Lagoon, the population has increased to about 24 birds.
c. The Marbled Godwit population on Bolinas Lagoon has increased from about 100 birds in 1971 to about 900 birds in 2005.

[8] The sentences in (1) and (2) are excerpts from the “Waterbird Census at Bolinas Lagoon, Marin County, CA” by Wetlands Ecology Division, Point Reyes Bird Observatory (PRBO) Conservation Science. The sentence (1a) excerpted from the “Gadwall” section, (1b) from the “Whimbrel”, (1c) from the “Marbled Godwit”, and the sentence (2) from the “Canada Goose” section of the report. http://www.prbo.org/cms/366#results; retrieved on October 11, 2009.

In (1a), the linguistic expression 'increase' induces information about the DIRECTION OF CHANGE. In (1b), the linguistic expression 'increase to about 24 birds' induces information about the *VALUE(END(INCREASE))* as well as the DIRECTION OF CHANGE, thus explicitly stating the *GOAL* argument by a *to*-PP. In (1c), the linguistic expression 'increase from about 100 birds … to about 900 birds' induces information about *VALUE(BEGIN(INCREASE))* as well as the *VALUE(END(INCREASE))* and the DIRECTION OF CHANGE, thus explicitly specifying the *GOAL* argument by a *to*-PP and the *SOURCE* argument by a *from*-PP. As stated by Eschenbach et al. (2000), directional PPs *to* and *from* can be specified as path functions, which have geometric properties in their lexical specifications. These geometric properties—induced by the lexical entities *to* and *from*—have reflexes in SpS, which construct the coreference relations with the geometric properties of graph line representations in SpS. In addition to directional PPs, non-directional (local) PPs, such as *at*, can also access SpS, as in 'the water level is at the top'.

Adverbial modifiers have several subclasses such as rotation, orientation (manner), and time adverbials. Adverbial modifiers introduce a modification to the concepts of spatial change, as exemplified in (2).

(2) Winter numbers on Bolinas Lagoon have increased sharply since 1998.

In (2), the linguistic expression 'sharply' induces information about the MANNER OF CHANGE. Other adverbial modifiers such as *up*, *down*, and *steadily*, modify different aspects of spatial change. In a similar manner, spatial adjectives modify nouns, as in the phrase 'the sharp increase'.[9]

In summary, shape nouns, verbs of change in space, spatial prepositional phrases, and spatial adverbials have reflexes both in CS and SpS. The concept of PATH is the common conceptual entity that is shared between CS and SpS. In the following part of the chapter, processing of graphical entities in graph-text constellations are discussed.

4.2.2 Processing of Graphical Entities in Graph-Text Constellations

Spatial representations are fundamental in comprehension of graphical entities, beginning with the visual encoding processes, leading to structured visual descriptions. Following Pinker's (1990) graph comprehension architecture, as introduced in Chapter 2, the visual processes transform external information induced by graphical entities—such as shape and position of graph line segments—into *visual*

[9] It is likely that for modification of the concepts of spatial change, shape adverbial modifies such as *sharply* are frequently used in graph-text constellations, whereas temporal modifiers such as *suddenly* are used in text-only description of domain entities. This is an open question left for further research.

array in the early stages of comprehension. Visual encoding processes (cf. visual routines, Ullman, 1984) transform the visual array into a *visual description* by extracting abstract geometric properties and spatial relations. The visual description is the structural, symbolic representation of the graph. The visual array is constructed solely based on the perceptually driven information, whereas the visual description is not determined by the visual input alone; the visual routines are selected to meet specific computational goals (Ullman, 1984).

Visual routines are high-level primitives that extract spatial information from visual input (viz., base representation) according to specified computational goals. Visual routines are composed of a set of elementary visual operators; Ullman (1984) exemplifies these visual operators in the following way: shifting the processing focus, indexing a salient item for further processing, spreading activation over an area delimited by boundaries, tracing boundaries, and marking a location or object for future reference. These visual operators are not applied uniformly over the entire visual field but to objects or areas specified by computational goals.

Visual routines approach has been applied to investigation of bar graphs (Simkin & Hastie, 1987), pie charts (Spence & Lewandowsky, 1991), and to broader level qualitative reasoning of topological relations by humans (Lovett & Forbus, 2009a, b). The specification of particular elementary visual operators in comprehension of line graphs is beyond the scope of this dissertation. In the analysis presented in this chapter, the system of spatial representations is considered a descriptional inventory, which is accessed by visual routines. Visual segmentation of the line graph in Figure 4.1a leads to the segmentation of the visually salient parts of the graph—inter alia—a line with a vertical overall direction possessing one local maximum of curvature.[10] Figure 4.2a depicts the correspondence between an idealized shape of this type and its structured description by a spatial representation. Figure 4.2b depicts the correspondence between a complex part of the line graph, namely a sequence of line segments, which has a horizontal overall orientation, and an abbreviated spatial description of that graphical constellation. The visual encoding processes that transform the visual array into a visual description are contributed by the perceptually induced Spatial Reference System (SRS_P), which is a means to determine the global orientation of lines, such as their horizontality and verticality. Accordingly, the VERTICALITY and HORIZONTALITY of the graphical entities in the descriptions are provided by the SRS_P. [11]

[10] As stated in Chapter 2, De Winter & Wagemans (2004, 2006) give a thorough overview about segmentation processes in perceiving line drawings. Hoffman and Singh (1997) and Cohen and Singh (2007) describe how geometric properties determine segmentation of contour lines. The results of both approaches are in agreement with the approach presented in this dissertation, which focuses on a specific types of lines, namely graph lines, in a specific comprehension context.

[11] Humans are able to conceptualize information induced by graphical entities at different levels of granularity. The INCREASE concept can be induced by graph lines with different shape properties such

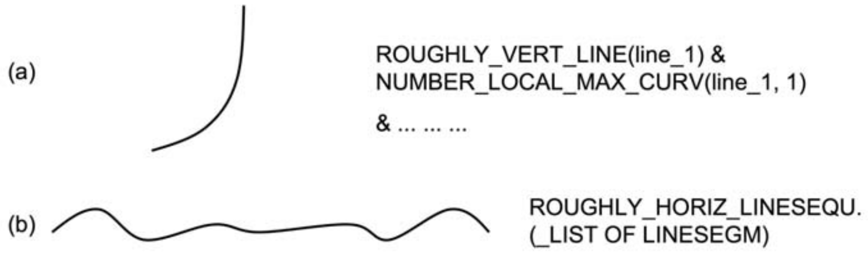


Figure 4.2. Graphical entities and their spatial descriptions.

To sum up, the core notions that characterize graphical entities in the initial stages of graph comprehension are the perceptually induced spatial reference system (SRS_P) and the visual description of object shape. The visual encoding processes that transform visual array into visual description are domain-independent processes. It is the *graph schema instantiation* that makes possible the processing of perceptual information provided by the lines on the paper or on screen as entities belonging to a line graph. In other words, in contrast to the domain independent visual encoding processes, graph schema instantiation has the domain-specific task to interpret elements as parts of graphs.

The graph schema is a long term memory structure that includes information for specifications of gestalt atoms in a specific type of graph. In line graphs, the basic gestalt atoms are the diagonal lines ‘ / ’ and ‘ \ ’ for the INCREASE and DECREASE concepts. The graph schema instantiation transforms visual description into internal conceptual and spatial representations by the contribution of two additional knowledge sources, namely the graphically induced Spatial Reference System (SRS_G) and Geometric Concept Specification (GCS). In particular, in the line graph presented in Figure 4.1a, the x-axis represents to the temporal domain, which possesses *natural order* of time, and the y-axis is in agreement with the natural order of the NUMBER-domain. The resulting SRS_G allows the interpretation of the vertically extreme POINT_OF_MAXIMA_CURVATURE, which is characterized as the connection point between two roughly vertically oriented lines, as a PEAK of the graph line. Moreover, the SRS_G leads to an inherent orientation of the line segments. In other words, by means of the SRS_G undirected lines are interpreted as PATHs, i.e. directed linear entities. As a result, the leftmost part of the graph is interpreted as an INCREASE, the following—after the PEAK—is interpreted as a DECREASE. Accordingly, the corresponding entities at conceptual level are INCREASE_P*(_PATH, SRSG) and*

---

as curvature, orientation, and straightness. Shape simplification algorithms may approximate these characteristics (e.g., Barkowsky, Latecki, & Richter, 2000 for simplification of geographic shapes). In comprehension of graph-text constellations, the alignment of granularity between linguistically induced representations and graphically induced representations is a research topic that is beyond the scope of this dissertation.

*DECREASE*_P($\_{PATH}$, $\_{SRSG}$) where the *PATH* argument is used for the graphical entities as directed linear entities (cf. Habel, 1990; Eschenbach et al., 2000).

In summary, in the visual description, the lines are undirected geometrical entities that are characterized by their vertical or horizontal orientation. After the graph schema instantiation, graph lines can be interpreted as directed linear entities (i.e., paths). In other words, the graph schema represents information about *directedness* of these graphical entities.

In conclusion, the *common conceptual basis* for the specification of graphical entities and the specification of textual entities in graph-text constellations characterizes the working assumption in this dissertation. Based on this working assumption and the modes of description introduced in 4.1, a cognitive architecture for multimodal comprehension of graph-text constellations is proposed in the following part of the chapter.

4.3 THE COGNITIVE ARCHITECTURE

In a coarse-grained cognitive architecture, modality-specific processing of linguistic entities and graphical entities can be schematically represented in an information flow diagram, as presented in Figure 4.3.

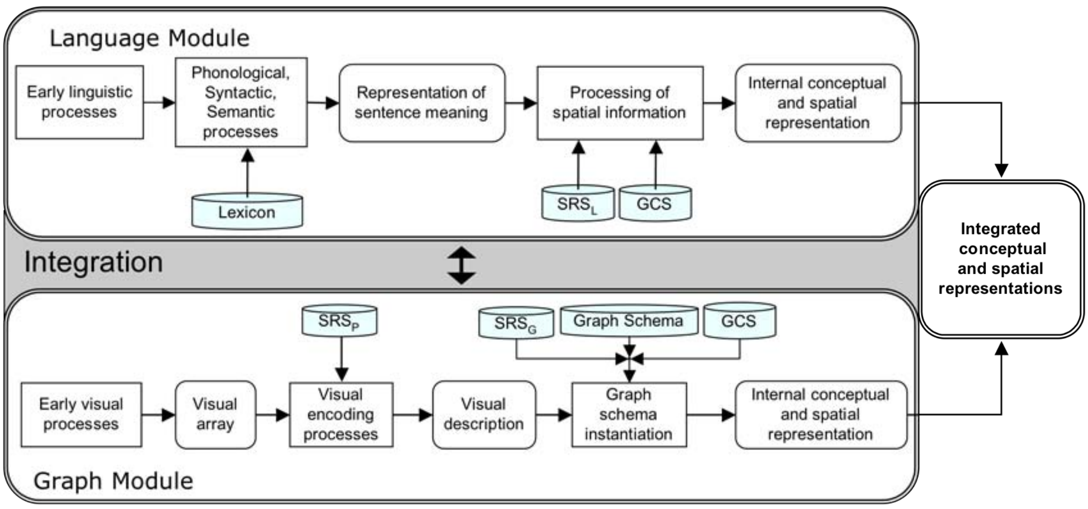


Figure 4.3. The computational architecture for multimodal comprehension of graph-text constellations.

In Figure 4.3, the modality-specific processors are shown by rectangles, the representations are shown by curved rectangles, and the knowledge sources are shown by cylindrical shapes. The processors are grouped under two major modules, namely the *language module* and the *graph module*. The arrows are used to show the information flow. The term *integration* is used for the flow of information contributed by the two representations, and the resulting representations are called *integrated conceptual and spatial representations*. In the region between the language module and the graph module, the information flow is shown with a double-sided arrow. The integrated conceptual and spatial representations—shortly the *integrated representations*—are also shown by a curved rectangle.

The language module is based on CRIL (Conceptual Route Instruction Language) proposed by Tschander et al. (2003). Their approach focuses on 'verbally instructed navigation', which is a language comprehension task, in which the interpretation of verbal route description is based on the knowledge of spatial language and temporal and spatial concepts. Therefore, specific components for processing spatial concepts and matching spatial representations with (idealized) visual percepts are foregrounded in that approach. In the architecture depicted in Figure 4.3, the language module transforms the information induced by external linguistic representations (i.e., words, phrases, sentences) into internal conceptual and spatial representations that are introduced in 4.2.1. The term *early linguistic processes* is used for pre-morphemic processes during the initial stages of language comprehension. The phonological, syntactic and semantic processes are governed by a set of rules and constraints—often called grammar—and a set of processes of memory retrieval and reasoning that employ knowledge about the world. The nature of the underlying mechanisms of phonological, syntactic and semantic processes constitutes subdomains of linguistics research, the details of which are beyond the scope of this dissertation.

The graph module is an elaborated version of Pinker's (1990) graph comprehension architecture. As described in 4.2.2, visual encoding processes transform the visual array into visual description, which is then transformed into internal spatial and conceptual representations by the graph schema instantiation (see Chapter 2 for Pinker's theory of graph comprehension).

The internal conceptual and spatial representations contributed by the linguistic entities and the graphical entities are integrated to form *integrated conceptual and spatial representations*. The idea of *integrated* representations is a consequence of the working assumption for the *common conceptual basis* for the specification of graphical and textual entities in graph-text constellations. The structural properties of integrated representations are discussed below.

4.3.1 Integrated Conceptual and Spatial Representations

Despite the differences in underlying theories, interdisciplinary research on graph comprehension, language comprehension, and comprehension of pictorial or diagrammatic illustrations with accompanying text has lead to the common view that a layer of common conceptual representations is the place where *coreference* links among conceptual entities that are contributed by various modalities are constructed

(Bodemer & Faust, 2006), and where *inter*- and *intra-representational coherence* (i.e., inter- and intra-modal coherence) is established (Seufert, 2003). For instance, in the cognitive theory of multimedia learning Mayer (2001/2009, 2005b) proposes that when reading picture-text constellations, the reader separately constructs a verbal representation of the text and a nonverbal representation of the picture, and then integrates the two representations; this integrated representation corresponds to the mental model of the reader. In addition, in multimodal generation systems, the constituents of multimodal presentations are generated from an underlying common representation; this is used as a basic principle of multimodal generation (e.g., WIP by Wahlster, André, Finkler, Profitlich, & Rist, 1993).

Nevertheless, the available models of picture-text constellations and diagram-text constellations do not specify the details of the structure of verbal, nonverbal, and integrated representations, largely due to the lack of grammatical structures of the empirically investigated depictive representations (e.g., the development of lightning storms, working principles of car brakes and bicycle pumps, mechanical pulley systems, and so on).

In the cognitive architecture presented in this chapter, following the previous findings in relevant research fields, it is assumed that the integration takes place at conceptual level. In particular, in comprehension of graph-text constellations, the common core of integration is based on the process of anchoring the semantic properties of linguistically induced and graphically induced information via the PATH concept. As discussed in 4.2.1, from the perspective of linguistics, the verbs *increase* and *decrease* are classified as motion verbs, when they are used in their spatial sense. These verbs exhibit the relationship between duration of movement and (change in) height (or vertical position in space), by their arguments. The process concepts INCREASE and DECREASE stand for a specification of a mapping from the temporal domain to the value domain, i.e. a PATH in the value space. In addition to motion verbs, shape nouns, spatial prepositional phrases, and adverbial modifiers can access the same PATH concept by providing the necessary arguments.

In graphs, the graph lines represent the mapping between temporal aspects of motion and value aspects of motion. The graph lines access the PATH concept at conceptual level, which can also be accessed by the linguistic entities such as *increase*, *decrease*, and *decline*. The correspondence between the INCREASE/DECREASE concepts and the graphical entities is provided by the graph schema. In other words, the relationship between graphical entities and linguistic entities is revealed by the concepts that are accessible both by textual entities and graphical entities by means of the graph schema. Within the framework of the representational modularity hypothesis (Jackendoff, 1996, 2002, see Chapter 2), the construction of referential links between graphical and linguistic entities is achieved by means of the graph schema, which has reflexes in both CS (conceptual structure) and SpS (spatial structure). In particular, the geometric characterizations of the atomic components—contributed by both linguistic and graphical entities—are specified by the SpS.

During the course of comprehension of the graph-text constellation shown in Figure 4.1a at the beginning of the chapter, the modality-specific comprehension modules contribute via the conceptual representations and the referential links build up during comprehension to an integrated and, in principle, coherent interpretation of the graph-text constellation: The verb *decline* provides DECREASE conceptualizations; the shape noun ‘peak’ is source of two referential links: a referential links to a domain-entity, namely an approximate number of birds, and a referential link to a graphical entity by means of the graph schema.

The information flow architecture presented in this chapter is a coarse-grained model in the sense that it is underspecified with respect to the details of processes in multimodal comprehension. The first simplification is the linear flow of information in modality-specific modules. Second, the model assumes that the processes of multimodal comprehension check all the time for an opportunity to shift between representational modalities to integrate the information induced by the different modalities. Finally, the interaction of the processes with memory systems (i.e., working memory and long term memory systems) is not included to the architecture. In its recent form, the purpose of the architecture is to provide the basis for fine-grained models that are devoted to simulate specific information processing tasks. In Chapter 9, a set of specific information processing tasks are investigated for the development of a GOMS-style processing model. In 4.3.2 and 4.3.3 below, two fundamental characteristics of the integration, *combinatoriality* and *incrementality* are presented respectively.

4.3.2 Combinatory Characteristic of Integration

The graph schema provides a syntactic and combinatory system at the interface between CS and SpS for the integration of information contributed by graphical and linguistic entities. The concatenations of geometrical entities have counterparts in CS, which can be accessed by linguistics entities. For instance, using an additional concatenation concept CONCAT$(_{PATH}, _{PATH}, _{SRSG})$, a V-structure in a line graph, built by two paths anchored in the same reference system, can be specified by

$$\text{V-STRUCTURE}(\text{CONCAT}(PATH_1, PATH_2, SRS_G)) \Leftrightarrow_{\text{def}} \text{DECREASE\_P}(PATH_1, SRS_G) \wedge \text{INCREASE\_P}(PATH_2, SRS_G)$$

The right hand side of this coreference relation can also be accessed by the lexical entities ‘retrace’ and ‘rebound’ (see the sample graph-text constellation in Figure 4.1b presented at the beginning of the chapter).

As another example, using an additional composition concept COMPOSE$(_{PATH}, _{PATH}, _{SRSG})$, a graphical structure, which is shown on the right of Figure 4.4 can be specified by

INCREASE_FLUCTUATE-STRUCTURE(COMPOSE$(PATH_1, PATH_2, SRS_G)$) $\Leftrightarrow_{def}$ INCREASE_P$(PATH_1, SRS_G)$ $\wedge$ FLUCTUATE_P$(PATH_2, SRS_G)$

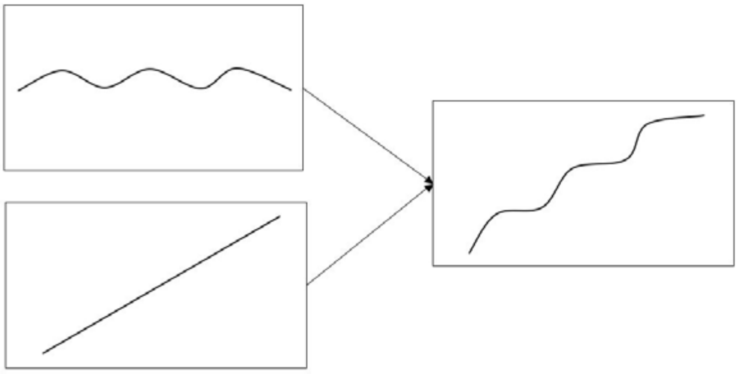


Figure 4.4. Graphical representation of the composition of FLUCTUATING INCREASE from FLUCTUATE_P and INCREASE_P.

Individuation of processes and events from graph lines is not a well-investigated domain, except for few experimental findings. In the literature on graph comprehension, the main determinant of graph complexity was proposed to be the number of trend reversals. According to Carswell, Emery, and Lonon (1993), the number of trend reversals in line graphs influences comprehension and verbal descriptions of graphs. However, research on graph comprehension is not at a mature state to explain the underlying nature of conceptualization of processes and events. Multimodal comprehension of graph-text constellation is a novel research domain, which may give insight for understanding of these conceptualizations. During the course of multimodal comprehension of a graph-text constellation, perceived entities are used to build up a hierarchical structure at the conceptual level, which can also be accessed by linguistic entities (cf. Guhe & Habel, 2001). Conceptualization of events and processes represented by graph lines, in other words the individuation of events that are represented by line segments in line graphs can be characterized by the *cut-hypothesis* (Avrahami & Kareev, 1994). The cut-hypothesis states that “a sub-sequence of stimuli is cut out of a sequence to become a cognitive entity if has been experienced many times in different contexts” (p. 239). In multimodal comprehension of graph-text constellations, the correspondence between the segmentation of graph lines in line graphs and the segmentation of processes and events that are represented by the graph line segments reflects the correspondence between our prior knowledge of segmentation of curves and our prior knowledge of segmentation of processes and events. In other words, the graph line segments—as perceived entities—can be cut out to become cognitive entities. In multimodal comprehension of graph-text constellations, these cognitive entities can also be accessed by comprehension of linguistic representations. Chapter 7 presents experimental investigations for conceptualization of events and processes represented by graph lines and their influence on causal attribution and temporal aspect concerning processes and events represented by verbal annotations.[12]

The second fundamental characteristic of integration in multimodal comprehension of graph-text constellations is incrementality, which is the subject of 4.3.3 below.

4.3.3 Incremental Characteristic of Integration

Most of the research in language comprehension and graph comprehension show that humans do language comprehension and graph comprehension incrementally (Carpenter & Shah, 1998, for incrementality in graph comprehension processes). The incremental processing is also the underlying assumption in language production (e.g., Guhe et al., 2004) and multimodal generation systems (e.g., Wahlster et al., 1993). Nevertheless, there is no systematic research on incremental characteristic of integration in multimodal comprehension of graph-text constellations.

The incremental characteristic of integration in comprehension of graph-text constellations is connected to four core research questions concerning the internal structure of the integration module: (a) which level of incremental entities are involved in integration?, (b) which types of representations are constructed by the modality-specific modules to be transferred to integration?, (c) how are these representations constructed by modality-specific modules be processed?, and (d) how do integrated representations influence modality-specific comprehension?.

The aspects of questions (b) and (c), in particular the ones about the construction of referential and coreferential links, with respect to the role of spatial and conceptual representations have been discussed in this chapter. These links are the backbone of inter-modal coherence in the resulting integrated conceptual and spatial representations that are built up interactively by the modality-specific comprehension modules. Theoretical investigations of the answers to the other two questions are left for further study. In Section 2 of the dissertation, the incremental characteristics of comprehension processes are experimentally investigated by the analysis of eye movements of participants during reading of graph-text constellations.

---

[12] In this dissertation, it is assumed that graph lines in line graphs represent process concepts such as INCREASE and DECREASE. Events, on the other hand, start or end processes. The difference between events and processes are beyond the scope of this dissertation. Therefore, the terms *event* and *process* are used for both graphical entities and verbal annotations.

4.4 SUMMARY

This chapter presented a coarse-grained cognitive architecture for multimodal comprehension of graph-text constellations. The main working assumption that underlies the infrastructure of the presented architecture is the common conceptual basis for the specification of linguistic and graphical entities in graph-text constellations. Coherent with the working assumption, the structural aspects of multimodal comprehension were investigated. In particular, the PATH concept was introduced as the basis of coreference construction between graphical and linguistic entities by means of graph schema instantiation.

The empirical studies reported in the following chapters (Section 2 of the dissertation) present empirical evidence for the construction of referential links between graphical and linguistic entities in graph-text constellations, based on the analysis of participants' eye movements during their reading of graph-text constellations. In particular, the empirical investigations present in Chapter 6, Chapter 7, and Chapter 8 investigate how readers integrate the information induced by the representations to reach a coherent interpretation and how the changes in one modality influence the comprehension of the other modality. Finally, Chapter 9 presents a GOMS-style cognitive processing model as a specific instance of the information processing tasks that are conducted within the framework of the introduced multimodal comprehension architecture.

Section 2

Construction of Inter-Modal Coreference and Coherence Relations: Empirical Investigations

5 Introduction to the Empirical Investigations of Graph-Text Comprehension

The empirical investigations reported in Section 2 of the dissertation are conceived as the analyses of multimodal comprehension of graph-text constellations in terms of the interaction between three groups of relevant factors: characteristics of users (i.e., participants), characteristics of the visual display of information (i.e., the presented stimuli and the presentation media), and the methods of investigation (e.g., eye tracking, posttest questions). The relevant factors are schematized in Figure 5.1. The items in the boxes exemplify the particular relevant factors in each group.[1]

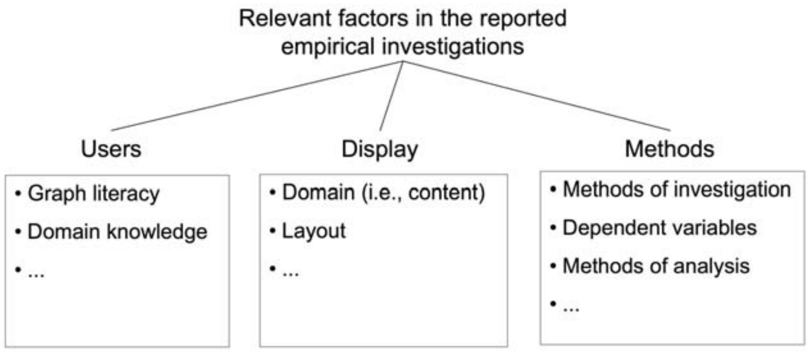


Figure 5.1. A schematic representation of the relevant factors in the reported empirical investigations.

These factors and their treatment in the reported empirical investigations are discussed in the following three parts of the chapter.[2]

---

[1] The terms *empirical* and *experimental* are used interchangeably in this dissertation.

[2] Kindred approaches in graph comprehension research emphasize the role of similar relevant factors in empirical investigations. For instance, according to Shah and Hoeffner (2002) the three main factors that influence the interpretation of graphs are the characteristics of the visual display, the reader's knowledge of graphical schemas and conventions, and the content of the graph and the reader's prior knowledge and expectations about the content. In this dissertation the reader's prior knowledge of the content of the graph is called *domain knowledge* and the reader's knowledge of the graph schema and conventions is called *graph literacy*.

5.1 USERS

The term *user* is used for the *participant* who attended to the experimental investigations. The terms *reader* and *comprehender* are also used interchangeably for the same meaning. Two major factors that are relevant to the empirical investigations of multimodal graph-text comprehension are the prior knowledge of the users and the skills and abilities of the users.

The prior knowledge of a user is a factor that influences his/her comprehension and learning (see Cook, 2006, for a review). Two types of prior knowledge are relevant to the study of comprehension of graph-text constellations: domain knowledge and graph literacy.

The term *domain knowledge* is used for the knowledge of the user about the content of a graph-text constellation. For instance, if the graph-text constellation is about the stock market, then the domain knowledge of the user is his/her knowledge of the stock market. It was stated in Chapter 1 that *experts* have a more abstract representation of domain-specific problems than *non-experts*, who mostly rely on problems' literal features (Chi, Feltovich, & Glaser, 1981). Accordingly, *non-expert* people have difficulties in integrating the information in different representational modalities compared to domain experts (e.g., Tabachneck-Schijf et al., 1997; Paoletti, 2004). In the domain of multimedia learning, the findings show that picture-text constellations are more useful for *low-knowledge* learners than *high-knowledge* learners, because high-knowledge learners more easily construct the necessary image representations out of text-only material than low-knowledge learners (cf. the individual differences principle, Mayer, 2001/2009).[3] Although the implications of the findings in multimedia learning are not clear at this stage, the reported differences between domain experts and non-experts suggest that domain knowledge is a factor that should be addressed in an empirical study multimodal comprehension of graph-text constellations.

One method to treat this issue is to test users for their knowledge of the domain, to define subgroups of users according to the test scores, and to compare the results obtained in the subgroups. However, the experimental investigations that are reported in this dissertation mostly have an exploratory character. The focus of experimental investigations is to find general patterns in participants' comprehension of graph-text constellations without a specific focus on the differences between subgroups of participants. Therefore, an alternative method was adopted in this dissertation, which was based on selecting the experimental stimuli such that the stimuli involve a variety of domains rather than a single domain. This method reduces the influence of the

[3] The terms *high-knowledge* and *low-knowledge* in the domain of multimedia learning research roughly correspond to the terms *expert* and *non-expert* in other domains, such as discourse comprehension.

domain knowledge effects by homogenizing it over the experimental stimuli, rather than fully treating them as a dependent variable. Sample domains that were selected in the experimental investigations included economics (Chapter 7, Study 1), wintering bird trends in a lagoon (Chapter 6, Chapter 9), and basic climate characteristics of cities (see Chapter 8). In addition, where experimenter-generated materials were used for a better control of experimental conditions, either a fictive made-up terminology (see Chapter 7, Study 3) or the terminology used in daily-life (e.g., change in temperature, see Chapter 7, Study 2) was employed to lessen the influence of prior domain knowledge of the participants on the experimental results. The issue of stimuli selection is discussed in more detail in part 5.2.

In addition, participants with different academic backgrounds were involved in the experiments. For this purpose, the experiments were announced at shared public locations such as libraries and cafeterias at the experiment site, the Middle East Technical University campus, Ankara. Participants from all major disciplines attended to the experiments, including engineering, natural and social sciences, educational sciences, economic and administrative sciences, and architecture. Consequently, the selected participants were non-experts (i.e., laypersons) with respect to the selected domains. In other words, the participants were not stock market specialists or tourism professionals but students from various disciplines.

Concerning the role of *graph literacy* in comprehension of graph-text constellations, as recognized by graph comprehension researchers the knowledge of graph schema (cf. Pinker, 1990) is a necessary knowledge source in graph comprehension. Therefore, graph literacy should be addressed as a factor in empirical investigations of multimodal comprehension of graph-text constellations. The participants of the reported experiments in this dissertation were selected from undergraduate or graduate university students. In most countries, students are formally taught about line graphs in the early middle school (e.g., Hardy, Schneider, Jonen, Stern, & Möller, 2005 for Germany). Novick (2006) states that "According to NCTM (2000), students are ready to begin learning about schematic diagrams in late elementary or middle school" (p. 9). In Turkey, where the experiments were conducted, the students are also given a formal education of graphs in the middle school. As a consequence, it was virtually guaranteed that the participants had formal education of line graphs. In addition, the participants gave verbal consent to being able to interpret the graphs that were presented in instruction session and the practice session of the experiments.[4]

[4] In addition, there is partial evidence that younger people, who lack formal education of graphs, are able to interpret graphs. For instance, Ainley, Nardi, and Pratt (1998) found that 8-9 years old children were able to engage in the task of using a data spreadsheet to produce a line graph despite the lack of explicit formal education. Mevarech and Kramarsky (1997) proposed that the eighth grade students did not start formal education of graphs as a *tabula rasa*, i.e. they had rather some knowledge of graphical representations before formal education. In short, the available findings show that people are exposed to graphs before formal education, and they could become graph-literate before the graduate-level education. The participants of the reported empirical investigations in this dissertation were university

The stimuli of all the experimental investigations reported in the dissertation included line graphs that have time as the independent parameter. As noted by Leinhardt et al. (1990) students more easily interpret line graphs that have time as the independent parameter compared to line graphs in non-temporal domains. This is possibly due to the more frequent use of time-domain line graphs than others in educational curricula. To sum up, the participants of the reported empirical investigations were non-experts with respect to the selected domains but they had a high level of graph literacy.

In addition to domain knowledge and graph literacy, users' skills and abilities (cf. cognitive styles) also comprise a group of factors that are relevant to the empirical investigation of graph-text constellations. However, in the current state of the art, the implications of the literature findings in cognitive styles are mixed, and their implications on graph comprehension and multimodal comprehension of graph-text constellations are not clear. Therefore, the investigation of the role of users' skills and abilities on multimodal comprehension of graph-text constellations is left for future research (see Chapter 11 for open questions and future research).

5.2 VISUAL DISPLAY OF INFORMATION

The term *visual display of information* is used for both the physical media of presentation (e.g., screen, paper) and the properties of graph-text constellations that are used as stimuli in the experimental investigations, such as the complexity of graphs.

Concerning the physical media of presentation, the graph-text constellations were presented on computer screen in all experiments, and the posttest questions were presented either on screen or on paper. In particular, a 50 Hz eye tracker recorded eye movements of the participants. The eye tracker was integrated into a 17" TFT monitor with a resolution of 1024 X 768 pixels. The participants were seated in front of the screen at an approximate viewing distance of 60 cm. Spatial resolution and accuracy of the eye tracker was about 0.25° and 0.50° respectively.

The stimuli selection in the reported empirical investigations of the dissertation reflects the method of experimental study in discourse comprehension research. At the one extreme, researchers use *naturalistic* expository or narrative text.[5] The use of

students. The investigation of multimodal comprehension of graph-text constellations in younger students is an open question for future research.

[5] The term *expository* is used in the sense in psychological research in text comprehension, where the term *expository text* is used for the major text type, which comprises textbooks, training manuals, software documentation, and so on. The other major type of text in text comprehension research is *narrative* text. The purpose of expository text is to communicate information for the purpose of informing the reader, whereas the purpose of narrative text is to entertain. On the other hand, the

naturalistic material introduces the advantage of ecological validity. However, the cost is the loss of experimental control over the stimulus material. At the other extreme, researchers use so-called *textoids*, which are experimenter-generated and crafted texts for controlling independent variables (Graesser et al., 1997). The use of textoids introduces the advantage designing controlled experiments with the cost of ecological validity (see Graesser et al., 1997, for a review). The approach towards the selection of stimuli for the reported empirical investigations was based on the selection of the stimuli from newspaper articles (Chapter 7, Study 1), travel guides (Chapter 8), and public reports (Chapter 6, Chapter 9). In other words, the stimuli that were used in the experimental investigations are naturalistic expository stimuli. In two studies, the stimuli were designed by the experimenter (Chapter 7, Study 2 & Study 3).[6]

The particular shape properties of graphical entities (e.g., the complexity of the graph line in terms of smoothness and granularity) were not evaluated as an independent parameter in the experimental studies that are reported in the dissertation. This issue is left for a further study, as discussed in Chapter 11.

5.3 METHODS

The third major group of relevant factors to the reported empirical investigations is the specification of the methods. In particular, three subfactors are discussed below: the selection of the empirical research paradigm, the specification of the dependent variables, and the specifications concerning the analysis of recorded data.

In Chapter 3, the eye tracking methodology was introduced as a research paradigm for empirical investigation of multimodal comprehension of graph-text constellations. Complementary techniques such as posttest questions and subjective judgment reports were also employed in the experiments. The dependent variables in the eye tracking methodology, the analyses methods, and the criteria for elimination noisy data are presented below.

The dependent variables in the eye tracking methodology can be classified into two major groups. The first group of variables is used in research on oculomotor system; these are motor variables such as velocity, acceleration and deceleration of saccades and other types of eye movement. The second group comprises time-related variables such as fixation duration, gaze time, and latency; and location-related variables such as fixation location, fixation count on a certain location, saccade amplitude, and

border between expository text and narrative text is not always clear (Weaver & Kintsch, 1991; Millis, Graesser, & Haberlandt, 1993).

[6] All the experimental material reported in this dissertation was translated into Turkish, the language of the experiment, and into English, where necessary, by the author of the dissertation. All the participants were native speakers of the experiment language.

saccade direction. The dependent variables in the second group are used in studies that use eye movements as a measure of cognitive processes. Accordingly, the experimental investigations of the dissertation employed the second group of eye movement variables for the analyses of the results. In the following paragraphs, these variables are called *eye movement parameters*.

The most frequently used eye movement parameters are *fixation count* (the frequency of the fixations), *fixation duration* (the mean duration of fixations), and *gaze time* (the sum of the fixation durations). A frequently used method for the analysis of eye movement data is to divide the stimuli into multiple *Areas of Interest* (AOIs, or Regions of Interest, ROIs) and to calculate the eye movement parameters for each AOI. The AOI specifications are presented in the chapters where experiments are reported.

A less frequently used method for the analysis of eye movement data is the analysis of gaze patterns, i.e. saccade-fixation-saccade sequences on the stimuli.[7] The analysis is often performed over spatial distribution of gaze patterns on the stimuli and more rarely, by their order of occurrence. The analysis of gaze patterns goes back to 1970s (cf. the scanpath theory and the string editing analysis by Noton & Stark, 1971; Brandt & Stark, 1997).[8]

Despite the early appearance of methods for the analysis of eye movement patterns, due to difficulties in quantification of eye movement patterns and the high variability of gaze patterns both between participants and between the gazes of the same participant on a scene at different times, the measurement of eye movement parameters—in particular, fixation count, fixation duration, and gaze time—found wider acceptance among the researchers who use eye tracking as research methodology. In the majority of the empirical investigations reported in the dissertation (Chapter 6, 7, and 8), these eye movement parameters were used. In the empirical investigations that are reported in Chapter 6 and Chapter 7 (Study 2) of the dissertation, gaze patterns are investigated in terms of the number of shifts of gaze between the text layer and the graph layer of the graph-text constellations. In Chapter 9, gaze patterns are used for the development of a GOMS-style computational model of graph-text comprehension. In particular, *fixation maps* are used depict transition eye movements by representing gaze patterns by arrows between the specified AOIs. The fixation maps are also represented as *transition matrices* (see Chapter 9 for

---

[7] The terms *scanpath*, *eye movement pattern*, and *fixation pattern* are also used interchangeably with the term *gaze pattern*.

[8] In the string editing analysis, the distribution of fixations is coded by letters where the letters are assigned to grid cells of the fixation distribution. The letter strings are then compared. In Levenshtein's string edit method, for example, the similarity between two strings is described in terms of the Levenshtein Distance (LD). The LD defines the number of insertions and deletions on one of the strings that is necessary to make that string the same as the other.

samples and for analyzed data, also see Ponsoda, Scott, & Finlay, 1995; Goldberg & Kotval, 1999; Webb & Renshaw, 2008, for the use of transition matrix in eye tracking research; see Henderson, 2003 for a review of methods for the analysis of human gaze control in scene viewing).

Depending on the technical specifications of the eye trackers, eye movement data may be noisy due to very short fixations, saccades, and blinks. Researchers specify a set of criteria to reduce the effect of noisy data. These criteria generally involve setting a lower limit for fixation duration or setting a lower limit for pixel radius that defines consecutive fixations that fall into the circle specified by the pixel radius as a single fixation. In some cases, these criteria can be applied by setting the filtering parameters of the integrated software tools; in others, filtering is done manually on the transcribed data.

The previous empirical findings about eye movement characteristics in reading, in visual scene viewing, and in multimodal comprehension provide the background knowledge for the specification of the elimination criteria (see Chapter 3). A short overview of previous research shows that researchers have used different elimination criteria to reduce the effect of noisy data. For instance, Hegarty and Just (1993) eliminated the fixations on the diagram components (in their experiments, mechanical pulley diagrams) with a duration of less than 100 ms. The criterion was based on Loftus (1981) who stated that humans are unable to decode new information from a visual display in less than 100 ms. They also eliminated fixations on the text with a duration of less than 250 ms (1993, p. 729). Carpenter and Shah (1998) eliminated gaze fixations on statistical line graphs that had a duration of less than 250 ms. Underwood et al. (2004) eliminated fixations that had a duration of less than 60 ms. Their stimuli included constellations of pictorial illustrations and text in a sentence verification task.

In the experimental investigations reported in the dissertation, the elimination criteria for fixations were set to a lower limit of 100 ms or below for duration, and 35 pixels for pixel radius. Although the implications of the literature findings about scene viewing and text-only reading on multimodal comprehension of graph-text constellations are not clear at this stage, these criteria are acceptable considering the minimum required fixation duration of 150 ms to process information in scene viewing in each fixation (Rayner et al., 2009) and an average of 250 ms to process information in text-only reading without time limitations (Rayner, 1998). Therefore, the elimination criterion of 100 ms for the empirical investigation of multimodal graph-text comprehension is acceptable with respect to the research findings in the current state of the art. More specific information about experimental methodology can be found in the following chapters where empirical investigations are reported.

5.4 SUMMARY

The empirical investigations in the dissertation are conceived as the interaction between three major groups of relevant factors, namely users, visual display of information, and methods. This chapter introduced those factors and their treatment in the reported experiments. In particular, concerning the influence of the domain knowledge (i.e., the knowledge of the content of graph-text constellations) and graph literacy (i.e., the knowledge of graphs) of users on experiment results, the participants of the experimental investigations were non-experts (i.e., laypersons) about the domains of the graph-text constellations, but they had high graph literacy since they were undergraduate or graduate university students. Concerning the visual display of information, the majority of the stimuli were selected from publicly available expository graph-text constellations, such as newspaper articles, travel guides, and public reports. Concerning the methods of investigation, the eye tracking technique was employed in the experiments, accompanied by a set of complementary techniques such as posttest questions and subjective judgment reports of the participants. The eye movement data were analyzed in terms of the eye movement parameters, namely fixation count, fixation duration, and fixation location, and in terms of gaze patterns. Specific information about the participants, experiment design, procedure, and analyses are presented in the chapters where the experiments are reported. Open questions and topics for future research are presented in Chapter 11.

6 Coreference Relations: Graph Schema and the Role of Syntactic Complexity[1]

This chapter presents five experimental studies that investigated the construction of coreferential relations between graphical and linguistic entities in graph-text constellations. The major goal of the empirical investigation was to investigate how changes in the text-layer of graph-text constellations influence comprehension of the other modality, i.e. comprehension of graphs. In particular, the following research questions were central to the study.

- What are the characteristics of the input increments (i.e., gaze fixations) in terms of fixation locations? In particular, during their reading of a graph-text constellation with a separate graph-layer—text-layer layout, readers may exhibit a “read-the-whole-text and inspect-the-whole-graph” strategy or they may integrate the information induced by the graphical and textual entities incrementally, i.e. by means of reciprocal gaze shifts between the text layer and the graph layer. The answer to this question was sought by investigating the influence of the number of graph-related sentences in the text layer of the graph-text constellations.
- How does syntactic complexity of the text in a graph-text constellation influence the integration of the information induced by graphical entities and textual entities? To answer this question, a special case of syntactic complexity in Turkish was used. In the experimental investigations reported in this chapter, suspended affixation in Turkish is considered an instantiation of a general phenomenon of syntactic complexity.
- How do readers of graph-text constellations employ the graph schema knowledge to access conceptual representations by means of different modalities? In particular, it was stated in Chapter 4 that readers access the concepts such as PEAK by reading the term ‘peak’ in the text and inspecting a peak in the graph. To answer this question, participants’ inspections of the graph during their reading of the sentences in the text were analyzed.

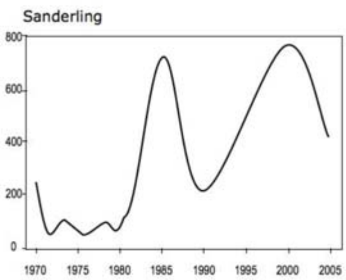
[1] The results of the experimental studies presented in this chapter were partially reported in Habel and Acarturk (2009).

In 6.1, a summary of the experimental investigations is presented in connection with the role of syntactic complexity in language comprehension and the potential influence of inconsistency between the order of the sentences in the text and the temporal order of the events and processes expressed by those sentences. The experiments are reported between 6.2 and 6.6. After an interim summary in 6.7, a more detailed analysis of participants' graph inspections is presented in 6.8.

6.1 OVERVIEW OF THE EXPERIMENTAL INVESTIGATIONS

In this part of the chapter, a summary of the reported experimental investigations is presented. The design of the experimental stimuli was based on the graph-text constellations presented in the waterbird census report that has already been excerpted in Chapter 1 and in Chapter 4. Figure 6.1a below shows a sample graph-text constellation from the experiment. Participants' eye movements were recorded by an eye tracker during their reading of the graph-text constellations. A sample eye movement protocol is presented in Figure 6.1b.

(a)



Sanderlings are small and plump species. They have a stout black bill, black legs and feet. Summer adults have a rufous head and breast, and a white belly. By winter, rufous areas have been replaced by pale gray, and birds look almost white and paler than any other shorebird. The number of birds in the lagoon peaked twice. The peaks occured at 1985 and 2000. The number of birds fell to about 200 in 1990. As of 2005, there are about 400 birds in the lagoon. Sanderlings have one of the most extensive winter ranges of all shorebirds in the world. They can be found worldwide on almost every sandy, coastal beach during the winter.

(b)

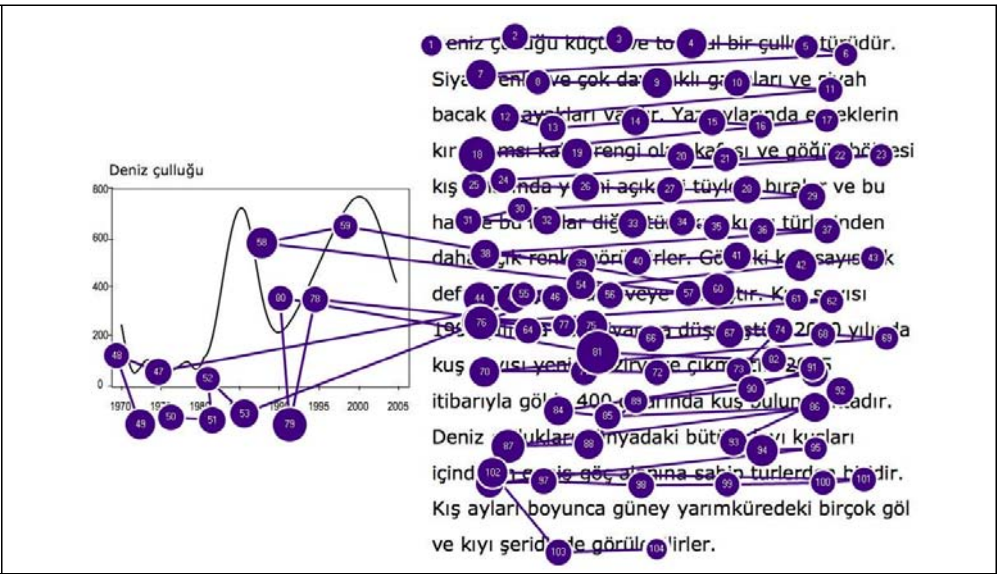


Figure 6.1. (a) A sample graph-text constellation as experimental stimulus, (b) a sample eye movement protocol from one participant. The experiment language was Turkish; all stimuli were translated by the author of this dissertation.

The eye movement recordings were analyzed in two dimensions with respect to the effort for the integration of information induced by the graphical entities and the linguistic entities. The first dimension was the number of shifts of gaze from the text layer (i.e., the paragraphs) to the graph layer (i.e., the graph) during participants' reading of the graph-related sentences in the text layer of the graph-text constellations (henceforth, gaze shifts). The results of the gaze shifts in Experiment 1 to Experiment 5 are presented in 6.2 to 6.6 respectively. The second dimension was the inspected locations on the graph layer during participants' reading of the textual entities in the graph-text constellations. The results of the inspected locations on graphs are presented in 6.8.

In all five experiments, the participants were presented graph-text constellations under different experimental conditions. A separate group of participants participated in each experiment. In Experiment 1, the text layer (i.e., the paragraphs) of the graph-text constellations involved one to four graph-related sentences (henceforth, target sentences). Three groups of graphs were used, each with a different shape property (see 6.2 for more information about the experimental conditions). In Experiment 2, another group of participants were presented the graph-text constellations that were used in Experiment 1. The difference between Experiment 2 and Experiment 1 was that in Experiment 2 the target sentences had a more complex syntax compared to the target sentences in Experiment 1. In particular, suspended affixation was used as the coordinate structure in Experiment 2 (see 6.1.1 and 6.1.2 for more information about syntactic complexity and suspended affixation respectively).

In both Experiment 1 and Experiment 2, there was no strict consistency between the order of the target sentences in the text and the temporal order of the events and processes expressed by the target sentences. The sentences in the source material (i.e., in the waterbird consensus report by PRBO) were used in their presented order, where the order of the sentences and the temporal order of occurrence of events were also not strictly consistent. On the other hand, as presented in 6.1.3, previous research on discourse comprehension shows that temporal order is a potential factor that may influence comprehension of language. Therefore, Experiment 3 and Experiment 4 were conducted with two different participant groups, as summarized below.

In Experiment 3, the target sentences were arranged such that the order of the target sentences and the temporal order of the events and processes expressed by the target sentences were strictly consistent. The rearrangement of the target sentences resulted in changes in the content of the target sentences as well as their order in the text compared to the stimuli used in Experiment 1 and Experiment 2. A fourth experiment was conducted with the stimuli of Experiment 3. The difference between Experiment 4 and Experiment 3 was that the target sentences of Experiment 4 had a more complex syntactic structure. In particular, suspended affixation was used as the coordinate structure in Experiment 4.

In a fifth experiment, the syntactic complexity of the target sentences was further altered. In Experiment 5, the syntactic complexity was reduced (compared to Experiment 4) by using a *partial* coordinate structure of suspended affixation in the target sentences. In Experiment 2 and Experiment 4, *fully coordinated* suspended affixation was used in the sense that all the verbs of the target sentences had suspended affixation, except for the last one as required by the suspended affixation coordination. In Experiment 5, the suspended affixation was *partially coordinated* in that not all but some of the target sentences involved suspended affixation. More information is presented in parts 6.3 and 6.6, where the experimental material for Experiment 2 and Experiment 5 are introduced respectively. Table 6.1 summarizes the design of the stimuli in the experiments.

Table 6.1. The design summary of the experiments.

	Suspended affixation	Temporal order
Experiment 1	no	not strict
Experiment 2	full	not strict
Experiment 3	no	strict
Experiment 4	full	strict
Experiment 5	partial	strict

A short review of the role of syntactic complexity in language comprehension is presented below.

6.1.1 A Short Review of the Role of Syntactic Complexity in Language Comprehension

As stated in Section 1 of the dissertation, language comprehension includes phonological, syntactic and semantic processes, which are governed by a set of rules and constraints—often called grammar—and processes of memory retrieval and reasoning that employ the knowledge about the world. The findings in the past several decades show that constructing interpretations for sentences involves moment-by-moment integration of different information sources that are constrained by available computational resources (Marslen-Wilson & Tyler, 1980; Tanenhaus et al., 1995; Gibson, 2000). An essential part of constructing the interpretation is the construction of the relationships between noun phrases and verb phrases in a sentence. In complex sentences, achieving the construction of these relationships can be difficult (Gordon, Hendrick, & Johnson, 2001).

Within the framework of the efforts towards a theoretical unification in psychology urged by Newell (1990), researchers have proposed theories and experimental investigations of the use of computational resources (e.g., working memory) in sentence processing, based on the investigation of comprehension difficulties in complex sentences (e.g., Lewis, 1996; Warren & Gibson 2002; Gordon et al., 2001; see Gibson & Pearlmutter, 1998 and Lewis, Vasishth, & van Dyke, 2006 for reviews). Most of this research has been concerned with comprehension difficulties in multiple center embedded object relative clauses in English, as can be exemplified by (1).

(1) The salmon that the man that the dog chased smoked fell.

After being introduced by Miller and Chomsky (1963) as nested dependency structures, center embedding has been a well-investigated topic in psycholinguistics. The nesting of dependencies occurs when a constituent X is embedded in another constituent Y, with material in Y to both the left and right of X (Lewis, 1996, p. 94).

Gibson's (1998) *syntactic prediction locality theory* proposed that both the cost of integrating syntactic and conceptual structures in memory and the memory cost associated with keeping track of obligatory syntactic requirements were influenced by locality, i.e., the proximity between verbs and arguments. According to Gibson "the longer a predicted category must be kept in memory before the prediction is satisfied, the greater is the cost for maintaining that prediction; and … the greater the distance between an incoming word and the most local head or dependent to which it attaches, the greater the integration cost" (p. 1, also see Vasishth & Lewis, 2006, and Konieczny, 2000 for counterevidence, viz. antilocality effects). Grodner and Gibson (2005) stated that in comprehension of linguistic entities with long distance dependency, a memory bottleneck arises because hierarchical relations must be extracted from a linear input stream (p.262).

Further research revealed the influence of similarity-based encoding and retrieval and storage-load effects besides the locality and antilocality effects in sentence processing (see Lewis et al., 2006, for a review).

In summary, although the debate on the relation between principles and resources of working memory and language comprehension has not been settled in the current state of the art, it is generally accepted that in language comprehension, computation of the linguistic relations between distal parts of the sentences requires working memory resources (Lewis et al., 2006). One instantiation of such a syntactic complexity is suspended affixation in Turkish. Suspended affixation is the elision of otherwise-repeated suffixes in coordinated constructions in Turkish (i.e., ellipsis). Suspended affixation allows the construction of a single long sentence with multiple coordinands instead of multiple sentences. A short review of suspended affixation is presented below.

6.1.2 A Short Review of Suspended Affixation in Turkish

The term *suspended affixation* is used for the elision of certain identical suffixes and/or clitics in coordinated constructions that involve a series of conjoined phrases or clauses in Turkish (Göksel & Kerslake, 2005, p. 534; Kornfilt, 1996, 1997; Kabak, 2007). The most common types of suspended affixation involve the omission of copular markers, person markers, generalizing modality markers, and nominal inflectional markers. For example, the copular marker *–(y)DI* can be omitted in the first two predicates in (2a). The omission of *–(y)DI* from *giyerdi* and *alırdı* leaves the predicates *giyer* and *alır* as incomplete forms—also called ‘small words’ by Kornfilt (1996)—with respect to the meaning expressed by these predicates. Similarly, in (2b), the person marker *–lAr* can be omitted in the first predicate, leaving the predicate *gitmiş* incomplete (the examples are from Göksel & Kerslake, p. 534).

(2) a. *Ziya her sabah paltosunu* ***giyer*** *(. . .), şemsiyesini* ***alır*** *(. . .) ve işe* ***gider*****di**.
‘Every morning Ziya *would put on* his coat, *pick up* his umbrella and *leave* for work.’
b. *Eskiden Kayseri’ye* ***gitmiş*** *(. . .), hatta orada* ***yaşamış*****lar**.
‘Appearantly at one time they *went* to Kayseri and even *lived* there.’

In a similar manner, the generalizing modality marker *–DIr* can be omitted in the first two predicates *düşmüş* and *çıkmış* in (3), leaving the predicates incomplete. The last predicate *görülmüş*, on the other hand, involves the modality marker *–DIr*.

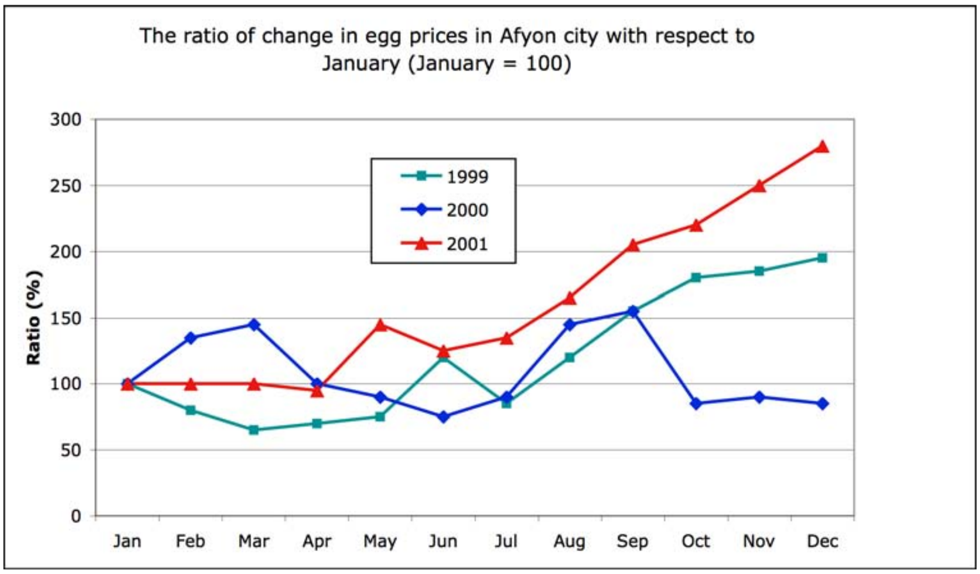
(3) *Göldeki kuş sayısı 1990 yılında 200 civarına* ***düşmüş*** *(. . .), 2000 yılında kuş sayısı yeniden zirveye* ***çıkmış*** *(. . .), 2005 itibarıyla gölde 100 civarında kuş* ***görülmüş*****tür**.
'The number of birds in the lagoon *fell* to about 200 in year 1990, in year 2000 the number of birds *peaked* again, as of 2005 about 100 birds were *seen* in the lagoon.'

In suspended affixation, coordination is not unbounded in terms of the distance (cf. long distance dependency) but unbounded in terms of the number of coordinands. However, suspended affixation is related to coordination and long distance dependency with respect to extraction and required memory resources for processing.

Suspended affixation in Turkish has been investigated with respect to the underlying phonological and morphological structures (see Kabak, 2007, and references therein). From the perspective of psycholinguistics, in the current state of the art, there is no research on suspended affixation in Turkish. On the other hand, suspended affixation is widely used in graph-text constellations both in daily communication such as graph-text constellations in newspaper and magazine articles, and in more formal reports that involve graph-text constellations.

As an example, Figure 6.2 shows an excerpt from the web portal of a government institution. The line graph represents the egg price for the 12 months of 1999, 2000, and 2001. The accompanying text describes the changes in the egg prices for the year 2000, using the information about the changes in prices represented by the graphical entities. It consists of two sentences. The first sentence involves two clauses coordinated by suspended affixation; one of the clauses involves a further embedded structure. The second sentence involves four clauses coordinated by suspended affixation. The translation of the text is presented in (4a–c).[2]

[2] The abbreviations used in literal translation (4b) are ACC: accusative, DAT: dative, GEN: genitive, AUX: auxiliary verb, PART: participle, ABL: ablative, SA: suspended affixation.



"... 2000 yılı Ocak ayında 29.500 TL. olan yumurtanın fiyatı Mart ayında 42.625 TL. ye yükselmiş, Nisan ayında 28.500 TL.ye düşmüştür. Haziran ayında ise 23.625 TL. [sic] *düşmüş,Eylül ayında 45.875 TL.ye çıkmış ve bu aydan sonra yıl sonuna kadar düşüşler olmuş ,Aralık ayında ise 35.264 TL.den kapanmıştır."*

Figure 6.2. A sample graph-text constellation with suspended affixation in the text layer. The graph was redrawn and the content was translated into English by the author of the dissertation (see below for the translation of the accompanying text). The original material was excerpted from a report of the Afyonkarahisar city Director's Office for Agriculture (a branch of the Ministry of Agriculture, Turkey); http://www.afyontarim.gov.tr/yapi/default.asp?sayfa=yumurta.htm, retrieved on June 26, 2009.

(4) a. *"... 2000 yılı Ocak ayında 29.500 TL. olan yumurtanın fiyatı Mart ayında 42.625 TL. ye yükselmiş, Nisan ayında 28.500 TL.ye düşmüştür. Haziran ayında ise 23.625 TL.* [sic] *düşmüş,Eylül ayında 45.875 TL.ye çıkmış ve bu aydan sonra yıl sonuna kadar düşüşler olmuş ,Aralık ayında ise 35.264 TL.den kapanmıştır."*

b. … 2000 year January month-LOC 29.500 TL AUX-PART- egg-GEN price-ACC March month-LOC 42.625 TL.-DAT increased(SA), April month-LOC 28.500 TL-DAT fell(SA). June month-LOC 23.625 TL[sic] fell(SA), September month-LOC 45.875 TL-DAT increased(SA) and this month-ABL after year end-DAT until decreases occurred(SA), December month-LOC 5.264 TL-ABL closed.

c. '… the price of egg which was about 29.500 TL in January 2000 increased to 42.625 in March, in April fell to 28.500 TL. In June (the price) decreased to 23625 TL, in September increased to 45.875 and after this month decreases occurred, in December (the price) closed by 35.264 TL.'

Based on the available experimental evidence in similar structures with syntactic complexity, it is assumed in this dissertation that high number of coordinands in suspended affixation construction results in a memory bottleneck. Mental representations of the coordinands degrade over time and with more input. As a result, a high number of target sentences with suspended affixation should result in more complex reading than the target sentences without suspended affixation. In graph-text constellations, the more complex reading of the target sentences should result in more gaze shifts because the comprehension of the graph-text constellations requires the integration of information induced by both modalities. In other words, since online integration processes require the integration of information induced by both modalities, the reading complexity should influence the integration characteristics in multimodal comprehension.

6.1.3 A Short Review of the Role of Temporal Order in Language Comprehension

The role of temporal order of events in comprehension has attracted the interest of discourse comprehension researchers mostly after the 1980s. In particular, researchers focused on the consistency between the organization of the events in the event world (viz., event structure) and the temporal arrangement of the events in the text (viz., discourse structure, Ohtsuka & Brewer, 1992)

Mandler (1986) found that the inconsistency between the chronological order of events and order of mention in the text results in long reading times compared to the text in which the chronological order of events and the order of mention in the text are consistent. Ohtsuka and Brewer (1992) reported similar findings by using a slightly different terminology: in canonical text—in which the event structure and the discourse structure were consistent—participants' comprehension of the text was superior to the text that involved backward flashback and embedded passages. In other words, the latter resulted in difficulties in comprehension. Zwaan (1996) observed that readers expect narrated events in adjacent clauses to follow a chronological order (viz., iconicity assumption). He found that the texts that violate the iconicity assumption led to difficulties in comprehension. More recent studies emphasize further aspects of temporal dimension such as the directionality and chronological distance between events (van der Meer, Beyer, Heinze, & Badel, 2002; Kelter, Kaup, & Claus, 2004), and introduce new experimental methodologies such as readers' monitoring of temporal order information in narrative texts by recording eye movements and by analyzing regressions (Rinck, Gámez, Díaz, de Vega, 2003).

In summary, previous research on discourse comprehension shows that temporal order is a potential factor that influences language comprehension. In particular, the inconsistency between event structure and discourse structure may lead to difficulties in comprehension. The implications for the findings in language comprehension to multimodal comprehension of graph-text constellations are not clear at this stage. However, the potential influence of temporal inconsistency between graphical representation of the processes and events and verbal representation of processes and events was considered in the design of materials for experimental investigation. In particular, the role of syntactic complexity in multimodal comprehension was investigated both with temporally not-strictly-consistent graph and text (cf. Experiment 1 and Experiment 2) and with temporally consistent graph and text (cf. Experiment 3, Experiment 4, and Experiment 5).

In the rest of this chapter the experimental investigations are reported.

6.2 EXPERIMENT 1

The purpose of Experiment 1 was to investigate the integration of the information induced by linguistic and graphical entities in graph-text constellations under different experimental conditions for the target sentences in the text layer and for the graphs in the graph layer of the constellations.

6.2.1 Participants, Materials, and Design

Thirty-six participants (mean age 23.3, SD = 4.39) were paid to participate in the experiment. The graph-text constellations used in Experiment 1 are exemplified in Figure 6.3. The accompanying text in each graph-text constellation is composed of three parts in the following order:

(1) *The pre-target sentences.* Each text started with several sentences that were not relevant to the information provided by the graph but about the appearance of the bird species, such as their shape of bills and their colors of plumage. These sentences are called pre-target sentences below. Accordingly, the reading of the pre-target sentences by the participants is called *pre-target reading phase*.

(2) *The target sentences.* After the pre-target sentences, each text included one, two, three, or four target sentences depending on the experimental condition (cf. the target sentence condition, see below).

(3) *The post-target sentences.* In each text, several post-target sentences followed the target sentences. Like the pre-target sentences, the post-target sentences were not relevant to the graph. The post-target sentences presented information such as diet characteristics and main food sources of the bird species, as well as migration and breeding characteristics.

Sample pre-target, post-target, and target sentences can be seen in Appendix B.1.1.

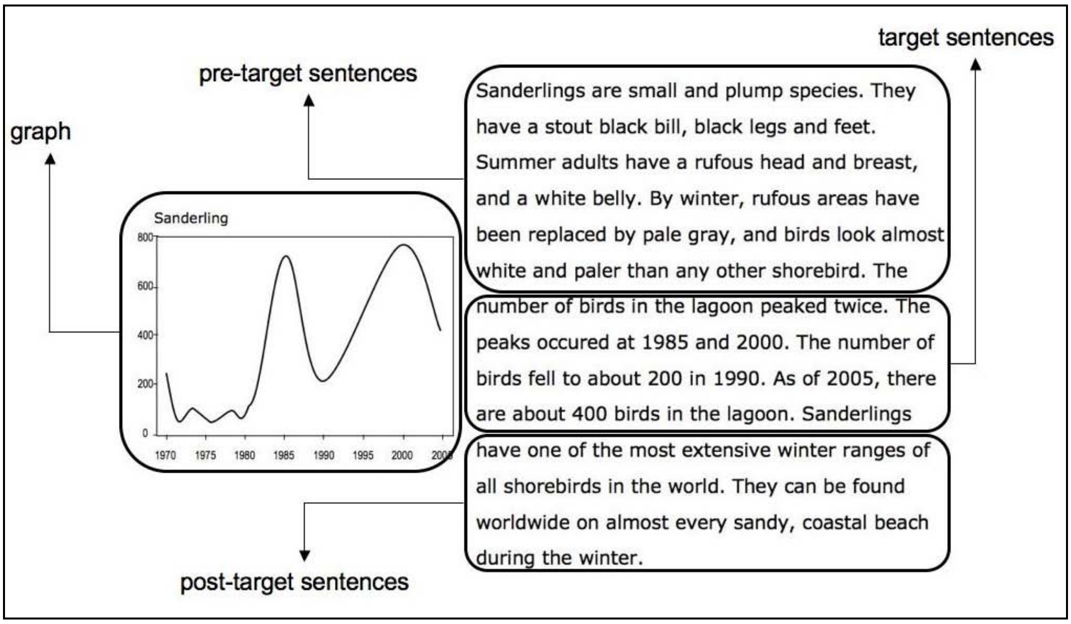


Figure 6.3. Sample pre-target, target, and post-target sentences in the exemplified stimuli. Approximate translations into Turkish and from Turkish into English are provided in the sample. A complete set of target sentences used in Experiment 3 and Experiment 4 can be found in part 6.8 of the chapter.

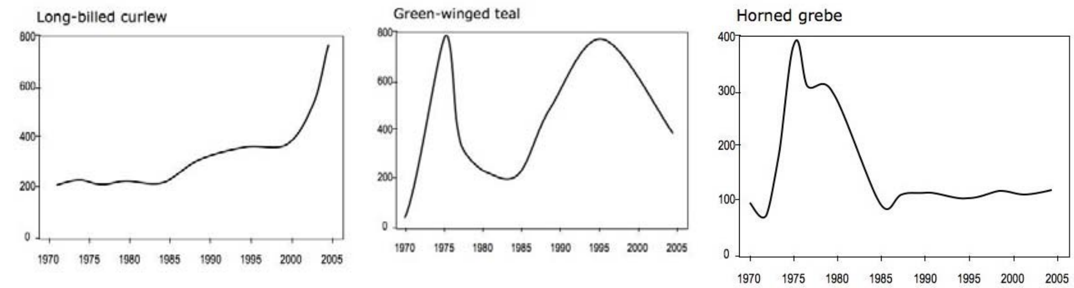
There were two within-subject factors in the experiment: the visual properties of the graph line (i.e., the *graph type* with three conditions) and the number of target sentences in the text (i.e., the *target sentence* with four conditions).[3] The three conditions of the graph type are presented below:

> *Condition 1 (GC-1)*. The graphs that were used in this condition involved an—almost—continuous curve that represented an increase or a decrease.
>
> *Condition 2 (GC-2)*. The graphs that were used in this condition involved a fluctuating curve with two-peaks. Accordingly, the graph lines were roughly composed of four or five line segments.
>
> *Condition3 (GC-3)*. The graphs that were used in this condition involved constellations of curves that represent decrease or increase and curves that represent constancy of value over time. Accordingly, the graph lines were roughly composed of two or three line segments.

For each graph type condition, four graphs were used to match the graph type conditions with the four target sentence conditions, as explained below. Where

[3] The term *visual property* is used as shorthand for the visually perceivable properties of the graph line. The terms such as *peak* were discussed in Chapter 4 from the perspective of being conceptual representations that are accessible by different modalities.

necessary, the graphs in the original sources were modified according experimental conditions specified by the target sentences. The graphs that were used in the three graph type conditions are exemplified in Figure 6.4.



(a) Condition 1 (GC-1) (b) Condition 2 (GC-2) (c) Condition 3 (GC-3)

Figure 6.4. Sample graphs that were used in three graph type conditions (a) GC-1, (b) GC-2, and (c) GC-3.

The target sentence conditions specified the number of target sentences in the text layer of the graph-text constellations. In the target sentence condition TC-1, the text included one target sentence. In the target sentence condition TC-2, the text included the sentence that was used in TC-1 and a second target sentence. In TC-3, a third target sentence was added to the previous two target sentences. In TC-4, four target sentences were used in the text. The distribution of target sentences according to target sentence condition is schematically presented in Figure 6.5.

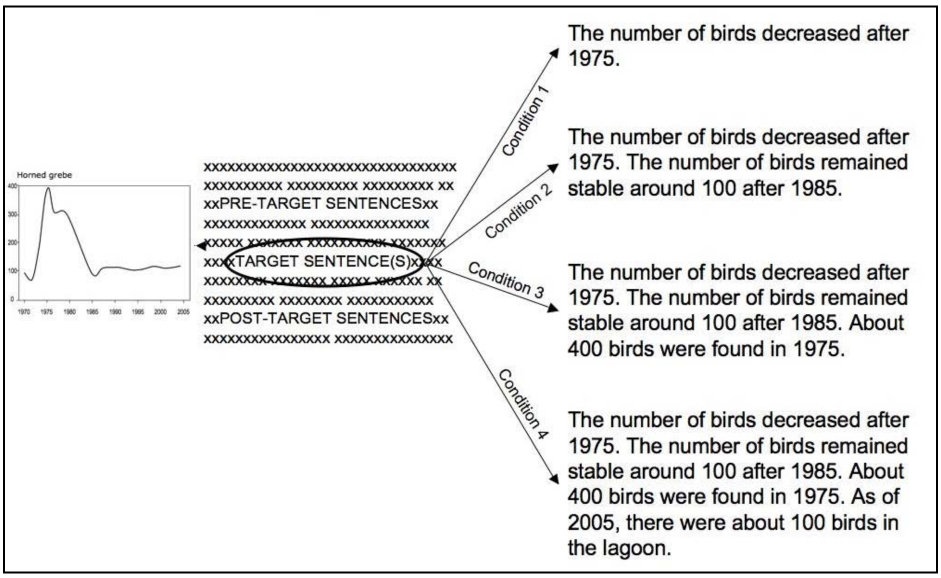


Figure 6.5. A schematic representation of the distribution of target sentences in the target sentence conditions TC-1, TC-2, TC-3, and TC-4.

The target sentences in the original source were modified for the purpose of experimental design, when necessary. In each of the three graph type conditions, four graphs that had similar visual properties were used.[4] Consequently, a total of 12 graphs were used in the experiment, four for each of the three graph conditions. Each graph was then matched with an accompanying text, which included one, two, three, or four target sentences (i.e., the target sentence conditions TC-1 to TC-4). As a result, a total of 48 graph-text constellations were prepared as experimental stimuli. The stimuli were equally distributed among the participants such that each participant was presented 12 graph-text constellations with three graph type conditions and four target sentence conditions. The presentation of the stimuli was randomized.

The stimuli included 12 multiple choice posttest questions. In the experiment, the presentation of each graph-text constellation was followed by a presentation of a posttest question. Six of the posttest questions were about the appearance of the bird species, such as their shape of bills and their colors of plumage; these questions are called *appearance questions* in the rest of this chapter. Since the appearance information was presented in the pre-target sentences in all graph-text constellations, these questions were at the same time about the information presented before the target sentences in the text. Three questions were about the graphs, asking whether the number of birds increased, decreased, remained constant, or peaked twice; these questions are called *graph questions*. The remaining three questions were about the main food sources of the bird species; these questions are called *feeding questions*. The diet characteristics of the bird species were presented in the post-target sentences in all graph-text constellations. Therefore, the feeding questions were the questions about the information presented after the target sentences.[5]

6.2.2 Procedure

The experiment was conducted in single sessions. The participants were first presented an *instructions* screen that included information about the experiment. The instructions can be seen in Appendix B.1.2. Briefly, the participants were informed that they were expected to investigate an inventory of 12 wintering bird species that migrate to Manyas Lagoon[6] for wintering since 1970, and answer the questions after

[4] In the graph condition GC-3, two types of graph lines were used; however, they were grouped as a single graph condition, namely GC-3, for simplicity. A more detailed analysis of the eye movement data is presented in 6.8.

[5] The posttest questions could further be classified according to the modality of information relevant to them: The appearance questions were based on the information represented in visual modality; the graph questions and the feeding questions were based on the information represented in graphical modality and propositional modality. However, this classification is not used in the analysis, because the order of information in the text layer of all graph-text constellations (i.e., information about appearance, information about graph, and information about feeding characteristics) was fixed, which prevents comparison among the modalities.

[6] Manyas Lagoon is a major side for breeding and wintering birds in northwest Anatolia.

the presentation of the inventory. They were also informed about the material, that they would see photos of the bird species and graphs that show population trends, i.e. the number of birds in years. The participants were asked to interpret the graphs given in the introduction session, and asked whether they had difficulties in interpreting the presented material, before the experiment session. There were no time limitations in the experiment. The entire session took approximately 20-25 minutes. Figure 6.6 shows the flow of stimuli as presented to the participants. Sample experimental stimuli (in Turkish) from Experiment 1 to Experiment 5 can be seen in Appendix B.1.3 to Appendix B.1.7 respectively.

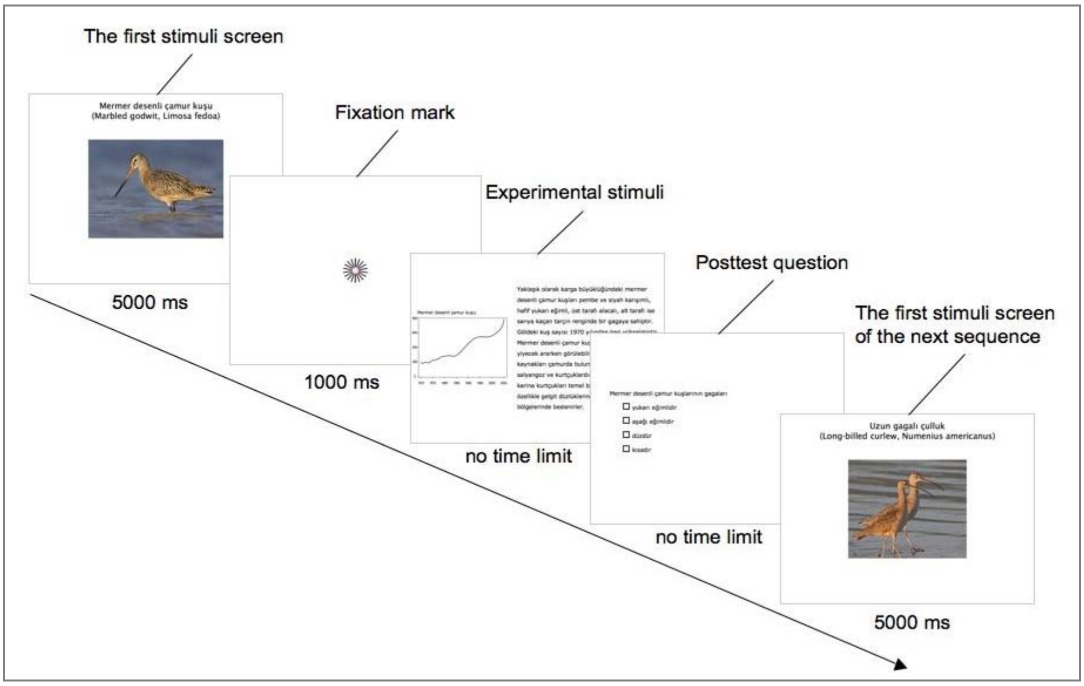


Figure 6.6. A schematic representation of the flow of stimuli as presented to the participants.

A 50 Hz. non-intrusive eye tracker recorded eye movements of the participants during their reading of the graph-text constellations (see 5.2 for the details of the eye tracker and the general description of the methodology). The results of Experiment 1 are presented below.

6.2.3 Results

The calibration of the eye tracker failed for three of the 36 participants. In addition, 7.32% of the data from the remaining 33 participants were not included to the analysis due to partial calibration problems. Fixations with duration of less than 60 ms were not included to the analysis. Concerning graph literacy, all the participants consented that they were able to interpret the exemplified graphs in the instructions session of the experiment.

The results are presented in three parts below. In the first part, the analysis of the shifts of gaze from the text layer to the graph layer of the graph-text constellations (henceforth, gaze shifts) is presented for the experimental conditions. In the second part, a more detailed analysis of gaze shifts is presented separately for each target sentence condition. In the third part, the analysis of the answers to the posttest questions is presented.

6.2.3.1 The Analysis of Gaze Shifts

In this part, the results of the analysis of participants' gaze shifts during their reading of the target sentences (henceforth, target-phase reading) are presented. The number of gaze shifts into the graph layer of the stimuli during target-phase reading was calculated for each graph condition and target sentence condition by the analysis of eye movement patterns. In the transcription of eye movements, a gaze shift was coded as belonging to a target sentence as long as the gaze shift occurred during the reading of the target sentence, during the reading of the last word of the target sentence, or during the reading of the first word of the next target sentence. In 55.3% of the presented graph-text constellations, the participants made at least one gaze shift into the graph layer during the target-phase reading. In 24.1% of the inspections with gaze shift, there was more than one gaze shift into the graph layer during reading the same target sentence (henceforth, regression).

The results showed that as the number of target sentences increased, the number of gaze shifts into the graph layer increased in all graph conditions. A two-way analysis of variance was conducted with two within-subject factors, graph type (graph conditions GC-1, GC-2, and GC-3) and the number of target sentences, (target sentence conditions TC-1, TC-2, TC-3, and TC-4). The target sentence main effect, the graph type main effect, and the Target sentence x Graph type interaction effect were tested using the multivariate criterion of Wilks's Lambda (Λ). The results of the analysis showed that the target sentence main effect was significant, $\Lambda = .30$, $F(3, 30) = 23.35$, $p < .01$, multivariate η^2=.70. The graph type main effect was significant, $\Lambda = .82$, $F(2, 31) = 3.42$, $p = .05$. The interaction effect Graph type x Target sentence was not significant, $\Lambda = .70$, $F(6, 27) = 1.95$, $p = .11$, multivariate η^2=.30.

Figure 6.7 shows the mean number of gaze shifts for the target sentence conditions and for the graph conditions.

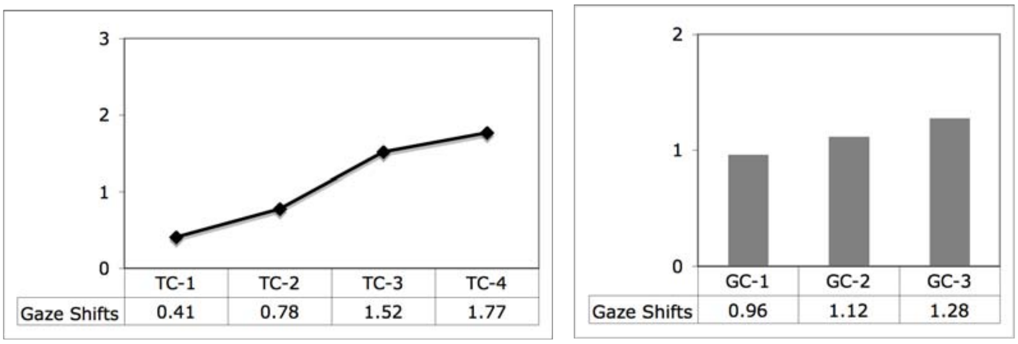


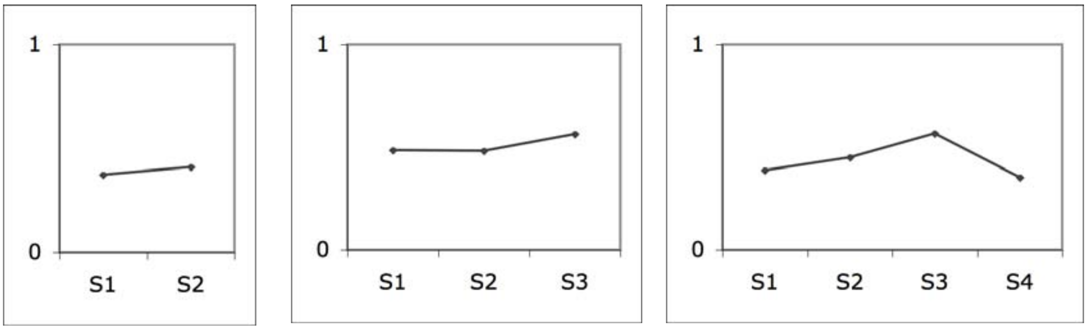
Figure 6.7. The mean number of gaze shifts into the graph layer in Experiment 1. The graphs show the results of the target sentence conditions (TC) and the graph conditions (GC).

Pairwise comparisons showed that all the differences between the target sentence conditions were significant except for the difference between TC-3 and TC-4. Concerning the graph conditions, only the difference between GC-1 and GC-3 was significant. Further pairwise comparisons can be seen in Appendix B.2.1.

In summary, the main finding of Experiment 1 was that as the number of target sentences in the text layer of the graph-text constellations increased the mean number of gaze shifts roughly increased. However, the increases exhibited different characteristics among the graph type conditions. Further analysis of gaze shifts is performed by calculating the mean number of gaze shifts during participants' reading of the target sentences in the target sentence conditions (henceforth, gaze shifts for the target sentences), as presented below.

6.2.3.2 Gaze Shifts for the Target Sentences

The focus of this analysis is the comparison of participants' mean number of gaze shifts during their reading of each target sentence. The results of the target sentence conditions TC-2, TC-3, and TC-4 are presented in Figure 6.8, without comparing the graph type conditions for brevity.



(a) Target Sentence Condition 2 (b) Target Sentence Condition 3 (c) Target Sentence Condition 4

Figure 6.8. The mean number of gaze shifts into graph layer during participants' reading of the target sentences (S) in Experiment 1.

In the target sentence conditions TC-2 and TC-3, there was no difference between the target sentences with respect to the mean number of gaze shifts into the graph layer. However, in TC- 4, the ANOVA test revealed a significant effect, $\Lambda = .74$, $F(3, 30) = 3.50$, $p < .05$. Further pairwise comparisons showed that in TC-4 (Figure 6.8c), the mean number of gaze shifts for the fourth target sentence was significantly lower than the mean for the third. However, the difference between Sentence 1 and Sentence 2 and the difference between Sentence 2 and Sentence 3 were not significant.

In summary, in the target sentence condition TC-2 the participants made almost equal number of gaze shifts during their reading of the first and the second target sentences. A similar finding was obtained for the first, second, and third target sentences in TC-3. In TC-4, however, the number of gaze shifts during participants' reading of the last sentence was lower than the ones during their reading of the third sentence. The findings will be discussed in comparison to the findings in Experiment 2.

It should be noted that the above analysis is an overall analysis for the target sentences, presented in this form for brevity. There were also differences among the graph type conditions, which can be seen the numerical details presented in Appendix B.2.1.

6.2.3.3 The Analysis of Answers to Posttest Questions

Each participant answered 12 posttest questions. A single posttest question was presented after each of the 12 graph-text constellations. Six questions were appearance questions, three were feeding questions and the remaining three were graph questions. The analysis of the answers to the posttest questions is presented below.

Answers from all 36 participants were included to the analysis. A total of 428 answers were obtained from the participants (two answers were not recorded due to a technical problem). For the analysis, each true answer was given the score *1*, each false answer was given the score *-1*. The scores for the two target sentence conditions TC-1 and TC-2 were combined, and the scores for the other two target sentence conditions TC-3 and TC-4 were combined to compare the low-number target sentence conditions with the high-number target sentence conditions. The graph condition was not analyzed as a factor in this analysis. Figure 6.9 shows the results. Numerical details can be seen in Appendix B.2.1.

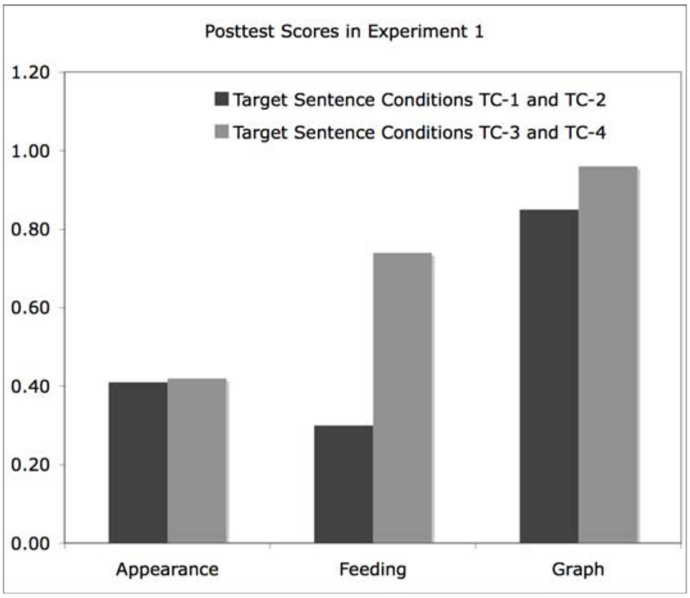


Figure 6.9. Posttest scores in Experiment 1.

A 3 x 2 ANOVA was conducted to evaluate the effects of three question types (namely, appearance questions, feeding questions, and graph questions) and two target sentence conditions (namely, TC-1/2 and TC-3/4) on posttest scores. The ANOVA indicated significant main effects for target sentence condition, $F(1, 424) = 5.34$, $p < .05$, multivariate $\eta^2 = .01$, and for question type, $F(2, 424) = 14.09$, $p < .05$, multivariate $\eta^2 = .06$, but no significant interaction between target sentence condition and question type, $F(1, 424) = 2.71$ $p = .07$, multivariate $\eta^2 = .01$. Follow-up tests were conducted to evaluate pairwise differences among the means. The analyses showed that there was no significant difference between the appearance scores and the feeding scores, but the difference between the graph scores and the appearance scores was significant; and the difference between the graph scores and the feeding scores was significant. Furthermore, the difference between the target sentence conditions was significant in the feeding scores, but not significant in the appearance scores, and in the graph scores.

In summary, the results showed that the answers to the graph questions received higher scores than the other two question types. The appearance scores were expectedly low, since the information about the appearance of the bird species were presented in the pre-target sentences. The target sentence conditions did not have influence on the appearance scores but had influence on the feeding questions. The results are further discussed in comparison with the results of the posttest scores in Experiment 2.

6.2.4 Summary of the Findings in Experiment 1

The first main finding of Experiment 1 was that as the number of target sentences increased the number of gaze shifts into the graph layer increased, though with different characteristics under different graphs type conditions. Moreover, the mean number of gaze shifts during participants' reading of each target sentence was almost equal, except for the decrease in the number of gaze shifts during the reading of the last target sentence when the text included four target sentences.

In Experiment 2, the graph-text constellations that were used in Experiment 1 were presented to another group of participants. The syntactic structure of the target sentences of the stimuli in Experiment 2 was changed to have a more complex syntactic structure compared to the ones in Experiment 1.

6.3 EXPERIMENT 2

The purpose of Experiment 2 was to investigate the integration of the information induced by graphical entities and linguistic entities in graph-text constellations when the graph-related sentences (viz., the target sentences) had increased syntactic complexity.

6.3.1 Participants, Materials, and Design

Twenty-five participants (mean age 22.2, SD = 2.24) were paid to participate in the experiment. The procedure in Experiment 2 was the same as the procedure in Experiment 1. The design of Experiment 2 was the same as the design of Experiment 1 with the difference being that the target sentences in Experiment 2 had increased the syntactic complexity. For this purpose, suspended affixation was used as the coordinate structure. The difference between the syntactic structures of the target sentences in the two experiments is schematically shown in Table 6.2.

Table 6.2. The target sentences in Experiment 1 and Experiment 2. The symbols S1 to S4 show the target sentences in each condition. The symbol (…) is used to mean that the generalizing modality marker *–DIr* was omitted in the predicate. In addition, in the target sentence conditions of Experiment 2 the non-final predicates were followed by a comma rather than a point.

Target sentence condition	Experiment 1	Experiment 2
TC-1	S1.	S1.
TC-2	S1. S2.	S1(…), S2.
TC-3	S1. S2. S3.	S1(…), S2(…), S3.
TC-4	S1. S2. S3. S4.	S1(…), S2(…), S3(…), S4.

It should be noted that in the sentences with suspended affixation, the *clauses* that are separated by comma are not complete sentences but incomplete clauses that are bound at the end of the sentence. Accordingly, the clause S1 in TC-2, the clauses S1 and S2 in TC-3, and the clauses S1, S2, and S3 in TC-4 are such incomplete clauses. On the other hand, the term *target sentence* will be used in the rest of this chapter to refer to these incomplete clauses, for the consistency of terminology.

6.3.2 Results

The calibration of the eye tracker failed for one of the 25 participants. In addition, 1.74% of the remaining data from 24 participants were not included to the analysis due to partial calibration problems. Fixations with duration of less than 60 ms were not included to the analysis. Concerning graph literacy, all the participants consented that they were able to interpret the exemplified graphs in the instructions session of the experiment.

The results are presented in three parts below. The first part presents the analysis of gaze shifts for the experimental conditions. The second part presents a more detailed analysis of the gaze shifts separately for the target sentence conditions. The third part presents the analysis of the answers to posttest questions.

6.3.2.1 The Analysis of Gaze Shifts

In 59.4% of the inspections of the presented graph-text constellations, the participants made at least one gaze shift into the graph layer during target-phase reading. In 32.7% of the inspections with gaze shift into the graph layer, there was more than one gaze shift during reading of the same target phrase (i.e., regression). Compared to the percentage of regressions in Experiment 1 (24.1%), the percentage of regressions was high in Experiment 2. This finding reflects the influence of using a relatively complex syntactic structure in the target sentences of Experiment 2.

The results of Experiment 2 are presented in Figure 6.10 below.

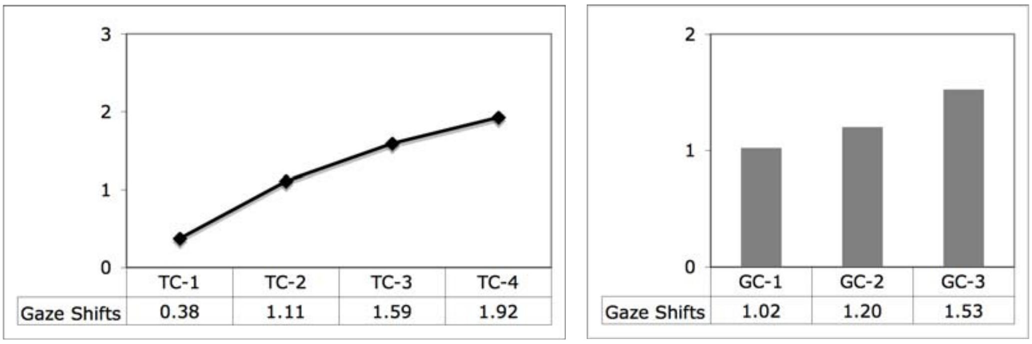


Figure 6.10. The mean number of gaze shifts into the graph layer in Experiment 2. The graphs show the results for the target sentence conditions (TC) on the left and for the graph conditions (GC) on the right.

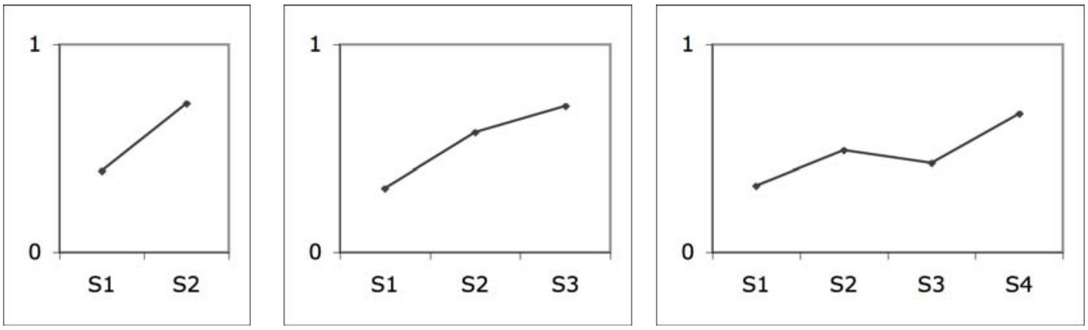
A two-way analysis of variance was conducted with two within-subject factors, graph type (graph conditions GC-1, GC-2, and GC-3) and the number of target sentences, (target sentence conditions TC-1, TC-2, TC-3, and TC-4). The results of the analysis showed that the target sentence main effect was significant, $\Lambda = .34$, $F(3, 21) = 13.79$, $p < .01$, multivariate η^2=.66. The graph type main effect was significant, $\Lambda = .70$, $F(2, 22) = 4.76$, $p < .05$, multivariate η^2=.30. The interaction effect Graph type x Target sentence was not significant, $\Lambda = .72$, $F(6, 18) = 1.60$, $p = .37$, multivariate η^2=.28.

Pairwise comparisons showed that all the differences between the target sentence conditions were significant except for the difference between TC-3 and TC-4. Concerning the graph conditions, only the difference between GC-1 and GC-3 was significant. Further pairwise comparisons can be seen in Appendix B.2.2.

In summary, the findings of Experiment 2 were parallel to the findings of Experiment 1 in that as the number of target sentences increased the number of gaze shifts into the graph layer increased. The characteristics of the increases were different among the graph type conditions. The patterns of overall difference were similar between Experiment 1 and Experiment 2. In both experiments, the difference in gaze shifts between TC-3 and TC-4 was not significant, and concerning the graph conditions, only the difference between GC-1 and GC-3 was significant in both experiments. Further analysis of the gaze shifts into the graph layer for the target sentences is presented below.

6.3.2.2 Gaze Shifts for the Target Sentences

In parallel to the analysis of the results in Experiment 1, the mean number of gaze shifts into the graph layer during participants' reading of the target sentences was calculated. Figure 6.11 shows the results of each target sentence condition. Numerical details can be found in Appendix B.2.2.



(a) Target Sentence Condition 2 (b) Target Sentence Condition 3 (c) Target Sentence Condition 4

Figure 6.11. The mean number of gaze shifts into graph layer during participants' reading of the target sentences (S) in Experiment 2.

In the target sentence condition TC-2, the difference in the mean number of gaze shifts between the two target sentences was significantly different, $t(23) = 2.51$, $p <$ 05. In TC-3, the ANOVA analysis revealed that there was significant difference among the means, $\Lambda = .64$, $F(2, 22) = 6.24$, $p < .05$. Further pairwise comparisons showed that in TC- 3, the mean number of gaze shifts for the second target sentence was significantly higher than the mean for the first target sentence. However, the difference between Sentence 2 and Sentence 3 was not significant. In the target sentence condition TC-4, the ANOVA analysis revealed that there was a significant difference among the means, $\Lambda = .71$, $F(3, 21) = 2.87$, $p = .06$. Further pairwise comparisons revealed a significant difference between the means for Sentence 3 and Sentence 4.

In summary, the mean number of gaze shifts for each target sentence was not equal in Experiment 2. Compared to the results obtained in Experiment 1, more significant increases were obtained in Experiment 2. In TC-2 and TC-3, the difference between the first and the second target sentence was significant. In TC-4, the difference between the third and last target sentence was significant. In particular, the mean number of gaze shifts was higher for S4 than S3, whereas in Experiment 3, S4 was lower than S3. Consequently, the results of Experiment 2 present evidence for the influence of syntactic complexity of the target sentences on the inspection of graphs.

As stated previously, the target sentences of Experiment 2—except for the last target sentence in each target sentence condition—were not complete sentences but incomplete clauses due to the use of suspended affixation as coordinate structure. These clauses were bound at the end of the last target sentence. In other words, in TC-2, the first target sentence was bound at the end of the target sentence, and in TC-

4, the first, the second, and the third target sentences were bound at the end of the fourth target sentence. The results of the mean number of participants' gaze shifts into the graph layer during their target-phase reading showed that the end of the last target sentence, where the previous clauses were bound, was the location where the participants made more gaze shifts into the graph layer, compared to the end of the clauses in the previous target sentences. On the other hand, in Experiment 1, where the syntactic complexity was low (i.e., there was no suspended affixation in the verbs), no such increase was observed.

As a result, the findings of the gaze shifts from the text layer to the graph layer of graph-text constellations in Experiment 2 shows the influence of syntactic complexity in the target sentences on inspection of graphs of the graph-text constellations.

6.3.2.3 The Analysis of Answers to Posttest Questions

As in Experiment 1, each participant responded to 12 posttest questions in Experiment 2. Answers from the 25 participants were included to the analysis. A total of 300 answers were recorded. Each true answer was given the score *1*, each false answer was given the score *-1*. The results are shown in Figure 6.12. Numerical details can be seen in Appendix B.2.2.

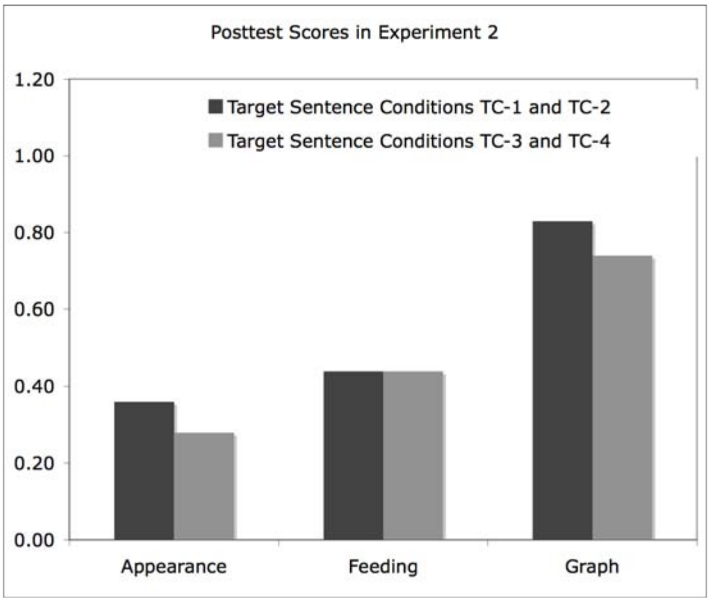


Figure 6.12. Posttest scores in Experiment 2.

A 3 x 2 ANOVA was conducted to evaluate the effects of three question types (namely, appearance questions, feeding questions, and graph questions) and two target sentence conditions (namely, TC-1/2 and TC-3/4) on posttest scores. The ANOVA indicated no significant main effect for target sentence condition, $F(1, 294)$

= .32, p = .57, multivariate η^2 = .00. However, the main effect for question type was significant, $F(2, 294) = 7.01$, $p < .05$, multivariate η^2 = .05. The interaction between target sentence condition and question type was not significant, $F(2, 294) = .35$ p = .95, multivariate η^2 = .00. Follow-up tests were conducted to evaluate the pairwise differences among the means. The analyses showed that there was no significant difference between the appearance scores and the feeding scores, but the difference between the graph scores and the appearance scores was significant; and the difference between the graph scores and the feeding scores was significant. Furthermore, the difference between the target sentence conditions was not significant in any of the three question types.

In summary, the results showed that, as in Experiment 1, the answers to the graph questions received higher scores than the other two question types. The appearance scores were expectedly low, since the information about the appearance of the bird species was presented in the pre-target sentences. The target sentence conditions did not have influence on any of the question types.

The difference in the feeding scores between the two target sentence conditions was not significant in Experiment 2, whereas the difference was significant in Experiment 1. The reason for the findings about the feeding scores and the decrease in the graph scores is not clear at this stage. On the other hand, compared to the appearance scores in Experiment 1, the appearance scores, the graph scores, and the feeding scores in Experiment 2 were low, in particular, in TC-3/TC-4. These findings indicate that the syntactic complexity of the target sentences resulted in participants' reduced recall of the content.

6.3.3 Summary of the Findings in Experiment 2

The results of Experiment 2 showed that, as found in Experiment 1, as the number of target sentences increased the number of gaze shifts increased, with different graph type conditions having different characteristics of increase. The overall patterns of differences in the target sentence conditions and the graph conditions were similar to the ones obtained in Experiment 2. On the other hand, further analysis of the gaze shifts showed that increasing the syntactic complexity of the target sentences resulted in a more complex inspection of the graphs as evidenced by the higher number of gaze shifts into the graph layer during participants' reading of the late target sentences compared to the early ones. In particular, the participants of Experiment 2 made the highest number of gaze shifts during reading of the last target sentence. This finding indicates the influence of the use of suspended affixation structure in the target sentences, in which all the previous target sentences were bound in the last target sentence. In summary, the results of Experiment 2 present evidence that the syntactic complexity of the graph-related (i.e., target) sentence in the text influences the integration characteristics of the information induced by linguistic and graphical modalities.

Although the results of the two experiments show the influence of syntactic complexity of the target sentences, as stated previously the order of the target sentences in the text was not strictly consistent with the temporal order of events and processes expressed by the target sentences. As a result, the order of the target sentences in the text was also not strictly consistent with the temporal order of the events and processes represented by the graphs. The previous studies in discourse comprehension show that this partial inconsistency may influence language comprehension, which may have implications for comprehension of graph-text constellations. Therefore, two more experiments (namely, Experiment 3 and Experiment 4) were conducted with two different participant groups to eliminate the potential influence of the temporal inconsistency in Experiment 1 and Experiment 2.

6.4 EXPERIMENT 3

The purpose of Experiment 3 was to investigate the integration of the information induced by linguistic and graphical entities in graph-text constellations. In Experiment 3, the temporal order of the events expressed by the target sentences was strictly consistent with the order of the target sentences in the text.

6.4.1 Participants, Materials, Design

Twenty-four participants (mean age 22.0, SD = 2.20) were paid to participate in the experiment. The graph-text constellations that were used in Experiment 1 and Experiment 2 were used in Experiment 3, though with changes in the order of the target sentences and where necessary in the content of the target sentences. For instance, the first target sentence in the second graph condition was ‘The number of birds peaked first in 1975’, whereas the first target sentence in Experiment 1 and Experiment 2 was ‘The number of birds peaked twice’. Overall, eight of the 48 target sentences were changed in the stimuli. The rest of the target sentences were the same. As a consequence, the results of Experiment 3 are not fully comparable with the results of Experiment 1. The procedure in Experiment 3 was the same as the procedure in the previous experiments.

6.4.2 Results

The calibration of the eye tracker failed for one of the 24 participants. In addition, 0.72% of the remaining data from 23 participants were not included to the analysis due to partial calibration problems. Fixations with duration of less than 60 ms were not included to the analysis. Concerning graph literacy, all the participants consented that they were able to interpret the exemplified graphs in the instructions session of the experiment.

The results are presented in three parts below. The first part presents the analysis of gaze shifts for the experimental conditions. The second part presents a more detailed analysis of gaze shifts separately for each target sentence condition. The third part presents the analysis of the answers to the posttest questions.

6.4.2.1 The Analysis of Gaze Shifts

In 65.7% of the inspections of the presented stimuli, the participants made at least one gaze shift into the graph layer during target-phase reading. In 15.0% of the inspections with gaze shift, there was more than one gaze shift into the graph layer during participants' reading of the same target phrase (i.e., regression). Compared to the percentage of regressions in Experiment 1 (24.1%) and in Experiment 2 (32.7%), the percentage of regressions was low in Experiment 3. The decrease in the percentage of regressions might have been due to the consistent temporal order in the text, as well as due to the changes in the target sentences. The results of Experiment 3 are presented in Figure 6.13.

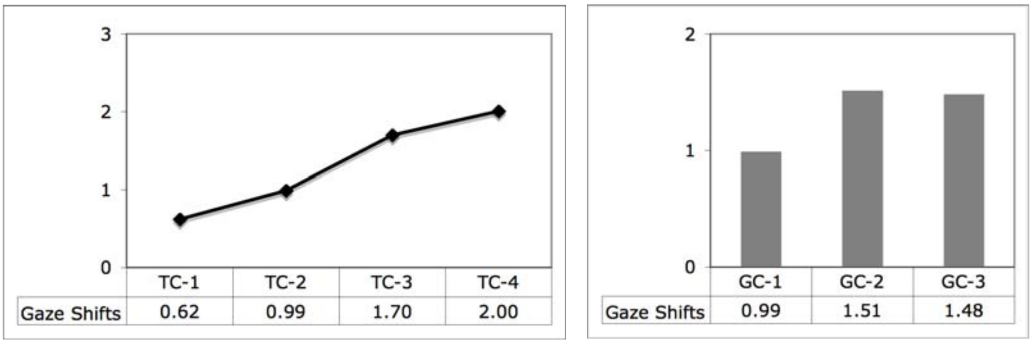


Figure 6.13. The mean number of gaze shifts into the graph layer in Experiment 3. The graphs show the results of the target sentence conditions (TC) and the graph conditions (GC).

A two-way analysis of variance was conducted with two within-subject factors, graph type (graph conditions GC-1, GC-2, and GC-3) and the number of target sentences, (target sentence conditions TC-1, TC-2, TC-3, and TC-4). The results of the analysis showed that the target sentence main effect was significant, $\Lambda = .34$, $F(3, 20) = 12.69$, $p < .01$, multivariate η^2=.66. The graph type main effect was significant, $\Lambda = .54$, $F(2, 21) = 8.80$, $p < .01$, multivariate η^2=.46. The interaction effect Graph type x Target sentence was not significant, $\Lambda = .85$, $F(6, 17) = .51$, $p = .79$, multivariate η^2=.15.

Pairwise comparisons showed that all the differences between the target sentence conditions were significant except for the difference between TC-3 and TC-4.

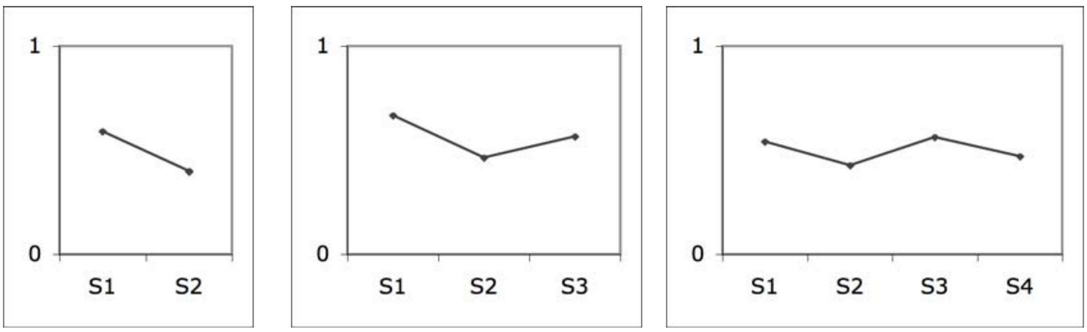
Concerning the graph conditions, the difference between GC-1 and GC-3 was significant, and the difference between GC-1 and GC-2 was significant. Further pairwise comparisons can be seen in Appendix B.2.3.

The percentage of the cases in which there was at least one gaze shift was higher in Experiment 3 (65.7%) compared to the Experiment 1 (55.3%) and Experiment 2 (59.4%). This resulted in a higher mean number of gaze shifts into the graph layer in Experiment 3 compared to the ones in Experiment 1 and Experiment 2. On the other hand, the low percentage of regressions in Experiment 3 (6.73%) compared to Experiment 1 (13.4%) and in Experiment 2 (19.4%) may reflect the influence of the consistent temporal order. Nevertheless, this decrease might also have been caused by the use of different target sentences in Experiment 3 and in the previous experiments. Although the number of the modified target sentences was low, this prevents a comparison of the results of Experiment 3 with the ones in Experiment 1 and Experiment 2.

A more detailed analysis of gaze shifts during participants' reading of the target sentences is presented below.

6.4.2.2 Gaze Shifts for the Target Sentences

As the next step of the analysis of the results, the mean number of gaze shifts during participants' reading of the target sentences was calculated. Figure 6.14 shows the results for each target sentence condition. Numerical details can be found in Appendix B.2.3.



(a) Target Sentence Condition 2 (b) Target Sentence Condition 3 (c) Target Sentence Condition 4

Figure 6.14. The mean number of gaze shifts into graph layer during participants' reading of the target sentences (S) in Experiment 3.

In the target sentence condition TC-2 there was no difference between the target sentences in the mean number of gaze shifts. However, in the target sentence condition TC-3, an ANOVA test revealed a significant difference among the means,

$\Lambda = .73$, $F(2, 21) = 3.93$, $p < .05$. Further pairwise comparisons showed that in TC-3, the mean number of gaze shifts for Sentence 2 was significantly lower then the mean for Sentence 1. However, the difference between Sentence 2 and Sentence 3 was not significant. In the target sentence condition TC-4, the differences between the means were not significant. The results are further discussed in comparison to the results of Experiment 4 in part 6.5. The analysis of answers to posttest questions is presented below.

6.4.2.3 The Analysis of Answers to Posttest Questions

As in the previous experiments, each participant responded to 12 posttest questions in Experiment 3. Answers from the 24 participants were included to the analysis. A total of 288 answers were obtained from the participants. Each true answer was given the score *1*, each false answer was given the score *-1*. The results are shown in Figure 6.15. Numerical details can be seen in Appendix B.2.3.

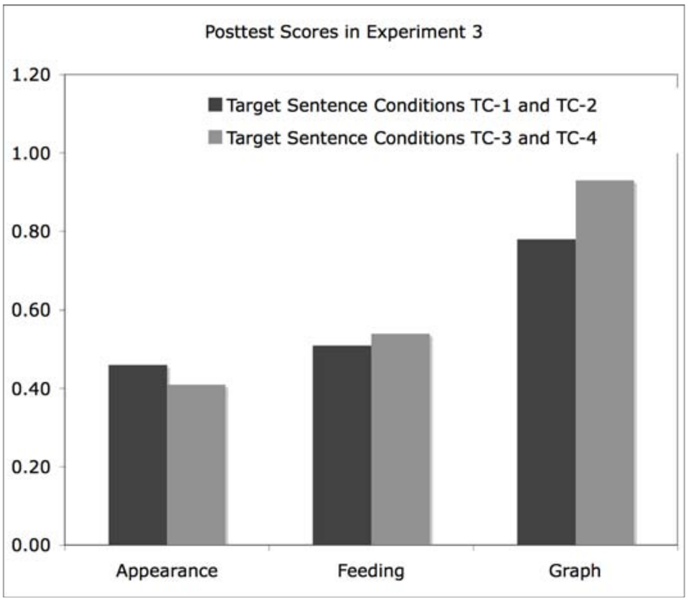


Figure 6.15. Posttest scores in Experiment 3.

A 3 x 2 ANOVA was conducted to evaluate the effects of three question types (namely, appearance questions, feeding questions, and graph questions) and two target sentence conditions (namely, TC-1/2 and TC-3/4) on posttest scores. The ANOVA indicated no significant main effect for target sentence condition, $F(1, 282) = .01$, $p = .94$, multivariate $\eta^2 = .00$. However, the main effect for question type was significant, $F(2, 282) = 4.92$, $p < .05$, multivariate $\eta^2 = .03$. The interaction between target sentence condition and question type was not significant, $F(2, 282) = .10$ $p = .90$, multivariate $\eta^2 = .00$. Follow-up tests were conducted to evaluate the pairwise differences among the means. The analyses showed that there was no significant difference between the appearance scores and the feeding scores, but the difference between the graph scores and the appearance scores was significant; and the difference between the graph scores and the feeding scores was significant. In addition, the difference between the target sentence conditions was not significant in any of the three question types.

In summary, the results showed that the answers to the graph questions received higher scores than the other two question types. The appearance scores were expectedly low. Moreover, the target sentence conditions did not have influence in any of the question types.

6.4.3 Summary of the Findings in Experiment 3

The results of Experiment 3 showed that, as found in the previous experiments, as the number of target sentences increased the number of gaze shifts into the graph layer increased, though with different characteristics under different graphs type conditions. The mean number of gaze shifts during participants' reading of the target sentences was either equal or lower in the late target sentences compared to the initial early (see Figure 6.14).

In a fourth experiment, presented below, the graph-text constellations that were used in Experiment 3 were presented to another group of participants. In Experiment 4, the syntactic structure of the target sentences of the stimuli was changed to have a higher syntactic complexity compared to the target sentences in Experiment 3.

6.5. EXPERIMENT 4

In Experiment 4, the integration of the information induced by graphical entities and linguistic entities in graph-text constellations were investigated under the condition that the target sentences had increased syntactic complexity compared to the target sentences in Experiment 3.

6.5.1 Participants, Materials, and Design

Twenty-five participants (mean age 22.2, SD = 1.23) were paid to participate in the experiment. The procedure in Experiment 4 was the same as the procedure in the previous experiments. The design of Experiment 4 was the same as the design of Experiment 3 with the difference being that the target sentences in Experiment 4 had increased the syntactic complexity. For this purpose, suspended affixation was used as the coordinate structure. The difference between the syntactic structures of the target sentences in the two experiments is schematically shown in Table 6.3.

Table 6.3. The target sentences in Experiment 3 and Experiment 4. The symbols S1 to S4 show the target sentences in each condition. The symbol (…) is used to mean that the generalizing modality marker *–DIr* was omitted in the predicate.

Target sentence condition	Experiment 3	Experiment 4
TC-1	S1.	S1.
TC-2	S1. S2.	S1(…), S2.
TC-3	S1. S2. S3.	S1(…), S2(…), S3.
TC-4	S1. S2. S3. S4.	S1(…), S2(…), S3(…), S4.

As stated in 6.3, in the sentences with suspended affixation, the clauses which are separated by comma are not complete sentences but incomplete clauses that are bound at the end of the sentence. Accordingly, the clause S1 in TC-2, the clauses S1 and S2 in TC-3, and the clauses S1, S2, and S3 in TC-4 are such incomplete clauses. On the other hand, the term *target sentence* will be used below for these clauses, to refer to these incomplete clauses, for the consistency of terminology.

6.5.2 Results

The results from all 25 participants were included to the experiment. Fixations with duration of less than 60 ms were not included to the analysis. Concerning graph literacy, all the participants consented that they were able to interpret the exemplified graphs in the instructions session of the experiment.

The results are presented in three parts below. The first part presents the analysis of gaze shifts for the experimental conditions. The second part presents a more detailed analysis of gaze shifts separately for each target sentence condition. The third part presents the analysis of the answers to the posttest questions.

6.5.2.1 The Analysis of Gaze Shifts

In 66.3% of the inspections of the presented stimuli, the participants made at least one gaze shift into the graph layer during target-phase reading. In 28.6% of the inspections with gaze shift, there was more than one gaze shift into the graph layer during reading of the same target phrase (i.e., regression). Compared to the percentage of regressions in Experiment 3 (15.0%), the percentage of regressions was high in Experiment 4. This finding reflects the influence of the increased syntactic complexity of the target sentences. The results of Experiment 4 are presented in Figure 6.16.

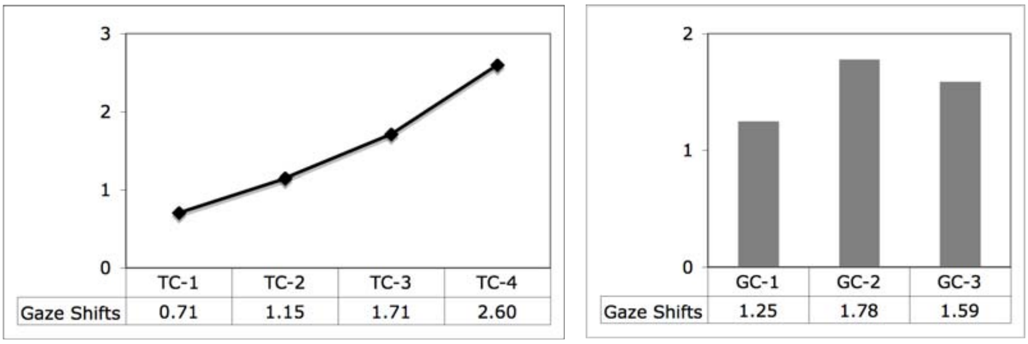


Figure 6.16. The mean number of gaze shifts into the graph layer in Experiment 4. The graphs show the results of the target sentence conditions (TC) and the graph conditions (GC).

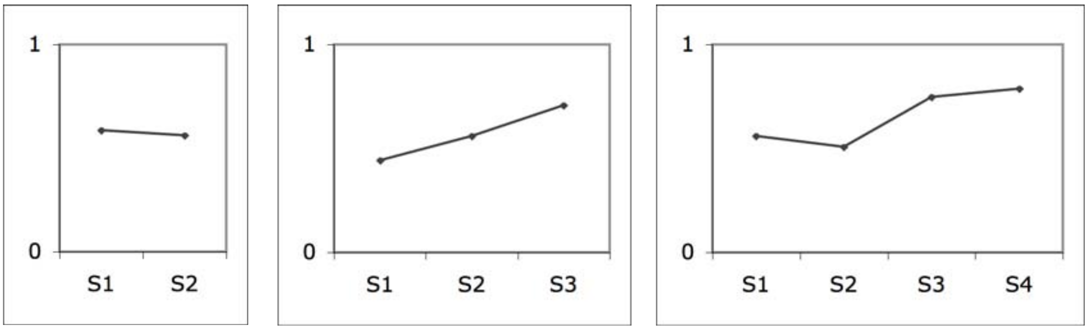
A two-way analysis of variance was conducted with two within-subject factors, graph type (graph conditions GC-1, GC-2, and GC-3) and the number of target sentences, (target sentence conditions TC-1, TC-2, TC-3, and TC-4). The results of the analysis showed that the target sentence main effect was significant, $\Lambda = .46$, $F(3, 22) = 8.58$, $p < .01$, multivariate $\eta^2=.54$. The graph type main effect was significant, $\Lambda = .54$, $F(2, 23) = 9.90$, $p < .01$, multivariate $\eta^2=.46$. The interaction effect Graph type x Target sentence was not significant, $\Lambda = .56$, $F(6, 19) = 2.52$, $p = .06$, multivariate $\eta^2=.44$.

Pairwise comparisons showed that all the differences between the target sentence conditions were significant, including the difference between TC-3 and TC-4. Concerning the graph conditions, only the difference between GC-1 and GC-3 was significant, and the difference between GC-1 and GC-2 was significant. Further pairwise comparisons can be seen in Appendix B.2.4.

The findings of Experiment 4 were parallel to the findings in the previous experiments in that as the number of target sentences increased the number of gaze shifts increased, with different characteristics of increase for different graph conditions. The difference in the mean number of gaze shifts between TC-3 and TC-4 was also significant, in contrast to Experiment 3. This finding indicates the influence of syntactic complexity of the target sentences in Experiment 4 compared to Experiment 3 on inspection of graphs. In addition, in Experiment 4, the percentage of regressions was 28.6%, which was high compared to the one obtained in Experiment 3 (15.0%). This difference reflects the influence of the increased syntactic complexity of the target sentences in Experiment 4. A more detailed analysis of gaze shifts is presented below.

6.5.2.2 Gaze Shifts for the Target Sentences

In this part the analysis of the mean number of gaze shifts into the graph layer during participants' reading of the target sentences is presented. Figure 6.17 shows the results of each target sentence condition. Numerical details can be found in Appendix B.2.4.



(a) Target Sentence Condition 2 (b) Target Sentence Condition 3 (c) Target Sentence Condition 4

Figure 6.17. The mean number of gaze shifts into graph layer during participants' reading of the target sentences (S) in Experiment 4.

In the target sentence condition TC-2 there was no difference between the target sentences in the mean number of gaze shifts into the graph layer. However, in the target sentence condition TC-3, the ANOVA test indicated significant difference among the means, $\Lambda = .75$, $F(2, 23) = 3.85$, $p < .05$. Further pairwise comparisons showed that in TC- 3, the mean number of gaze shifts was significantly different between Sentence 1 and Sentence 3. In the target sentence condition TC-4, there was significant difference among the means, $\Lambda = .63$, $F(3, 22) = 4.24$, $p < .05$. Further pairwise comparisons showed that in TC-4, the mean number of gaze shifts was significantly different between Sentence 2 and Sentence 3.

Compared to the results obtained in Experiment 3, the results of Experiment 4 revealed the differences in the overall patterns of the mean number of gaze shifts in the target sentence conditions. The target sentence TC-4 had a significantly higher mean number of gaze shifts compared to TC-3 in Experiment 4, whereas the difference was not significant in Experiment 3. Moreover, comparison of the mean number of gaze shifts during participants' target-phase reading showed that in TC-2, there was no significant difference between S1 and S2. In Experiment 3, the mean number of gaze shifts for Sentence 2 was significantly lower than the mean for Sentence 1 in TC-2. In other words, compared to Experiment 3, the gaze shifts into the graph layer during reading of the second target sentence increased in TC-2. Furthermore, in Experiment 3, in TC-3 the difference between Sentence 2 and Sentence 3 was not significant whereas the difference was significant in Experiment 4.

In summary, the results of Experiment 4 partially revealed the influence of syntactic complexity in the target sentences. In the target sentence condition TC-2, the decreases in the mean number of gaze shifts for the second target sentence (compared to the first target sentence) in Experiment 3 were not observed in Experiment 4. In TC-2 and TC-3, the highest number of gaze shifts was obtained for the last target sentence, in which the previous target sentences with suspended affixation were bound.

6.5.2.3 The Analysis of Answers to Posttest Questions

As in the previous experiments, each participant responded to 12 posttest questions in Experiment 4. Answers from the 25 participants were included to the analysis. A total of 300 answers were recorded. Each true answer was given the score *1*, each false answer was given the score *-1*. The results are shown in Figure 6.18. Numerical details can be seen in Appendix B.2.4.

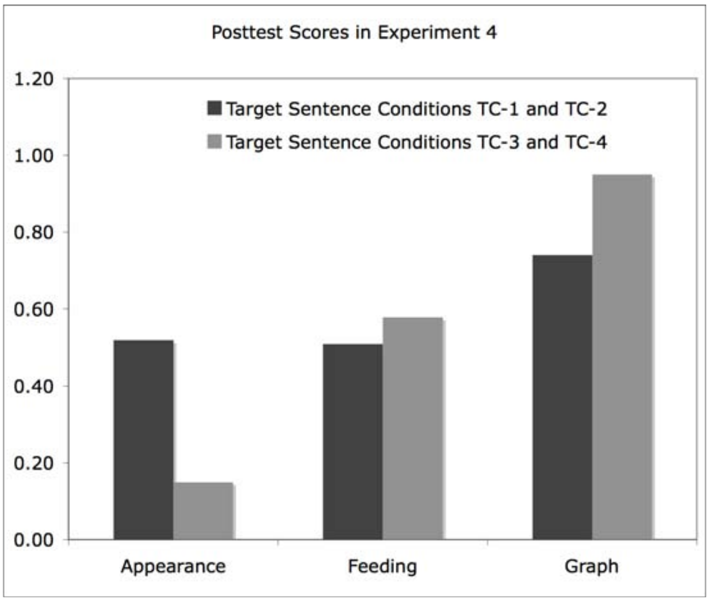


Figure 6.18. Posttest scores in Experiment 4.

A 3 x 2 ANOVA was conducted to evaluate the effects of three question types (namely, appearance questions, feeding questions, and graph questions) and two target sentence conditions (namely, TC-1/2 and TC-3/4) on posttest scores. The ANOVA indicated no significant main effect for target sentence condition, $F(1, 294) = .11$, $p = .74$, multivariate $\eta^2 = .00$. However, the main effect for question type was significant, $F(2, 294) = 9.47$, $p < .05$, multivariate $\eta^2 = .06$. The interaction between target sentence condition and question type was also significant, $F(2, 294) = 3.70$ $p < .05$, multivariate $\eta^2 = .03$. Follow-up tests were conducted to evaluate the pairwise differences among the means. The analyses showed that there was no significant difference between the appearance scores and the feeding scores, but the difference between the graph scores and the appearance scores was significant; and the difference between the graph scores and the feeding scores was significant. Furthermore, the difference between the target sentence conditions was not significant in feeding scores and in graph scores. However, the difference in the appearance scores between the two target sentence conditions was significant.

In summary, as in the previous experiments, the results showed that the answers to the graph questions received higher scores than the other two question types. The most important finding in Experiment 4 was that the difference between the target sentence conditions was significant in the appearance scores. This finding reflects the influence of the reduced recall possibly due to processing difficulties in particular in the target sentence conditions with higher number of target sentences.

6.5.3 Summary of the Findings in Experiment 4

The results of Experiment 4 showed that, as found in the previous experiments, as the number of target sentences increased the number of gaze shifts increased, with different graph type conditions having different characteristics of increase. Moreover, in the target sentence condition TC-4, the mean number of gaze shifts was significantly higher than the ones in TC-3, an effect which was not obtained in Experiment 3. In addition, increasing the syntactic complexity of the target sentences resulted in a more complex inspection of the graphs as evidenced by the higher number of gaze shifts for the late target sentences compared to the early ones. In summary, the results of Experiment 4 present experimental evidence that the syntactic structure of the target sentences in the text influences the integration characteristics of the information induced by linguistic and graphical modalities.

A further difference between Experiment 4 and Experiment 3 was obtained in the answers to posttest questions. In Experiment 4, the participants had a reduced recall for the pre-target appearance questions i.e. in the target sentences TC-3/4, which had more syntactic complexity compared to TC-1/2. This finding presents support for the cognitive load introduced by the relatively complex coordinate structures with suspended affixation. A similar significant effect, i.e. decrease in appearance scores, was not obtained in Experiment 2, compared to Experiment 1. Further research is needed to understand the underlying reason.

The influence of syntactic complexity was further investigated in a fifth experiment, as presented in the following part of the chapter.

6.6 Experiment 5

The purpose of Experiment 5 was to investigate the integration of the information induced by graphical entities and linguistic entities in graph-text constellations under the condition that the target sentences had syntactic complexity between the ones in

Experiment 3 and Experiment 4. In particular, a partially coordinated suspended affixation was used instead of a fully coordinated suspended affixation.

6.6.1 Participants, Materials, and Design

Twelve participants (mean age 22.1, SD = 3.23) were paid to participate in the experiment. The procedure in Experiment 5 was the same as the procedure in the previous experiments. The design of Experiment 5 was the same as the design of Experiment 4 with the difference being that the target sentences in Experiment 5 had syntactic complexity between the ones used in Experiment 3 and Experiment 4. In other words, the syntactic complexity of the target sentences in Experiment 5 was higher than the syntactic complexity of the target sentences in Experiment 3; but it was lower than the syntactic complexity of the target sentences in Experiment 4. For this purpose, a partially coordinated suspended affixation was used, as schematically represented in Table 6.4.

Table 6.4. The target sentences in Experiment 3, Experiment 4 and Experiment 5. The symbols S1 to S4 show the target sentences in each condition. The symbol (…) is used to mean that the generalizing modality marker *–DIr* was omitted in the predicate.

Target sentence condition	Experiment 3	Experiment 4	Experiment 5
TC-1	S1.	S1.	S1.
TC-2	S1. S2.	S1(…), S2.	S1. S2.
TC-3	S1. S2. S3.	S1(…), S2(…), S3.	S1. S2(…), S3.
TC-4	S1. S2. S3. S4.	S1(…), S2(…), S3(…), S4.	S1. S2(…), S3(…), S4.

The suspended affixation in the target sentences of Experiment 5 was partial in the sense that the first target sentence did not have suspended affixation in any of the target sentence conditions. In TC-2, the second target sentence had suspended affixation. TC-2 was the same as the TC-2 in Experiment 3. In TC-3, the second and the third target sentences had suspended affixation. In Experiment 5, the syntactic structure of the target sentences in TC-3 and TC-4 was different from the ones used in the previous experiments. In other words, the target sentences in Experiment 5 did not have as high syntactic complexity as the ones in Experiment 4, but they had more syntactic complexity than the target sentences in Experiment 3.

6.6.2 Results

The calibration of the eye tracker failed for one of the 12 participants. In addition, 0.76% of the remaining data from 11 participants were not included to the analysis due to partial calibration problems. Fixations with duration of less than 60 ms were not included to the analysis. Concerning graph literacy, all the participants consented that they were able to interpret the exemplified graphs in the instructions session of the experiment.

The results are presented in three parts below. The first part presents the analysis of gaze shifts for the experimental conditions. The second part presents a more detailed analysis of gaze shifts for each target sentence condition. The third part presents the analysis of the answers to the posttest questions.

6.6.2.1 The Analysis of Gaze Shifts

In 65.6% of participants' inspections of the presented graph-text constellations, the participants made at least one gaze shift into the graph layer during target-phase reading. In 22.4% of the inspections with gaze shifts, there was more than one gaze shift into the graph layer during reading of the same target phrase (i.e., regression). Compared to the percentage of regressions in Experiment 3 (15.0%) and Experiment 4 (28.6%), the percentage of regressions was between the two previous experiments. This finding reflects the influence of the use of partially coordinated suspended affixation in the text. The results of Experiment 5 are presented in Figure 6.19.

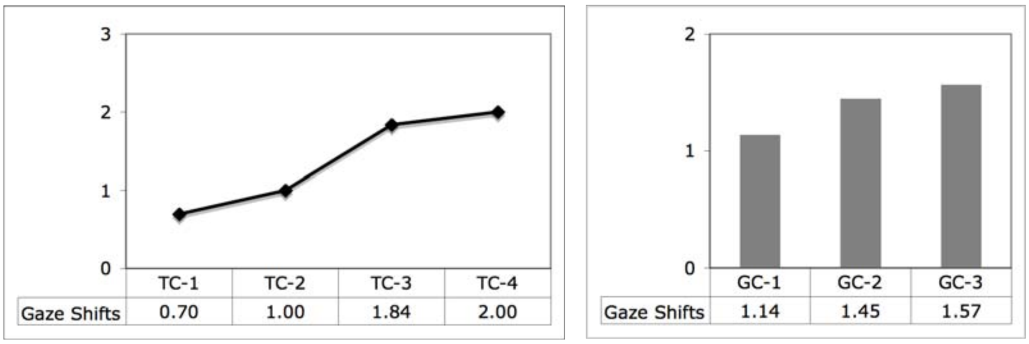


Figure 6.19. The mean number of gaze shifts into the graph layer in Experiment 5. The graphs show the results of the target sentence conditions (TC) and the graph conditions (GC).

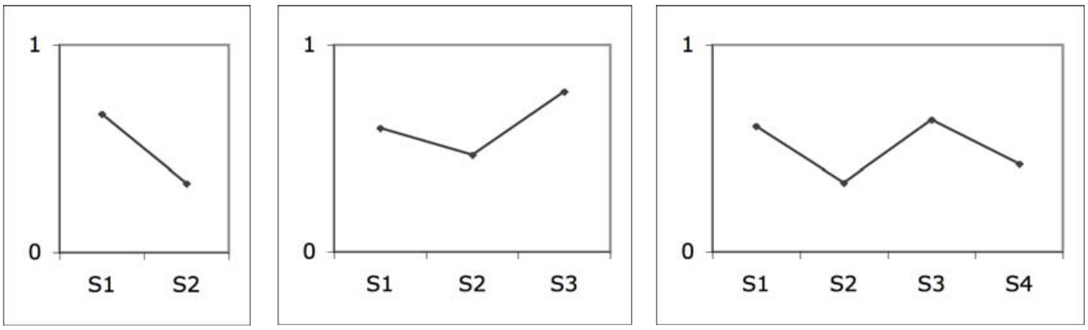
A two-way analysis of variance was conducted with two within-subject factors, graph type (graph conditions GC-1, GC-2, and GC-3) and the number of target sentences, (target sentence conditions TC-1, TC-2, TC-3, and TC-4). The results of the analysis showed that the target sentence main effect was significant, $\Lambda = .36$, $F(3, 8) = 4.79$, $p < .05$, multivariate η^2=.64. The graph type main effect was not significant, $\Lambda = .55$, $F(2, 9) = 3.70$, $p = .07$, multivariate η^2=.45. The interaction effect Graph type x Target sentence was not significant, $\Lambda = .51$, $F(6, 5) = .80$, $p = .61$, multivariate η^2=.49.

Pairwise comparisons showed the difference between the target sentence conditions TC-2 and TC-3 was significant, whereas the difference between TC-1 and TC-2 and the difference between TC-3 and TC-4 were not significant. Concerning the graph conditions, only the difference between GC-1 and GC-3 was significant. Further pairwise comparisons can be seen in Appendix B.2.5.

The findings of Experiment 5 parallel the findings in the previous experiments in that as the number of target sentences increased the number of gaze shifts roughly increased. However, the patterns of increase in Experiment 5 were more similar to the patterns of increase in Experiment 3 than the ones in Experiment 4. In addition, in Experiment 5, the percentage of regressions was 22.4%, which was high compared to the one obtained in Experiment 3 (15.0%) but low compared to the one obtained in Experiment 4 (28.6%). This finding reflects the influence of the use of partially coordinated suspended affixation in Experiment 5. A more detailed investigation of gaze shifts is presented below.

6.6.2.2 Gaze Shifts for the Target Sentences

As the next step of the analysis of the results, the mean number of gaze shifts during participants' reading of the target sentences was analyzed. Figure 6.20 shows the results of each target sentence condition. Numerical details can be found in Appendix B.2.5.



(a) Target Sentence Condition 2 (b) Target Sentence Condition 3 (c) Target Sentence Condition 4

Figure 6.20. The mean number of gaze shifts into graph layer during participants' reading of the target sentences (S) in Experiment 5.

In the target sentence condition TC-2, the difference between the two clauses in the mean number of gaze shifts was significantly different, $t(10) = 2.80$, $p < 05$. In the target sentence conditions TC-3 and TC-4, there were no significant differences among the means.

The lack of significance among the means in TC-3 and TC-4 prevents further comparison of the results of Experiment 5 with the previous experiments. The number of participants in Experiment 5 was relatively low, which resulted in high variability and reduced the statistical reliability of the results.

6.6.2.3 The Analysis of Answers to Posttest Questions

As in the previous experiments, each participant responded to 12 posttest questions in Experiment 4. Answers from the twelve participants were included to the analysis. A total of 144 answers were obtained from the participants. Each true answer was given the score *1*, each false answer was given the score *-1*. A 3 x 2 ANOVA was conducted to evaluate the effects of three question types (namely, appearance questions, feeding questions, and graph questions) and two target sentence conditions (namely, TC-1/2 and TC-3/4) on posttest scores. The ANOVA indicated no significant main effects for target sentence condition, $F(1, 138) = .22$, $p = .64$, multivariate $\eta^2 = .00$, for question type, $F(2, 138) = 1.67$, $p = .19$, multivariate $\eta^2 = .02$, and no significant interaction between target sentence condition and question type $F(2, 138) = .32$ $p = .73$, multivariate $\eta^2 = .01$. Therefore, the results of the posttest scores in Experiment 5 are not further discussed. Numerical details can be seen in Appendix B.2.5.

6.6.3 Summary of the Findings in Experiment 5

Although the percentage of regressions in Experiment 5—which was higher than the percentage of regressions in Experiment 4 and lower than the ones in Experiment 3—indicated the influence of the use of reduced syntactic complexity compared to the one in Experiment 4, the non-significant differences in statistical analyses prevent further interpretation of the results. In Experiment 5, the number of participants was low compared to the previous experiments. As a result, high variability among the participants reduced the reliability of the results, as evinced by statistical analyses. A replication of the experiment with more participants is needed for more reliable results.

6.7 INTERIM SUMMARY OF THE EXPERIMENTS

The results of the five experiments reported above show that the number of graph-related sentences (i.e., the target sentences) in the text layer of a graph-text constellation influences the number of gaze shifts during readers' reading of these sentences. Higher number of target sentences results in higher number of gaze shifts. This finding reflects the incremental characteristic of the integration of information in terms of the input increments, i.e. gaze fixations. However, this finding does not make a claim about complexity of the sentences in the text layer of a graph-text constellation. The reported empirical investigations present evidence for the influence of syntactic complexity on multimodal comprehension by showing that the use of suspended affixation coordinate structure in the target sentences results in a higher number of gaze shifts, specifically in relatively complex cases where the coordinate structure includes high number coordinands, compared to the number of gaze shifts in the material where suspended affixation is not used as the coordinate structure.

In the analyses presented so far, the characteristics of multimodal integration were investigated in terms of the gaze shifts during participants' reading of the target sentences in the text layer of the graph-text constellations. Although this analysis provides evidence for the influence of changes in syntactic structure of the graph-related sentences on multimodal comprehension, it does not provide information about whether the participants integrated the information induced by textual and graphical entities. The following part of the chapter investigates the instantiation of the graph schema knowledge by the participants, which is based on the analysis of the inspected locations on the graph during their reading of the target sentences.

6.8 ANALYSIS OF INSPECTIONS ON THE GRAPH AOIs

This part of the chapter presents the analyses of participants' inspections on the graph layer, during their reading of the target sentences. The analyses are based on the comparison of the mean gaze times on the specified graph layer AOIs (Areas of Interest). The specified AOIs covered the numerical values on the axes and the graph line segments. Each numerical entity in the graph layer (namely, *x-axis labels* and *y-axis labels*) was specified as an AOI. In addition, the AOIs on the graph lines were specified separately for each graph condition and for each sentence, as "hit" AOIs and "nonhit" AOIs. The term *hit* was used for the AOIs on the graph layer (including both the axis AOIs and the graph line AOIs) that corresponded to the information presented in the target sentence under investigation. The term *nonhit* was used for the remaining AOIs. More information on the specific AOIs on the graph lines is presented in the following paragraphs. For brevity, only the results of Experiment 3 and Experiment 4 are presented, and only the statistically significant results are reported; non-significant differences are stated where necessary.

The structure of the target sentences in Experiment 3 and Experiment 4 were previously shown in Table 6.4, partially repeated below as Table 6.5.

Table 6.5. The target sentences in Experiment 4 and Experiment 5. The symbols S1 to S4 show the target sentences in each condition. The symbol (…) is used to mean that the generalizing modality marker *–DIr* was omitted in the predicate.

Target sentence condition	Experiment 3	Experiment 4
TC-1	S1.	S1.
TC-2	S1. S2.	S1(…), S2.
TC-3	S1. S2. S3.	S1(…), S2(…), S3.
TC-4	S1. S2. S3. S4.	S1(…), S2(…), S3(…), S4.

In Experiment 3, the target sentences were used as complete sentences, whereas in Experiment 4, the sentences had an increased syntactic complexity due to the suspended affixation coordinated structure. In the target sentence condition TC-1, there was one target sentence (namely, the target sentence S1), which was the same in both experiments. In the target sentence condition TC-2, there were two target sentences (namely, S1 and S2) in Experiment 3; and a point separated these two sentences. In Experiment 4, however, the first target sentence had suspended affixation at the end; and a comma separated these two sentences. The second sentence had the point at the end, without suspended affixation. Similar situations are pertinent to TC-3 and TC-4 as shown in the table.

In the analyses below, the inspections on the graph AOIs are presented, independent of the target sentence conditions for brevity. In other words, the gaze shifts from the text to the graph during participants' reading of the first target sentence in all target sentence conditions are grouped as the "inspections for the first target sentence", without taking into account whether the first target sentence was the single sentence of the TC-1 condition, or the first sentence of one of the TC-2, TC-3, or TC-4 conditions. Considering the limited number of data points for the inspections for the target sentence per graph type and target sentence condition, this assumption was necessary for a valid statistical comparison of the results. As stated previously, in the transcription of eye movements, a gaze shift was coded as belonging to a target sentence as long as the gaze shift occurred during the reading of the target sentence, during the reading of the last word of the target sentence, or during the reading of the first word of the next target sentence.

In the analyses, the mean gaze times of participants' inspections on the graph layer AOIs during their reading of target sentences were calculated. Not all participants inspected the graph layer during their reading of each target sentence. The analyses below include the inspections from the participants who inspected the graph layer at least once, during target-phase reading. In addition, the AOIs that were inspected by these participants were assigned the corresponding gaze times, whereas the AOIs that were not inspected by the participants were assigned zero gaze time. In other words, all the data from the participants who inspected the graph layer at least once were included to the analysis. As a result, the AOIs that were inspected by most of the participants—who inspected the graph layer at least once during the reading of the corresponding target sentence—had higher scores than the AOIs which were inspected by less number of participants. Therefore, the mean gaze times presented in this part of the chapter do not show the mean gaze time *average per participant*, but the average values of the gaze times on the AOIs.

In summary, the main goal of the analyses presented below is to understand whether the participants integrated the information induced by the target sentences and the information induced by the entities in the graph layer during their target-phase reading. In addition, the inspected AOIs on the graph layer in Experiment 3 and

Experiment 4 were compared to understand the influence of increased syntactic complexity in the target sentences of Experiment 4.

6.8.1 Analysis of Inspections in the Graph Condition GC-1

The four graphs that were used in the first graph type condition (GC-1) are shown in Figure 6.21.

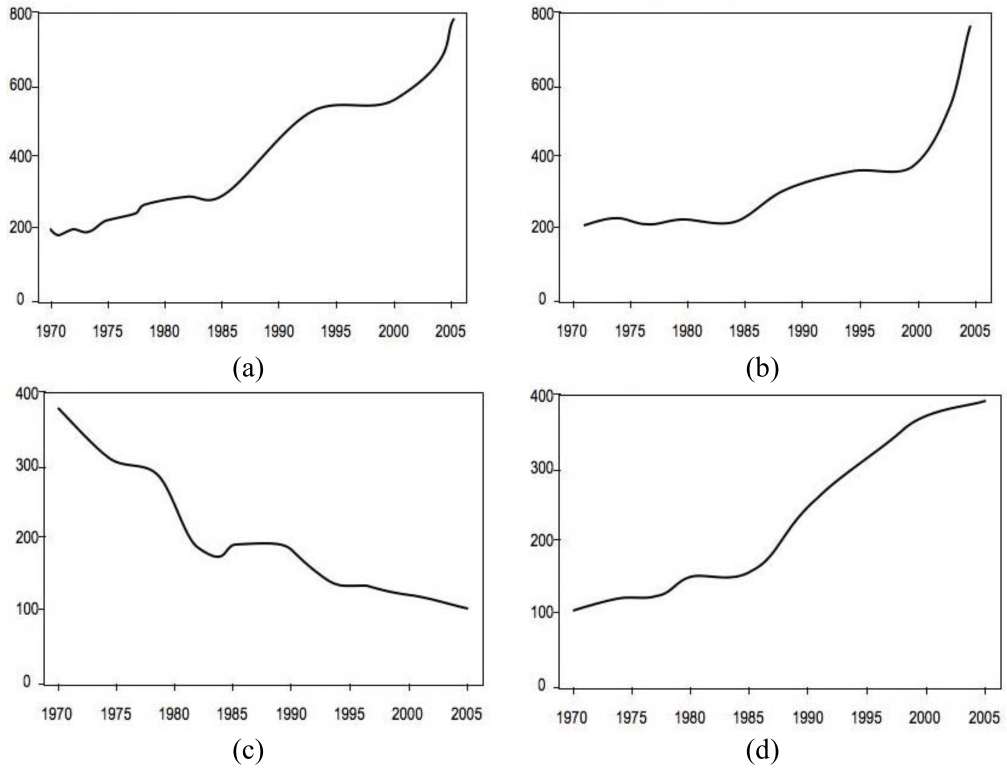


Figure 6.21. The four graphs used in the first graph type condition (GC-1).

6.8.1.1 Inspections for the First Target Sentence

In the graph condition GC-1, the following sentences were used as the first target sentence in the text layer of the stimuli, as shown in (5).

(5) a. *1970 yılında gölde 200 civarında kuş tespit edilmiştir.*
'In year 1970, in the lagoon, about 200 birds were detected.'
b. *1970 yılında gölde 200 civarında kuş tespit edilmiştir.*
'In year 1970, in the lagoon, about 200 birds were detected.'
c. *1970 yılında gölde 400 civarında kuş tespit edilmiştir.*
'In year 1970, in the lagoon, about 400 birds were detected.'
d. *1970 yılında gölde 100 civarında kuş tespit edilmiştir.*
'In year 1970, in the lagoon, about 100 birds were detected.'

Table 6.6 shows the mean gaze times on the graph layer AOIs during participants' reading of the target sentence S1 in Experiment 3 and Experiment 4. The ranking order of the AOIs was the same in both experiments. The *X axis label (hit)* AOI covered the x-axis label "1970", whereas the *X axis label (nonhit)* AOI covered all the other x-axis labels, from "1975" to "2005". The *Graph line + Y axis label (hit)* AOI was the region that covered the start point of the graph line and the y-axis label next to it. The *Graph line (nonhit)* AOI covered the remaining parts of the graph line. The *Y axis label (nonhit)* AOI covered the y-axis labels which were not covered in the *Graph line + Y axis label (hit)* AOI.

Table 6.6. Mean gaze times on the graph layer AOIs for the target sentence S1 in the graph condition GC-1. SD is used for standard deviation.

Target Sentence S1, Exp. 3	**Gaze Time**		**Target Sentence S1, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
X axis label (nonhit)	498	546	X axis label (nonhit)	527	538
Graph line + Y axis label (hit)	376	228	Graph line + Y axis label (hit)	437	352
Graph line (nonhit)	306	487	Graph line (nonhit)	281	638
X axis label (hit)	187	148	X axis label (hit)	211	199
Y axis label (nonhit)	75	184	Y axis label (nonhit)	28	79

The *X axis label (hit)* AOI, which was the x-axis label "1970" in this case, had a mean gaze time of 187 ms. The sum of the means of remaining x-axis labels—the seven x-axis labels from "1975" to "2005"—was 498 ms. Although this sum is high, the mean gaze time per nonhit x-axis label was approximately 70 ms and 75 ms in Experiment 3 and 4 correspondingly. Compared the x-axis label "1970", which had the mean gaze time of 187 ms, these values are low. In other words, compared to the other x-axis labels, the hit x-axis label (i.e., "1970") was inspected longer. Concerning the graph line inspections, the nonhit inspections of the graph line were as high as the inspections on the *Graph line + Y axis label (hit)* AOI. Compared to the mentioned AOIs, less number of participants inspected the nonhit y-axis labels.

In summary, the participants inspected the relevant AOIs on the graph layer (i.e., the hit AOIs), as well as the nonhit AOIs. Since this was the participants' first linguistically guided inspection of the graph layer during reading of the text, the inspection of the nonhit AOIs as well as the hit AOIs was expected.

6.8.1.2 Inspections for the Second Target Sentence

In graph condition GC-1, the target sentence (6a) was used for the graph 6.21a, 6.21b, and 6.21d, and the target sentence (6b) was used for the graph 6.21c as the second target sentence in the text layer of the stimuli.

(6) a. *Kuş sayısı 1970 yılından beri yükselmiştir.*
‘The number of birds has increased since 1970.’
b. *Kuş sayısı 1970 yılından beri düşmüştür.*
‘The number of birds has fallen since 1970.’

Table 6.7 shows the mean gaze times on the graph layer AOIs for the target sentence S2 in the graph condition GC-1. The *X axis label (hit)* AOI was “1970” for all the graphs used in the graph condition GC-1. The *X axis label (other)* AOI covered the other x-axis labels. The term “other” is used here instead of “hit” because of the use of the term *beri* ‘since’ in the target sentences. The *Y axis label (S1)* AOI was the y-axis value of the first sentence. This AOI was separated from the other y-axis labels to reveal the potential influence of the previous target sentence on the graph layer inspections for the target sentence S2.

Table 6.7. Mean gaze times on the graph layer AOIs for the target sentence S2 in the graph condition GC-1.

Target Sentence S2, Exp. 3	**Gaze Time**		**Target Sentence S2, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
X axis label (other)	554	474	Graph line	591	685
Graph line	450	294	X axis label (other)	500	423
X axis label (hit)	112	131	X axis label (hit)	188	281
Y axis label (S1)	65	166	Y axis label (nonhit)	40	90
Y axis label (nonhit)	43	88	Y axis label (S1)	32	90

In both experiments, the participants inspected both the graph line and the x-axis labels. Considering that the *X axis label (other)* AOI covered seven x-axis labels, the mean gaze time on the *X axis label (hit)* AOI was higher than the means for the other single x-axis labels. The ranking order of the mean gaze times for the second target sentence was slightly different between the two experiments. Although the *X axis label (other)* AOI had a lower mean gaze time than the *Graph line* AOI in Experiment 4, the difference between the *X axis label (other)* AOI and the *Graph line* AOI was not different in either of the experiments.

6.8.1.3 Inspections for the Third Target Sentence

The third target sentence that was used in the graph condition GC-1 for the four graphs is shown in (7). The corresponding region on the graph line was the short horizontal segment of the graph line.

(7) a. *Kuş sayısı 1995 ve 2000 yılları arasında sabit kalmıştır.*
'The number of birds remained constant between years 1995 and 2000.'
b. *Kuş sayısı 1995 ve 2000 yılları arasında sabit kalmıştır.*
'The number of birds remained constant between years 1995 and 2000.'
c. *Kuş sayısı 1985 ve 1990 yılları arasında sabit kalmıştır.*
'The number of birds remained constant between years 1985 and 1990.'
d. *Kuş sayısı 1980 ve 1985 yılları arasında sabit kalmıştır.*
'The number of birds remained constant between years 1980 and 1985.'

Table 6.8 shows the mean gaze times on the relevant parts of the graph layer during reading of the third target sentence S3. The *X axis label (hit)* AOI covered the two x-axis labels that were used in the target sentence. The *Graph line (hit)* AOI was the corresponding horizontal segment of the graph line. The nonhit AOIs of the x-axis label and the graph line were correspondingly the remaining x-axis labels and the remaining parts of the graph line. The *X axis label (S2)* AOI was the x-axis label "1970", which was used in the second target sentence S2. The *Y axis label (nonhit)* AOI covered all the y-axis labels.

Table 6.8. Mean gaze times on the graph layer AOIs for the target sentence S3 in the graph condition GC-1.

Target Sentence S3, Exp. 3	**Gaze Time**		**Target Sentence S3, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
X axis label (hit)	564	342	X axis label (hit)	733	449
Graph line (hit)	533	468	Graph line (hit)	659	489
X axis label (nonhit)	199	374	X axis label (nonhit)	379	409
Graph line (nonhit)	194	472	Graph line (nonhit)	190	245
X axis label (S2)	7	32	Y axis label (nonhit)	65	161
Y axis label (nonhit)	6	29	X axis label (S2)	8	42

The mean gaze times on the graph layer AOIs showed that in both experiments the relevant x-axis labels and the relevant graph line segment were the longest inspected regions on the graph. The mean gaze times for the hits were significantly higher than the nonhit gaze times for both x-axis labels and graph line, in both experiments. In addition, the mean values in Experiment 4 were relatively high compared to the mean values in Experiment 3. This difference might have been due to the increased syntactic complexity of the target sentences in Experiment 4. Furthermore, the nonhit inspections on the x-axis labels and the nonhit inspections on the y-axis labels—those were all nonhits in this case, because no y-axis value was mentioned in the target sentences—were higher in Experiment 4 than the ones in Experiment 3.

In Experiment 4, the participants made relatively longer inspections on the nonhit x-axis labels compared to the hit x-axis label inspections and hit graph line inspections.

However, in Experiment 3, the nonhit x-label inspections were not as long as the ones in Experiment 4, compared to the hit x-axis label inspections and hit graph line inspections. This leads to the conclusion that in Experiment 3 the participants made more *isolated* inspections than the participants in Experiment 4. A similar finding was obtained in some of the graph conditions and target sentence conditions in the following analyses.

6.8.1.4 Inspections for the Fourth Target Sentence

The last target sentence S4 that was used in the graph condition GC-1 for the four graphs is shown in (8). The target sentence S4 was the last target sentence in the stimuli.

(8) a. *2005 itibarıyla gölde 800 civarında kuş görülmüştür.*
‘As of 2005, about 800 birds were seen in the lagoon.’
b. *2005 itibarıyla gölde 800 civarında kuş görülmüştür.*
‘As of 2005, about 800 birds were seen in the lagoon.’
c. *2005 itibarıyla gölde 100 civarında kuş görülmüştür.*
‘As of 2005, about 100 birds were seen in the lagoon.’
d. *2005 itibarıyla gölde 400 civarında kuş görülmüştür.*
‘As of 2005, about 400 birds were seen in the lagoon.’

Table 6.9 shows the mean gaze times on the graph layer AOIs for the fourth target sentence S4. The *X axis label (hit)* AOI covered the x-axis label “2005”, and the *Y axis label (hit)* AOI covered the y-axis value given in the target sentence. Correspondingly, the remaining x-axis labels and y-axis labels were covered in the *X axis label (nonhit)* and the *Y axis label (nonhit)* AOIs, with the exception that the x-axis label that was used in the previous target sentence S3 was specified as the *X axis label (S3)*. The end of the graph line was specified by the *Graph line (hit)* AOI, and the remaining part of the graph line was specified by the *Graph line (nonhit)* AOI.

Table 6.9. Mean gaze times on the graph layer AOIs for the target sentence S4 in the graph condition GC-1.

Target Sentence S4, Exp. 3	**Gaze Time**		**Target Sentence S4, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
X axis label (hit)	275	172	Y axis label (nonhit)	401	481
Graph line (nonhit)	270	355	Graph line (hit)	264	357
X axis label (nonhit)	257	245	Graph line (nonhit)	257	340
Graph line (hit)	208	279	X axis label (S3)	202	434
X axis label (S3)	77	119	Y axis label (hit)	197	222
Y axis label (hit)	73	147	X axis label (hit)	187	189
Y axis label (nonhit)	35	106	X axis label (nonhit)	115	260

In both experiments, there was no difference between the gaze times on the *Graph line (hit)* AOI and the *Graph line (nonhit)* AOI. Similarly, there was no difference between the gaze times on the *X axis label (hit)* AOI and the *X axis label (nonhit)* AOI. However, in Experiment 3, the four graph-layer AOIs shown in the table received most of the inspections, whereas the inspections in Experiment 4 were distributed among the AOIs. This provides further evidence that in Experiment 3 the participants made more isolated inspections on the graph layer. Also, higher priority of y-axis labels in Experiment 4 might have been due to the fact that in Experiment 4, the target sentences with suspended affixation were bound at the end of the fourth sentence.

6.8.2 Analysis of Inspections in the Graph Condition GC-2

The analysis of the mean gaze times in the second graph condition GC-2 is presented below. The four graphs used in GC-2 are shown in Figure 6.22.

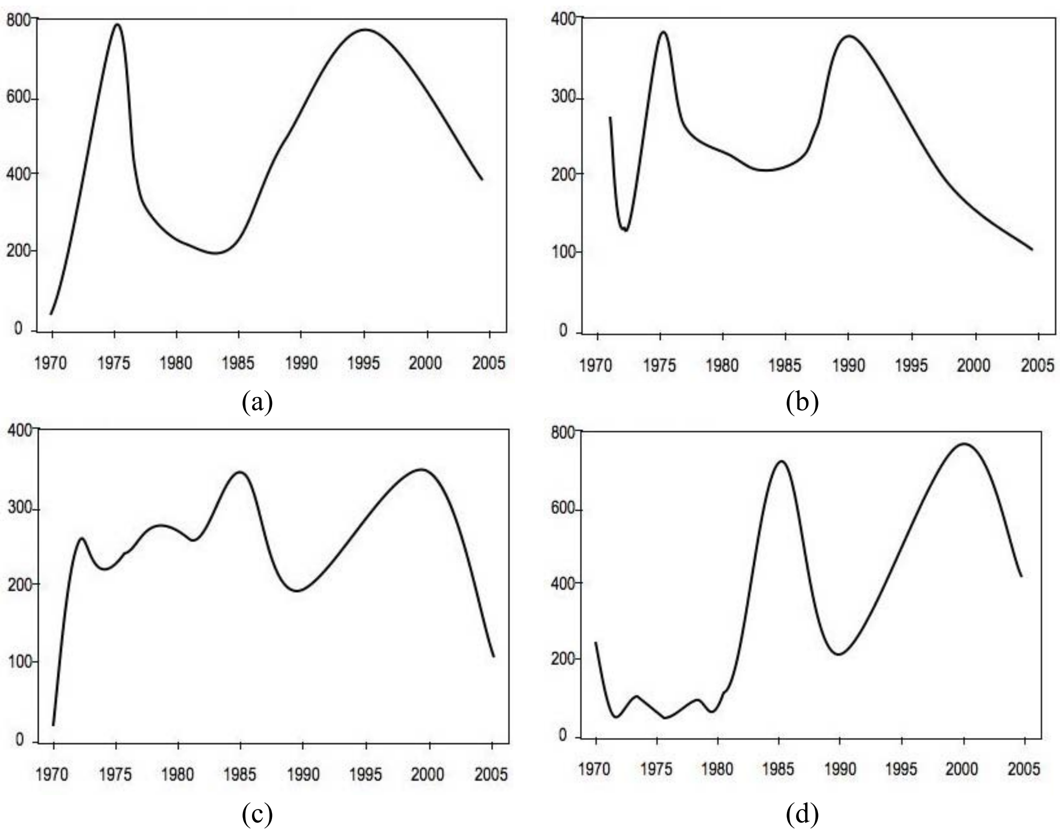


Figure 6.22. The four graphs used in the second graph type condition (GC-2).

6.8.2.1 Inspections for the First Target Sentence

In GC-2, the following sentences were used in the text layer of the stimuli as the first target sentence, as shown in (9).

(9) a. *Göldeki kuş sayısı ilk defa 1975 yılında zirveye çıkmıştır.*
‘The number of birds peaked the first time at year 1975.’
b. *Göldeki kuş sayısı ilk defa 1975 yılında zirveye çıkmıştır.*
‘The number of birds peaked the first time at year 1975.’
c. *Göldeki kuş sayısı ilk defa 1985 yılında zirveye çıkmıştır.*
‘The number of birds peaked the first time at year 1985.’
d. *Göldeki kuş sayısı ilk defa 1985 yılında zirveye çıkmıştır.*
‘The number of birds peaked the first time at year 1985.’

The mean gaze times on the AOIs during participants’ reading of the first target sentence are shown in Table 6.10. In all four graphs, the *Graph line peak (hit)* AOI covered the left peak, and the *Graph line peak (nonhit)* AOI covered the right peak. The *Graph line local peaks (nonhit)* AOI and the *Graph line local bottoms (nonhit)* AOI covered the local peaks in Figure 6.22c and the local bottoms in Figure 6.22d respectively. The *Graph line bottom (nonhit)* AOI covered the bottom part of the graph between the two peaks.

Table 6.10. Mean gaze times on the graph layer AOIs for the target sentence S1 in the graph condition GC-2.

Target Sentence S1, Exp. 3	**Gaze Time**		**Target Sentence S1, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
Graph line peak (hit)	524	578	Graph line peak (hit)	476	340
X axis label (nonhit)	366	638	X axis label (nonhit)	332	368
X axis label (hit)	341	255	X axis label (hit)	322	205
Graph line local peaks (nonhit)	322	655	Graph line peak (nonhit)	318	398
Graph line peak (nonhit)	271	567	Graph line local bottoms (nonhit)	169	255
Y Axis label (nonhit)	129	434	Graph line local peaks (nonhit)	108	181
Graph line local bottoms (nonhit)	110	179	Graph line bottom (nonhit)	78	168
Graph line bottoms (nonhit)	54	199	Y Axis label (nonhit)	69	184

The results showed that the ranking order of the inspected regions were similar between Experiment 3 and Experiment 4 with minor differences. The relevant peak of the graph line (i.e., the left peak in this case) was the longest inspected AOI, which was significantly different than all the other AOIs in both experiments; the right peak and the bottom region of the graph lines were inspected for a shorter time. There was no significant difference between the nonhit x-axis labels and the x-axis labels that were used in the sentence. However, considering that the *X axis label (nonhit)* AOI covers seven x-axis labels, the mean gaze times on the *X axis label (hit)* AOI were high compared to the nonhit ones in both experiments. The difference between the two experiments was that the inspection of the local peaks in 6.22c was longer in Experiment 3 than the inspections in Experiment 4.

6.8.2.2 Inspections for the Second Target Sentence

The sentences that were used as the second target sentence in GC-2 are shown in (10).

(10) a. *Kuş sayısı 1985 yılında 200 civarına düşmüştür.*
'The number of birds fell to about 200 in year 1985.'
b. *Kuş sayısı 1985 yılında 200 civarına düşmüştür.*
'The number of birds fell to about 200 in year 1985.'
c. *Kuş sayısı 1990 yılında 200 civarına düşmüştür.*
'The number of birds fell to about 200 in year 1990.'
d. *Kuş sayısı 1990 yılında 200 civarına düşmüştür.*
'The number of birds fell to about 200 in year 1990.'

The mean gaze times of the inspections for the second target sentences in the graph condition GC-2 are shown in Table 6.11. The *Graph line bottom (hit)* AOI covered the bottom parts of the graph line that were between the two peaks. The *Graph line peak (nonhit)* AOI covered the right peak whereas the *Graph line peak (S1)* AOIs covered the left peak, which was also used in the previous target sentence. The naming of the other AOIs were similar to the ones presented above.

Table 6.11. Mean gaze times on the graph layer AOIs for the target sentence S2 in the graph condition GC-2.

Target Sentence S2, Exp. 3	**Gaze Time**		**Target Sentence S2, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
X axis label (hit)	466	273	X axis label (hit)	328	294
Graph line bottom (hit)	326	251	Graph line bottom (hit)	317	290
X axis label (nonhit)	254	344	X axis label (nonhit)	253	361
Graph line peak (nonhit)	123	200	Graph line peak (nonhit)	145	190
Graph line local peaks (nonhit)	94	240	Graph line peak (S1)	125	231
Graph line peak (S1)	75	167	Y Axis label (hit)	91	187
Y Axis label (hit)	39	103	Graph line local bottoms (nonhit)	84	188
X axis label (S1)	35	152	X axis label (S1)	66	130
Y Axis label (nonhit)	22	102	Graph line local peaks (nonhit)	58	146
Graph line local bottoms (nonhit)	0	0	Y Axis label (nonhit)	42	152

The results showed that the ranking order of the inspected regions were similar between the two experiments with slight differences. The bottom parts of the graph lines and the corresponding x-axis labels were inspected longest during reading of the target sentence S2. Although the difference between *X axis label (hit)* and the *X axis label (nonhit)* was significant in Experiment 3, the difference was not significant in Experiment 4; the participants made less inspections on the hit x-axis label in the latter.

6.8.2.3 Inspections for the Third Target Sentence

The sentences that were used as the third target sentence in GC-2 are shown in (11).

(11) a. *1995 yılında kuş sayısı yeniden zirveye çıkmıştır.*
'The number of birds peaked again at year 1995.'
b. *1990 yılında kuş sayısı yeniden zirveye çıkmıştır.*
'The number of birds peaked again at year 1990.'
c. *2000 yılında kuş sayısı yeniden zirveye çıkmıştır.*
'The number of birds peaked again at year 2000.'
d. *2000 yılında kuş sayısı yeniden zirveye çıkmıştır.*
'The number of birds peaked again at year 2000.'

The mean gaze times are shown in Table 6.12. In the AOI specification, the *Graph line peak (hit)* AOI covered the right peak of the graph lines, whereas the *Graph line peak (nonhit)* AOI covered the left peak of the graph lines. The other AOI specifications were parallel to the ones presented previously.

Table 6.12. Mean gaze times on the graph layer AOIs for the target sentence S3 in the graph condition GC-2.

Target Sentence S3, Exp. 3	**Gaze Time**		**Target Sentence S3, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
Graph line peak (hit)	457	307	X axis label (hit)	469	321
X axis label (hit)	267	443	Graph line peak (hit)	459	459
X Axis label (nonhit)	89	266	X Axis label (nonhit)	284	451
X axis label (S2)	52	201	Graph line peak (nonhit)	131	221
Graph line bottom (S2)	35	91	X axis label (S2)	82	167
Graph line peak (nonhit)	31	81	Graph line bottom (S2)	55	134
Y Axis label (nonhit)	27	103	Graph line local peaks (nonhit)	48	107
Graph line local peaks (nonhit)	0	0	Y Axis label (nonhit)	42	98

The results showed that the right peak of the graph lines and the corresponding x-axis label received most of the inspections in Experiment 3, whereas the results were mixed in Experiment 4. The mean gaze time on the *X axis label (hit)* AOI was longer in Experiment 4 compared to Experiment 3. This finding presents further support to the previous findings for the graph condition GC-1 in the sense that the participants of Experiment 3 made more isolated inspections on the graph layer compared to the participants of Experiment 4.

6.8.2.4 Inspections for the Fourth Target Sentence

The sentences that were used as the third target sentence in GC-2 are shown in (12).

(12) a. *2005 itibarıyla gölde 400 civarında kuş görülmüştür.*
‘As of 2005, about 400 birds were seen in the lagoon.’
b. *2005 itibarıyla gölde 100 civarında kuş görülmüştür.*
‘As of 2005, about 100 birds were seen in the lagoon.’
c. *2005 itibarıyla gölde 100 civarında kuş görülmüştür.*
‘As of 2005, about 100 birds were seen in the lagoon.’
d. *2005 itibarıyla gölde 400 civarında kuş görülmüştür.*
‘As of 2005, about 400 birds were seen in the lagoon.’

The mean gaze times for S4 are shown in Table 6.13.

Table 6.13. Mean gaze times on the graph layer AOIs for the target sentence S4 in the graph condition GC-2.

Target Sentence S4, Exp. 3	**Gaze Time**		**Target Sentence S4, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
X Axis label (nonhit)	446	702	Graph line (hit)	219	184
Graph line local peaks (nonhit)	445	771	X axis label (hit)	168	184
X axis label (hit)	230	195	X Axis label (nonhit)	151	368
Graph line peak (S3)	221	266	Graph line peak (S3)	120	185
Graph line peak (nonhit)	185	393	Y Axis label (hit)	95	199
Y Axis label (nonhit)	165	457	Graph line peak (nonhit)	88	195
X axis label (S3)	138	220	X axis label (S3)	79	142
Graph line bottom (nonhit)	132	277	Graph line bottom (nonhit)	77	142
Graph line (hit)	118	119	Y Axis label (nonhit)	61	168
Graph line local bottoms (nonhit)	60	103	Graph line local bottoms (nonhit)	57	138
Y Axis label (hit)	47	111	Graph line local peaks (nonhit)	40	89

In Experiment 3, the results were partially unexpected, specifically for the *Graph line (hit)* AOI, which received relatively low inspections. In Experiment 4, the relevant part of the graph line, which was the endpoint of the graph line in this case, and the corresponding x-axis label were the longest inspected regions on the graph line. In addition, the nonhit x-axis labels, as well as the graph line peak and the x-axis labels that were used in the previous target sentence were inspected. The comparatively low gaze times in Experiment 4 were not expected. Furthermore, the inspections of the participants in Experiment 3 were not isolated, in the sense of the use of the term “isolated” in the previous analyses. The reason is not clear at this stage. This points to the need for further research with more participants and with different types of target sentences.

6.8.3 Analysis of Inspections in the Graph Condition GC-3

The analysis of the mean gaze times in the third graph condition GC-3 is presented below. The four graphs used in GC-3 are shown in Figure 6.23.

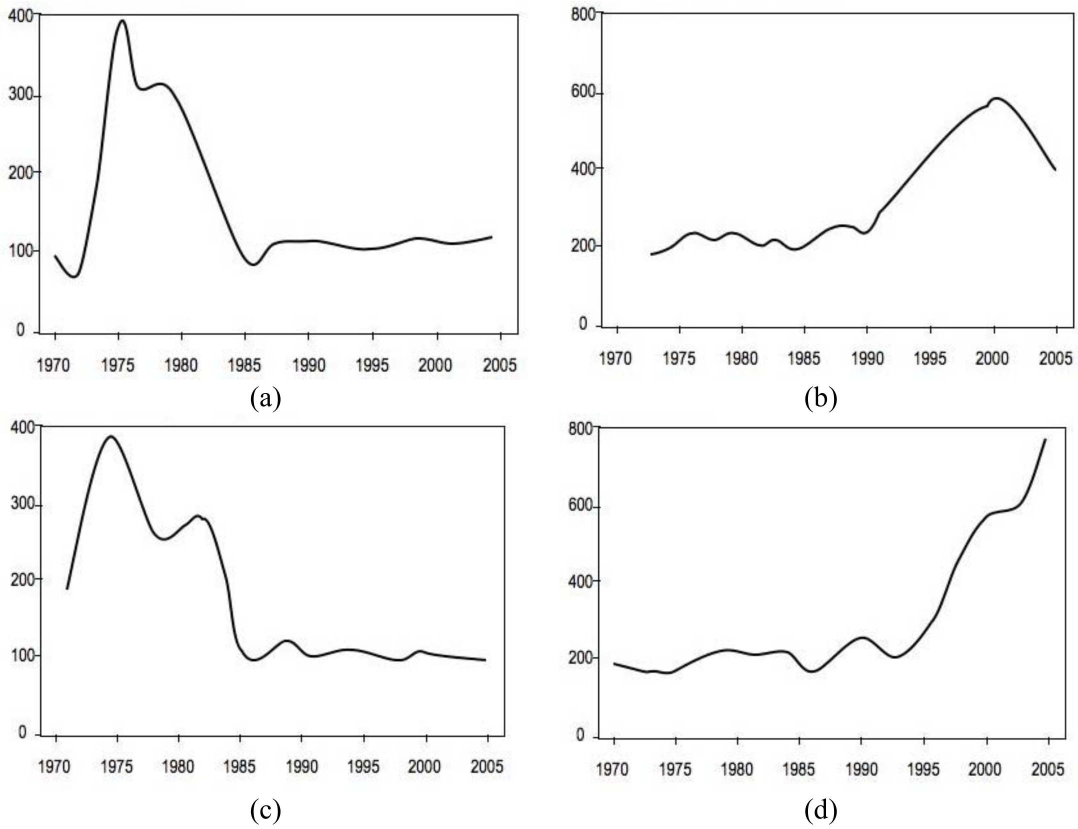


Figure 6.23. The four graphs used in the third graph type condition (GC-3).

In GC-3, the graphs 6.23a and 6.23c were used with different target sentences than the graphs 6.23b and 6.23d. For this reason, the inspected regions are analyzed separately for the graphs GC-3ac and GC-3bd.

6.8.3.1 Inspections for the First Target Sentence in GC-3ac

The following sentence was used in the text layer of the stimuli as the first target sentence in GC-3ac, as shown in (13).

(13) a. *1975 yılında gölde yaklaşık 400 kuş tespit edilmiştir.*
'In year 1975 about 400 birds were detected in the lagoon.'

Table 6.14 shows the mean gaze times for the first target sentence in GC-3ac. The relevant region on the graph line was the peak, which was specified by the *Graph line peak (hit)* AOI in the table. The remaining regions of the graph line were nonhits; these were the *Graph line decrease (nonhit)* AOI, the *Graph line increase (nonhit)* AOI, and the *Graph line remain (nonhit)* AOI. The pivot region between the decrease line and the remain line were included to the *Graph line remain (nonhit)* AOI.

Table 6.14. Mean gaze times on the graph layer AOIs for the target sentence S1 in the graph condition GC-3ac.

Target Sentence S1, Exp. 3	**Gaze Time**		**Target Sentence S1, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
Graph line peak (hit)	369	260	X axis label (hit)	374	369
X axis label (hit)	305	327	Graph line peak (hit)	357	365
X Axis label (nonhit)	218	380	Graph line decrease (nonhit)	255	287
Y Axis label (hit)	155	196	Graph line increase (nonhit)	222	300
Graph line decrease (nonhit)	114	202	X Axis label (nonhit)	169	225
Y Axis label (nonhit)	74	235	Y Axis label (nonhit)	140	428
Graph line remain (nonhit)	71	251	Y Axis label (hit)	120	154
Graph line increase (nonhit)	50	138	Graph line remain (nonhit)	78	240

The results showed that, the participants of Experiment 4 made longer inspections on the graph layer AOIs. Concerning the comparison of the inspected regions, the mean gaze time for the inspections on the peak region was significantly longer than the other graph region AOIs in Experiment 3, whereas the mean gaze time for the inspections on the peak region was significantly longer than the *Graph line remain (nonhit)* AOI but not than the *Graph line decrease (nonhit)* AOI and the *Graph line increase (nonhit)* AOI. This finding shows that in Experiment 3, the participants made more isolated fixations on the graph line. Similar findings were found for the previous graph conditions.

6.8.3.2 Inspections for the Second Target Sentence in GC-3ac

The following sentence was used in the text layer of the stimuli as the second target sentence in GC-3ac, as shown in (14).

(14) a. *Kuş sayısı 1975 yılından sonra düşmüştür.*
'The number of birds fell after 1975.'

Table 6.15 shows the mean gaze times on the AOIs for the second target sentence in GC-3ac. For the target sentence S2, the decrease part of the graph line was specified as the *Graph line decrease (hit)* AOI, and the other parts as nonhits.

Table 6.15. Mean gaze times on the graph layer AOIs for the target sentence S2 in the graph condition GC-3ac.

Target Sentence S2, Exp. 3	**Gaze Time**		**Target Sentence S2, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
Graph line decrease (hit)	388	423	X Axis label (nonhit)	438	564
X axis label (hit)	243	291	X axis label (hit)	341	428
Graph line remain (nonhit)	151	205	Graph line decrease (hit)	317	293
X Axis label (nonhit)	116	191	Graph line remain (nonhit)	145	235
Graph line peak (S1)	65	127	Graph line peak (S1)	71	148
Graph line increase (nonhit)	33	97	Graph line increase (nonhit)	39	105
Y axis label (S1)	20	57	Y Axis label (nonhit)	29	92
Y Axis label (nonhit)	20	82	Y axis label (S1)	0	0

The results showed that the mean gaze times on the decrease segment of the graph line were longer than the other parts of the graph line in both experiments. However, in Experiment 3, the hit x-axis inspections were significantly longer than the nonhit x-label inspections, whereas the difference between the two in Experiment 4 was not significant. Both the hit and nonhit x-axis label inspections were longer in Experiment 4 compared to the ones in Experiment 3.

6.8.3.3 Inspections for the Third Target Sentence in GC-3ac

The following sentence was used in the text layer of the stimuli as the third target sentence in GC-3ac, as shown in (15).

(15) a. *1985 yılından sonra kuş sayısı 100 civarında kalmıştır.*
'After year 1985 the number of birds remained about 100.'

Table 6.16 shows the mean gaze times for the target sentence S3 in GC-3ac.

Table 6.16. Mean gaze times on the graph layer AOIs for the target sentence S3 in the graph condition GC-3ac.

Target Sentence S3, Exp. 3	**Gaze Time**		**Target Sentence S3, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
Graph line remain (hit)	503	604	Graph line remain (hit)	340	146
X axis label (hit)	367	261	Graph line decrease (nonhit)	248	512
X Axis label (nonhit)	197	225	X Axis label (nonhit)	231	359
Y Axis label (hit)	186	225	X axis label (hit)	201	239
Graph line decrease (nonhit)	153	208	Y Axis label (hit)	96	145
Graph line peak (nonhit)	44	175	Y Axis label (nonhit)	28	98
Y Axis label (nonhit)	30	90	Graph line peak (nonhit)	0	0
X axis label (S2)	29	115	Graph line increase (nonhit)	0	0
Graph line increase (nonhit)	14	55	X axis label (S2)	0	0

The results showed that in Experiment 3, the hit x-axis label inspections were significantly higher than the nonhit x-axis label inspections; and the mean gaze time for the *Graph line remain(hit)* AOI was significantly longer than the nonhit graph line AOIs. However, in Experiment 4, the difference between the hits and nonhits were not significant both for the x-axis labels and the graph line segments. This shows that the participants of Experiment 4 inspected the hit AOIs as well as the nonhit AOIs. On the other hand, the mean gaze times for the hit locations were shorter in Experiment 4 compared to Experiment 3; this was contrary to the expectation. This result points to the need for further research with more participants.

6.8.3.4 Inspections for the Fourth Target Sentence in GC-3ac

The following sentence was used in the text layer of the stimuli as the last target sentence in GC-3ac, as shown in (16).

(16) a. *2005 itibarıyla gölde 100 civarında kuş görülmüştür.*
‘As of 2005, about 100 birds were seen in the lagoon.’

Table 6.17 shows the mean gaze times for the last target sentence in GC-3ac.

Table 6.17. Mean gaze times on the graph layer AOIs for the target sentence S4 in the graph condition GC-3ac.

Target Sentence S4, Exp. 3	**Gaze Time**		**Target Sentence S4, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
X axis label (hit)	263	243	Graph line (hit)	370	236
Graph line (hit)	259	179	X Axis label (nonhit)	217	241
X Axis label (nonhit)	76	169	Graph line (nonhit)	119	190
Y Axis label (hit)	48	107	X axis label (hit)	113	141
X axis label (S3)	0	0	Y Axis label (hit)	42	90
Graph line (nonhit)	0	0	X axis label (S3)	31	93

The results showed that in Experiment 3, the two hit AOIs received most of the inspections on the graph layer, whereas in Experiment 4, the inspections were distributed also to the nonhit AOIs. The inspections in Experiment 3 were more isolated than the inspections in Experiment 4. Similar findings were obtained in the previous analyses.

The analysis of the mean gaze times for the remaining two graphs GC-3bd (see Figure 6.23) is presented below.

6.8.3.5 Inspections for the First Target Sentence in GC-3bd

The following sentences were used in the text layer of the stimuli as the first target sentence in GC-3bd, as shown in (17).

(17) a. *Kuş sayısı 1990 yılına kadar 200 civarında kalmıştır.*
‘The number of birds remained about 200 until year 1990.’
b. *Kuş sayısı 1995 yılına kadar 200 civarında kalmıştır.*
‘The number of birds remained about 200 until year 1995.’

Table 6.18 shows the mean gaze times for the first target sentence in GC-3bd. the *Graph line remain (hit)* AOI covered the roughly horizontal segment of the graph line, as well as the pivot region between the horizontal and the diagonal segment. The other parts of the graph line were specified as the *Graph line (nonhit) AOI*.

Table 6.18. Mean gaze times on the graph layer AOIs for the target sentence S1 in the graph condition GC-3bd.

Target Sentence S1, Exp. 3	**Gaze Time**		**Target Sentence S1, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
X Axis label (nonhit)	619	615	X Axis label (nonhit)	652	399
Graph line remain (hit)	458	390	Graph line remain (hit)	543	333
X axis label (hit)	395	314	X axis label (hit)	373	267
Graph line (nonhit)	206	364	Graph line (nonhit)	175	234
Y Axis label (hit)	199	231	Y Axis label (hit)	166	222
Y Axis label (nonhit)	65	158	Y Axis label (nonhit)	55	181

The results were similar between the two experiments. The mean gaze times on the hit graph line AOIs were significantly longer than the mean gaze times on the nonhit graph AOIs. Although the nonhit x-axis labels were inspected longer than the hit x-axis labels, considering that the *X axis label (hit)* AOI covered a single value and the *X axis label (nonhit)* AOI covered the remaining seven values, the mean gaze times on the hit x-axis label was longer than the mean gaze times for the single nonhit x-axis values.

6.8.3.6 Inspections for the Second Target Sentence in GC-3bd

The following sentences were used in the text layer of the stimuli as the second target sentence in GC-3bd, as shown in (18).

(18) a. *Kuş sayısı 1990 yılından sonra yükselmiştir.*
'The number of birds increased after 1990.'
b. *Kuş sayısı 1995 yılından sonra yükselmiştir.*
'The number of birds increased after 1995.'

Table 6.19 shows the mean gaze times on the graph layer AOIs for the second target sentence in GC-3bd.

Table 6.19. Mean gaze times on the graph layer AOIs for the target sentence S2 in the graph condition GC-3bd.

Target Sentence S2, Exp. 3	**Gaze Time**		**Target Sentence S2, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
Graph line increase (hit)	406	637	X axis label (hit)	279	192
X axis label (hit)	261	280	X Axis label (nonhit)	272	219
X Axis label (nonhit)	209	367	Graph line increase (hit)	219	273
Y Axis label (nonhit)	178	562	Graph line remain (S1)	141	168
Graph line remain (S1)	82	140	Y Axis label (nonhit)	98	199
Y axis label (S1)	0	0	Y axis label (S1)	67	151

In Experiment 3, the mean gaze time on *Graph line increase (hit)* AOI was longer than the mean gaze time on the *Graph line remain (S1)* AOI, which was also a nonhit AOI in this analysis. The hit x-axis label was inspected longer, as expected. In Experiment 4, the mean gaze time for the inspections on the graph line were shorter compared to the one in Experiment 3.

6.8.3.7 Inspections for the Third Target Sentence in GC-3bd

The following sentence was used in the text layer of the stimuli as the third target sentence in GC-3bd, as shown in (19).

(19) a. *2000 yılında gölde yaklaşık 600 kuş tespit edilmiştir.*
'In year 2000 about 600 birds were detected in the lagoon.'

Table 6.20 shows the mean gaze times on the graph layer AOIs for the third target sentence in GC-3bd.

Table 6.20. Mean gaze times on the graph layer AOIs for the target sentence S3 in the graph condition GC-3bd.

Target Sentence S3, Exp. 3	**Gaze Time**		**Target Sentence S3, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
Graph line (hit)	406	475	Graph line (hit)	367	407
X axis label (hit)	307	309	X axis label (hit)	314	274
X Axis label (nonhit)	151	216	X Axis label (nonhit)	173	328
Y Axis label (hit)	102	166	X axis label (S2)	160	200
Y Axis label (nonhit)	74	127	Y Axis label (hit)	131	170
X axis label (S2)	62	132	Graph line remain (nonhit)	66	100
Graph line remain (nonhit)	34	107	Y Axis label (nonhit)	27	80

The results were similar between the two experiments except for a few differences. The mean gaze times on the relevant part of the graph lines and the corresponding x-axis labels were significantly higher than the corresponding nonhit AOIs.

6.8.3.8 Inspections for the Fourth Target Sentence in GC-3bd

The following sentences were used in the text layer of the stimuli as the last target sentence in GC-3bd, as shown in (20).

(20) a. *2005 itibarıyla gölde 400 civarında kuş görülmüştür.*
'As of 2005, about 400 birds were seen in the lagoon.'
b. *2005 itibarıyla gölde 800 civarında kuş görülmüştür.*
'As of 2005, about 800 birds were seen in the lagoon.'

Table 6.21 shows the mean gaze times on the graph layer AOIs for the fourth target sentence in GC-3bd.

Table 6.21. Mean gaze times on the graph layer AOIs for the target sentence S4 for the graph condition GC-3bd.

Target Sentence S4, Exp. 3	**Gaze Time**		**Target Sentence S4, Exp. 4**	**Gaze Time**	
AOI	**M**	**SD**	**AOI**	**M**	**SD**
X axis label (hit)	319	54	Graph line (nonhit)	498	633
Graph line (hit)	124	87	X axis label (S3)	199	188
X axis label (S3)	105	210	X axis label (hit)	128	175
X Axis label (nonhit)	90	180	Y Axis label (hit)	84	187
Y Axis label (hit)	65	130	Graph line (hit)	80	109
Y Axis label (nonhit)	0	0	Y Axis label (nonhit)	40	89
Graph line (nonhit)	0	0	X Axis label (nonhit)	0	0

The hit x-axis label and the corresponding region of the graph line received most of the inspections in Experiment 3 whereas the results were not that isolated in Experiment 4. In Experiment 4, the nonhit regions of the graph line received the highest mean gaze time, which were followed by the x-axis labels that were used in the previous target sentence.

6.8.4 Summary of the Analysis of Graph Inspections

The main goal of the analysis of graph layer inspections during participants' reading of the target sentences was to understand whether the participants integrated the information induced by the target sentences and the information induced by the entities in the graph layer by means of the knowledge of the graph schema. For this purpose, participants' inspections on the graph layer during their target-phase reading were analyzed in terms of their locations (i.e., the AOIs) and gaze times. For brevity, only the gaze patterns in Experiment 3 and Experiment 4 were analyzed.

The results show that the participants inspected the locations on the graph layer, which were described by the target sentences. The inspected locations included the graphical entities such as peaks and bottoms, as well as the numerical values on the axes. In the graph comprehension architecture presented in Chapter 4, it was stated that conceptual representations such as peaks could be accessed by different modalities, i.e. the linguistic entities and the graphical entities. The results of the analyses present empirical evidence for how the same information was accessed consecutively via two different modalities.

In addition, differences were found in the distribution of inspected locations on the graph layer between Experiment 3 and Experiment 4. The participants of Experiment

4 made inspections that were distributed over more AOIs compared to the participants of Experiment 3. This finding presents support for the influence of the increased syntactic complexity in the target sentences of Experiment 4 compared to the target sentences in Experiment 3. However, there were cases contrary to the expectations as well as the expected ones, which point to the need for further investigation with alternative sets of target sentences and with a higher number of participants.

6.9 SUMMARY

This chapter presented five experimental case studies that investigated the construction of coreferential relations between graphical and linguistic entities in graph-text constellations. In particular, the influence of syntactic complexity on the integration of the information induced by the two modalities was investigated. The design of the stimuli was based on the naturalistic stimuli, i.e. the graph-text constellations presented in a waterbird census report, as exemplified in Chapter 1 and under the multimodal graph-text comprehension architecture presented in Chapter 4.

In Experiment 1, following the source material in the waterbird census report, the text layer of the graph-text constellations involved graph-related sentences (i.e., target sentences) with no strict mentioning of the events and processes with their temporal order of occurrence. In Experiment 2, the same graph-text constellations with increased syntactic complexity of the target sentences were presented to the participants. In Experiment 3, the target sentences were arranged such that the order of the target sentences and the temporal order of the events and processes expressed by the target sentences were strictly consistent. In Experiment 4, the stimuli of Experiment 3 were used; the difference between Experiment 4 and Experiment 3 was that the target sentences of Experiment 4 had a more complex syntactic structure. In Experiment 5, the syntactic complexity of the target sentences was further altered such that the syntactic complexity of the target sentences in Experiment 5 was higher than the syntactic complexity of the target sentences in Experiment 3 but it was lower than the syntactic complexity of the target sentences in Experiment 4.

For the analysis of the results, the shifts of gaze from the text layer to the graph layer of the graph-text constellations (i.e., gaze shifts) were analyzed. The results showed that in all experiments the number of target sentences in the text layer (i.e., in the paragraphs) of graph-text constellations influenced the number of gaze shifts into the graph layer during participants' reading of the target sentences: As the number of the target sentences increased the number of gaze shifts into the graph layer increased. Concerning the first research question presented at the beginning of the chapter, the findings show that multimodal comprehension of graph-text constellations did not have the "read-the-whole-text and inspect-the-whole-graph" characteristics, but rather reciprocal shifts between the text and the graph were observed.

More detailed analyses of gaze shifts were conducted for the comparison of the gaze shifts during participants' reading of the particular target sentences, which in turn provided data about the influence of syntactic complexity of the target sentences (see Figure 6.8 and Figure 6.11 for Experiment 1 and Experiment 2). Compared to the results obtained in Experiment 1, in Experiment 2 more significant increases in the mean number of gaze shifts were obtained, particularly during participants' reading of the late target sentences (i.e., the second, third, and the fourth target sentences rather than the first target sentence). Consequently, the results of Experiment 2 presented empirical evidence for the influence of syntactic complexity of the target sentences on multimodal comprehension of graph-text constellations.

A parallel finding was obtained for Experiment 3 and Experiment 4 (see Figure 6.14 and Figure 6.17 for Experiment 3 and Experiment 4): The increased complexity of the target sentences resulted in higher number of gaze shifts in Experiment 4, compared to the same conditions in Experiment 3. On the other hand, the results of Experiment 5 did not exhibit statistically significant differences among the means for the number of gaze shift, possibly due to relatively lower number of participants compared to the previous experiments.

The analysis of posttest scores provided partial evidence for the influence of syntactic complexity on recall. In Experiment 4, the posttest scores for the pre-target posttest questions were lower in the experimental conditions that had target sentences with higher syntactic complexity. However, no statistical significance in the differences for the pre-target questions was achieved among the experimental conditions in the other experiments. This indicates the need for further investigation of the relationship between the content of the graph-text constellation and the posttest questions (see Chapter 11 for open questions and future research).

Further analysis of the inspected locations on the graph layer during participants' reading of the particular target sentences was conducted for Experiment 3 and Experiment 4 in 6.8, to understand whether the participants integrated the information induced by the target sentences and the information induced by the entities in the graph layer by means of the graph schema knowledge. The results showed that the participants inspected the graphical entities as well as numerical entities on the graph, which were described by the target sentences. This finding reflects participants' cognitive effort for the integration of information induced by the two modalities. In addition, the differences in the distribution of inspections among the target sentences between Experiment 3 and Experiment 4 provided partial support for the influence of syntactic complexity on the characteristics of the integration of information induced by different modalities.

In summary, the empirical investigations presented in this chapter showed how changes in one modality influence comprehension of the other modality, which, in turn, influence multimodal comprehension. In particular, the empirical investigations provided answers for the following three research questions posited at the beginning of the chapter:

- Multimodal comprehension of graph-text constellations exhibits an incremental character, as indicated by the reciprocal shifts of the input increments between the text and the graph. This finding supports the view that online integration processes require the integration of information induced by both modalities.
- Syntactic complexity of the text influences integration of information induced by graphical entities and textual entities. In particular, the assumption that the increased number of coordinands in suspended affixation of construction results in a memory bottleneck is supported. Since mental representations of the coordinands degrade over time, a high number of target sentences with suspended affixation resulted in more complex readings of the graph-text constellations (i.e., higher number of gaze shifts), compared to the target sentences without coordinate structure. Since the online integration processes require the integration of information induced by both modalities, the complexity of the target sentences influence the integration characteristics.
- The findings presented in the last part of the analysis, i.e. in 6.8, showed that the participants inspected the graph locations which were described in the text. The comparison of the inspected locations that were inspected during participants' reading of particular target sentences showed that the integration processes could be investigated in terms of the analysis of eye movement recordings.

7 Coherence Relations: Causal Attribution and Aspectual Characteristics in Graph-Internal Multimodality[1]

This chapter presents three empirical studies that investigate a specific type of means for providing interaction between text layer and graph layer of a graph-text constellation, namely *verbal annotations*. The term *verbal annotation* is used for textual entities that annotate parts of graphical entities within the graph layer of a graph-text constellation. The Study 1 investigates verbal annotations with respect to their bridging role between the graph layer and the text layer of a graph-text constellation, as well as between the sublayer of graphical entities and the sublayer of textual entities within the graph layer. The other two studies investigated the construction of inter-modal coherence relations during the course of comprehension of verbally annotated graphs. In particular, they investigated the conceptual structures and the contents that are communicated by verbal annotations and graphical entities: the processes and events represented by verbal annotations and graph lines, their temporal properties and their causal relations. The Study 2 is an experimental investigation of the interaction between verbal annotations—in particular, their positions with respect to the graph lines—and readers' causal attributions concerning the events and processes represented by the verbal annotations and the processes and events represented by the graphical entities.[2] The Study 3 investigates the interaction between temporal aspect of the events and processes represented by verbal annotations and graphical entities. In 7.1, verbal annotations in graph-text constellations are introduced with respect to their functional similarities to other annotating acts in communication, such as gestures. The Study 1, Study 2, and Study 3 are presented in 7.2, 7.3, and 7.4 respectively.

---

[1] The results of the experimental studies presented in this chapter were partially reported in Acartürk, Habel, Çağıltay, and Alaçam (2008), Acartürk, Habel, and Çağıltay (2008), and Acartürk and Habel (2009).

[2] The focus in this dissertation is the psychological phenomenon of causal attribution (or causal inference) rather than the philosophical aspects of causal reasoning. In other words, the focus is the investigation of how people make causal inferences rather than the investigation of whether the presented relations are causal or not.

7.1 VERBAL ANNOTATIONS IN GRAPH-TEXT CONSTELLATIONS

Verbal (or linguistic) annotations are widely used in graph-text constellations both in printed media and in electronic media. Comprehension of verbally annotated graphs presents a specific subcase of multimodal graph-text comprehension in written settings (cf. graph-internal multimodality, see Chapter 1).[3] In order to comprehend a verbally annotated graph, the reader has to integrate the information induced by the annotation text (i.e. the sublayer of textual entities in the graph layer) with the relevant parts of the graph (proper), e.g. the relevant parts of the graph line in line graphs (i.e., the sublayer of graphical entities in the graph layer), to reach a coherent interpretation. In particular, coreference links are constructed between the information induced by the text sublayer of verbal annotations (namely, *annotation label* or *annotation text*) and the information induced by a certain part of the graph determined by visual or spatial properties, such as a peak of the graph line. In most cases, graphical means such as straight lines are used between annotation text and the relevant part of the graph line to facilitate the construction of referential links. In this dissertation, these graphical means are called *annotation icons*. If there is an accompanying text with a separate layout than the graph such as a set of paragraphs (i.e., in a graph text constellation with a separate layout of the text layer and the graph layer), than multimodal comprehension processes additionally involve the integration of the information induced by those textual entities in the text layer of the graph-text constellation.

A sample graph-text constellation with a verbally annotated graph is shown in Figure 7.1. The graph layer includes the graph (proper) and graph-related textual information (e.g., graph title, captions, verbal annotations, etc.).[4] In the verbally annotated graph-text constellation sample, the graph layer on the left involves the graph (proper), the graph title ‘New High for the Dow’, and two verbal annotations ‘Previous high–Jan. 14, 2000–11,722.98’ and ‘Tuesday’s close–11,727.34’. Attached to the annotation texts, annotation icons are connected to circles, which highlight the two peaks of the graph line.

---

[3] The term *annotation* in this dissertation has a *descriptive* aspect rather than *instructive* aspect according to Bernard’s (1990) classification of annotations. In addition, the annotations investigated in this dissertation correspond to *integration-level* annotations in Wallen, Plass and Brünken’s (2005) classification.

[4] In this dissertation, *verbal annotations* and *captions* are used to mean different parts of graph-text constellations. Captions are verbal descriptions of the figure; they are generally located below the figure and they start with a phrase like ‘Figure 1’. Although both verbal annotations and captions enrich an illustration in a way that labels and legends cannot achieve (Preim, Michel, Hartmann, & Strothotte, 1998), figure captions often refer to a whole illustration, while verbal annotations refer to parts of a figure.

STOCKS & BONDS

Dow Jones Index Hits a New High, Retracing Losses

By VIKAS BAJAJ
Published: October 4, 2006

The best-known measure of the stock market, the Dow Jones industrial average of 30 major stocks, rose 0.49 percent yesterday to squeak past a closing high that was set in January 2000 amid a technology-driven market boom.

In contrast to those heady days, though, investors and market professionals are greeting the current rally with more relief than euphoria, noting that the broader stock market has yet to find its way back to previous highs.

"I am happy that it has now happened so that we can move onto something else," said Liz Ann Sonders, chief investment strategist at Charles Schwab.

Stocks have been climbing without fanfare since late in July, bolstered by a decline in energy prices and by mounting signs that the Federal Reserve will not raise interest rates again this year.

The spark for yesterday's gains was a 4 percent drop in oil prices that pushed the next-month futures price of crude oil below $60 a barrel for the first time since March.

Figure 7.1. A sample graph-text constellation with a verbally annotated graph. (© The New York Times. Dow Jones Index Hits a New High, Retracing Losses, by Vikas Bajaj, published on October 4, 2006).

A sample of a verbally annotated graph is shown in Figure 7.2.

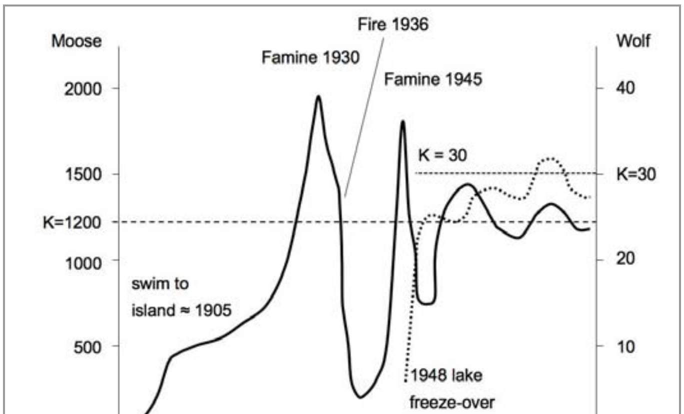


Figure 7.2. A sample verbally annotated line graph: Isle Royale Moose/Wolf Progression. The figure is redrawn based on N. C. Heywood's course material in Biogeography. http://www.uwsp.edu/geo/faculty/heywood/Geog358/Population/Populate2.htm, retrieved on October 11, 2009). The figure goes back to Harris, A. and Tuttle, E. Geology of National Parks. ISBN 0-8403-2810-9.

The sample verbally annotated graph in Figure 7.2 has been previously presented in Chapter 2. The graph shows the population trend of moose and wolf species in the Isle Royale from the 1900s until the end of 1970s. The population of the moose species in the island, which is represented by the solid graph line in the graph, increased after they swam to the island in 1905. In 1939, a famine broke up and the moose population started to decrease. The decrease speeded up after a fire in 1936. In the graph, the annotation text ‘Fire 1936’ is accompanied by an annotation icon, i.e. the straight line between the text and the graph (proper), whereas the annotation text “Famine 1930” is not accompanied by an annotation icon. The moose population decreased after a second famine in 1945. In 1948, the wolves arrived at the island after a lake freeze-over. The two populations then reached to an equilibrium state, which is characterized by the two S-shaped curves. The verbally annotated graph in Figure 7.2 exemplifies the representation of temporal and causal relations between the processes and events represented by means of graphical entities and verbal annotations in the graph layer.

Verbal annotations are used to refer to various types of entities in graph-text constellations. Annotation texts may represent information about the time which designates when the events occurred, or they may represent numerical information of the value of the domain variable. Moreover, verbal annotations may be used to name the graphical entities in graphs with multiple graph lines. They may also be used to annotate events and processes represented by graphical entities, as well as to annotate external events and their relation to the events and processes represented by the graph lines, as exemplified in Figure 7.2.[5] In all those uses of verbal annotations, they *foreground* a graphical entity as a whole or a certain part of the graphical entity, such as the peak of a graph line. By foregrounding graphical entities (and/or real-world entities referred to by the graphical constituents such as event occurring at different times), verbal annotations provide graph-internal information, which is necessary for graph-internal multimodal integration.[6]

In addition to providing graph-internal information, verbal annotations provide information that facilitates the integration between graphical elements in the graph

[5] In this dissertation, this type of annotation is called *event annotation*; it is assumed that graph lines in line graphs represent process concepts such as INCREASE and DECREASE, as stated in Chapter 4. Events, on the other hand, start or end processes. The difference between events and processes are beyond the scope of this dissertation. Therefore, the terms *event* and *process* are used for both graphical entities and verbal annotations.

[6] The term *foregrounding* has been used by mental model researchers who propose that cognitive representation of text is the representation of objects, events, and processes that are described by the text, rather than the description of the text (see Glenberg, Meyer, & Lindem, 1987, for contribution of mental models in foregrounding concepts during text comprehension). In this dissertation, the term *foregrounding* is used in the same sense, to mean activation of certain elements (tokens) and retention of this information foregrounded during comprehension (also see Zwaan & Radvansky, 1998).

layer and accompanying text in the text layer of graph-text constellations; the content of annotation text usually corresponds to graphical or real-world referents of the graph mentioned in the accompanying text. Therefore, annotations may have the role of bridging the information induced by the graphical entities and information induced by the accompanying text.

In multimodal comprehension of graph-text constellations, verbal annotations have similar roles with other referential entities used in communication, such as gestures and expressions with deictic reference. In a classroom setting or in a presentation setting, it is quite a common observation that the presenter uses gestures to annotate graphical entities. In particular, a set of pointing gestures is used which are characterized by hand and arm movements, and body posture. Those pointing gestures unfold in time synchronously with the verbal description of the domain events. Not only in one-to-many meeting settings such as classroom and presentation settings, but also in one-to-one discussions with graphs, humans use those gestures to attract the attention of the communication partner to the relevant parts of graphical entities. As stated by Hill and Hollan (1991), the use of visualizations in spoken multimodal communication settings always involves *pointing*. Moreover, the act of referring to can take different forms in communication (Clark, 2003; Brennan, 2005; Heer & Agrawala, 2008; Kong & Agrawala, 2009). The specific pointing act in multimodal communication settings is called *deixis*: It aims to attract visual attention of the communication partner and it refers to an entity in the environment.

Jackendoff (2002) states that the processing of purely referential deictic linguistic expressions requires "going out of the language faculty and making use of the visual system" (p. 307). Within the framework of the representational modularity hypothesis (see Chapter 2, Chapter 4), the necessary construction is a *visual percept*, which is linked to the deictic expression via the SpS-CS interface. Gestures, like linguistic expressions with a deictic reference, can access entities in the conceptualized world. The conceptualized entities are induced not only by objects but also by events and locations. In other words, the common underlying correspondence to conceptual entities allows gestures accompanying linguistic expression with a deictic reference. Jackendoff's approach and the findings in the visualization literature have implications for multimodal comprehension of graph-text constellations that involve verbally annotated graphs: Verbal annotations in graphs may facilitate the construction of referential links between the representations, in a similar way to a pointing gesture that facilitates the construction of the referential link between a deictic expression and its reference. The absence of annotations is similar to the case in which the deictic expression is uttered but not accompanied by a pointing gesture (or eye gaze). As a result, the absence of an annotation icon results in an increase in cognitive effort in construction of referential links. Moreover, the working assumption of this dissertation for common conceptual basis for multimodal representations (see Chapter 4) suggests that humans integrate the information induced by verbal annotations and graphical entities to reach a coherent interpretation of the representations in different modalities. The experimental investigations that are reported in this chapter focus on those implications of verbal annotations in multimodal comprehension of graph-text constellations. The Study 1, reported below, investigates the influence of presence and absence of verbal annotations in graph-text constellations.

7.2 STUDY 1: THE BRIDGING ROLE OF ANNOTATIONS IN GRAPH-TEXT CONSTELLATIONS

The purpose of the Study 1 was to investigate the influence of verbal annotations in graph-text constellations on participants' inspection of graphs. In addition, facilitatory and inhibitory effects of verbal annotation on comprehension and recall were investigated. The closest investigations of the role of verbal annotations in multimodal constellations—in particular, in picture-text constellations—have been conducted in the domain of multimedia learning. Although implications of the findings for picture-text constellations to graph-text constellations are not clear at this stage, a set of initial predictions can be made based on the previous findings in multimedia learning research. In particular, following the *multimedia effect* proposed by Mayer (2001/2009), which states that adding pictures to text may facilitate learning, an initial prediction could be proposed for graph-text constellations: adding graphs to text may have facilitatory effects in comprehension and recall of graph-text constellations, which could result in facilitatory effects on learning with graph-text constellations. In addition, following Mayer's *spatial contiguity effect*, it could be predicted that using verbal annotations in the graph layer of a graph-text constellation might introduce further facilitatory effects on comprehension and recall. In other words, verbal annotations might have a bridging role in graph-text constellations during the course of the integration of information contributed by graphical entities and linguistic entities. On the other hand, the presentation of information with a separate layout of the graph-layer and the text-layer may result in inhibitory effects due to the attention split between the graph and the text (cf. the *split attention effect* proposed by Chandler & Sweller, 1992). An experimental study was conducted to test these predictions, which is reported below.

7.2.1 Participants, Materials and Design

Thirty-two participants were paid to participate in the experiment. The mean age of the participants was 22.0 (SD = 3.68).

The experiment had a within-subject design. Each participant was presented three newspaper articles. Each article was presented to the participants in one of the following three conditions:

Condition 1 (Text plus annotated graph condition, shortly, *annotated-graph condition*): In this condition, a verbally annotated graph and accompanying text (i.e., the paragraphs) were presented to the participants.

Condition 2 (Text plus non-annotated-graph condition, shortly, *non-annotated-graph condition*): In this condition, the text was presented together with a non-annotated graph.

Condition 3 (Text-only condition): In this condition, the text was presented without an accompanying graph.

In the annotated-graph condition and the non-annotated-graph condition, the graphs were placed to the left of the paragraphs, as exemplified in Figure 7.3. In the text-only condition, the graph layer was left blank.

The stock market domain was selected as the domain of experimental investigation. Graph-text constellations with verbally annotated graphs appear frequently in newspaper and magazine articles, as well as in scientific articles. In the stimuli presented to the participants, the text layer was based on the source newspaper articles; part of the content was omitted for the purpose of reducing the size of the material. The main protagonists in the source articles (e.g., the Turkish stock market IMKB and the Dow Jones) were changed to less known alternatives (namely, Taiwan, Hungary, Budapest, and Singapore stock markets), to prevent interference with previous knowledge of the participants. Since the graphs in the source articles were different from each other in design layout, the graphs were redrawn to maintain a similar layout. All experimental stimuli were in Turkish, the native language of the participants (see Figure 7.3 for sample stimuli).[7]

[7] The original articles were: ©The New York Times. "Dow Jones Index Hits a New High, Retracing Losses", by Vikas Bajaj, published on October 4, 2006 (translated into Turkish by the experimenter); ©Radikal. "Endeks 30 bin sınırını aştı [The Index exceeded the 30 thousand limit" (translation by the experimenter)], published on August 5, 2005; ©Sabah. "Borsa, Mayıs 2006'ya geri döndü [The stock market retraced to May 2006" (translation by the experimenter)], published on January, 25, 2007.

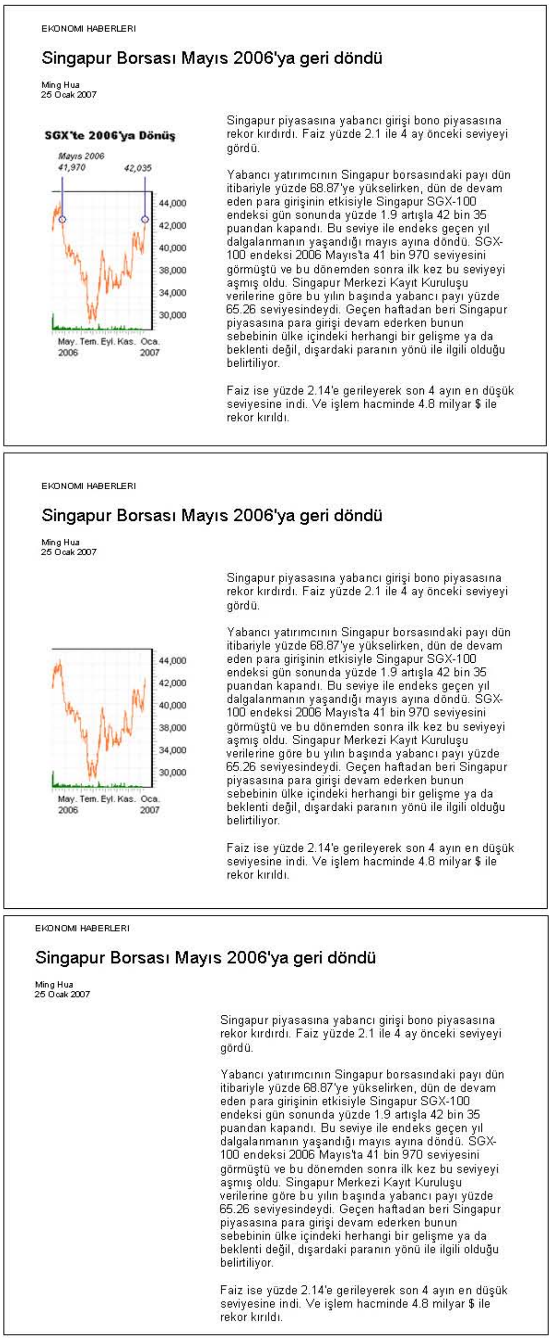


Figure 7.3. Sample stimuli from the experiment. The figure at the top shows the annotated-graph condition; the middle figure shows the non-annotated-graph condition; and the figure at the bottom shows the text-only condition.

7.2.2 Procedure

The participants were tested in single sessions. After calibration of the eye tracker, the participants read the instructions. They were informed that they were expected to investigate three newspaper articles and after the presentation of the three newspaper articles, they were expected to answer a set of posttest questions. After reading the instructions at their own pace, the participants moved to the next screen by pressing a key on the keyboard. There were no time limitations in the experiment. The entire session took approximately 10-15 minutes. After the experiment session, the participants answered the posttest questions on paper. The material was presented to the participants in random order. A 50 Hz. non-intrusive eye tracker recorded eye-tracking data (see 5.2 for the details of the eye tracker and the general description of the methodology).

7.2.3 Results

Eye tracker calibration partially or totally failed in seven cases either due to inaccurate calibration at the beginning or loss of calibration in later stages of the experiment. The data from one participant was also excluded after a self-assessment of experience in stock market graphs. The data from 24 participants were included to the analysis.

The three dependent variables of the experiment were fixation count, fixation duration, and gaze time. The Areas of Interest (AOIs) were specified as shown Figure 7.4. The non-annotated-graph condition and the text-only condition did not have all the AOIs since they did not have the corresponding parts of the stimuli.

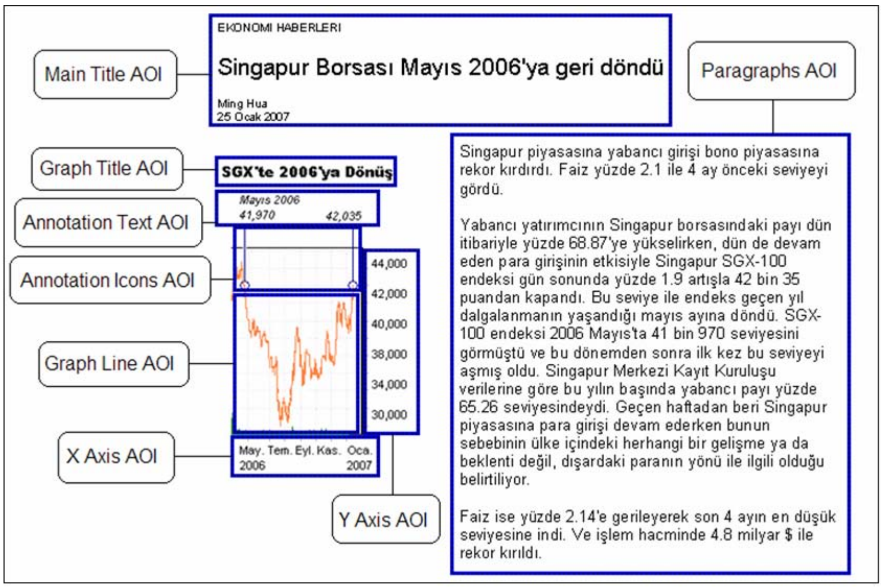


Figure 7.4. Sample Area of Interest (AOI) specification.

The results of the experiment are presented in three parts below. The first part presents the analysis of the eye movement parameters. The second part presents findings concerning the characteristics of eye movement patterns (i.e., scanpaths) on the stimuli. The third part presents the analysis of answers to posttest questions.

7.2.3.1 The Analysis of Eye Movement Parameters

In the analysis of eye movement parameters, fixation count, fixation duration, and gaze time values were calculated for the specified AOIs. Fixations with duration of less than 100 ms were not included to the analysis (see Chapter 5 for the criteria for elimination of eye movement data). The results of the mean fixation counts on the AOIs under the three conditions are shown in Table 7.1.

Table 7.1. Mean fixation counts on the AOIs. The numbers in parentheses show standard deviation. The symbol "–" means that the analysis was not applicable to the AOI.

AOI	Annotated-graph	Non-annotated-graph	Text-only
Main Title	14.8 (8.01)	12.0 (5.47)	14.3 (8.10)
Paragraphs	164.8 (90.5)	169.5 (74.4)	158.4 (62.3)
Graph Line	1.38 (1.53)	3.62 (4.15)	–
X-Axis	1.62 (1.02)	1.95 (2.50)	–
Y-Axis	1.33 (1.66)	1.24 (1.97)	–
Graph Title	2.05 (2.22)	–	–
Annotation Text	3.38 (5.20)	–	–
Annotation Icons	1.52 (3.06)	–	–

A two-way analysis of variance was conducted with two within-subject factors, AOI (*main title* versus *paragraphs*) and condition (*annotated-graph*, *non-annotated graph*, and *text-only*). The AOI main effect, the condition main effect, and the AOI x Condition interaction effect were tested using the multivariate criterion of Wilks's Lambda (Λ). The AOI main effect was significant, $\Lambda = .13$, $F(1, 20) = 137.28$, $p < .01$. The condition main effect was not significant, $\Lambda = .99$, $F(2, 19) = 0.13$, $p = .87$. Also, the interaction effect AOI x Condition was not significant, $\Lambda = .96$, $F(2, 19) = 0.37$, $p = .87$. The results showed that the use of graph in the stimuli, either in annotated form or in non-annotated form, did not affect mean fixation counts on the *main title* AOI and the *paragraphs* AOI, compared to the text-only condition. In other words, there was no significant difference between the three conditions with respect to fixation counts on the title and paragraphs of the presented graph-text constellations.

A second analysis of variance was conducted with two within-subject factors, AOI (*graph line, x-axis*, and *y-axis*) and condition (*annotated-graph* versus *non-annotated-graph*). The results of the test showed that the AOI main effect was significant, $\Lambda = .62$, $F(2, 19) = 5.80$, $p < .01$; the condition main effect was significant, $\Lambda = .70$, $F(1, 20) = 8.58$, $p < .01$; and the interaction effect AOI x Condition was not significant, $\Lambda = .85$, $F(2, 19) = 1.71$, $p = .21$. The follow up pairwise comparisons showed that the difference in mean fixation counts on the *graph line* AOI was significant between the two conditions, $t(20) = 2.66$, $p < .05$. The difference between the conditions was also significant for *x-axis* AOI, $t(20) = 2.75$, $p < .05$. Nevertheless the difference was not significant for *y-axis* AOI, $t(20) = 1.96$, $p = .06$.

In summary, mean fixation counts on the paragraphs were not different between the three conditions. However, mean fixation counts on the graph region, in particular, on the *graph line* AOI and on the *x-axis* AOI (but not on the *y-axis* AOI), were higher in the absence of annotations. The results of mean fixation durations are presented below.

The results of the mean fixation durations on the AOIs under the three conditions are shown in Table 7.2.

*Table 7.2***.** Mean fixation durations on the specified AOIs. All numbers are in ms. The numbers in parentheses show standard deviation. The symbol "–" means that the analysis was not applicable to the AOI.

AOI	Annotated-graph	Non-annotated-graph	Text-only
Main Title	274 (81)	258 (79)	272 (68)
Paragraphs	288 (79)	302 (89)	293 (73)
Graph Line	220 (49)	322 (184)	–
X-Axis	248 (73)	390 (186)	–
Y-Axis	336 (230)	370 (104)	–
Graph Title	315 (202)	–	–
Annotation Text	327 (155)	–	–
Annotation Icons	209 (126)	–	–

A two-way analysis of variance was conducted with two within-subject factors, AOI *(main title* versus *paragraphs*) and condition (*annotated-graph*, *non-annotated graph*, and *text-only*). The results showed that the AOI main effect was significant, $\Lambda = .79$, $F(1, 20) = 5.37$, $p < .05$; the condition main effect was not significant, $\Lambda = .99$, $F(2, 19) = 0.04$, $p = .96$; and the interaction effect AOI x Condition was not significant, $\Lambda = .87$, $F(2, 19) = 1.44$, $p = .26$. The results of the test revealed that adding the graph to the text, either in annotated form or in non-annotated form, did not affect mean fixation durations on *main title* AOI and the *paragraphs* AOI.

A second analysis of variance was conducted with two within-subject factors, AOI (*graph line*, *x-axis*, and *y-axis*) and condition (*annotated-graph* versus *non-annotated-graph*). The test indicated a significant condition main effect, $F(1, 54) = 5.30$, $p < .01$. The AOI main effect was not significant, $F(2, 54) = 1.54$, $p = .22$. The Condition x AOI interaction effect was also not significant, $F(2, 54) = 0.51$, $p = .60$. Further pairwise comparisons showed that mean fixation duration on the *graph line* AOI was significantly different between the two conditions, $t(16) = -2.05$, $p < .05$. Furthermore, mean fixation duration on the *x-axis* AOI was significantly different between the two conditions $t(12) = -2.20$, $p < .05$. Nevertheless, the difference was not significant for the y-axis AOI between the two conditions, $t(5) = 0.31$, $p = .38$.

In summary, the difference between the mean fixation durations on the title and the paragraphs was significant. The mean fixation durations on the paragraphs were longer than the mean fixation durations on the title. However, mean fixation durations on the titles and the paragraphs were not significantly different among the three conditions. On the other hand, mean fixation durations on the graph, in particular on the *graph line* AOI and the *x-axis* AOI, but not on the mean fixation durations on the *y-axis* AOI, were higher in the absence of annotations than the mean fixation durations on these AOIs in the annotated-graph condition. This reflects higher cognitive effort of the participants in the non-annotated-graph condition. The results of mean gaze times are presented below.

The results of the mean gaze times on the AOIs under the three conditions are shown in Table 7.3.

Table 7.3. Mean gaze times on the specified AOIs. All numbers are in ms. The numbers in parentheses show the standard deviation. The symbol "–" means that the analysis was not applicable to the AOI.

AOI	Annotated-graph	Non-annotated-graph	Text-only
Main Title	4125 (2263)	3205 (1749)	3883 (2164)
Paragraphs	48313 (30102)	52162 (30054)	46943 (26212)
Graph Line	329 (396)	983 (1145)	–
X-Axis	172 (337)	757 (1149)	–
Y-Axis	110 (244)	487 (848)	–
Graph Title	755 (1139)	–	–
Annotation Text	1168 (1983)	–	–
Annotation Icons	529 (1709)	–	–

A two-way analysis of variance was conducted with two within-subject factors, AOI (*main title* versus *paragraphs*) and condition (*annotated-graph*, *non-annotated graph*, and *text-only*). The results showed that AOI main effect was significant, $\Lambda = .12$, $F(1, 20) = 69.71$, $p < .01$; condition main effect was not significant, $\Lambda = .96$, $F(2, 19) = 0.38$, $p = .69$; and interaction effect AOI x Condition was also not significant, $\Lambda = .92$, $F(2, 19) = 0.85$, $p = .44$. The result showed that the addition of a non-annotated-graph or the addition of an annotated-graph to the text did not influence the mean gaze time on the title or on the paragraphs.

A two-way ANOVA was conducted with two within-subject factors, AOI (*graph line*, *x-axis*, and *y-axis*) and condition (*annotated-graph* versus *non-annotated-graph*). The results showed that the AOI main effect was significant, $\Lambda = .66$, $F(2, 19) = 4.85$, $p < .05$. The condition main effect was also significant, $\Lambda = .68$, $F(1, 20) = 9.26$, $p < .01$. The interaction effect AOI x Condition was not significant, $\Lambda = .89$, $F(2, 19) = 1.15$, $p = .34$. The follow-up pairwise comparisons showed that the differences in mean gaze times between the two conditions were significantly different between the *graph line* AOIs, $t(20) = 2.80$, $p < .05$; between the *x-axis* AOIs, $t(20) = 2.54$, $p < .05$; and between the *y-axis* AOIs, $t(20) = 2.06$, $p = .05$.

In summary, the addition of the graph, either in annotated form or in non-annotated form, did not influence the mean gaze time on the paragraphs. Nevertheless, mean gaze time on the graph region, in particular in the *graph line* AOI, the *x-axis* AOI, and the *y-axis* AOI, was longer in the absence of annotations than in mean gaze times in the annotated-graph condition.

7.2.3.2 The Analysis of Scanpath Characteristics

The quantitative analysis of scanpaths is technically difficult, particularly in naturalistic and complex stimuli, as the ones used in the Study 1. Therefore, the general characteristics of scanpaths are verbally described below. The focus of the description is the gaze shifts, i.e., the shifts of eye movements from the text layer (i.e., the paragraphs) to the graph layer of the graph-text constellations. The graph layer covers all the AOIs on the graph layer, i.e., the AOIs on the graph (proper), axes, graph title, and annotations.

The analysis of the eye movement protocols showed that text was attended before the graphs; in other words, the paragraphs were usually read before the graph was inspected. This finding is parallel to the findings in previous research on multimodal comprehension of text and pictures or diagrams (Hegarty et al., 1991; Carroll et al., 1992; Rayner, et al., 2001).

The results of the gaze shifts showed that the participants made more gaze shifts in the annotated-graph condition compared to the gaze shifts in the non-annotated-graph condition. Table 7.4 shows the percentage of participants who made two gaze shifts, and the percentage of participants who made more than two gaze shifts in the annotated-graph condition and in the non-annotated graph condition. For both, the percentages in the annotated-graph condition were higher than the percentages in the non-annotated-graph condition.

Table 7.4. The percentages of the participants who made two gaze shifts and the percentages of the participants who made more than two gaze shifts during their inspection of the stimuli.

	Annotated-graph	Non-annotated-graph
Two gaze shifts	66.7	42.9
More than two gaze shifts	45.8	23.8

The analysis of inspection characteristics showed that after the gaze shifts in the annotated-graph condition, the participants inspected the annotations more than they inspected the graph line. In other words, the shifts between the graph layer and the text layer of the graph-text constellations in the annotated-graph condition were mainly between the annotations on the graph and the paragraphs in the text layer. Since there were no annotations in the non-annotated-graph condition, the gaze shifts in the non-annotated-graph condition were between the graph (proper), i.e. the graph line, in the graph layer and the paragraphs in the text layer of the graph-text constellations. On the other hand, the results of the fixation count analysis showed that the mean number of fixations on the graph (proper) was higher in non-annotated-graph condition than in the annotated-graph condition. These findings show that the participants inspected the stimuli in the non-annotated-graph condition by fixating consecutively within the graph layer; whereas, the stimuli in the annotated-graph condition were inspected by more fixations between the text layer (i.e., the paragraphs) and annotations, rather than by consecutive fixations within the graph region.

7.2.3.3 The Analysis of Answers to Posttest Questions

The participants answered a posttest questionnaire after the presentation of the experimental stimuli. The questionnaire included nine multiple-choice recall questions about the content of the graph-text constellations. Each correct answer was scored by *1*, each incorrect answer was scored by *-1*. The answers from all 32 participants were included to the analysis. The outlier data, which were above or below the mean plus/minus two standard deviations, were excluded from the analysis.

A one-way analysis of variance was conducted with a within-subject factor condition (annotated-graph, non-annotated graph, and text-only). The dependent variable was the test score; the independent variable was condition. The results showed that the condition main effect was significant, $\Lambda = .79$, $F(2, 25) = 3.32$, $p < .05$. Further pairwise comparisons showed that posttest scores in the non-annotated-graph condition were significantly higher than the posttest scores in the annotated-graph condition, $t(26) = -2.21$, $p < .05$. The posttest scores in the non-annotated-graph condition were also significantly higher than the posttest scores in the text-only condition, $t(26) = 2.36$, $p < .05$. On the other hand, the difference between the posttest scores in text-only condition and annotated-graph condition was not significant, $t(26) = 0.00$, $p = 1.00$.

In summary, the results of the posttest scores showed that participants' recall was better if a non-annotated graph was presented together with the paragraphs. However, there was no difference if an annotated graph was presented together with the paragraphs.

7.2.4 Summary of the Findings in Study 1

The results of the Study 1 were presented in three parts. The first part of the results, i.e., the analysis of fixation counts, fixation durations, and gaze times on the specified AOIs under the three experimental conditions showed that the presence of the graph, either in the annotated form or in the non-annotated form, did not influence eye movements on the title and the paragraphs in the text-layer of the graph-text constellations. On the other hand, higher mean fixation counts and longer gaze times in the non-annotated-graph condition were obtained compared to the annotated-graph condition. This finding shows that the participants spent more cognitive effort in the non-annotated-graph condition than they did in the annotated-graph condition. In particular, on the graph-related AOIs, fixation counts were higher, fixation durations were longer, and gaze times were longer in the non-annotated-graph condition compared to the ones in the annotated-graph condition. Although the analysis of fixation counts and fixation durations supported this finding partially (i.e., the difference between the two conditions was significant for the *graph line* AOI and the *x-axis* AOI but not for the *y-axis* AOI), the analysis of total gaze times resulted in significant differences between the two conditions for all AOIs specified for the graph layer of the graph-text constellations.

In the second part of the results, the analysis of gaze pattern characteristics was conducted. The results showed that in the annotated-graph condition more participants made gaze shifts from the text layer to the graph layer of the graph-text constellations compared to the non-annotated-graph condition. In addition, in the annotated-graph condition, more inspections of the verbal annotations were observed after the gaze shifts compared to the inspections of the graph line (proper). A possible reason for this is that the annotation texts included specific date information and value information, which resulted in a reduced need to extract date information from the x-axis and value information from the y-axis. On the other hand, the inspections in the non-annotated-graph condition consisted of consecutive fixations within the graph region. These findings support the view that verbal annotations might have the role of bridging the information induced by graphical entities and linguistic entities within the graph region, as well as the information induced by graphical entities and information induced by the accompanying text.

Nevertheless, the use of verbally annotated graphs in graph-text constellations also introduced inhibitory effects. The third part of the analysis, i.e., the analysis of posttest scores showed that, first, the posttest scores in the non-annotated-graph condition were higher than the posttest scores in the text-only condition. This finding provides empirical support for the multimedia effect (Mayer, 2001/2009), which claims that students learn better with pictures and text than they learn with text. However, the posttest scores in the annotated-graph condition were not higher than posttest scores in the text-only condition. Moreover, the posttest scores in the non-annotated-graph condition were higher than the ones in the annotated-graph condition. These findings show that more frequent gaze shifts between the paragraphs and the graph layer of the graph-text constellations in the annotated-graph condition may result in a split-attention effect (Chandler & Sweller, 1992). Another reason may be that the longer gaze times on the graph time might have improved subsequent recall as early research on recall stated (e.g., Craik and Lockhardt, 1972).

In summary, the analyses of eye movements provided empirical evidence for the role of annotations as bridging the information contributed by different representational modalities in graph-text constellations. However, the results of the posttest scores show that the gaze shifts between different representational modalities may also result in inhibitory effects. Therefore, more research is needed for the specification of conditions under which the use of verbal annotations fulfill the bridging function in graph-text constellations.

In the following parts of the chapter, the role of verbal annotations is investigated in terms of the interaction between conceptual structures and contents communicated by graphical entities and verbal annotations. In particular, Study 2 investigates causal attribution in verbally annotated graphs.

7.3 STUDY 2: CAUSAL ATTRIBUTION IN VERBALLY ANNOTATED GRAPHS

As stated in Chapter 2, humans are able to make causal attributions (or causal inferences) using the information induced by linguistic entities and graphical entities in graph-text constellations. The processes and events represented by graphical entities, such as increases and decreases, can also be referred to by verbal annotations, thus contributing to readers' construction of coherent interpretation of graph-text constellations. In multimodal comprehension of graph-text constellations, the construction of coherence relations between graphical entities and linguistic entities by means of causal attribution, which is a relation among conceptual representations (Jackendoff, 1983), supports the view that the integration of information induced by the two modalities takes place at conceptual level, as presented in the architecture for multimodal comprehension of graph-text constellations (see Chapter 4). An empirical study of the causal attribution in verbally annotated graphs (cf. graph-internal multimodality, see Chapter 1) is presented below. In particular, the influence of the position of verbal annotations on participants' judgments concerning the causal attributions between the events and processes represented by verbal annotations and the graphical entities is investigated. In addition, based on the empirical findings, a set of guidelines is devised for appropriate design of graph-text constellations.

7.3.1 Participants, Materials, and Design

Twenty-nine participants (mean age of 23.4, SD = 3.5) were paid to take part in the experiment. Verbally annotated line graphs that represented change of a domain value in time prepared by the experimenter. The design was 3 x 3 x 2, with three independent parameters (two within-subject parameters and one between-subject parameter).

The first within-subject parameter was the constellation of line segments to form the graph line, with three conditions (henceforth, *graph condition*). The conditions were specified by the binary constellations of two graph lines in different order. The three conditions are presented below (see Figure 7.5 below for sample stimuli):

Condition 1 (GC-1): The graphs in this condition had a straight diagonal line or a smooth curve that represented an increase in the canonical reading direction of graphs in Cartesian frame. This was followed by a horizontal line that represented constancy in the domain value.

Condition 2 (GC-2): The graphs in this condition had a horizontal line that represented constancy in the domain value, which was followed by a straight diagonal line or a smooth curve that represented an increase of the domain value in time.

Condition 3 (GC-3): The graphs in this condition had a straight diagonal line or a smooth curve that represented an increase in the canonical reading direction of the graphs. This was followed by another straight diagonal line or smooth curve that represented a decrease.

The second within-subject parameter was concerned with the use of verbal annotations in the graph with three conditions (henceforth, *annotation condition*). The three conditions are presented in the following:

Condition 1 (Left-Annotation Condition): In this condition, there was a single verbal annotation in the graph. The annotation was positioned at the leftmost end of the graph line.

Condition 2 (Middle-Annotation Condition): In this condition, there was a single verbal annotation in the graph. The annotation was positioned at the middle, where the two segments of the graph lines were connected.

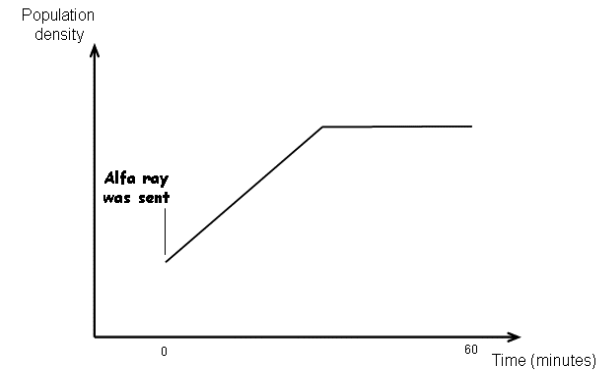
Condition 3 (Double-Annotation Condition): In this condition, there were two verbal annotations in the graph. One of the annotations was positioned at the leftmost end of the graph line; the other annotation was positioned at the middle, where the two segments of the graph lines were connected.

The between-subject parameter was the smoothness of the graph line with two conditions (henceforth, *smoothness condition*). The two conditions are explained below:

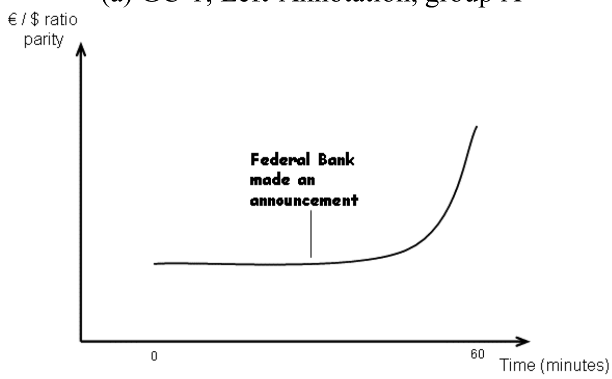
Condition 1 (Group A): The participants of group A (15 participants) were presented graphs that had graph lines in the form of binary constellations of straight lines. Particularly, the graphs in the GC-1 of group A had a straight diagonal line that represented an increase, which was followed by a horizontal line that represented constancy in the domain value. The graphs in the GC-2 had a horizontal line that represented constancy in the domain value, which was followed by a straight diagonal line that represented an increase. The graphs in the GC-3 had a straight diagonal line that represented an increase, which was followed by another straight diagonal line that represented a decrease.

Condition 2 (Group B): The participants of group B (14 participants) were presented graphs that had graph lines in the form of binary constellations of smooth curves and straight lines. Particularly, the graphs in the GC-1 of group A had a smooth curve that represented an increase, which was followed by a horizontal line that represented constancy in the domain value. The graphs in the GC-2 had a horizontal line that represented constancy in the domain value, which was followed by a smooth curve that represented an increase. The graphs in the GC-3 had a smooth curve that represented an increase, which was followed by another smooth curve that represented a decrease.

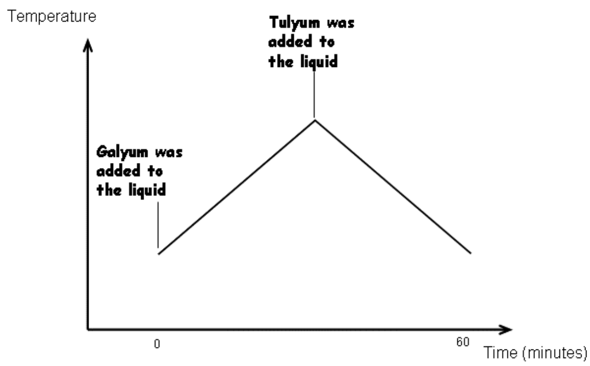
Sample material from the experiment is shown in Figure 7.5.



(a) GC-1, Left-Annotation, group A



(b) GC-2, Middle-Annotation, group B



(c) GC-3, Double-Annotation, group A

Figure 7.5. Sample material from the experiment (translated into English by the author of the dissertation; the font sizes in the figure were enlarged for better visibility). A more complete set of experimental stimuli can be seen in 7.3.3, where the results of the experiment are presented.

Three domains were used (chemistry, finance and biology) in the stimuli. The domain was not evaluated as an independent parameter in this study. The participants answered a set of questions concerning the causal relation between the processes and events represented by the verbal annotations, and the processes and events represented by the graphical entities, i.e. the process-lines and state-lines. Each participant answered a total of 30 questions. Each question was located under the graph, in the same screen. There were nine questions in the GC-1 (three questions per annotation condition), nine questions in the GC-2 (three questions per annotation condition), and twelve for the GC-3 (four questions per annotation condition). The questions were in the form as shown in (1a), and exemplified in (1b–c). A schematic representation of the questions used in the experiment is shown in Table 7.5.

(1) a. Did <*the event represented by the left/middle annotation*> cause <*the domain value*> to <*increase/decrease/stay constant*>?
b. Did sending the alpha ray cause population density to increase?
c. Did adding Tulyum to the liquid cause temperature to decrease?

Table 7.5. A schematic representation of the questions used in the experiment. E is used as an abbreviation for the event or process represented by the verbal annotations. Accordingly, E-left is abbreviation for the left-annotations, and E-middle is the abbreviation the middle annotations. V is used as an abbreviation for the domain value of the graph.

	GC-1	GC-2	GC-3
Left-annotation	Q-1: Did E cause V to increase?	Q-1: Did E cause V to remain constant?	Q-1: Did E cause V to increase?
	Q-2: Did E cause V to stop increasing?	Q-2: Did E cause V to stop remaining constant?	Q-2: Did E cause V to stop increasing?
	Q-3: Did E cause V to remain constant?	Q-3: Did E cause V to increase?	Q-3: Did E cause V to decrease?
Middle-annotation	Q-1: Did E cause V to increase?	Q-1: Did E cause V to remain constant?	Q-1: Did E cause V to increase?
	Q-2: Did E cause V to stop increasing?	Q-2: Did E cause V to stop remaining constant?	Q-2: Did E cause V to stop increasing?
	Q-3: Did E cause V to remain constant?	Q-3: Did E cause V to increase?	Q-3: Did E cause V to decrease?
Double-annotation	Q-1: Did E-left cause V to increase?	Q-1: Did E-left cause V to remain constant?	Q-1: Did E-left cause V to increase?
	Q-2: Did E-left cause V to remain constant?	Q-2: Did E-left cause V to increase?	Q-2: Did E-left cause V to decrease?
	Q-3: Did E-middle cause V to remain constant?	Q-3: Did E-middle cause V to increase?	Q-3: Did E-middle cause V to decrease?
	Q-4: Did E-middle cause V to stop increasing?	Q-4: Did E-middle cause V to stop remaining constant?	Q-4: Did E-middle cause V to stop increasing?

The participants reported their answers by clicking on a Visual Analog Scale (VAS) that was located under the question (see Figure 7.6). A visual analog scale is a horizontal or vertical straight line that is used to measure subjective judgments (see Wewers & Lowe, 1990, for an overview of types and historical development of visual analog scales). The horizontal VAS in the experiment had the end anchors *No with high probability* and *-100* at the left end, the end anchors *Yes with high probability* and *-100* at the right end, and the anchors *I am in the middle* and *0* at the middle. The participants were informed that they could use not only the anchor points but also the points in between the two to report their judgments.

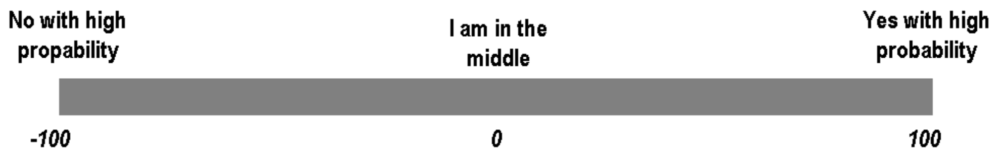


Figure 7.6. The visual analog scale (VAS) used in the experiment.

7.3.2 Procedure

The participants participated in the experiment in single sessions. They were first introduced preliminary information about statistical information graphics with samples. After the introduction, a practice session was presented for explaining the task and the use of the visual analog scale (VAS). In the experiment session, before the presentation of the stimuli, a story that included the necessary context information was presented to the participants. A sample context story can be seen in Appendix C.1. The participants were then presented the graph on the screen, without question and the VAS. In the next screen, the question, together with the VAS appeared below the graph. The questions were presented to the participants in their labeled order. However, the order of presentation of the graph conditions (GCs) was randomized. In other words, the question Q-1 was followed by Q-2, and then Q-3. The questions appeared below the graph, so the graph was kept intact during the presentation of the questions. A schematic representation of the flow of stimuli can be seen in Figure 7.7.

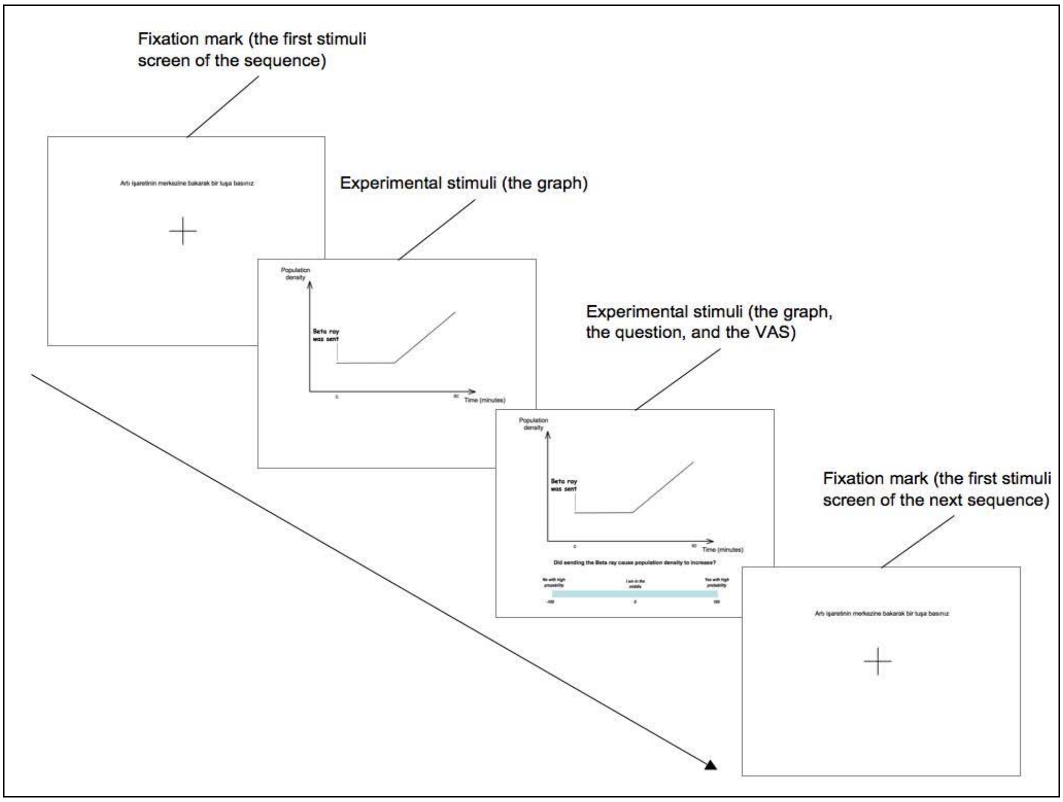


Figure 7.7. The flow of the stimuli in the experiment.

The experiment was self-paced; the participants used the keyboard to proceed to the next screen without time limitation. The experiment took about 30-35 minutes to complete. The results of the experiment are presented below.

7.3.3 Results

The results of the experiment are presented in two parts. First, the results of gaze times of the participants are presented; then, the subjective judgment scores are presented.

7.3.3.1 The Analysis of Gaze Times

For the analysis of gaze times, two Areas of Interest (AOI) were specified as shown by the dotted lines in Figure 7.8. The same AOI template was used for all experimental stimuli.

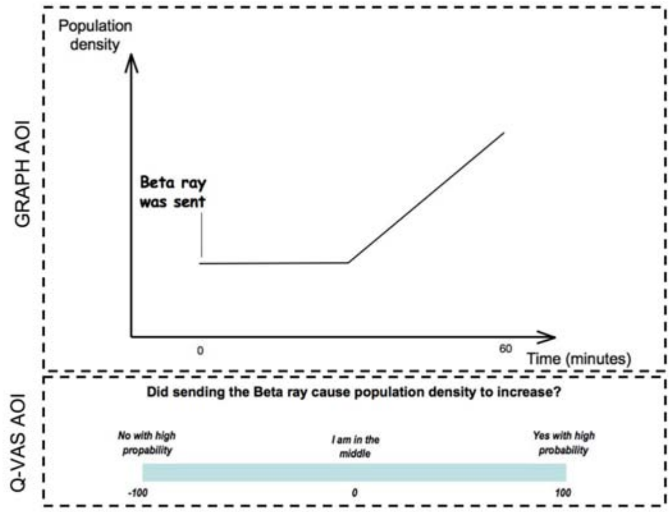


Figure 7.8. AOI specification for the analysis of gaze times.

The graph AOI covered the graph, including the axes and axis labels. The Q-VAS AOI covered the question phrase and the visual analog scale (VAS).

The calibration failed for two participants in group B. All the data from the remaining participants were included to the analysis. The gaze times of the participants on the graph AOI were analyzed for the purpose of comparison between the three annotation conditions and between the two groups of participants. An analysis of variance was conducted with one within-subject factor (annotation condition with three conditions, namely left-annotation, middle-annotation, and double-annotation) and one between-subject factor (group with two conditions, namely group A (the straight line group) and group B (the smooth line group)). The results indicated a significant main effect for the annotation condition, Wilks's $\Lambda = .55$, $F(2, 24) = 9.74$, $p < .05$, multivariate $\eta^2 = .45$. The difference between the two groups was also significant, $F(1, 25) = 9.19$, $p < .05$, multivariate $\eta^2 = .27$. The mean gaze time in group B was longer than the mean gaze time in group A. Further pairwise comparisons revealed that the longest mean gaze time was obtained for the left-annotation condition in both groups.

The gaze time of the participants on the Q-VAS AOI were analyzed for comparison between the three annotation conditions and between the two groups of participants. An analysis of variance was conducted with one within-subject factor (annotation condition with three conditions, namely left-annotation, middle-annotation, and double-annotation) and one between-subject factor (group with two conditions, namely group A (the straight-line group) and group B (the smooth-line group)). The results indicated a significant main effect for the annotation condition, Wilks's $\Lambda = .23$, $F(2, 24) = 22.74$, $p < .05$, multivariate $\eta^2 = .65$. The difference between the two groups was not significant, $F(1, 25) = 0.09$, $p = .77$, multivariate $\eta^2 = .04$. Further pairwise comparisons revealed that the longest mean gaze time was obtained for the left-annotation condition in both groups.

The results of the mean gaze time of the participants show that the participants spent more time to inspect the graph and the question/the visual analog scale in the left-annotation condition compared to the other two annotation conditions. In other words, the difficulties in interpreting the position of the verbal annotation in the left-annotation condition resulted in longer inspection of both the graph and the question/VAS. Moreover, the smooth-line group (i.e., group B) made longer inspections compared to group A. This finding indicates that both the annotation position and the graph type are factors that influence multimodal comprehension of verbal annotations and graphical entities.

A more detailed presentation of gaze times and gaze distributions according to the experimental conditions is provided in Appendix C.2. The analysis of participants' subjective judgment scores is presented below.

7.3.3.2 The Analysis of Subjective Judgment Scores

The participants' subjective judgment scores are analyzed according to the graph conditions (GC-1, GC-2, and GC-3), smoothness conditions (group A and group B) and the questions, as presented in three parts for the three annotation conditions below (i.e., left-annotation, middle-annotation, and double annotation).

7.3.3.2.1 Left-Annotation Condition

The stimuli used in the left-annotation condition are exemplified in Figure 7.9. The top row exemplifies the stimuli for group A, and the bottom row exemplifies the stimuli for group B.

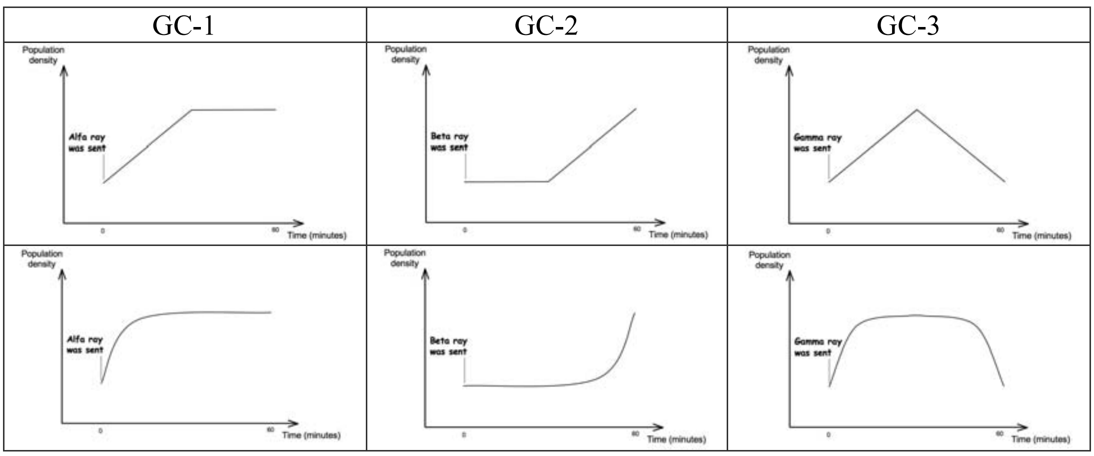


Figure 7.9. Sample stimuli in the left-annotation condition. The upper row exemplifies the stimuli for group A, and the bottom row exemplifies the stimuli for group B.

In the left-annotation condition, the participants answered nine questions (three questions for each graph condition). The first question in each graph condition was about the relation between the events represented by the annotation (henceforth, annotation event), and the processes or states represented by the adjacent graph line. The second question was about the end of the processes or states. The third question was about the processes or states represented by the next line segment.

The judgment scores of the participants were analyzed for comparison between the three graph conditions, the three question conditions, and the two groups of participants. A three-way analysis of variance was conducted with two within-subject factors (graph type with three conditions, namely GC-1, GC-2, and GC-3 and question type with three conditions, namely Q-1, Q-2, and Q-3) and one between-subject factor (group with two conditions, namely group A and group B). The results indicated a significant main effect for question type, Wilks's $\Lambda = .54$, $F(2, 26) = 11.1$, $p < .05$, multivariate $\eta^2 = .46$, and a Graph type x Question type interaction effect, $\Lambda = .40$, $F(4, 24) = 8.91$, $p < .05$, multivariate $\eta^2 = .60$. There was no difference between the mean scores of the two groups. The results are presented in more detail below. Only the statistically significant differences are reported without statistical details for brevity; non-significant differences are stated where necessary.

The questions and the mean judgment scores in the graph condition GC-1 are shown in Table 7.6 below.

Table 7.6. The questions and the mean scores in GC-1, in the left-annotation condition. E is used as an abbreviation for the annotation events; V is used as an abbreviation for the domain value of the graph.

		Group A		Group B	
	Question	M	SD	M	SD
GC-1	Q-1: Did E cause V to increase?	88	18	93	13
	Q-2: Did E cause V to stop increasing?	-35	86	-47	77
	Q-3: Did E cause V to remain constant?	-4	82	-21	75

In the graph condition GC-1 of the left-annotation condition, the participants reported high scores for Q-1. The questions Q-2 and Q-3 received low scores with very high deviation from the mean scores. This shows that agreement among participants was low for Q-2 and Q-3. Pairwise comparisons showed that the mean score for Q-1 was higher than both the mean score for Q-2 and the mean score for Q-3, in both group A and group B. The difference between Q-2 and Q-3 was not significant for either of the groups. In summary, the participants reported in favor of a causal relation between the annotation event and the adjacent graphical entity that represented an increase. On the other hand, the participants rejected causal relation between the annotation event and the graphical entity that represented constancy. A possible explanation for this result is that the temporal distance between the annotation event and the end of the increase process, which was represented between the spatial distances in the graph space, violated the temporal order of occurrence of cause and effect, which in turn prevented the participants to make a causal attribution.

The questions and the mean judgment scores in the graph condition GC-2 are shown in Table 7.7.

Table 7.7. The questions and the mean scores in GC-2, in the left-annotation condition. E is used as an abbreviation for the annotation events; V is used as an abbreviation for the domain value of the graph.

		Group A		Group B	
	Question	M	SD	M	SD
GC-2	Q-1: Did E cause V to remain constant?	11	79	1	85
	Q-2: Did E cause V to stop remaining constant?	15	85	37	80
	Q-3: Did E cause V to increase?	41	81	54	69

In GC-2, the means of the reported scores for the three questions were close to the middle part of the visual analog scale (VAS), with high standard deviations. This shows that the agreement among the participants was low, which indicates that the relation between the annotation event and the processes and events represented by the graphical entities was ambiguous in GC-2. There were no significant differences between the means in any of the questions, as well as between the two groups due to high variability of the scores. On the other hand, the mean score for Q-3 in group A points to the possible interpretation for a *causal delay*.

The literature on causal inference shows that humans are to able make different temporal assumptions about causal delays during the course of dynamic comprehension of events (Hagmayer & Waldmann, 2002; Buehner & May, 2003). The representation of events and processes by line graphs presents a kindred conceptualization of delayed cause. In the current state of the art, this aspect in graph comprehension has not been investigated. Further research is needed to have a better understanding of the underlying mechanisms of delayed cause in line graphs.

The questions and the mean judgment scores in the graph condition GC-3 are shown in Table 7.8.

Table 7.8. The questions and the mean scores in GC-3, in the left-annotation condition. E is used as an abbreviation for the annotation events; V is used as an abbreviation for the domain value of the graph.

		Group A		Group B	
	Question	M	SD	M	SD
GC-3	Q-1: Did E cause V to increase?	50	53	50	66
	Q-2: Did E cause V to stop increasing?	-5	66	10	76
	Q-3: Did E cause V to decrease?	14	70	8	73

In GC-3, the participants reported higher scores for Q-1 than both Q-2 and Q-3. On the other hand, the mean score for Q-1 in GC-3 is lower than the mean score for Q-1 in GC-1. This shows that the presence of the decrease-line after the increase line influenced participants' judgments concerning the causal relation between annotation event and the adjacent increase line. This points to the presence of ambiguity for the causal relation between the annotation event and the processes and events represented by the graphical entities in GC-3.

In summary, the results of the left-annotation condition showed that the position of the annotation resulted in ambiguities in most of the cases, as indicated by the disagreement among the participants in the judgment scores. In group B, in GC-1 and in GC-3, the lack of temporal contiguity between the annotation event, and the events and processes represented by the graphical entities resulted in low judgment scores for causal attribution. In GC-2, in group A, a possible influence of delayed cause was observed. However, more research is needed to have a better understanding of the implications. Those ambiguities in the left-annotation condition are also observed in gaze times of the participants as presented previously. The longer gaze times in the left-annotation condition, compared to the middle-annotation condition and the double-annotation condition indicate that the participants had difficulties in the interpretation of the relation between verbal annotations and graphical entities with respect to causal attribution.

From the perspective of verbally annotated graph design, the findings for the left-annotation condition show that the use of annotations at the left end of the graph line results in ambiguities in the interpretation of causal relations between the annotation events, and the processes and events represented by the graphical entities. One possible explanation is that the participants needed information about the domain value before the occurrence of the annotated event, as well as the information about the domain value after the occurrence of the annotation event. In the middle-annotation condition, this aspect is investigated.

7.3.3.2.2 Middle-Annotation Condition

The stimuli used in the middle-annotation condition are exemplified in Figure 7.10. The top row exemplifies the stimuli for group A, and the bottom row, for group B.

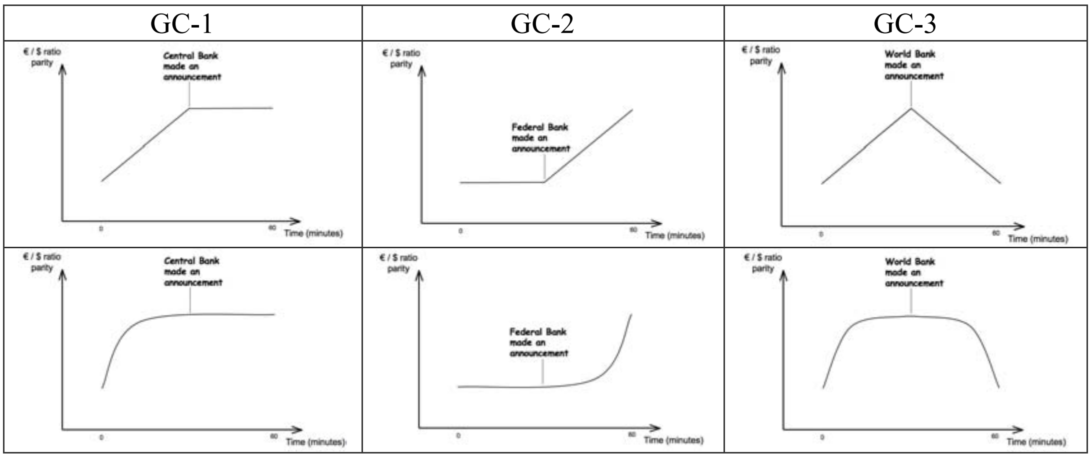


Figure 7.10. Sample stimuli in the middle-annotation condition. The top row exemplifies the stimuli for group A, the bottom row exemplifies the stimuli for group B.

In the middle-annotation condition, the participants answered nine questions (three per graph condition). The first question in each graph condition was about the relation between the annotation events, and the processes and states represented by the left graph line. The second question was about the end of the processes or states. The third question was about the processes or states represented by the next line segment.

The judgment scores of the participants were analyzed for a comparison between the three graph conditions, the three question conditions, and the two groups of participants. A three-way analysis of variance was conducted with two within-subject factors (graph type with three conditions, namely GC-1, GC-2, and GC-3 and question type with three conditions, namely Q-1, Q-2, and Q-3) and one between-subject factor (group with two conditions, namely group A and group B). The results indicated significant main effect for graph type, Wilks's $\Lambda = .46$, $F(2, 26) = 15.2$, $p < .05$, multivariate $\eta^2 = .54$, significant main effect for question type, $\Lambda = .06$, $F(2, 26) = 193.8$, $p < .05$, multivariate $\eta^2 = .94$, and Graph type x Question type interaction effect, $\Lambda = .63$, $F(4, 24) = 3.58$, $p < .05$, multivariate $\eta^2 = .37$. In addition, the mean scores were different between the two groups. The results are presented in more detail below.

The questions and the mean judgment scores in the graph condition GC-1 are shown in Table 7.9.

Table 7.9. The questions and the mean scores in GC-1, in the middle-annotation condition. E is used as an abbreviation for the annotation events; V is used as an abbreviation for the domain value of the graph.

		Group A		Group B	
	Question	M	SD	M	SD
GC-1	Q-1: Did E cause V to increase?	-94	10	-61	72
	Q-2: Did E cause V to stop increasing?	84	50	-19	85
	Q-3: Did E cause V to remain constant?	95	6	40	85

In graph condition GC-1, in the middle-annotation condition, both Q-2 and Q-3 received high scores in group A with no significant difference between the two. The participants in both groups rejected Q-1; the difference between the two groups was not significant. The participants in group B rejected also Q-2, and rarely accepted Q-3. This is possibly due to the lack of temporal contiguity between the annotation event, and events and processes represented by the graphical entities, in particular, the decelerating increase before the annotation event. Moreover, the high standard deviations in group B points to low agreement among the participants, which points to ambiguities in the interpretation of the causal relations between annotation events and the processes and events represented by the graphical entities.

The questions and the mean judgment scores in the graph condition GC-2 are shown in Table 7.10.

Table 7.10. The questions and the mean scores in GC-2, in the middle-annotation condition. E is used as an abbreviation for the annotation events; V is used as an abbreviation for the domain value of the graph.

		Group A		Group B	
	Question	M	SD	M	SD
GC-2	Q-1: Did E cause V to remain constant?	-97	5	-64	66
	Q-2: Did E cause V to stop remaining constant?	83	51	80	27
	Q-3: Did E cause V to increase?	97	4	85	17

In graph condition GC-2, the participants rejected Q-1; the difference between the two groups was not significant. Both Q-2 and Q-3 received high scores in both groups. Furthermore, the mean score for Q-3 was higher in group A than the one in group B. In other words, the causal relation between the annotation event and the increase represented by the graphical entity was stronger in group A than the one in group B. This result indicates that participants' interpretation of the graphs with smooth line segments is different than the interpretation of the graphs with straight line segments, in terms of the relationships between annotation events the processes and events represented by graph lines.

The questions and the mean judgment scores in the graph condition GC-3 are shown in Table 7.11.

Table 7.11. The questions and the mean scores in GC-3, in the middle-annotation condition. E is used as an abbreviation for the annotation events; V is used as an abbreviation for the domain value of the graph.

		Group A		Group B	
	Question	M	SD	M	SD
GC-3	Q-1: Did E cause V to increase?	-94	9	-94	14
	Q-2: Did E cause V to stop increasing?	83	50	-49	81
	Q-3: Did E cause V to decrease?	85	42	68	46

In GC-3, the participants in both groups strongly rejected Q-1. In group A, the difference in the mean scores between Q-2 and Q-3 was not significant; in group B, the difference in the mean scores between Q-1 and Q-2 was not significant. In other words, in group B, the participants rejected the causal relation suggested by Q-2, as well as the one suggested by Q-1. This is possibly due to the lack of the temporal contiguity between the annotation event, and the events and processes represented by the graphical entities. Furthermore, the high variability of the mean scores (i.e., standard deviations) for Q-2 and Q-3 shows disagreement among the participants. Although the agreement among the participants was higher in the middle-annotation condition than in the left-annotation condition, ambiguities remained for GC-3.

In summary, the results of the middle-annotation condition showed that the position of the annotation did not result in ambiguities, in particular in group A. On the other hand, in most of the cases in group B, the agreement among the participants was low. This shows that the graphs in group B were ambiguous with respect to the relation between the annotation events, and the events and processes represented by graphical entities. The lack of temporal contiguity in GC-1 and GC-3 in group B resulted in the rejections of the possibility of causal relation between the annotation events, and the events and processes represented by graphical entities. The findings reveal that the interpretation of the graphs with smooth line segments is different compared to the interpretation of the graphs with straight line segments and corners, with respect to the conceptual structures and content communicated by graphical entities and verbal annotations.

Consequently, the results of the middle-annotation condition showed that, some of the ambiguities that were observed in the left-annotation condition were resolved in the middle-annotation condition. Nevertheless, in group B, some of the ambiguities remained unresolved. Post-experiment reviews with the participants showed that the participants in group B had difficulties in constructing the relation between the annotation event, and the processes and events represented by the graphical entities. This issue is further investigated by Study 3, presented in 7.4. One possibility to facilitate the construction of the coreference relations between linguistically induced and graphically induced entities is to use more than one annotation. The results of the double-annotation condition are presented below.

7.3.3.2.3 Double-Annotation Condition

The stimuli used in the double-annotation condition are exemplified in Figure 7.11. The bottom row exemplifies the stimuli for group B, and the upper row, for group A.

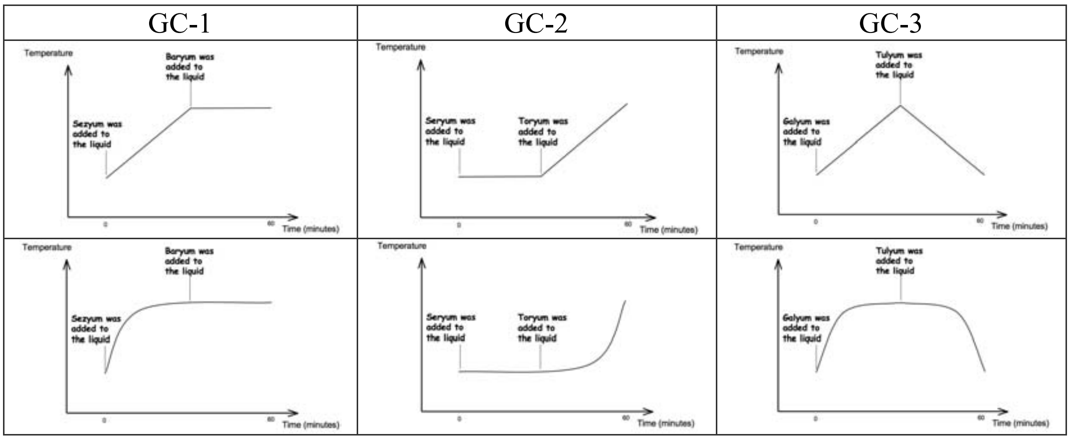


Figure 7.11. Sample stimuli in the double-annotation condition. The upper row exemplifies the stimuli for group A, the bottom row exemplifies the stimuli for group B.

In the double-annotation condition, the participants answered twelve questions (four per graph condition). The first question in each graph condition was about the relation between the left annotation event, and the processes or states represented by the left graph line. The second question was about the relation between the left annotation event, and the processes or states represented by the right graph line. The third question was about the relation between the middle annotation event, and the processes or states represented by the right graph line. The fourth question was about the relation between the middle annotation event, and the processes or states represented by the left graph line.

The judgment scores of the participants were analyzed for comparison between the three graph conditions, the four question conditions, and the two groups of participants. A three-way analysis of variance was conducted with two within-subject factors (graph type with three conditions, namely GC-1, GC-2, and GC-3 and question type with four conditions, namely Q-1, Q-2, Q-3, and Q-4) and one between-subject factor (group with two conditions, namely group A and group B). The results indicated significant main effect for graph type, Wilks's $\Lambda = .61$, $F(2, 26) = 8.28$, $p < .05$, multivariate $\eta^2 = .39$, significant main effect for question type, $\Lambda = .08$, $F(3, 25) = 92.9$, $p < .05$, multivariate $\eta^2 = .92$, and Graph type x Question type interaction effect, $\Lambda = .38$, $F(6, 22) = 6.04$, $p < .05$, multivariate $\eta^2 = .52$. In addition, the mean scores were different between the two groups. The results are presented in more detail below.

The questions and the mean judgment scores in the graph condition GC-1 are shown in Table 7.12.

Table 7.12. The questions and the mean scores in GC-1, in the double-annotation condition. E-left is abbreviation for the annotation events represented by the left-annotations, and E-middle is an abbreviation for the annotation events represented by the middle annotation. V is used as an abbreviation for the domain value of the graph.

		Group A		Group B	
	Question	M	SD	M	SD
GC-1	Q-1: Did E-left cause V to increase?	94	6	77	51
	Q-2: Did E-left cause V to remain constant?	-59	72	-30	77
	Q-3: Did E-middle cause V to remain constant?	88	19	21	83
	Q-4: Did E-middle cause V to stop increasing?	91	14	-15	76

In GC-1, in the double-annotation condition, the participants of group A rejected Q-2. They reported judgments in favor of a causal relation between the left annotation event and the increase process. They also reported that the middle annotation event caused the increase process to stop, and caused the domain value to remain constant. In the mean scores for group A, the standard deviations for the three questions were low, which shows agreement among the participants. In other words, compared to the previously analyzed annotation conditions, the ambiguities were low in the double-annotation condition. The participants of group B reported high scores for Q-1; the difference in Q-1 scores between the two groups was not significant. They rejected Q-2; the difference in Q-2 scores between the two groups was not significant. They also rejected Q-4. There was a weak acceptance for Q-3 in group B. In summary, using double annotations in the graphs with straight line segments resolved the ambiguities that were observed in the left-annotation condition. On the other hand, ambiguities partially remained in the graphs with smooth line segments.

The questions and the mean judgment scores in the graph condition GC-2 are shown in Table 7.13.

Table 7.13. The questions and the mean scores in GC-2, in the double-annotation condition. E-left is abbreviation for the annotation events represented by the left-annotations, and E-middle is an abbreviation for the annotation events represented by the middle annotation. V is used as an abbreviation for the domain value of the graph.

		Group A		Group B	
	Question	M	SD	M	SD
GC-2	Q-1: Did E-left cause V to remain constant?	83	23	45	68
	Q-2: Did E-left cause V to increase?	-85	30	-63	56
	Q-3: Did E-middle cause V to increase?	90	15	70	52
	Q-4: Did E-middle cause V to stop remaining constant?	91	14	86	18

In GC-2, the means scores for Group A were parallel to the mean scores of group A in GC-1. Compared to the GC-2 in the left-annotation condition, the attribution of causality to the left-annotation was different in the double annotation condition, as shown by the mean scores for Q-2, as well as Q-1. In the double-annotation condition, the participants of group A reported that the middle annotation event caused the constancy state to reach to an end, and it caused the domain value to increase, as shown by the high mean scores for Q-3 and Q-4. The mean scores reported by group B were parallel to the ones reported by group A, though with less agreement among the participants in group B. In group B, Q-3 and Q-4 received high scores with no difference between the two. The mean score for Q-1 was low, being different than the Q-1 score in group A.

The questions and the mean judgment scores in the graph condition GC-3 are shown in Table 7.14.

Table 7.14. The questions and mean scores in GC-3, in the double-annotation condition. E-left is abbreviation for the annotation events represented by the left-annotations, and E-middle is an abbreviation for the annotation events represented by the middle annotation. V is used as an abbreviation for the domain value of the graph.

		Group A		Group B	
	Question	M	SD	M	SD
GC-3	Q-1: Did E-left cause V to increase?	96	6	81	50
	Q-2: Did E-left cause V to decrease?	-85	45	-87	26
	Q-3: Did E-middle cause V to decrease?	87	26	56	55
	Q-4: Did E-middle cause V to stop increasing?	95	6	-30	75

In GC-3, the participants of group A reported high scores for Q-1, Q-3, and Q-4, with no significant difference among the three. The mean scores in group B were parallel to the mean scores in group A, except for Q-4. The participants of group B had a higher indeterminacy in their judgment scores compared to the participants of group A.

In summary, the results of the double-annotation condition showed that, some of the ambiguities that were observed in the middle-annotation condition as well as the left-annotation condition were resolved in the middle-annotation condition. Nevertheless, in group B, some of the ambiguities remained unresolved. As previously mentioned, these were ambiguities due to the difficulties in constructing the relation between the annotation event, and the processes and events represented by the graphical entities. This issue is further investigated in Study 3 presented in 7.4.

7.3.4 Summary of the Findings in Study 2

The purpose of the Study 2 was to investigate the construction of coherence relations in verbally annotated graphs. In particular, the interaction between verbal annotations and graphical entities was investigated with respect to the attribution of causal relations between the annotation events and events and processes represented by the graph lines. The mean gaze times and the mean judgment scores concerning the likelihood of a causal relation between annotation event and the events and processes represented by graphical entities showed that different experimental conditions resulted in differences in participant's interpretations. The results of the study are summarized below.

The findings of Study 2 showed that the left-annotation condition resulted in ambiguities in interpretation of the causal relation between annotation events, and the events and processes represented by the graph lines. A possible reason for these ambiguities may be the lack of information about the pre-event conditions, i.e., the information about the domain value before the occurrence of annotation event. In addition, the findings showed that not only the lack of information about the pre-event conditions but also the post-process or post-state conditions influenced the judgment scores; the presence of a state after an increase process resulted in stronger causal inferences than the inferences in the presence of a decrease process after the increase process (cf. the judgment scores in conditions GC-1 and GC-3 in the left-annotation condition). Nevertheless, the addition of a second annotation resolved the ambiguity in most of the cases (cf. the double-annotation condition).

These findings suggest preliminary guidelines on the use of verbal annotations in line graphs that represent events and processes. The use of a verbal annotation at the left end of the graph line may result in ambiguities in interpretations concerning the relationship between the event represented by the annotation, and the events and processes represented by the graph lines. Therefore, the single use of verbal annotations at this position should be avoided. Alternatively, the pre-event conditions could be made clear to prevent ambiguous interpretations, for example, by extending the graph line to the left of the annotation, if the data are available. Another alternative is to provide the necessary information in the accompanying text, if available.

The findings also showed that an interpretation of delayed cause might occur in certain constellation of graphical entities (cf. GC-2 in the left-annotation condition of group A). This may lead to a misinterpretation of the relation between annotation event and the events and processes represented by the graph lines. Although it seems explicit that events generally cause processes rather than states, it does not always have to be the case. This finding leads to a design guideline which states that: if it is not the intention of the author, an annotation which is placed at the beginning of a graph line which represents a state (i.e., the constancy of the domain value for a period of time) may result in misinterpretations about the relation between the annotation event and the processes that follow the state. Therefore, annotations in such cases should be cautiously used.

The findings exhibited differences in interpretation of the causal relations in the smooth graphs (cf. group B) compared to the ones in the straight graphs (cf. group A). The mean judgment scores in group B were lower than the ones in group A in most of the conditions, with high variability among the participants. This points to differences in interpretation of causal relations between annotation events, and events and processes represented by graph lines. In addition, the post-experiment reviews by the participants showed that some of the participants in group B made different assumptions about the temporal aspect of the processes and events represented by the curved graph lines compared to the participants in group A. In particular, the participants of group B assumed that there was a causal relation between the annotation event and the whole process represented by the graphical entities. In other words, the participants of group B interpreted the graph line representing a single process rather than the composition of two distinct processes.

As presented in the theoretical framework presented in Chapter 4, multimodal comprehension of the information induced by the graphical entities and linguistic entities is achieved by the construction of different types of referential links by the reader. In the verbally annotated graphs that were used in the Study 2, the reader should firstly construct the referential link between the linguistic expression in the annotation (e.g., ‘Barium added to the liquid’) and the domain of the event (i.e., the event of addition of a substance into the liquid). Second, the reader should construct the referential link between the graph lines and the domain information (e.g., the temperature). Third, the referential link between annotation text and a point on the graph line should be constructed via the thin line between the two (namely, annotation icon), which is achieved possibly via Gestalt principles (Pinker, 1990) or a similar underlying visual/perceptual mechanism. These referential links at different levels are employed to construct the inferences concerning the relationship between the annotation events, and the events and processes represented by the graph lines.

Therefore, interpretation of the graph line as a single unit, rather than as the composition of two segments, results in differences in inferences concerning the relationship between the annotation event, and the events and processes represented by the graph lines.

The findings about the interpretation ambiguities in the smooth graphs suggest a guideline for designing verbally annotated graphs: the standard use of annotation icons (i.e., the annotation icon in the form of a straight line, as the ones used in the experiment) results in an ambiguity on one of the ends of the line, which points a specific location on the graph. The ambiguity occurs between the two interpretations. The first interpretation is that the annotation may be attributed to a specific point in time; in other words, the annotation event may be interpreted as a *point-like* event. The second interpretation is that the annotation event may be assumed to occur in a time interval, which corresponds to the whole or part of the events and processes represented by the graph lines; in other words, the event may be interpreted as a *durative* event. On the one hand, the knowledge of the event domain may determine the aspectual characteristics of the event. On the other hand, there may be cases in which the domain knowledge allows the use of both durative and point-like interpretations of the event. In such cases, the ambiguity may not be resolved by graphical means and the information about the aspectual characteristics of the event has to be stated explicitly by linguistic statements, either on the graph or in the accompanying text in a graph-text constellation. This issue was further investigated by a third empirical study, presented in the following part of the chapter.

7.4 STUDY 3: TEMPORAL ASPECT IN VERBALLY ANNOTATED GRAPHS

The Study 3 had two major goals: The first goal was to further investigate the issue of temporal ambiguity in graphs with smooth line segments, which was observed in the Study 2. The second goal was to investigate the role of the presence and absence of an annotation icon on verbally annotated graphs.

7.4.1 Participants, Materials, and Design

Thirty-six participants (mean age 22.8, SD = 2.49) were paid to participate in the experiment. The experimental stimuli consisted of verbally annotated line graphs that represented the change of a domain value in time. The annotations represented information about occurrence of a fictitious event, as exemplified in Figure 7.12.

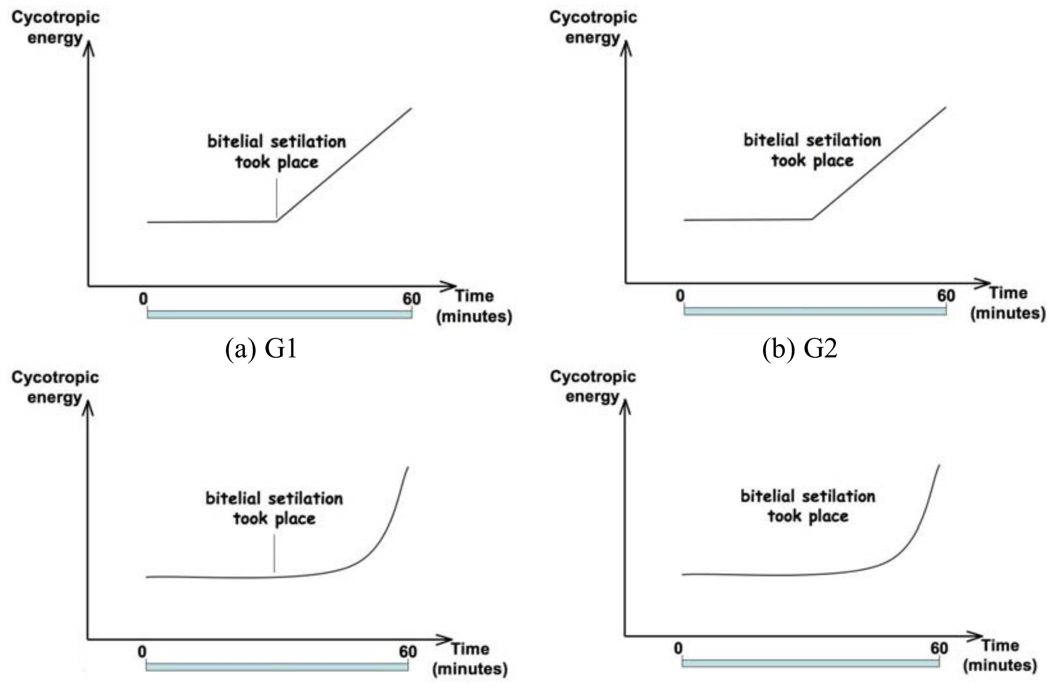
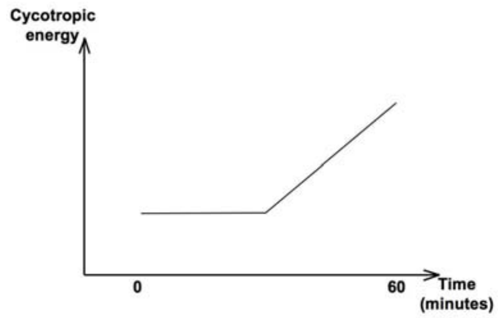


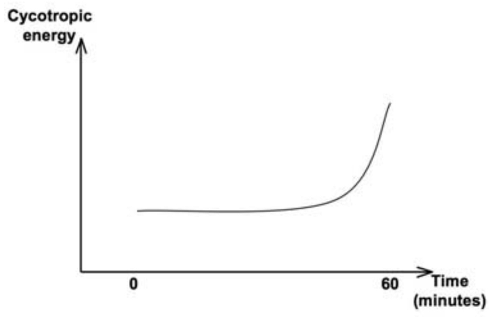
Figure 7.12. Four verbally annotated graphs that show sample material for the four experimental conditions G1 to G4 (translated into English by the author of the dissertation; the font sizes in the figure were enlarged for better visibility).

The design was 2 x 2, with two independent parameters (two within-subject parameters), and one dependent parameter. The first independent parameter was the *graph format*. The graph was either a constellation of two straight line segments, as shown in 7.12a and 7.12b, or a constellation of an almost straight line segment with a smooth line segment, as shown in 7.12c and 7.12d. The second independent parameter was the *annotation icon*. The graph either included an annotation icon, as shown in 7.12a and 7.12c, or it did not include an annotation icon, as shown in 7.12b and 7.12d. As a result, each participant was presented four experimental conditions. The four conditions are named G1 to G4 below. The dependent parameter was participants' reports of interpretations concerning the duration of the annotated event. The participants reported their judgments by clicking on a horizontal visual analog scale (VAS), which was presented below the graphs (see Figure 7.12), as explained in 7.4.2 below. The order of presentation of the material was randomized.

In addition to the four experimental conditions, a non-annotated graph with straight line segments and a non-annotated graph with smooth line segments (i.e., C1 and C2 as shown in Figure 7.13a and 7.13b respectively) were presented to the participants after the presentation of the experimental stimuli, without a specific given task. They were presented for the purpose of comparison with the experimental conditions.



(a) C1



(b) C2

Figure 7.13. The two graphs presented after the presentation of the experimental stimuli in the four experimental conditions.

The graphs were prepared by the experimenter. The domain was not evaluated as an independent parameter in this study. The domain value labels, as well as the annotated events were prepared for four different fictional domains. The participants were informed that the graphs were excerpted from lecture notes in medicine, which was an unfamiliar domain for the participants of the study (i.e., there was no medicine student who participated in the experiment).

7.4.2 Procedure

The participants attended the experiment in single sessions. They were first presented preliminary information about line graphs with samples. After the presentation of the preliminary information, a practice session was presented to explain the task and the use of the visual analog scale (VAS) for the responses. In the experiment session, in each screen, the participants investigated the graphs and reported their interpretation by clicking on the VAS, which was presented below the x-axis in the same screen (see Figure 7.12). The VAS was a horizontal bar extending from the beginning to the end of the horizontal time axis of the graph. If the participant interpreted the event as *point-like*, i.e., the event occurred in a specific point in time, then he/she clicked the occurrence time of the event on the scale (this type of interpretation is called *point interpretation* below). If the participant interpreted the event as *durative*, i.e., the event occurred in a time interval rather than a specific point in time, then he/she clicked the starting time of the event and clicked the ending time of the event on the scale (this type of interpretation is called *interval interpretation* below). The experiment was self-paced, and it took an average of 10 minutes to complete.

7.4.3 Results

The distribution of the clicks on the response scale is shown in Figure 7.14. The figures show histograms for the number of clicks on the VAS.

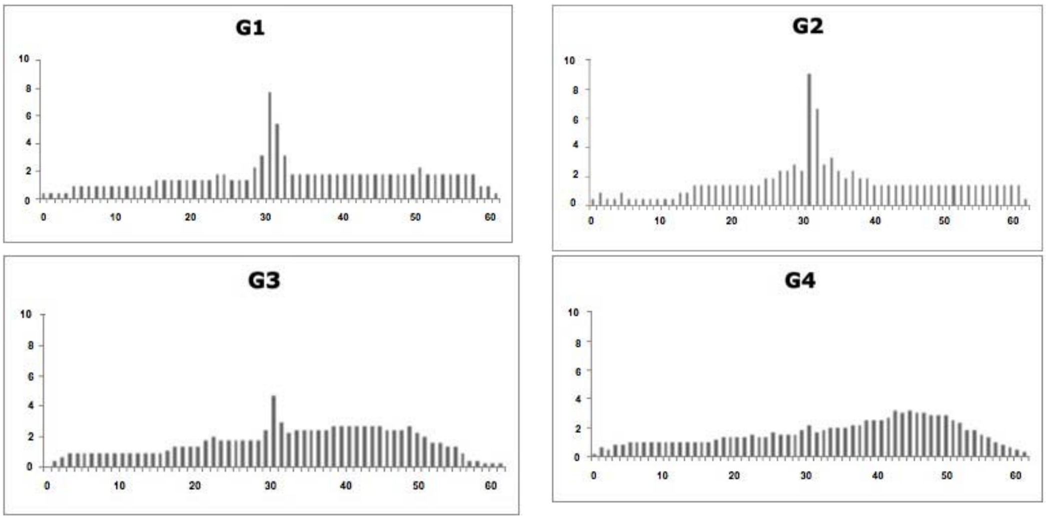


Figure 7.14. The distribution of the number of clicks on the visual analog scale for the four experimental conditions.

The distribution of the number of participants who reported a point interpretation and the number of participants who reported an interval interpretation shows that the main determinant for the interpretations was the graph format. Although the histograms show the influence of annotation icon in interpretations (i.e., G3 vs. G4), the presence or absence of the annotation icon did not exhibit significant difference between the interpretations.

A Cochran test was conducted to evaluate differences between related proportions. The test was significant, $\chi 2$ (3, N = 36) = 20.47, $p < .01$, Kendall coefficient of concordance was .19. Follow-up pairwise comparisons were conducted using a McNemar's test. The results showed that the number of participants who made point interpretation was significantly higher in conditions G1 and G2 than the number of participants who made point interpretation in conditions G3 and G4. Correspondingly, the number of participants who made interval interpretation (almost half of the participants) was significantly higher in conditions G3 and G4 than the number of participants who made interval interpretation in conditions G1 and G2.

Further analysis of the role of the annotation icon, as well as the graph format, was investigated by the analysis of eye movement parameters, namely fixation count, gaze time and fixation duration. For the analysis, the region covered by the graph line was divided into 15 rectangle AOIs (Areas of Interest), namely AOI 1 to AOI 15, as shown in Figure 7.15. The same AOI template was used in the analyses of all four experimental conditions.

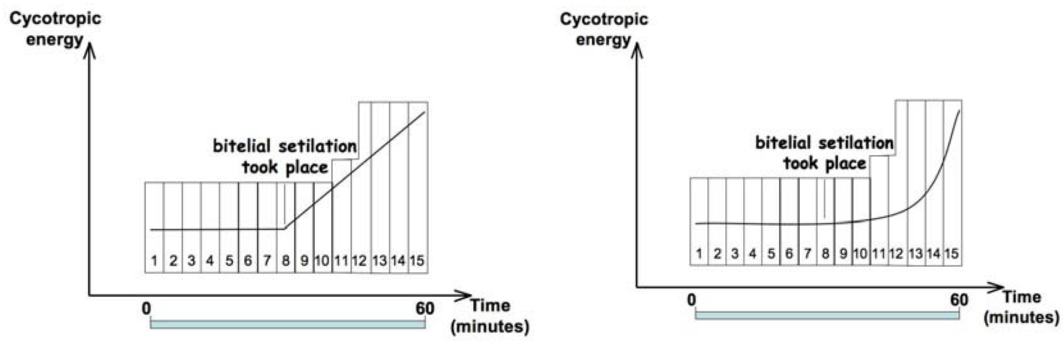


Figure 7.15. The specified AOIs (Area of Interest) for the analysis of eye movement parameters.

Mean fixation counts on the AOIs were calculated, and z-score normalization was used for the analysis. An analysis of variance was conducted with the factors being the conditions and mean fixation counts on the 15 AOIs. The results indicated a significant condition effect, Wilks's $\Lambda = .18$, $F(5, 31) = 28.95$, $p < .01$, a significant AOI effect, $\Lambda = .04$, $F(14, 22) = 35.53$, $p < .01$, and a significant interaction between the conditions and the AOIs. The distribution of the mean fixation counts for C1 (the straight graph without annotations) and C2 (the smooth graph without annotations), revealed information about the salient regions on the graph lines. As shown in Figure 7.16, in the C1 graph the salient region was AOI 8, whereas in the C2 graph the salient region was AOI 12.

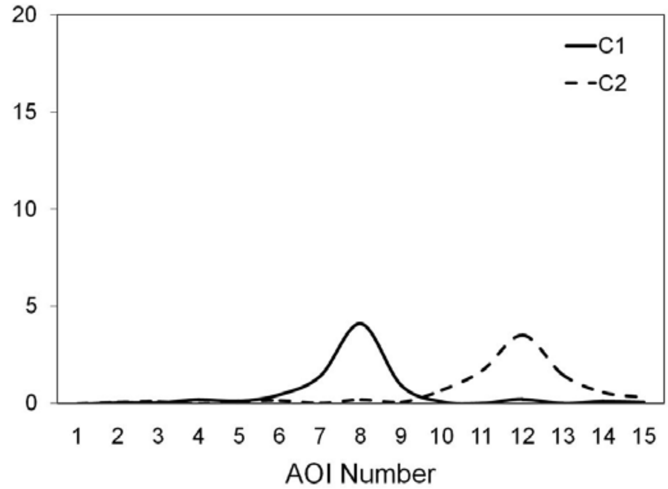


Figure 7.16. The distribution of mean number of fixations (i.e. fixation counts) on the AOI 1 to AOI 15 in C1 and C2 graphs.

The distributions of the mean fixations on the AOIs in the experimental conditions are presented in Figure 7.17.

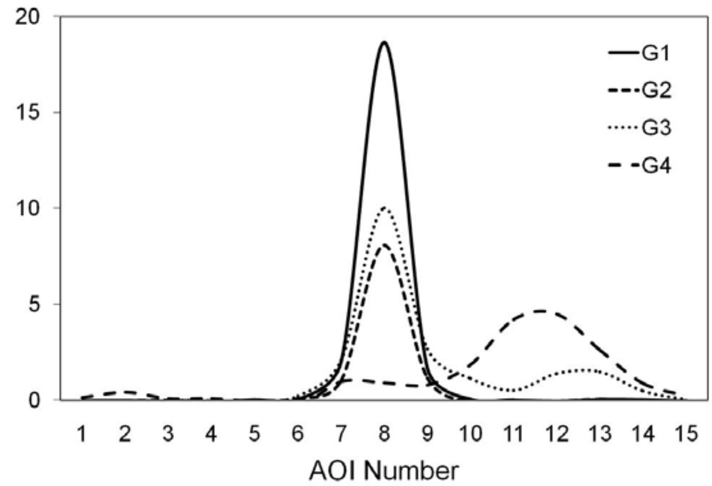


Figure 7.17. The distribution of mean number of fixations (i.e., fixation counts) on the AOI 1 to AOI 15 in G1, G2, G3, and G4.

The graphs used in G1 were straight-line graphs with an annotation icon. These graphs received highest number of mean fixations. The salient AOI was AOI 8 in G1, the same as the salient AOI in C1. The absence of annotation icon in G2 resulted in reduced mean number of fixations, but the distribution was similar to the one obtained in G1. In G3, where the graph had a smooth shape and an annotation icon was present, the participants inspected the AOI 12 and AOI 13, and AOI 8 with fewer fixations compared to the ones on AOI 12 and AOI 13. The absence of an annotation icon resulted in the distribution of fixations in G4, which was similar to the distribution in C2. In other words, the salient region was shifted from AOI 11 and AOI 12 (in G4) to AOI 8 (in G3) with the addition of the annotation icon. The comparison of mean fixation counts for the C1-C2 graphs and the G2-G4 graphs shows that in G2 and G4, the addition of the annotation text resulted in an overall increase in the mean number of fixations on the AOI 8, which was the region below the annotation text.

The analysis of gaze times on the specified AOIs revealed similar distributions to the ones for mean fixation counts. The results of mean fixation durations were calculated for the AOIs that had an average number of fixations of one or greater than one. Accordingly, mean fixation durations were calculated for AOI 7 to AOI 9 in G1 and G2 conditions and for AOI 7 to AOI 13 in G3 and G4. An analysis of variance was conducted for the differences between condition G3 and G4. The results of the ANOVA indicated a significant condition effect, $F(1, 296) = 6.09$, $p < .05$ and a significant AOI effect, $F(6, 296) = 3.15$, $p < .01$. The most important finding from the perspective of this study is that the mean fixations in the G4 condition were longer than the ones in the G3 condition, which indicate processing difficulties in G4 compared to G3.

7.4.4 Summary of the Findings in Study 3

The participants' reports of interpretations concerning the temporal properties of the annotation events show that the graph format, rather than the presence or absence of an annotation icon, is the main determinant for the interpretation of temporal aspect of the annotated events (cf. the situation aspect, i.e. Aktionsarten, see Pustejovsky, 1995; Vendler, 1957; Verkuyl, 1972, 1993; Jackendoff, 2002). Most of the participants reported that the annotation event occurred at a specific point in time in the conditions where the graph lines were composed of straight line segments. On the other hand, almost half of the participants reported that the annotation event occurred in a time interval in the smooth-graph conditions. Although a slight difference was observed in the distribution of mouse clicks on the response scale between the two annotation conditions of the smooth graphs, the role of the annotation icon on participants' subjective judgment reports was not significantly different between the conditions.

Further analyses of eye movement parameters revealed more detailed information about the differences between the experimental conditions and the role of the presence or absence of an annotation icon. The results showed that on the one hand, the addition of a verbal annotation to the graphs (without an annotation icon) did not reveal major changes in the distribution of the mean number of fixations and gaze times on the AOIs, compared to the distributions on the non-annotated graphs (C1-C2 vs. G2-G4). This implies that the annotation text without an annotation icon did not strongly influence the salient regions on the graphs. On the other hand, mean fixation counts and mean gaze times increased with the addition of the verbal annotation (G2-G4 vs. G1-G3). The increase in the mean fixation counts and the mean gaze times on the graph lines indicates participants' effort to integrate the information induced by the verbal annotations and graphical entities.

Moreover, the addition of an annotation icon (together with the addition of the annotation text) resulted in major changes in the distribution of mean fixation counts and in the distribution of mean gaze time values in the smooth graphs. In other words, the presence of an annotation icon shifted the visually/informationally salient regions in the smooth graphs. This is expected, since the addition of a new graphical element, i.e. the annotation icon attracted participants' attention to this region. In addition, the mean number of fixations and gaze times further increased with the addition of the annotation icon. This increase indicates participants' further effort to integrate the information induced by the annotation text and the annotation icon, as well as the annotation icon and the relevant part of the graph.

The participants experienced difficulties in determining the relevant part of the graph in the absence of an annotation icon in smooth graphs. The results of the analysis for mean fixation durations showed that the absence of an annotation icon in smooth graphs resulted in longer fixations on the salient regions of the graph lines. This finding indicates that the lack of an annotation icon may result in ambiguities in constructing the coherence relations between annotation events and events and processes represented by graph lines, particularly in graphs with smooth line segments.[8] The findings also indicate that annotations icons belong to those top-down factors which interact with geometric factors in visual segmentation of curves (see Chapter 2).

[8] This is kindred to the *attachment ambiguity* in Futrelle's (1999) classification of diagram parsing ambiguities.

7.5 SUMMARY

Multimodal comprehension of graph-text constellations is based on the construction of coreference relations between the sublayer of graphical entities (i.e., graph (proper)) and sublayer of textual entities (e.g., verbal annotations) in the graph layer of a graph-text constellation, as well as the construction of coreference relations between the graph layer and the text layer (e.g., paragraphs in the accompanying text). Moreover, beyond the construction of the inter-modal coreference links, readers construct coherence links between the events and processes represented by verbal entities and events and processes represented by graphical entities. Comprehension of verbally annotated graphs presents a specific subcase of those constructions during the course of multimodal comprehension of graph-text constellations.

The three empirical studies that are reported in this chapter investigated the interaction between the information induced by graphical entities and verbal annotations in various dimensions. The findings of Study 1 showed that verbal annotations might serve the bridging role between the two representational modalities in graph-text constellations not only with respect to the referential link constructions between graphical entities and accompanying text but also with respect to the referential link constructions between graphical and linguistic entities within the graph layer, viz. graph-internal multimodality. The Study 2 and the Study 3 investigated the construction of coherence relations in verbally annotated graphs. In particular, the Study 2 investigated graph-internal multimodality with respect to the attribution of causal relations between annotation events, and events and processes represented by graphical entities in verbally annotated graphs. The Study 3 investigated the interaction between the interpretations concerning the temporal aspect of annotation events and the processes and events represented by graphical entities.

In summary, the findings of the experimental investigations show that changes in one modality influence comprehension of the other modality. From the perspective of HCI, poorly designed combinations of verbal annotations and graphical entities may create misconceptions, deceive or confuse an issue; they may complicate the reader's task and hinder comprehension and learning. Therefore, for an appropriate design of verbally annotated graphs, the interaction between verbal annotations and graphical entities should be taken into account rather than verbal and graphical entities in isolation. In the current state of the art, the design and production of verbal annotations is based on the experience and practice of the designers, rather than theory, guidelines or systematic empirical research. On the application side, recent data visualization components of statistical and mathematical software tools offer limited capacity for annotation design. Preparing effective diagrams requires both practice and also evidence from empirical research studies. The implications of the findings of the empirical studies that are reported in this chapter are discussed in

Chapter 10, where a set of principles and guidelines are presented for the design of graph-text constellations and for the design of artificial systems that have the capability to support end-users during their course of designing and producing graph-text constellations and verbally annotated graphs.

8 Coherence Relations: Scalar Concepts in Language and Graphs

This chapter presents two empirical studies that investigated the interaction between linguistic expressions and graphical representation of scalar concepts in graph-text constellations. As stated in Chapter 2, the term *scale* is used for abstract representations of measurement that map objects to degrees (Kennedy, 1999; Cresswell, 1977; von Stechow, 1984; Stevens, 1975/1986; Kosslyn, 1989). The concept of scale can be regarded as a conceptual entity by which coherence relations between textual and graphical entities are constructed during the course of multimodal comprehension of graph-text constellations. Within the theoretical framework of common conceptual basis for multimodal representations (see Chapter 4) the concept of scale is accessible by representations in different modalities. In particular, scalar expressions in language, such as *high*, *tall*, *low*, and so on, access concepts that are also accessible by spatial configurations of graphical entities in the Cartesian space. From the perspective of HCI, this aspect is important for the design of graph-text constellations because the vagueness of language gives the opportunity to designer to select appropriate linguistic expressions that accompany the graph and to adjust the ranges of the axes so that a coherent interpretation of linguistic expressions and graphical entities is reached.

On the other hand, poorly designed graphs may lead to conceptualizations that are not compatible with the linguistic entities in graph-text constellations. The empirical studies reported in this chapter investigate how different design of graphical entities influences comprehension of linguistic entities in graph-text constellations. In particular, Study 1, reported in 8.1, investigates the interaction between scalar adjectives (e.g., *rainy*, *windy*) as linguistic expressions of scale and the positions of graphical entities in the graph scale as graphical representations of scale. Study 2, reported in 8.2, investigates the interaction between scalar adverbs (e.g., *fast*, *slowly*) and the steepness of graph lines in the graph space.

Experimentally, the literature on eye movement behavior in scene viewing states that incongruous (i.e., inconsistent) objects in scenes are fixated longer than congruous objects (see Rayner, 1998, for a review). This implies that the interaction between linguistic expressions of scale and graphical representations of scale can be investigated by measuring the eye movement behavior.[1] In particular, in the two empirical investigations presented in this chapter, the inconsistency between the linguistic expressions and graphical representations of scale is measured by the analysis of gaze times on linguistic expressions.

8.1 STUDY 1: SCALAR CONGRUENCY IN GRADABLE ADJECTIVES

The purpose of the Study 1 was to investigate the interaction between gradable adjectives and the position of graphical entities in graph space in graph-text constellations with a separate graph-layer–text-layer layout.

8.1.1 Participants, Materials, and Design

Fifty-two university students or graduates from various academic departments were paid to participate in the experiment. The mean age of the participants was 21.0 (SD = 2.00). The graph-text constellations used in the experiment are exemplified in Figure 8.1.[2] Each graph-text constellation provided information about a city and its characteristics such as its geographic location, its climate condition, and its transportation and accommodation facilities. The graphs presented information about one of the four climate parameters: the amount of rainfall, the wind speed, the number of sunny days, and the number of foggy days. Each participant was presented four graph-text constellations.

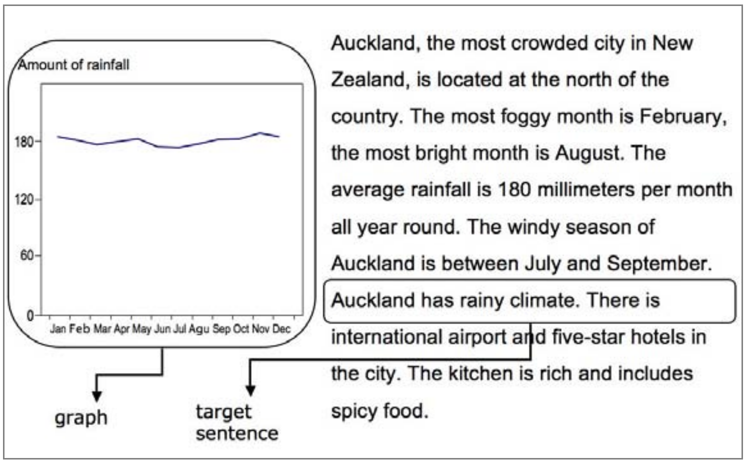


Figure 8.1. Sample stimuli from Study 1. The rectangular regions show the specification of the two Areas of Interest (AOIs). They were not shown to the participants.

[1] The terms *congruency* and *consistency* are used interchangeably in the dissertation.

[2] The text content of the graph-text constellations was prepared by the experimenter, based on the city travel guides presented in the Lonely Planet web portal: http://www.lonelyplanet.com (retrieved on October, 20, 2008). The graphs were redrawn and redesigned, and the text material was summarized and edited for the experimental conditions when necessary. The material was translated into Turkish, the language of the experiment, by the experimenter.

There were two independent variables in the experiment (one within-subject factor and one between-subject factor). The within-subject factor was the *graph type* with four conditions (shortly, the *graph condition*). The four conditions are presented in below:

Condition 1 (High-Graph Condition): The graphs in this condition involved a graph line, which was located at the upper side of the graph frame, as exemplified in Figure 8.1.
Condition 2 (Low-Graph Condition): The graphs in this condition involved a graph line, which was located at the lower side of the graph frame.
Condition 3 (Middle-Graph Condition): The graphs in this condition involved a graph line, which was located at the middle of the graph frame.
Condition 4 (Text-Only Condition): This condition did not involve a graph. Only the paragraphs were presented to the participants.

Figure 8.2 shows sample stimuli for the four graph conditions. The numerical value of the presented climate parameter (e.g., the amount of rainfall in the exemplified stimuli) was the same in the four conditions. The range of the y-axis values was arranged accordingly.

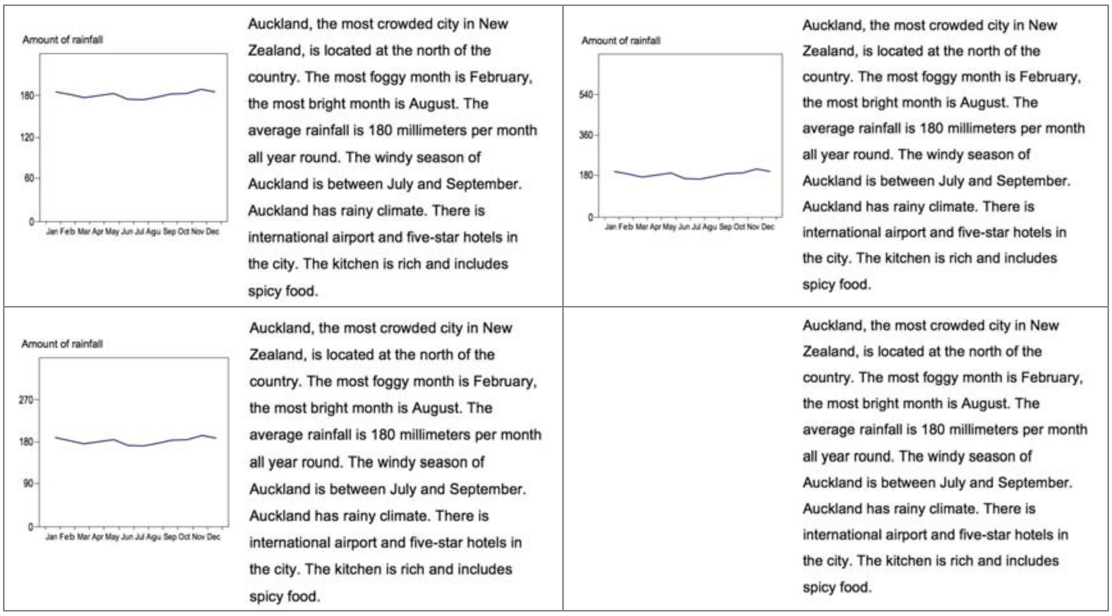


Figure 8.2. Sample stimuli for the four graph conditions (high-graph, low-graph, middle-graph, and text-only conditions).

In each graph-text constellation, there was a sentence about the climate parameter that was also presented by the graph. For instance, in the sample stimuli in Figure 8.2, this sentence was “The average rainfall is 180 millimeters per month all year round”. The purpose of this sentence was to attract the participants' attention to the climate parameter that was presented by the graph. Following this sentence, a filler sentence about another climate parameter was used, such as "The windy season of Auckland is between July and September". After this sentence, another sentence about the climate parameter was used, the content of which was also referred to by the graph. For instance, in the exemplified stimuli this sentence was "Auckland has rainy climate". This sentence is called the *target sentence* below. The target sentence involved a gradable adjective, which was in the focus of the present experiment. More information about target sentences is presented below.

The between-subject parameter was the *adjective type* with two conditions. The participants were divided into two groups and each group consisted of 26 participants. The first group (group A, or the positive-adjective group) was presented a positive adjective in the target sentence. The second group (group B, or the negative-adjective group) was presented a negative adjective in the target sentence. The four gradable adjectives used in the experiment were *rainy*, *windy*, *sunny*, and *foggy*. Sample material for the two groups is presented in Table 8.1.[3]

Table 8.1. Sample target sentences for group A and group B.

Group A (the positive adjective group)	*Group B (the negative-adjective group)*
Auckland has rainy climate.	Auckland has non-rainy climate.
Manila has windy climate.	Manila has non-windy climate.
San Juan has sunny climate.	San Juan has non-sunny climate.
La Ceiba has foggy climate.	La Ceiba has non-foggy climate.

The material consisted of three posttest question sets. Posttest questions are discussed in 8.1.3 below.

8.1.2 Procedure

The experiment was conducted in single sessions. The participants were informed that they were expected to inspect the presented city guides before a business trip. In the business trip, they would visit four cities in different locations, having different climate conditions. The participants were also provided with line graphs, as well as sample graphs that presented climate parameters such as temperature and the number

[3] The adjective pairs used in the experiment were noun modifiers. These adjectives were derived from nouns by means of two suffixes, namely the *–lI* suffix and the *–sIz* suffix. The adjectives derived by means of the former suffix mean possessing the quality expressed by the basic morpheme (Kornfilt, 1997, p. 457). The latter suffix, on the other hand, means 'without'. The adjective pairs derived by means of the two suffixes do not necessarily have to be antonyms on the same scale. They rather have overlapping multiple scales. The use of multiple scales by adjectives is a phenomenon observed in many languages (Cruse, 1986). Accordingly, those adjectives are used in the stimuli: *güneşli* 'sunny', *güneşsiz* 'non-sunny', *yağışlı* 'rainy', *yağışsız* 'non-rainy', *sisli* 'foggy', *sissiz* 'non-foggy', *rüzgarlı* 'windy', and *rüzgarsız* 'non-windy'.

of dry days per month. The sample graphs were similar to the graphs used in the middle-graph condition of the experiment. The participants were asked to interpret the graphs given in the introduction session, and asked for their consent for not having difficulties in interpretation of the presented material, before the experiment session. The presentation of the material was randomized. There was no time limitation in the experiment. The entire session took approximately 15-20 minutes. A 50 Hz. non-intrusive eye tracker recorded eye-tracking data (see 5.2 for the details of the eye tracker and the general description of the methodology).

8.1.3 Results

The results are presented in two parts below. The first part presents the results of gaze times; the second part presents the results of answers to the posttest questions.

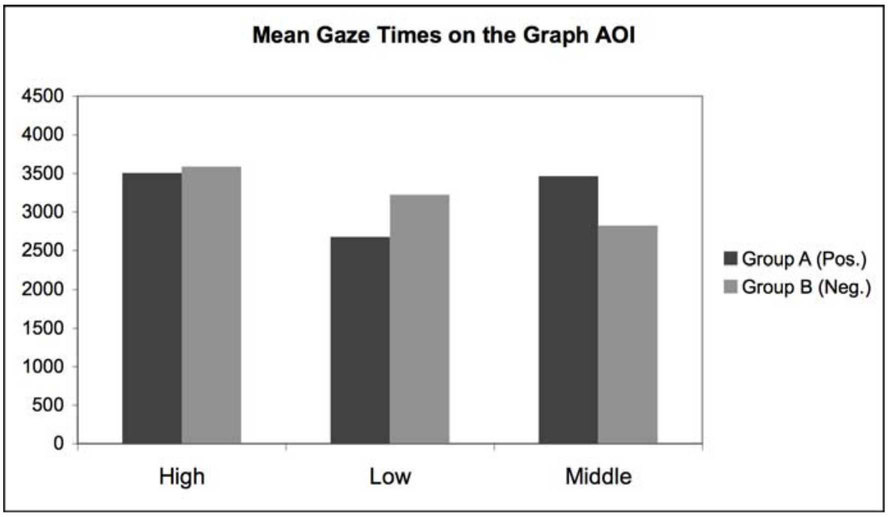
8.1.3.1 The Analysis of Gaze Times

The calibration of the eye tracker failed for four participants. Data from the remaining 48 participants were included to the analysis. Fixations with duration of less than 60 ms were not included to the analysis.

Two AOIs (Areas of Interest) were specified for the analysis. The first AOI covered the target sentence and the word after the target sentence.[4] The second AOI covered the graph region, including the graph label and the axis labels (i.e., numerical values, see Figure 8.1 above). The mean gaze times on the graph AOI and the target sentence AOI are shown in Figure 8.3.

[4] The literature on eye movement characteristics in reading shows that when a word is fixated for a long time, the word that follows the fixated word is also fixated long. This is called the *spillover effect* (Rayner, 1998). The purpose of including the word after the target sentence to the AOI was due to the potential spillover effect.

(a)



(b)

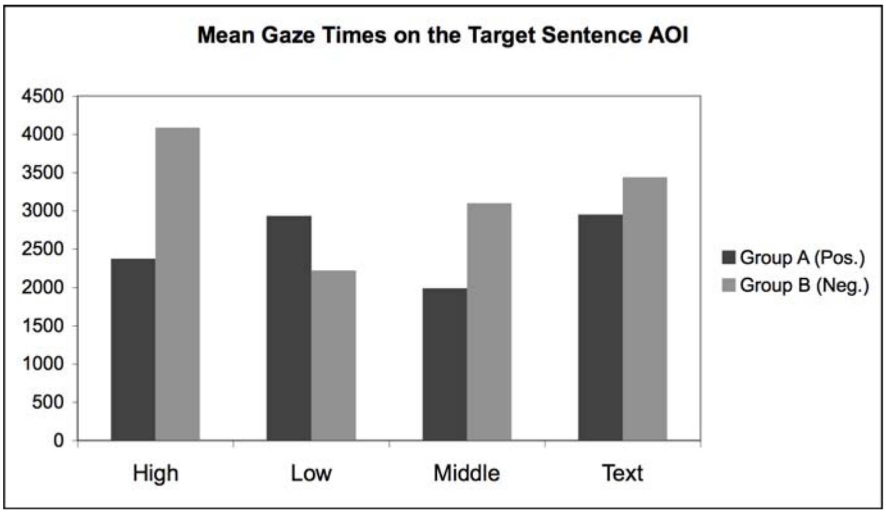


Figure 8.3. Mean gaze times on (a) the graph AOI and, (b) the target sentence AOI. The numbers are in ms.

For the analysis of the mean gaze times on the graph AOI an analysis of variance was conducted with one within-subject factor (graph type with three conditions, namely high-graph, low-graph, middle-graph) and one between-subject factor (group with two conditions, namely group A (the positive adjective group) and group B (the negative adjective group). The results indicated no significant difference between the two groups, Wilks's $\Lambda = .98$, $F(2, 45) = .46$, $p = .64$, multivariate $\eta^2 = .02$.

For the analysis of the mean gaze times on the target sentence AOI an analysis of variance was conducted with one within-subject factor (graph type with four conditions, namely high-graph, low-graph, middle-graph, and text-only) and one between-subject factor (group with two conditions as described above). The results indicated a significant difference between the groups, Wilks's $\Lambda = .81$, $F(3, 44) = 3.52$, $p < .05$, multivariate $\eta^2 = .19$. Further pairwise comparisons showed that the mean gaze times on the target sentence AOI were significantly different between the groups in the high-graph condition, and in the low-graph condition. However, the mean gaze times were not significantly different between the two groups in the middle-graph condition, and in the text-only condition.[5]

8.1.3.2 The Analysis of Answers to the Posttest Questions

After the presentation of material, the participants were presented three posttest question sets.

The first set included open questions about participants' numerical predictions of climate parameters, and it will not be analyzed here.[6] In the second question set, the questions were about the level of the climate parameter, such as, "the wind speed in Manila was high / middle level / low." The purpose of the second posttest question set was to measure the recall of the participants by using the terms *high*, *middle level*, and *low* in the question phrase. These terms are verbal descriptions of the position of the graph line in the graph space. The maximum number of "high" answers is expected in the high-graph condition, and the maximum number of "low" answers is expected in the low-graph condition. Moreover, a difference between group A and group B in total scores would reveal the influence of the verbal entities (i.e., positive and negative adjectives) in the text-layer of the graph-text constellations. In the text-only condition, a higher number of "high" answers is expected in group A, whereas a higher number of "low" answers is expected in group B. One answer from one of the participants was not recorded due to a technical problem; the remaining 207 answers from 2 participants are presented in Table 8.2.

[5] The frequency of use of positive adjectives is often higher than the frequency of use of negative adjectives in daily language. It has been also known that during reading fixation durations are longer on the words with lower frequency of use compared to the words with higher frequency of use. These findings suggest that readers made longer fixations on the negative adjectives that were used in the experiment. This bias can be observed in the difference between slightly longer gaze time on the sentences that involved negative adjectives compared to the gaze time on the sentences that involved positive adjectives. However, the difference was not statistically significant. Therefore, word frequency was not taken into account as an independent parameter in the studies reported in this chapter.

[6] Participants' world knowledge about numerical values of climate parameters, such as the amount of rainfall per month, is a factor that influences comprehension of graphical and linguistic entities. However, the purpose of the studies reported in this chapter is to analyze whether incongruency between the representations of the scale concept by textual and graphical entities lead to processing difficulties during the course of multimodal comprehension. The investigation of the influence of world knowledge is left for future research in this study.

Table 8.2. Answers to the second posttest question set. The numbers show the number of answers.

Group A	High	Low	Middle	No Answer
High-Graph Condition	12	3	10	1
Middle-Graph Condition	9	7	7	3
Low-Graph Condition	11	4	7	4
Text-Only Condition	12	5	6	3
TOTAL	*44*	*19*	*30*	*11*

Group B	High	Low	Middle	No Answer
High-Graph Condition	7	8	7	3
Middle-Graph Condition	10	6	6	4
Low-Graph Condition	8	9	7	2
Text-Only Condition	9	10	3	4
TOTAL	*34*	*33*	*23*	*13*

The results show that in group A, the number of "high" answers was higher than the number of "low" answers in all four conditions, whereas in group B the number of "high" answers and the number of "low" answers were almost equal. In other words, the group A participants showed a tendency to give "high" answers in all four conditions, whereas the tendency to give "high" answer in group B participants was not as strong as the tendency in group A participants. In group A, the maximum number of "high" answers was obtained for the high-graph condition. In contrast to the prediction, the maximum number of "high" answers was obtained for the low-graph condition. On the other hand, almost equal numbers of "low" and "high" answers were obtained in high-graph condition and in low-graph condition in Group B. In summary, the results of the experiment show that the gradable adjectives in the text-layer of graph-text constellations influence the recall of the participants.

In the third posttest question set, participants' recall for the linguistic phrases that were used in the text-layer of the graph-text constellations (i.e., the target sentences) was measured. The participants answered questions such as "Manila had windy / non-windy / calm / stormy climate." The results are presented in Table 8.3.

Table 8.3. Answers to the third posttest question set. The numbers show the number of answers. The term *other answers* is used for the answers that were not either the positive or the negative adjective but a third one that was presented in the multiple-choice test.

Group A	Positive Adjective	Negative Adjective	Other answers	No Answer
High-Graph Condition	22	1	2	1
Middle-Graph Condition	9	4	6	7
Low-Graph Condition	20	2	0	4
Text-Only Condition	13	6	3	4
TOTAL	*64*	*13*	*11*	*16*

Group B	Positive Adjective	Negative Adjective	Other answers	No Answer
High-Graph Condition	12	9	2	3
Middle-Graph Condition	9	7	4	6
Low-Graph Condition	15	6	2	3
Text-Only Condition	12	8	1	5
TOTAL	*48*	*40*	*9*	*17*

The results show that both the participants of group A and the participants of group B gave more positive-adjective answers compared to the negative-adjective answers. However, the tendency for giving positive-adjective answers in group B was not as strong as the tendency in group A. However, the influence of the graph conditions on participants' recall is not clear at this stage. The graph, whether it was congruent or incongruent with the adjective in the target sentence, resulted in increase in recall in group A; compared to the text-only condition and the middle-graph condition, the number of positive adjective answers was strikingly higher than the number of negative adjective answers in group A. However, this finding was not pertinent to the results in Group B. These findings indicate that further research is needed to have a better understanding of negative adjectives in graph-text constellations.

8.1.4 Summary of the Findings in Study 1

The results of the Study 1 presented partial experimental evidence for the interaction between the representations of scalar concepts by graphs; in particular, the position of the graph line in the graph frame, and the linguistic expressions of scalar concepts by gradable adjectives. The longer gaze times on the target sentences in the case of incongruency between textual and graphical entities show that the participants experienced difficulties during the course of multimodal comprehension. On the other hand, the analysis of posttest questions presented partial support for the interaction between graphical and verbal entities since the participants showed tendency to give answers in favor of positive gradable adjectives. In 8.2, two experimental investigations on gradable adverbs are presented.

8.2 STUDY 2: SCALAR CONGRUENCY IN GRADABLE ADVERBS

The purpose of the Study 2 was to investigate the interaction between gradable adverbs as linguistic expressions of scale and spatial configurations of graphical entities—in particular, the steepness of the graph lines—in graph-text constellations. Two experimental investigations that investigated the interaction are presented below.

8.2.1 Experiment 1

8.2.1.1 Participants, Materials, and Design

Thirty-three university students or graduates from various departments were paid to participate in the experiment. The mean age of the participants was 23.6 (SD = 4.58). Each participant was presented four graph-text constellations. The material presented information about different domains such as job security, natural resources, and research and development expenses. There was one between-subject factor (namely, graph type) with two conditions. The participants were divided into two groups. The first group (i.e., group A with 17 participants) received the stimuli which included a line graph with a graph line that occupied the graph space such that the upper and lower limits of the graph line covered approximately the whole range of the y-axis. The second group (i.e., group B with 16 participants) received the stimuli which included a line graph with a graph line that occupied less graph space compared to the stimuli of group A. The stimuli of group A and group B are exemplified in Figure 8.4.[7]

[7] The text content of the graph-text constellations was prepared by the experimenter, based on the information provided in the following articles: Job insecurity still high in Spain by Eurofound, the European Foundation for the Improvement of Living and Working Conditions, European Working Conditions Observatory (EWCO),
http://www.eurofound.europa.eu/ewco/2004/11/ES0411NU01.htm (retrieved on December 5, 2008);
El Nino Southern Oscillation Educational Material by National Oceanic and Atmospheric Administration of the U.S. Department of Commerce (NOAA),
http://www.srh.noaa.gov/mlb/enso/Enso-season-forecast.html (retrieved on December 5, 2008); The Portneuf River by the Portneuf River Ecosystem Project,
http://www.isu.edu/bios/prep/groundwater.htm, (retrieved on December 5, 2008); and Research as a Vital Foundation for Society by Erich Bloch, Council on Competitiveness, National Science Foundation (NSF), http://www.nsf.gov/pubs/1998/nsb97150/bloch.htm (retrieved on December 5, 2008). The graphs were redrawn and redesigned, and the text material was summarized and edited for the experimental conditions where necessary. The material was translated into Turkish, the language of the experiment, by the experimenter.

Figure 8.4. Sample stimuli from the Experiment 1. The graph above exemplifies the stimuli presented to the participants of group A, the graph below exemplifies the stimuli presented to the participants of group B.

Each graph-text constellation involved a *target sentence* that conveyed the information represented by the graph. For instance, in the sample stimuli in Figure 8.4, this sentence was "According to the report, the job security increased fast between years 2002-2003". The target sentence involved a gradable adverb, which was in the focus of the present experiment.[8]

8.2.1.2 Procedure

The experiment was conducted in single sessions. The participants were informed that they were expected to investigate four newspaper articles, and at the end, they were expected to answer a question about the content. The purpose of the question was to keep the focus of the participants on the material and it will not be analyzed

[8] In Experiment 1, for all the four graph-text constellations that were presented to the participants, the gradable adverb that was used in the target sentence was *hızla* 'fast, rapidly'.

below. The participants were also asked to give verbal consent that they were able to interpret exemplified line graphs in the introductory screens. The presentation of the material was randomized. There were no time limitations in the experiment. The entire session took approximately 10 minutes.

8.2.1.3 Results

The calibration of the eye tracker failed for one participant in Group B. Data from the remaining 32 participants were included to the analysis. Fixations with a duration of less than 60 ms were not included to the analysis. Two AOIs (Areas of Interest) were specified for the analysis. The first AOI covered the target sentence and the word after the target sentence.[9] The second AOI covered the graph region, including the graph label and the axis labels (i.e., numerical values). The mean gaze times on the graph AOI and the target sentence AOI are shown in Figure 8.5.

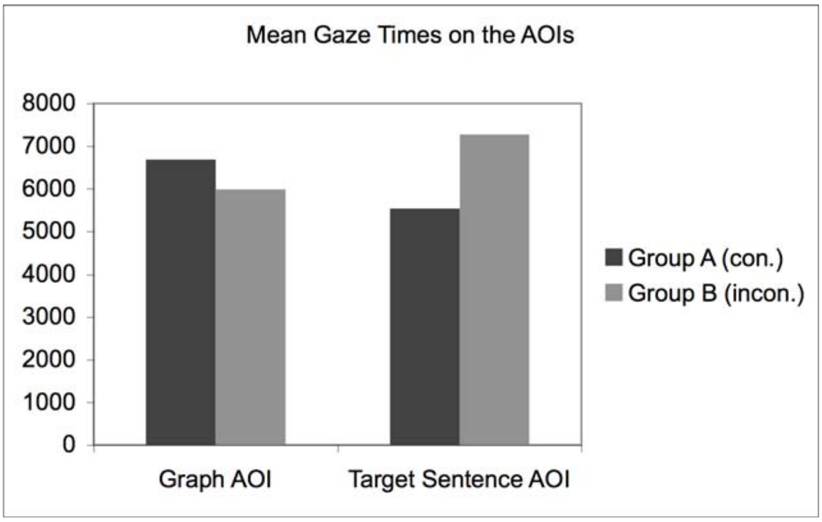


Figure 8.5. Mean gaze times on the graph AOI and the target sentence AOI in Experiment 1. The numbers are in ms. The term *con* is used as an abbreviation for *congruent*, and the term *incon* is used as an abbreviation for *incongruent*.

Pairwise comparisons revealed no significant difference between the two groups in the mean gaze times on the target sentence AOI, $t(30) = 1.87$, $p = .70$, and on the graph AOI, $t(30) = .59$, $p = .56$. Since the results did not reveal a significant difference, a second experiment was conducted with another gradable adverb, *yavaşça* 'slowly', which is the opposite of the gradable adverb used in Experiment 1, i.e. *hızla* 'fast, rapidly'.

---

[9] As stated above, the purpose of including the word after the target sentence to the AOI was due to the potential spillover effect (Rayner, 1998).

8.2.2 Experiment 2

8.2.2.1 Participants, Materials, and Design

Forty-eight university students or graduates from various academic departments were paid to participate in the experiment. The mean age of the participants was 22.0 (SD = 1.43). The graph-text constellations that were used in Experiment 1 were also used in Experiment 2, the only difference being that in the target sentences, the adverb *yavaşça* 'slowly' is used instead of *hızla* 'rapidly, fast'. There were 25 participants in group A, and 23 participants in group B.

8.2.2.2 Procedure

The procedure in Experiment 2 was the same as the procedure in Experiment 1.

8.2.2.3 Results

The calibration of the eye tracker failed for three participants in group A, and one participant in group B. Data from the remaining 44 participants were included to the analysis. Fixations with a duration of less than 60 ms were not included to the analysis. The same AOI specifications of Experiment 1 were used in Experiment 2. The mean gaze times on the graph AOI and the target sentence AOI are shown in Figure 8.6.

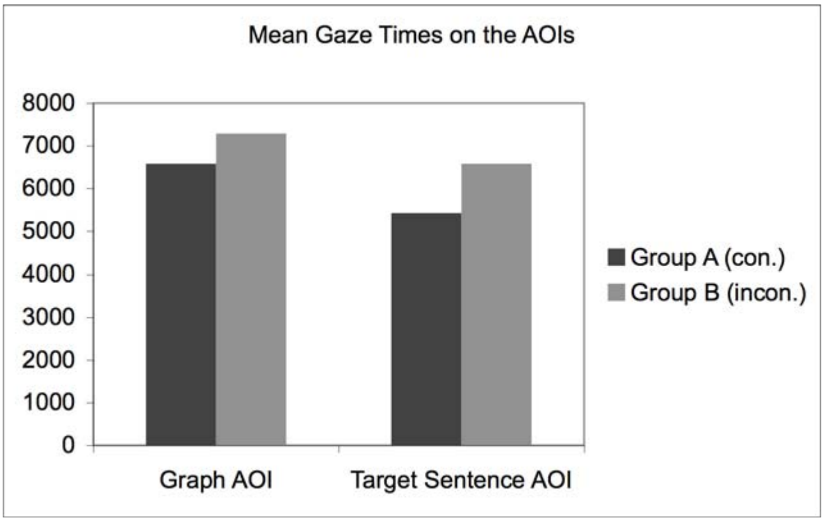


Figure 8.6. Mean gaze times on the graph AOI and the target sentence AOI in Experiment 2. The numbers are in ms.

Pairwise comparisons revealed no significant difference between the two groups in the mean gaze times on the graph, target sentence AOI, $t(42) = .64$, $p = .53$. However, the mean gaze time on the target sentence was significantly different between the two groups, $t(42) = 2.01$, $p < .05$. The mean gaze time was higher in the group B, where the participants received incongruent graph-text constellations, compared to the mean gaze time in group A, where the participants received congruent graph-text constellations.

8.2.3 Summary of the Findings in Study 2

The results of the two experimental studies in the Study 2 presented partial empirical evidence for the interaction between the representations of scalar concepts by graphs—in particular, the steepness of the graph line—and the linguistic expressions of scalar concepts by gradable adverbs, in the case of the gradable adverb *yavaşça* 'slowly'. The longer gaze times on the target sentences in the case of incongruency between graphical and textual entities show that the participants experienced difficulties during the course of multimodal comprehension.

8.3 SUMMARY

This chapter presented the investigation of a specific case of the interaction between graphical entities and linguistic entities in graph-text constellations: the interaction between the representation of scalar concepts by graphical entities and the representation of scalar concepts by linguistic expressions. In the theoretical framework presented in Chapter 4, it was stated that the information induced by graphical entities and linguistic entities is integrated at a conceptual level, based on the common conceptual basis for representations in different modalities; in other words, the integrated representations are accessible by different modalities. The findings of the two studies in this chapter present experimental evidence for the theoretical framework.

From the perspective of HCI, the findings suggest that congruency-checkers, like the spell-checkers in word processing software could be developed for the design of graph-text constellations by non-experienced users. In particular, the congruency between gradable adjectives in the text-layer of a graph-text constellation and the position of the graph line in the graph frame can be analyzed by the system. In a similar way, the congruency between gradable adverbs in the text-layer of a graph-text constellation and the steepness of the graph line can be analyzed by the system. Implications of the empirical findings for the development of design guidelines and principles are discussed in Chapter 10.

Section 3

HCI Aspects of the Study of Multimodal Graph-Text Comprehension

9 A GOMS-Style Cognitive Model for Multimodal Comprehension of Spatial Prepositional Phrases in Graph-Sentence Verification Tasks

This chapter presents a cognitive model for multimodal comprehension of graph-text constellations, based on the GOMS analysis technique. The processing model introduced in this chapter presents a complementary analysis of the cognitive architecture for multimodal comprehension of graph-text constellations presented in Chapter 4. The cognitive architecture highlighted the major components of the underlying infrastructure of multimodal comprehension of graph-text constellations. The model presented in this chapter analyzes the information processing tasks of the readers during their course of integration of the information induced by linguistic and graphical entities. In particular, the task analysis of the GOMS modeling approach is employed for making predictions about gaze patterns. The sentences were decomposed into tasks and subtasks, which are used for predicting readers' inspection of graphs after their reading of the sentences. An empirical investigation was conducted to test model predictions.

In 9.1, the origins of cognitive modeling in HCI and cognitive science, and the GOMS task analysis are introduced. In 9.2, the specification of the model components and predictions are presented within the framework of the cognitive architecture for multimodal graph-text comprehension presented in Chapter 4. In 9.3, the empirical investigation is introduced, which provided the data to test the predictions of the model. In 9.4, the model is evaluated within the framework of the relevant graph-based reasoning models proposed in the literature on HCI and cognitive science.

9.1 AN OVERVIEW OF COGNITIVE MODELING IN HCI AND COGNITIVE SCIENCE

Experimental investigation—in particular, empirical user testing—has been accepted as a standard technique for developing a usable human-computer interface (Kieras, 1997; Landauer, 1995). Nevertheless, user testing is not always the most cost-effective method for evaluating the usability of an interface, in terms of the time and effort required for conducting experiments. The fast changes in the development of interface technologies, such as the development of graphical user interfaces in the

1990s, points to the need for low-cost methods for usability testing. The inspection methods such as heuristic evaluation (Nielsen, 1994) and cognitive walkthrough (Lewis, Wharton et al., 1990; Wharton et al., 1994) were developed as a reaction to this need (Kieras, 1997; Nielsen & Mack, 1994; see Chapter 3 of the dissertation for a review of research methods in HCI and cognitive science).

Another consequence was the development of cognitive engineering models and cognitive architectures, which were mostly originated from the seminal work Card et al. (1983). Cognitive modeling was shortly introduced as one of the research methods in HCI and cognitive science, in Chapter 3. The cognitive models of human information processing range from descriptions of general steps to complete a particular task to simulation models that predict user behavior during task performance with an interface (Cox & Peebles, 2008). Cognitive models are specifications of particular human information processing tasks in the course of interaction with complex information displays.

Card et al. (1983) presented a modeling approach to HCI, which was based on a basic model of human information processing and a task analysis method. This approach has been called the GOMS family of analysis techniques. The term GOMS is an acronym that stands for Goals, Operators, Methods, and Selection Rules. Kieras (1997) defines the GOMS model as

> a description of the knowledge that a user must have in order to carry out tasks on a device or system; it is a representation of the 'how to do it' knowledge that is required by a system in order to get the intended tasks accomplished. (p. 734)

A GOMS model is then a description of the Methods to accomplish specific Goals, where the Methods are composed of the Operators that the user has to perform. Selections Rules choose the appropriate method, depending on the context, if there is more than one Method to accomplish the Goal (Kieras, 1997). The models that are based on the GOMS technique are predictive engineering models since they make predictions about human information processing tasks during the course of interaction with an interface, before the interface is actually presented to the users (Cox & Peebles, 2008).

Since the 1980s, several types of GOMS modeling technique have been developed. The earliest GOMS model, the Keystroke-Level Model (KLM) is a specification of the operators (such as pressing a key on the keyboard, clicking a mouse button, pointing the mouse pointer to a target on the display etc.) and the required methods for performing a specific task (Card, Moran, & Newell, 1983). The main feature of the KLM was the prediction of task completion times, based on the completion times of the operators. Another type of GOMS model, the CPM-GOMS was developed by Gray, John, and Atwood (1993), which is a complex model compared to the KLM. The CPM-GOMS model outlines the sequential dependencies between a user's cognitive, perceptual, and motor (abb., CPM) processes in a schedule chart, where the execution time is predicted by the critical path, i.e. the shortest path composed of the necessary steps for completion of a task (see John & Kiearas, 1996a, b, for a review of GOMS family of analysis techniques; see Gray et al., 1997; Cox & Peebles, 2008; Langley et al., 2009; and Sun, 2009 for reviews of cognitive models and architectures in HCI).

The core of the GOMS modeling techniques is the *task analysis* method, which is based on the breaking down of a task into a set of subtasks (John & Kieras, 1996a, b; Kieras, 1997, 2004). Kieras (2004) provides the following definitions for the task analysis methods:

> *Task analysis* is the process of understanding the user's task thoroughly enough to help design a computer system that will effectively support users in doing the task. By **task** is meant the user's job or work activities, what the user is attempting to accomplish. By *analysis* is meant a relatively systematic approach to understanding the user's task that goes beyond unaided intuitions or speculations, and attempts to document and describe exactly what the task involves (p. 46-1, emphasis in original)

There are many aspects of the interaction of a user with human-computer interfaces, which include aspects of multimodal comprehension of graph-text constellations. Therefore, user testing is still a necessary component of the development of cognitive models (Kieras, 1997). The analysis of a user's activity in the context of the interaction with an interface usually precludes the aspects of human information processing that require a fine-grained level of analysis. Therefore, task analysis for system design has a rather heuristic character (Kieras, 2004; Annett, 2003). Both formalized and informal methods are used for task analysis.

In the model presented in this chapter, a quasi-formal approach is used for the decomposition of tasks specified by linguistic entities. In particular, sentences are analyzed in terms of the decomposition into subtasks of the specified task. The subtasks that are derived from the sentences are then used to predict gaze patterns of the readers on a set of stimulus graphs after their reading of a set of stimulus sentences. Eye tracking data obtained from an experimental investigation with human participants are used to test model predictions. In the next part of the chapter below, the model components and the predictions for gaze patterns are presented.

9.2 SPECIFICATION OF THE MODEL COMPONENTS AND PREDICTIONS

As stated in the discussion of the cognitive architecture for multimodal graph-text constellations in Chapter 4, the idealized shape for the graph lines used in this experimental investigation and its abbreviated spatial description can be shown as presented in Figure 9.1.

ROUGHLY_VERT_LINE(line_1) &
NUMBER_LOCAL_MAX_CURV(line_1, 1)
& … … …

Figure 9.1. The idealized graphical entity and its abbreviated spatial description.

The system of these structured representations is considered a descriptional inventory that is accessed by visual routines. Further contribution of the knowledge source SRS_G and the instantiation of the graph schema lead into the conceptual representation INCREASE_P($_{PATH}$, $_{SRSG}$) where the path argument is used for the graphical entities as directed linear entities (Habel, 1990; Eschenbach et al., 2000). The *PATH* argument of the INCREASE concept is the conceptual entity shared by language module and the graph module, which leads to the integration of the information induced by linguistic and graphical entities. Before further discussion of the INCREASE concept, the specification of the most basic concept, namely the state concept BE($_{TEMP}$, $_{VALUE}$) is introduced below.

9.2.1 The Specification of the State Concept BE($_{TEMP}$, $_{VALUE}$)

The most basic task in graph comprehension is to find the mapping between the time domain and the value domain for a certain point in time. This task is specified by the sentences that involve the 'be' predicate, as exemplified in (1), which is also one of the stimulus sentences used in the experimental investigation.

(1) *Kuş sayısı 2005 yılında 180 civarındadır.*
The number of birds in the year 2005 about 180 is.
'The number of birds is about 180 in the year 2005.'

The linguistic entities 'is about 180' and 'in the year 2005' in (1) correspond to entities in the domain of discourse. The lexical entity 'is' involves the state concept BE($_{TEMP}$, $_{VALUE}$) provided by the lexicon in its conceptual structure (CS). The *TEMP* argument stands for the specified time of the state and the *VALUE* argument stands for the value at that state. In (1) the *TEMP* and the *VALUE* arguments are explicitly specified. The BE($_{TEMP}$, $_{VALUE}$) state concept then stands for the specification of a mapping from the temporal domain to the value domain. In the terminology of topology, this is the specification of a mapping from the temporal domain to a 'position' in the value space.

In line graphs, positions in the value space are referred to by the points in the Cartesian space. In particular, for line graphs, the position corresponds to a point on a path. This correspondence can be specified by the conceptual entity BE_P($_{PATH}$, $_{SRSG}$), which is induced by a point in the graph space.

Geometrically, any point in the Cartesian space is specified in terms of a vertical location and a horizontal location. The vertical location of the point specifies the value of the *VALUE* argument in the specification of the BE($_{TEMP}$, $_{VALUE}$) state concept; the horizontal location of the point specifies the value of the *TEMP* argument. In graph comprehension, humans construct the correspondence between the *VALUE* argument and the TEMP argument by inspecting the location of the point with respect to the horizontal and vertical axes. This construction requires an execution of a saccade in the vertical direction (henceforth, vertical projection) and an execution of a saccade in the horizontal direction (henceforth, horizontal projection). Tick marks and numerical values on the axes facilitate these projections.

In multimodal comprehension of a graph-text constellation that involves the sentence (1) in the text layer, the reader should use the information induced by the lexical entity 'is' and the information extracted from the graph to construct the mapping from the temporal domain to the value domain. The construction of the mapping based on the state concept BE($_{TEMP}$, $_{VALUE}$) that is induced by the lexical entity 'is' can then be considered the task specified by (1). The task of the mapping construction can be decomposed into two subtasks.

- The construction of the mapping between the time domain and the position in the value space, which should be realized by making a vertical projection between the horizontal axis and the corresponding point on the path.
- The construction of the mapping between the value domain and the position in the value space, which should be realized by making a horizontal projection between the vertical axis and the corresponding point on the path.

The reader has the freedom to inspect the relevant entities on the graph at any order; he/she may first search for the value of the *TEMP* argument on the horizontal axis and then make a vertical projection to the path, or he/she may first search the value of the *VALUE* argument on the vertical axis and then make a horizontal projection to the path. In both cases, the reader should make at least one horizontal projection and at least one vertical projection during graph inspection, after reading the sentence (1). These predictions are tested in 9.3, where the results of the experimental investigation are presented.

The outcomes of the analysis presented above are used as the basic components of the task analysis presented below, where the INCREASE($_{TEMP}$, $_{VALUE}$) process concept is discussed.

9.2.2 The Specification of the Process Concept INCREASE($_{TEMP}$, $_{VALUE}$) by the Prepositional Phrases 'from' and 'to' as Path Functions

The process concept INCREASE_OF_VALUE($_{TEMP}$, $_{VALUE}$), henceforth shortly INCREASE($_{TEMP}$, $_{VALUE}$), was introduced in Chapter 4 as the conceptual entity that is induced by lexical entities such as 'increase' and 'rise', which are frequently used terms in graph-text constellations. The INCREASE concept is the conceptual entity that is contributed by the language module during comprehension of multimodal graph-text constellations. The lexical entity 'increase' is often accompanied by prepositional phrases (PPs) *from* and *to* (henceforth, 'from' and 'to'), which are the path functions that specify the *SOURCE* and *GOAL* for the *TEMP* argument and the *VALUE* argument. The graph module contributes to the integration module the conceptual representation INCREASE_P($_{PATH}$, $_{SRSG}$). As stated above, *PATH* is the conceptual entity shared by language comprehension module and the graph comprehension module, which leads to the integration of the information induced by linguistic and graphical entities.

The sentence in (2) below exemplifies the use of the 'increase' lexical entity accompanied by the 'to' PP.

(2) *Kuş sayısı 2005 yılında 180 civarına yükselmiştir.*
The number of birds in the year 2005 to about 180 increased.
'The number of birds in the year 2005 increased to about 180.'

In multimodal comprehension of a graph-text constellation, where the text layer of the constellation involves a sentence such as (2), the reader should construct the mapping between the time domain and value domain by using the information induced by the lexical entities in the sentence. As previously stated, the construction of the mapping can be considered the task specified by (2). The task of the mapping construction can be decomposed into three subtasks for inspection of graphs.

- The construction of the mapping between the time domain and the position in the value space, which should be realized by making a vertical projection between the horizontal axis and the corresponding point on the path.
- The construction of the mapping between the value domain and the position in the value space, which should be realized by making a horizontal projection between the vertical axis and the corresponding point on the path.
- The construction of the mapping between the two arguments by means of the spatial properties of the INCREASE process concept. In particular, the INCREASE concept holds the necessary condition *VALUE(END(INCREASE))* > *VALUE(BEGIN(INCREASE))* where the former is the *GOAL* argument of the path and the latter is the *SOURCE* argument of the path. In (2), only the *GOAL* argument is explicitly stated.

Since the sentence (2) involves the process concept increase, the differences in the gaze patterns of readers on graphs after their reading of the sentence (1) and after their reading of the sentence (2) should reflect the differences between the state concept BE(_$_{TEMP}$, _$_{VALUE}$) and the process concept INCREASE(_$_{TEMP}$, _$_{VALUE}$).

The components of the analyses presented so far can be used to predict gaze patterns of readers on graphs after their reading of more complex sentences, as exemplified in (3) below.

(3) *Kuş sayısı 1970 yılında 30 civarından 2005 yılında 180 civarına yükselmiştir.*
The number of birds in the year 1970 from about 30 in the year 2005 to about 180 increased.
'The number of birds increased from about 30 in the year 1970 to about 180 in the year 2005.'

The sentence in (3) induces the process concept INCREASE(_$_{TEMP}$, _$_{VALUE}$) which specifies the mapping between the two arguments by stating the beginning and the end values explicitly. In other words, the value of the *VALUE(BEGIN(INCREASE))* and the value of the *VALUE(END(INCREASE))* are explicitly given in the sentence; the *TEMP* argument is given by the values of the *TEMP(BEGIN(INCREASE))* and the *TEMP(END(INCREASE))*. Accordingly, the sentence (3) suggests the following three subtasks of mapping construction between the arguments in graph comprehension.

- The construction of the mapping between the *TEMP(BEGIN(INCREASE))* and the *VALUE(BEGIN(INCREASE))* arguments. This subtask can be divided into two further subtasks of graph inspections: the construction of the correspondence between the value of the *TEMP(BEGIN(INCREASE))* and the point on the path by means of a vertical projection and the construction of the correspondence between the point on the path and the value of the *VALUE(BEGIN(INCREASE))* by means of a horizontal projection. Consequently, the construction between the *TEMP(BEGIN(INCREASE))* and the *VALUE(BEGIN(INCREASE))* arguments should be similar to the constructions in (1), as discussed above.
- The construction of the mapping between the *TEMP(END(INCREASE))* and the *VALUE(END(INCREASE))* arguments. In a similar way to the construction of the mapping between the *TEMP(BEGIN(INCREASE))* and the *VALUE(BEGIN(INCREASE))* arguments described above, this construction can be further divided into the vertical and horizontal projection subtasks. Eventually, the construction between the *TEMP(END(INCREASE))* and the *VALUE(END(INCREASE))* arguments should be similar to the constructions in (1), as discussed above.
- The construction of the mapping between the arguments by means of the spatial properties of the INCREASE process concept. In particular, the INCREASE concept holds the necessary condition *VALUE(END(INCREASE))* >

VALUE(BEGIN(INCREASE)) where the former is the *GOAL* argument of the path and the latter is the *SOURCE* argument of the path. This specification, as discussed in the predictions of inspections of the graphs after (2), should lead to multiple fixations on the graph line, i.e. the tracing of the graph line by the readers.

In addition to the PPs 'from' and 'to', the lexical entity 'increase' is often accompanied by 'since', 'until', and 'between', which are also path functions that specify the *SOURCE* and *GOAL* for the *TEMP* argument and the *VALUE* argument. Those PPs are discussed below.

9.2.3 The Specification of the Process Concept INCREASE($_{TIME}$, $_{VALUE}$) by the Prepositional Phrases 'Since', 'Until', and 'Between' as Path Functions

In this part of the task analysis, the specification of the mapping from the temporal domain to the value domain by the process concept INCREASE($_{TEMP}$, $_{VALUE}$) is investigated with respect to the specification by the prepositional phrases 'since', 'until', and 'between'. The sentences presented in (4a–c) exemplify the uses of those PPs.

(4) a. *Kuş sayısı 1970 yılından beri yükselmiştir.*
The number of birds 1970 the year since increased.
'The number of birds increased since year 1970.'
b. *Kuş sayısı 2005 yılına kadar yükselmiştir.*
The number of birds 2005 the year until increased.
'The number of birds increased until the year 2005.'
c. *Kuş sayısı 1970 ve 2005 yılları arasında yükselmiştir.*
The number of birds 1970 and 2005 years between increased.
'The number of birds increased between the years 1970 and 2005.'

The sentence (4a) explicitly states the value of the *TEMP* argument for the beginning of the increase, i.e. *TEMP(BEGIN(INCREASE))*; there is no other explicitly stated argument value by the sentence.[1] The modification by the PP *beri* 'since, for', sets the argument as the *SOURCE* argument. In other words, the 'since' PP has a similar function to the 'from' PP. Therefore, the decomposition of the task specified by (4a) suggests that the task will be composed of the subtasks exemplified by (1) and the additional task of the construction of the mapping specified by the INCREASE concept.

[1] The language of the exemplified sentences, Turkish, does not have prepositions but postpositions that follow their complements. The postpositional phrases headed by the term *beri* 'since, for' always have temporal meaning in Turkish. In the use of *beri* in its sense of 'since', the complement noun phrase (e.g., 'the year 1970') takes the ablative case and it expresses the time at which the process or state began. (Göksel & Kerslake, 2005, p. 247). In (4a), the 'since' sense of *beri* is used.

Moreover, the mapping should be further characterized by the specification provided by the 'since' PP in the sentence.

The PP 'until' is used for the specification of the *TEMP(END(INCREASE))* argument in (4b).[2] The specification by *kadar* 'until' sets the argument as the *GOAL* argument. In other words, 'until' in (4b) has a similar function to 'to' in (2) and in (3). Therefore, the subtasks that are obtained by the decomposition of the task specified by (4b) should be a combination of the subtasks specified by (1) and the additional task of the construction of the mapping specified by the INCREASE concept. In addition, the mapping should be further characterized by the specification provided by the 'until' PP in the sentence.

In particular, the PPs 'since' and 'until' place a temporal boundary on the INCREASE concept. From the perspective of aspectual semantics, the PPs 'since' and 'until' are aspectual modifiers that modify the act of increase (cf. the situation aspect, i.e. Aktionsarten, see Pustejovsky, 1995; Vendler, 1957; Verkuyl, 1972, 1993; Jackendoff, 2002). Therefore, more frequent inspections of the graph line are expected after readers' reading of the sentences as exemplified in (4a) and (4b) compared to the inspections after their reading of (2) and (3). This prediction is tested in the succeeding part of the chapter, where the experimental investigation is reported.

The PP 'between' has a more specific use in the time domain, in contrast to its use in the spatial domain. In both uses, the PP 'between' has two arguments. However, in the use of the PP 'between' in the time domain, the time specifies one argument as the *SOURCE* argument and the other argument as the *GOAL* argument; whereas in the use of the PP 'between' in the spatial domain the two arguments are not specified as *SOURCE* and *GOAL*. Therefore, the lexical entity 'between' in (4c) can be seen as a combination of (4a) and (4b), in which the not-stated *SOURCE* argument in (4a) and the not-stated *GOAL* argument in (4b) are explicitly stated in (4c). Consequently, the fixation map for (4c) should be a combination of the fixation maps obtained in (4a) and (4b).

The experimental investigation is presented in 9.3 below.

[2] The postpositional phrases headed by the term *kadar* 'until, so far as' may have a temporal or spatial meaning in Turkish, in both cases involving a terminal point. The complement noun phrase (e.g., 'the year 2005') takes the dative case and it expresses the time at which the process or state end. (Göksel & Kerslake, 2005, p. 245). In (4b), the 'until' sense of *kadar* is used.

9.3 THE EXPERIMENTAL INVESTIGATION

The experimental investigation was conducted with 36 participants (mean age 21.8, SD = 1.62), who were paid to participate in the experiment.

9.3.1 Materials and Design

The experimental material was composed of 12 stimulus sentences and 12 line graphs that depicted population trends of waterbird species. The stimulus sentences were excerpted from the waterbird consensus report that was used as experimental stimuli in the experiments reported in Chapter 6. The sentences in the source material were used either without modification or they were regenerated by modifying the original sentences for the purpose of the systematic analysis aimed in the experimental investigation. The twelve stimulus sentences are presented below.

	Stimulus Sentences
S01	The number of birds is about 180 in the year 2005.
S02	The number of birds is about 30 in the year 1970.
S03	The number of birds in the year 2005 increased to about 180.
S04	The number of birds increased to about 180 in the year 2005.
S05	The number of birds increased from about 30 in the year 1970 to about 180 in the year 2005.
S06	The number of birds increased between the years 1970 and 2005 from about 30 to about 180.
S07	The number of birds increased from about 30 to about 180 between the years 1970 and 2005.
S08	The number of birds increased from about 30 to about 180.
S09	The number of birds increased since year 1970.
S10	The number of birds increased until the year 2005.
S11	The number of birds increased until about 180.
S12	The number of birds increased between the years 1970 and 2005.

The 12 graphs that were used in the experiment were redrawn and regenerated based on the four graphs excerpted from the waterbird consensus report. In particular, four graphs were redrawn based on the original, four graphs were generated by constructing the horizontal mirror images from the four original graphs and by keeping the numerical values on the axes the same, and the final four graphs were generated by keeping the graph line the same and by changing the value of the numerical axis values in the original. Consequently, 12 graphs were obtained. Two sample graphs are shown in Figure 9.2, the whole set of graph stimuli can be seen in Appendix D.1.

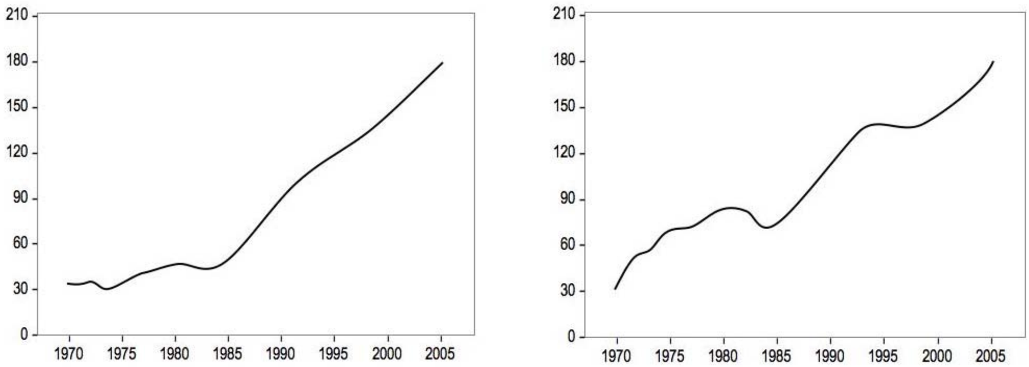


Figure 9.2. Two sample graph stimulus from the experiment.

All of the 12 graphs used in the experiment were approximate representations of an *increase* process, though with different granularity. The particular spatial properties of the graph lines, such as the presence of local maxima and minima, were not evaluated as a dependent parameter. The purpose of using different types of increase graphs was to minimize the potential influence of the spatial properties of a specific graph line.

9.3.2 Procedure

The experiment was conducted in single sessions. The following procedure was used in the experiment: The participants were informed that they were expected to evaluate whether the graphs were correct or incorrect according to the stimulus sentences. Accordingly, 12 sentence-graph pairs were presented to the participants in 24 consecutive screens: A stimulus sentence was presented first, and when the participant decided that he/she understood the sentence, by proceeding to the next screen by pressing a key, he/she reported whether the presented graph in the succeeding screen was true or false. In other words, a sentence-graph verification paradigm was used in the experiment. Eight of the twelve graphs were congruent with all the stimulus sentences, whereas the remaining four graphs were not congruent with any of the stimulus sentences. The presentation of the material was randomized. There were no time limitations in the experiment. The entire session took approximately 5-10 minutes.

The sentence-graph verification paradigm corresponds to flipping the page in a newspaper or magazine article, or clicking a link for the pop-up window, for the purpose of investigating the graph, after reading a sentence in real-world settings. A 50 Hz. non-intrusive eye tracker recorded eye-tracking data (see 5.2 for the details of the eye tracker and the general description of the methodology).

9.3.3 Analysis of the Data

The recorded eye movement data were transcribed manually based on the AOI (Area of Interest) specification shown in Figure 9.3.

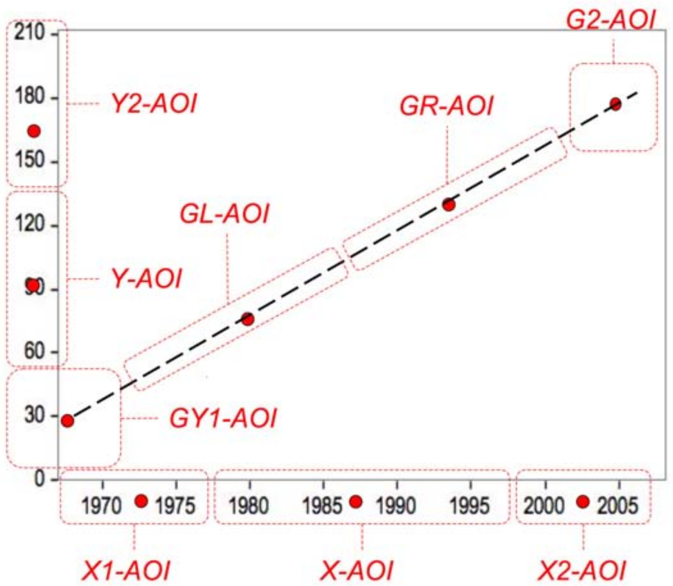


Figure 9.3. The representation AOI specification for the analysis of eye movement data.

The x-axis was divided into three AOIs, the y-axis was divided into three AOIs, and the graph line was divided into four AOIs in each graph. The locations of the x-axis AOI and the y-axis AOI were kept the same in all graph stimuli, whereas the graph line AOIs (i.e., the GY1-AOI, the GL-AOI, the GR-AOI, and the G2-AOI) were specified separately for each graph type. For the specification of the graph line AOIs, the GY1-AOI and the G2-AOI were assigned two rectangular regions, as shown in Figure 9.3. The start point and the end point of the graph lines were set as the center of the rectangular region to compose the GY1-AOI and the G2-AOI. The GY-AOI covered the start point of the graph line and the corresponding y-axis value because of the close distance between the two in all graphs. The remaining line segment was divided into two diagonally equal-length segments; the segment on the left is named the GL-AOI and the segment on the right is named the GR-AOI. The width of the GL-AOIs and the GR-AOIs were determined by the 2° visual angle around the graph line. In the following parts of the chapter, the nine AOIs are represented by the nine circles as shown in Figure 9.3.

Since the focus of the model is the processing perspective, the transition of eye movements between the AOIs was analyzed rather than the total number of fixation counts in the AOIs. The eye movement data are presented in two forms: fixation map and transition matrix. The fixation maps used in this chapter depict transition eye movements by representing gaze patterns by arrows between the AOIs. The thickness of an arrow represents the frequency of the gaze pattern between two AOIs or within and AOI. As an example consider the AOI specification on the left in Figure 9.4, which includes eight AOIs from A to H. A sample fixation map for the eight AOIs is shown on the right. The gaze patterns are represented by the arrows. Accordingly, there are seven gaze patterns in the figure. There is one within-AOI gaze pattern in the D-AOI. This shows that there was at least one consecutive fixation pair within that AOI. The gaze pattern with the circled number two is the most frequently recorded gaze pattern in the fixation map, because is has the highest thickness. The gaze pattern number 3 is also a frequent gaze pattern, but its frequency (or its percentage in the total number of recorded gaze patterns) is less than the gaze pattern number 2 and higher than all the others.

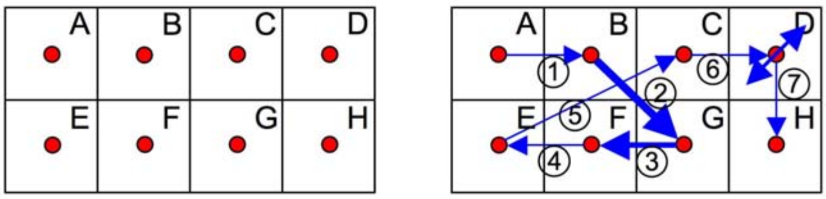


Figure 9.4. Sample AOI specification (left) and fixation map (right).

The fixation map can be represented as a transition matrix as shown below (see Ponsoda et al., 1995; Goldberg & Kotval, 1999; Webb & Renshaw, 2008, for the use of transition matrix in eye tracking research).

to / from	A	B	C	D	E	F	G	H	*SUM (from)*
A		1							1
B							1		1
C				1					1
D				1				1	2
E			1						1
F					1				1
G						1			1
H									0
SUM (to)	0	1	1	2	1	1	1	1	

In the following matrix, the values are shown in terms of their weights; in other words, each number in the cell is divided by the total value of all the numbers in the matrix. The resulting transition matrix of the exemplified gaze patterns is shown below.

to / from	A	B	C	D	E	F	G	H	*SUM (from)*
A	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.13
B	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.13
C	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.13
D	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.13	0.25
E	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.13
F	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.13
G	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.13
H	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SUM (to)	0.00	0.13	0.13	0.25	0.13	0.13	0.13	0.13	

In the following part of the chapter, eye tracking data are evaluated to test the predictions concerning the tasks specified by the stimulus sentences. As stated above, each participant was presented 12 sentences and 12 graphs in the experiment. Eight of the graphs were congruent with the presented sentences, whereas the remaining four graphs were not congruent. The gaze patterns on the congruent graphs comprised 36 x 8 (= 288) of the total 36 x 12 (= 432) responses. In the following analysis, 17 responses for congruent graphs were not included to the analysis due to calibration problems in the eye tracker; the remaining 271 responses were included to the analysis.

9.3.4 The Fixation Maps

The stimulus sentence S01 is repeated as (5) below.

(5) *Kuş sayısı 2005 yılında 180 civarındadır.*
The number of birds in the year 2005 about 180 is.
'The number of birds is about 180 in the year 2005.'

Figure 9.5 shows the fixation map obtained from the specified AOIs after the participants' reading of S01. As described above, thicker arrows represent higher frequency of gaze patterns. One-way arrows show the gaze patterns between two AOIs, whereas two-way arrows show gaze patterns within a single AOI.

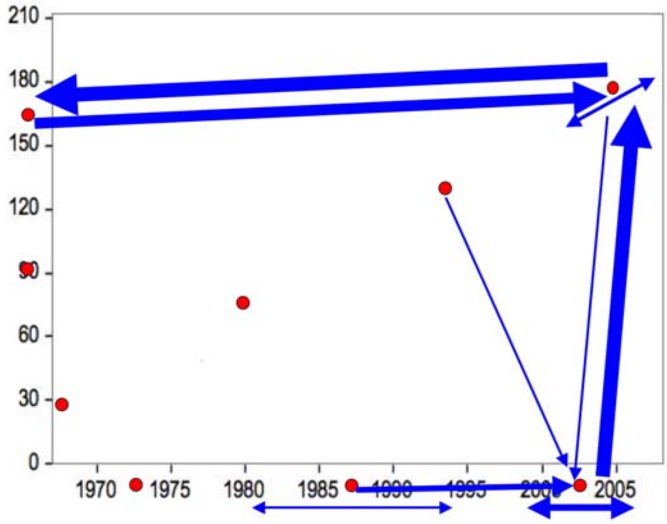


Figure 9.5. The fixation map after the stimulus sentence S01: 'The number of birds is about 180 in the year 2005.'

The fixation map shown in Figure 9.5 is a partial depiction of the transition matrix from the specified AOIs, shown below; only the values that had the weight value of .03 and more (shown in italics in the transition matrix) are the ones that are depicted in the fixation map.

to / from	X1	X	X2	Y	Y2	GY1	G2	GL	GR	*SUM (from)*
X1	.00	.01	.01	.00	.00	.01	.00	.00	.00	.02
X	.00	***.03***	***.07***	.00	.00	.00	.00	.00	.00	.10
X2	.01	.01	***.08***	.00	.00	.00	***.17***	.00	.01	.28
Y	.00	.00	.00	.00	.01	.00	.00	.00	.00	.01
Y2	.00	.00	.02	.01	.01	.00	***.13***	.01	.01	.19
GY1	.00	.00	.00	.00	.00	.00	.01	.00	.00	.01
G2	.01	.01	***.03***	.00	***.17***	.00	***.04***	.00	.02	.26
GL	.00	.01	.01	.00	.00	.00	.00	.00	.01	.03
GR	.00	.02	***.03***	.01	.02	.00	.02	.01	.01	.11
SUM (to)	.01	.08	.25	.01	.20	.01	.35	.02	.05	

The sum of the weight of the four horizontal and vertical gaze patterns is .50; in other words, the horizontal and vertical gaze patterns constitute 50% of the total number of gaze patterns on the graphs after the stimulus sentence S01. This shows that the model, which is based on the decomposition of the task specified by the stimulus sentence into two subtasks, predicts 50% of gaze patterns. Part of the remaining 50% reflects the participants' effort to reach the value of the *TEMP* argument; in particular, the gaze patterns from the X1-AOI, the X-AOI, the Y2-AOI, the GL-AOI, and the GR-AOI, to the X2-AOI, and the other two gaze patterns on the x-axis (the gaze patterns within the X-AOI and the X2-AOI, which are shown by the two two-way arrows) reflect the readers' effort to reach the value of the *TEMP* argument; these gaze patterns constitute 25% of the total number of gaze patterns. Moreover, the within-AOI gaze patterns in the X2-AOI with relatively higher weights compared to other within-AOI fixations in the axis AOIs reflect the participants' effort to find the value of the *TEMP* argument.

The fixation map also reveals readers' tendencies about the directionality of the projections: a stronger tendency to make a vertical projection from the x-axis to the graph line compared to making a vertical projection from the graph line to the x-axis, and a slightly stronger tendency to make a horizontal projection from the graph line to the y-axis compared to making a horizontal projection from the y-axis to the graph line. Moreover, although the order of occurrence of gaze patterns is not shown in the fixation map, a more detailed investigation of the order of fixation locations shows that the majority of the horizontal projections from the y-axis to the graph line take place after the horizontal projection from the graph line to the y-axis, which is preceded by the vertical projection from the x-axis to the graph line. In other words, the order of occurrence of the frequently seen gaze pattern is: a vertical projection from the x-axis to the graph line, a horizontal projection from the graph line to the y-axis, and a second horizontal projection from the way back from the y-axis to the graph line.

The tendencies about the directions of the gaze patterns show that the directionality in the specification of the mapping *from* the temporal domain *to* the value space, i.e. the BE($_{TEMP}$, $_{VALUE}$) state concept induced by the 'is' lexical entity, can be captured by the directions of vertical and horizontal projections, which are reflected by the directions of gaze patterns. This is also evidenced by participants' higher effort to find the value of the *TEMP* argument (25% of the gaze patterns, as stated above) compared to the effort to find the value of the *VALUE* argument; except for the horizontal projection from the point on the graph line to the Y2-AOI, the sum of the gaze patterns to the Y2-AOI is 4% of the total.

The projections do not depend on the position of the point on the path (as represented by the graph line); whether the point is a starting point, final point or a point between the starting and final point of the path, the projections are necessary for the construction of the mapping from the temporal domain to the value space. The stimulus sentence S02, shown in (6) below, exemplifies a case where the position of the point on the path is the starting position.

(6) *Kuş sayısı 1970 yılında 30 civarındadır.*
The number of birds in the year 1970 about 30 is.
'The number of birds is about 30 in the year 1970.'

The fixation map, shown in Figure 9.6 shows the projections to the axes and the projections from the axes, and the other major gaze patterns recorded from the participants after their reading of the stimulus sentence S02. As stated above, since the starting point of the graph line was close to the y-axis values, an AOI that covered the starting point and the relevant y-axis value was specified for the analysis, namely the GY1-AOI. Therefore, vertical and horizontal gaze patterns are not separately shown in the figure but represented by L-shape arrows.

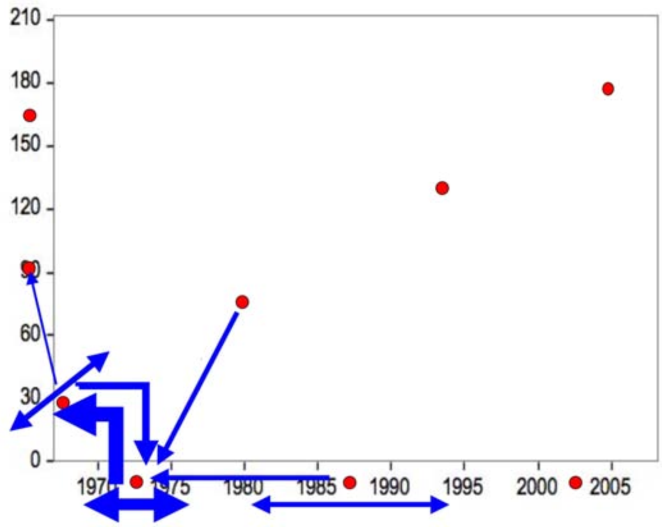


Figure 9.6. The fixation map after the stimulus sentence S02: 'The number of birds is about 30 in the year 1970.'

The transition matrix from the specified AOIs is presented in Appendix D.2. As previously stated, the fixation map shown in Figure 9.6 is a partial depiction of the transition matrix from the specified AOIs: Only the values that had the weight value of .03 and more (shown in italics) are the ones that are depicted in the fixation map.

The projections, which are shown by the two L-shape arrows and the one two-way arrow in the GY1-AOI constitute 34% of the total number of gaze patterns recorded after the participants' reading of the stimulus sentence S02, as can be calculated from the transition matrix. The percentage of the projections is low compared to the percentage of the projections after S1, which was 50%, as reported above. The reason for this difference is the close distance between the starting point of the graph line and the numerical value on the x-axis: The participants were able to attend to both in a single fixation, without making a saccade.

The directionality of the specification of the mapping between the temporal domain and the value domain, i.e. the BE($_{TEMP}$, $_{VALUE}$) state concept induced by the 'is' lexical entity, is reflected by the larger thickness of the arrow that originates from the x-axis compared to the thickness of the arrow that originates from the y-axis. As stated in the analysis of the post-S01 gaze patterns on the graphs, the directionality of the mapping from the temporal domain to the value domain can be captured by the directionality in participants' gaze patterns on the graphs after their reading of S02.

In summary, the analysis of the two stimulus sentences S01 and S02 and the succeeding gaze patterns on the graphs after participants' reading of S01 and S02 reveal that the specification of the mapping from the time domain to a position in the value space, i.e. the state concept BE(*_TEMP*, *_VALUE*) that is induced by the 'is' lexical entity, can be described in terms of a task, which can be decomposed into two subtasks, as described in 9.2.1 above. In other words, the language module of the multimodal graph-text comprehension architecture provides the two subtasks to the integration module. The graph comprehension module of the architecture provides the conceptual representation BE_P(*_PATH*, *_SRSG*) for integration where the position in the value space corresponds to a point on the path. The tests for the predictions concerning the process concept INCREASE are presented below (see 9.2.2 for the predictions).

The two stimulus sentences S03 and S04 repeated below as (7a–b) exemplify the use of the 'increase' lexical entity accompanied by the 'to' PP.

(7) a. *Kuş sayısı 2005 yılında 180 civarına yükselmiştir.*
The number of birds in the year 2005 to about 180 increased.
'The number of birds in the year 2005 increased to about 180.'
b. *Kuş sayısı 180 civarına 2005 yılında yükselmiştir.*
The number of birds to about 180 in the year 2005 increased.
'The number of birds increased to about 180 in the year 2005.'

The two sentences convey similar information with a different emphasis. The stimulus sentence S03 (in 7a) emphasizes the increase in the value domain, whereas the stimulus sentence S04 (in 7b) emphasizes the time of the increase in the value domain, i.e. the year 2005. The investigation of the influence of the word order (thus, the emphasis on the temporal or the value domain) is an open question that is left for future research. The major goal of this study is to analyze the differences between the gaze patterns that are obtained after participants' reading of S01 and the gaze patterns that are obtained after participants' reading of S03/S04. Accordingly, the differences in the gaze patterns between S01 and S03/S04 should reflect the differences between the state concept BE(*_TEMP*, *_VALUE*) and the process concept INCREASE(*_TEMP*, *_VALUE*).

Figure 9.7 shows the fixation map based on the gaze patterns in the graphs after participants' reading of the stimulus sentence S03.

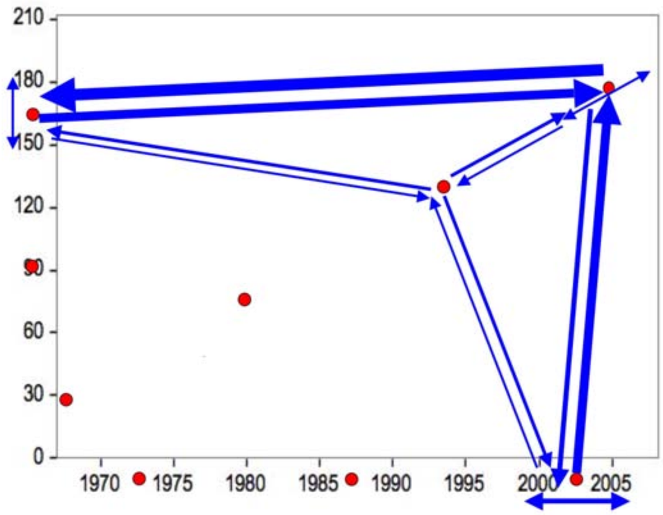


Figure 9.7. The fixation map after the stimulus sentence S03: 'The number of birds in the year 2005 increased to about 180.'

The distribution of horizontal and vertical projections is quite similar to the distribution obtained for S01 (see Figure 9.5). Moreover, the readers made projections from other parts of the graph line—in particular, from the graph line segments that are covered by the GR-AOI—to both axes, and projections from the axes to the GR-AOI. In addition to the gaze patterns between the GR-AOI and the axes, the two gaze patterns between the GR-AOI and the G2-AOI show that the participants traced these segments of the graph line by fixating back and forth between the two AOIs. This tracing behavior is particularly important for the difference between the state concept BE($_{TEMP}$, $_{VALUE}$) and the process concept INCREASE($_{TEMP}$, $_{VALUE}$), which will be further discussed later in this part of the chapter.

The transition matrix from the specified AOIs is shown Appendix D.2. As can be calculated from the transition matrix, the four projection gaze patterns between the axes and the G2-AOI constitute 42% of the total number of gaze patterns. The six gaze patterns that are connected to the GR-AOI, which reflect the difference between the use of the state concept BE($_{TEMP}$, $_{VALUE}$) and the process concept INCREASE($_{TEMP}$, $_{VALUE}$), constitute 21% of the total gaze patterns.

The major difference between the gaze patterns obtained after S01 and the gaze patterns obtained after S03 is the difference between the weights of the gaze patterns connected to the GR-AOI in both directions. (This difference can be seen by comparing the *SUM (to)* values of the GR-AOI (.05 in S01 matrix vs. .14 in S03 matrix) and by the *SUM (from)* value of the GR-AOI (.11 in S01 matrix vs. .18 in S03 matrix), see Appendix D.2).

Figure 9.8 shows the fixation map after participants' reading of S04. As stated above, the difference between S03 and S04 is the word order, which results in more emphasis on the time domain in S04 compared to S03. The sentence S04 can be rephrased as "It is the year 2005 when the number of birds increased to 180", where the sentence S03 can be rephrased as "It is the value of 180 where the number of birds increased to in 2005". The difference between S03 and S04 resulted in a different distribution of projections, in particular, vertical projections in S04 (Figure 9.8 below) compared to S03 (Figure 9.7 above).

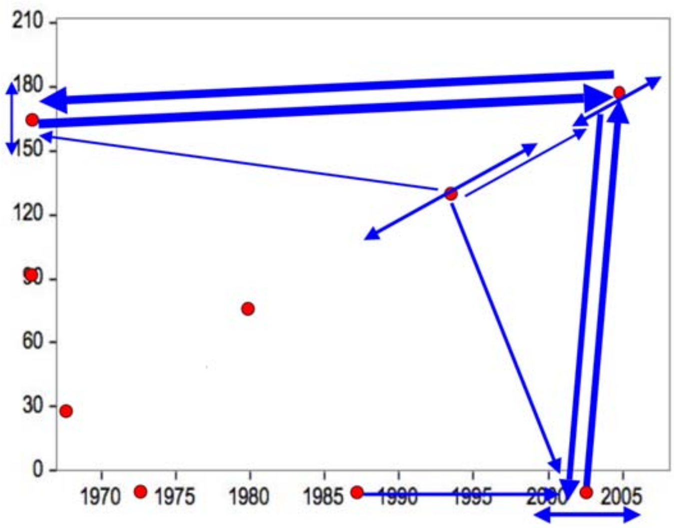


Figure 9.8. The fixation map after the stimulus sentence S04: 'The number of birds increased to about 180 in the year 2005.'

The transition matrix for the post-S04 gaze patterns is shown in Appendix D.2. The comparison of the gaze patterns between S03 and S04 show that the distributions of gaze fixations over the AOIs were similar with a few differences. In addition, the directionality of the mapping, which is specified by the syntactic order of lexical entities, is reflected in the weights of the projection gaze patterns (see Appendix D.2 for a more detailed comparison).

In summary, the lexical entity 'increase' in the sentences S03 and S04 introduced the subtask of graph line inspection in addition to the subtasks specified by S01 and S02.

As stated in 9.2.2, the findings obtained from the stimulus sentences S03 and S04 can be used for predicting gaze patterns after participants' reading of more complex sentences. The stimulus sentences S05, S06, S07, and S08 exemplify those relatively complex sentences. The stimulus sentence S05, S06, and S07 are repeated as (8a–c) below. All the three stimulus sentences are grammatical in the language of the experiment.

(8) a. *Kuş sayısı 1970 yılında 30 civarından 2005 yılında 180 civarına yükselmiştir.*
The number of birds in the year 1970 from about 30 in the year 2005 to about 180 increased.
'The number of birds increased from about 30 in the year 1970 to about 180 in the year 2005.'

b. *Kuş sayısı 1970 ve 2005 yılları arasında 30 civarından 180 civarına yükselmiştir.*
The number of birds 1970 and 2005 years between from about 30 to about 180 increased.
'The number of birds increased between the years 1970 and 2005 from about 30 to about 180.'

c. *Kuş sayısı 30 civarından 180 civarına 1970 ve 2005 yılları arasında yükselmiştir.*
The number of birds from about 30 to about 180 between the years 1970 and 2005 increased.
'The number of birds increased from about 30 to about 180 between the years 1970 and 2005.'

The predicted fixation map on the graphs after the participants' reading of the stimulus sentence S05 is a combination of the gaze patterns obtained in S01 (Figure 9.5), in S02 (Figure 9.6) and S03/S04 (Figure 9.7 and 9.8). The fixation map obtained by the recorded eye movements of the participants on the graphs after their reading of S05 is shown in Figure 9.9.

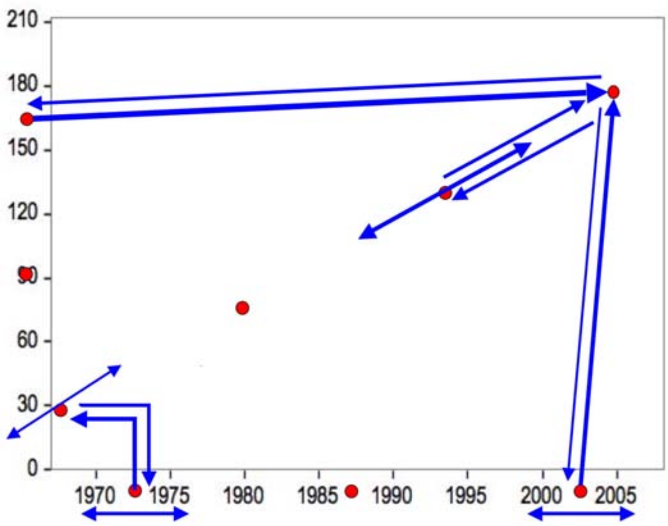


Figure 9.9. The fixation map after the stimulus sentence S05: 'The number of birds increased from about 30 in the year 1970 to about 180 in the year 2005.'

The obtained gaze patterns are quite close to the predicted gaze patterns. The vertical and horizontal projections are obtained for both the *TEMP(BEGIN(INCREASE))* - *VALUE(BEGIN(INCREASE))* argument pair and the *TEMP(END(INCREASE))* - *VALUE(END(INCREASE))* argument pair. The projections between the values of the arguments and the two points of the path constitute 39% of the total number of gaze patterns. The back and forth fixations between the G2-AOI and the GR-AOI, as well as the within-AOI fixations in the GR-AOI indicate participants inspection of the graph line for the INCREASE process concept. The transition matrix, presented in Appendix D.2, provides the data for all gaze patterns.

The stimulus sentence S08 (8b), is similar to S05 with respect to the explicitly stated argument values, but the word order is different. The word order in S06 emphasizes the increase in the value domain more than the duration of the increase. As described in 9.2.2, the decomposition of the task specified by S06 results in three subtasks, similar to the decomposition of the task specified by S05: the projections concerning the *TEMP(BEGIN(INCREASE))* - *VALUE(BEGIN(INCREASE))* argument pair, the projections concerning the *TEMP(END(INCREASE))* - *VALUE(END(INCREASE))* argument pair, and the construction of the mapping between the arguments via the spatial properties of the INCREASE process concept. The expected fixation map should then be similar to the one obtained in S06 except for the differences due to the different word order in S06 compared to S05. The fixation map obtained by the recorded eye movements of the participants on the graphs after their reading of S06 is shown in Figure 9.10.

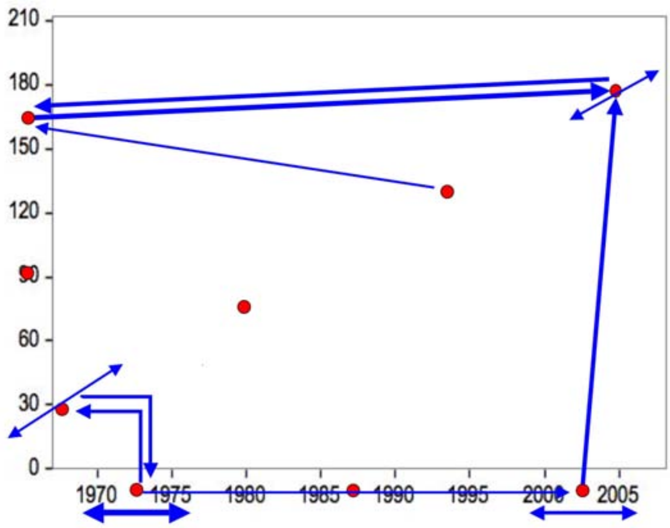


Figure 9.10. The fixation map after the stimulus sentence S06: 'The number of birds increased between the years 1970 and 2005 from about 30 to about 180.'

The obtained gaze patterns for S06 are close to the prediction. The major differences between the fixation maps of S05 and S06 are the reduced weight of the projections (25% in S07 vs. 39% in S06) and the reduced weight of the gaze patterns between the

GR-AOI and the G2-AOI. The gaze pattern from the X1-AOI to the X2-AOI and the gaze pattern from the GR-AOI to the Y2-AOI reflect the difference between S05 and S06. However, as stated previously, the focus of the analysis presented in this chapter is the comparison of gaze patterns on the graphs obtained after the reading of different stimulus sentences by the readers, in particular the comparison between simple sentences such as S01–S04 and more complex sentences such as S05–S08, rather than a detailed investigation of the influence of the word order on task decompositions and their corresponding gaze patterns. Therefore, the most important finding of the analysis is the presence of the projection gaze patterns for the *TEMP(BEGIN(INCREASE))* - *VALUE(BEGIN(INCREASE))* argument pair, the projections for the *TEMP(END(INCREASE))* - *VALUE(END(INCREASE))* argument pair, and the presence of the fixations on the graph line.

The transition matrix, presented in Appendix D.2, provides the data for all gaze patterns. The projections constitute 25% of the total gaze patterns. The within-AOI gaze patterns for the GY1-AOI and the G2-AOI show the fixations on the graph line. The gaze patterns between the GR-AOI and the G2-AOI in S06 have a lower weight (4%) compared to the one obtained in S05 (8%).

The analysis of the third of the four relatively complex sentences is presented below. The stimulus sentence S07, presented in (8c) above, is repeated below as (9).

(9) *Kuş sayısı 30 civarından 180 civarına 1970 ve 2005 yılları arasında yükselmiştir.*
The number of birds from about 30 to about 180 between the years 1970 and 2005 increased.
'The number of birds increased from about 30 to about 180 between the years 1970 and 2005.'

The word order in S07 emphasizes the duration of the increase more than the increase in the value domain. A similar gaze pattern on the graphs is predicted after the participants' reading of S07, though with differences that reflect the influence of the different word order. In other words, the predicted gaze pattern for S07 is a combination of the gaze patterns obtained in S01 (Figure 9.5), in S02 (Figure 9.6) and S03/S04 (Figure 9.7 and 9.8). Figure 9.11 below shows the gaze pattern obtained from the recorded eye movements after the participants' reading of the stimulus sentence S07.

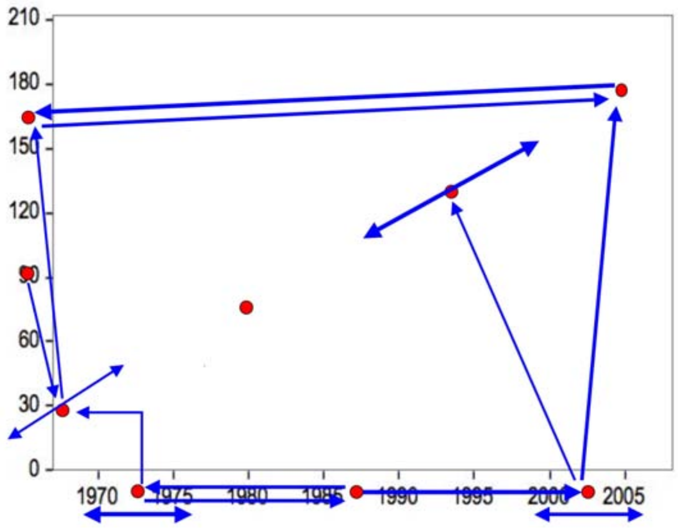


Figure 9.11. The fixation map after the stimulus sentence S07: 'The number of birds increased from about 30 to about 180 between the years 1970 and 2005.'

The obtained gaze patterns from the recorded eye movements are close to the expected gaze patterns. The transition matrix, presented in Appendix D.2, provides the data for all gaze patterns. As can be seen from the transition matrix, the projections for the *TEMP(BEGIN(INCREASE))* - *VALUE(BEGIN(INCREASE))* argument pair and the projections for the *TEMP(END(INCREASE))* - *VALUE(END(INCREASE))* argument pair constitute 19% of the total gaze patterns (not all of them are shown in the figure). The participants fixated on the graph line, as indicated by the within-AOI gaze patterns in the GR-AOI (shown in the figure), G2-AOI and the GL-AOI (not shown in the figure). The major difference in gaze patterns after participants' reading of S07, compared to S05 and S06 is the increased weight of the gaze patterns between the AOI on the x-axis and the gaze patterns between the GY1-AOI and the y-axis AOIs.

The analysis of the last of the four relatively complex sentences is presented below. The stimulus sentence S08 is shown as (10) below.

(10) *Kuş sayısı 30 civarından 180 civarına yükselmiştir.*
The number of birds from about 30 to about 180 increased.
'The number of birds increased from about 30 to about 180.'

Figure 9.12 shows the fixation map after participants' reading of S08.

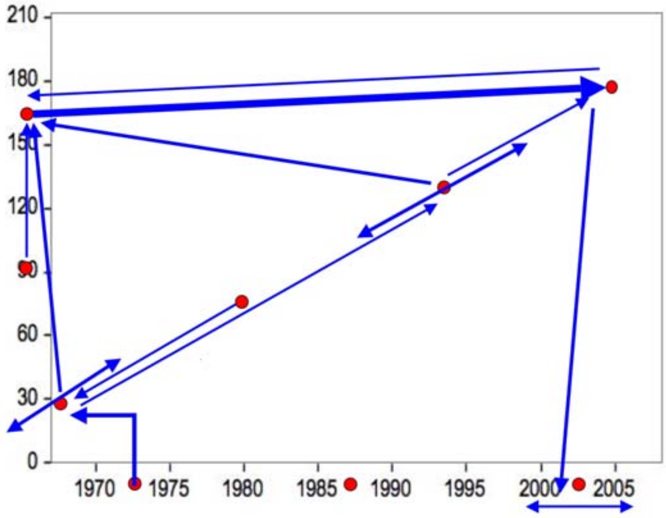


Figure 9.12. The fixation map after the stimulus sentence S08: 'The number of birds increased from about 30 to about 180.'

The fixation map obtained after participants' reading of S08 is similar to the fixation maps obtained for S05, S06, S07; it is a combination of the fixation maps presented above for S01 to S04, though with more inspections on the y-axis values compared to the x-axis values. The transition matrix for all gaze patterns is presented in Appendix D.2.

In summary, the most important finding obtained from the analysis of the four stimulus sentences, S05, S06, S07, and S08 is that the decompositions of the task specified by the stimulus sentences into subtasks exhibit a *combinatory* characteristic. The findings show that the basic gaze patterns obtained in the analysis of S01–S04 can be used to predict the gaze patterns in more complex cases, as stated in 9.2.2. In addition, the emphasis on certain entities over the others, which is specified by word order, led to differences between gaze patterns. For instance, the horizontal and vertical projections constituted the highest percentage in S05 (39%), a lower percentage in S06 (25%), and the lowest percentage in S07 (19%). A more detailed investigation of the influence of word order is left for a further study.

In the rest of this part of the chapter, the predictions of the task analysis approach presented in 9.2.3 are tested for the prepositional phrases 'since', 'until', and 'between'.

The stimulus sentence S09 is repeated as (11) below.

(11) a. *Kuş sayısı 1970 yılından beri yükselmiştir.*
The number of birds 1970 the year since increased.
'The number of birds increased since year 1970.'

Figure 9.13 shows the fixation map obtained from the recorded eye movements of the participants after their reading of the stimulus sentence S09.

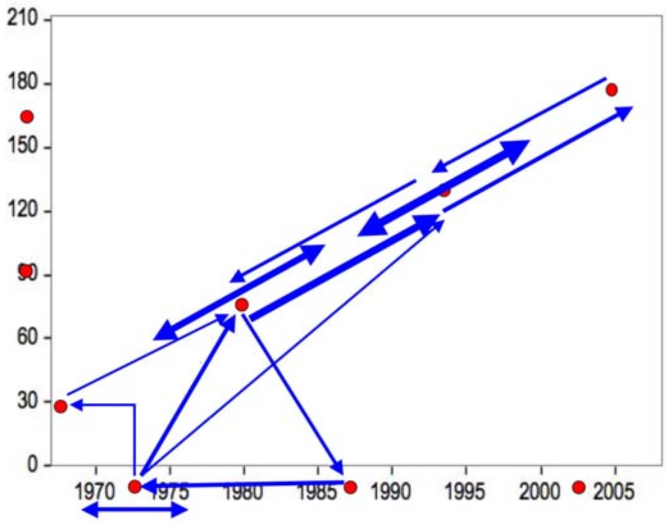


Figure 9.13. The fixation map after the stimulus sentence S09: 'The number of birds increased since year 1970.'

The obtained gaze patterns show the characteristics of the specification by 'since'. The gaze patterns along the graph line and the gaze patterns between the graph line and the x-axis, in particular the gaze patterns between the GL-AOI and the X1- and the X-AOIs, show that the 'since' PP resulted in a closer inspection of the graph line by the readers compared to the use of PPs 'from' and 'to'.[3] The projections between the value of the *TEMP(BEGIN(INCREASE))* argument on the x-axis and the GY1-AOI (which covers the starting point of the path and the corresponding y-axis value) are as expected (one of them shown in the figure). The transition matrix for the gaze patterns obtained after participants' reading of the stimulus sentence S09 is presented in Appendix D.2.

In the next step of the analysis, the stimulus sentence S10 is investigated, which is repeated below as (12).

---

[3] The prepositional phrase 'from' is not acceptable alone in a similar way to 'since', as exemplified in (i) below.

(i) *The number of birds increased from the year 1970.

Therefore, a comparison of the use of the lexical entities 'from' and 'since' is not possible. However, the available data from the analysis of the previous cases suggest that the closer inspection of the graph line in S09, compared to the cases where the 'from' and the 'to' PPs are used, is due to the use of 'since'.

(12) *Kuş sayısı 2005 yılına kadar yükselmiştir.*
The number of birds 2005 the year until increased.
'The number of birds increased until the year 2005.'

The fixation map obtained from participants' eye movement recordings is shown in Figure 9.14.

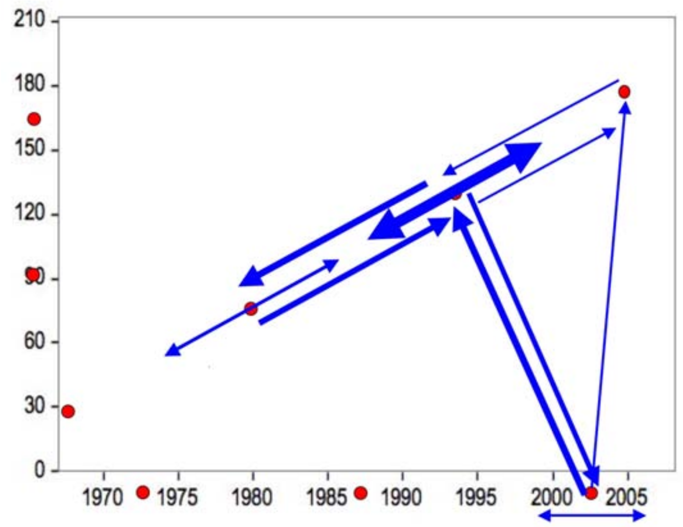


Figure 9.14. The fixation map after the stimulus sentence S10: 'The number of birds increased until the year 2005.'

The gaze pattern is similar to the one obtained for 'since' with the exception that the projections are shifted from the beginning of the increase to the end of the increase, as specified by the *TEMP(END(INCREASE))* argument. The gaze patterns on the graph line and the gaze patterns between the graph line and the axes, in particular the gaze patterns between the GR-AOI and the X2-AOI, are also shifted towards the right end of the graph line, i.e. the end point of the path. The transition matrix, presented in Appendix D.2, shows all the data obtained from the recordings.

The stimuli sentence S11 is repeated as (13) below.

(13) *Kuş sayısı 180 civarına kadar yükselmiştir.*
The number of birds 180 until increased.
'The number of birds increased until about 180.'

The fixation map obtained from participants' eye movement recordings is shown in Figure 9.15.

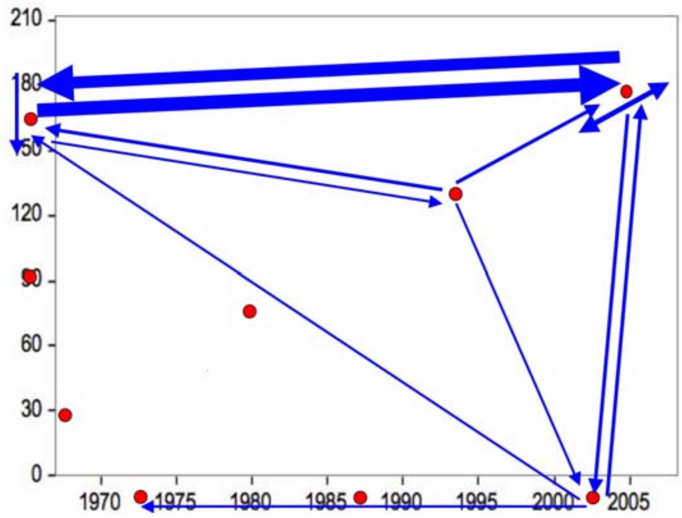


Figure 9.15. The fixation map after the stimulus sentence S11: 'The number of birds increased until about 180.'

The prepositional phrase 'until' in S11 does not place a temporal boundary to the process represented by the verb 'increase'. In other words, the 'until' PP in S11 does not modify aspectual characteristics of the verb; it is used in the sense 'so far as' rather than its temporal sense 'until'. Therefore, the 'until' PP is S11 has a similar function to the 'to' PP in S03/S04, rather than the 'until' PP in S10. Consequently, the fixation map that is obtained after participants' reading of S11 is similar to the fixation maps that are obtained after participants' reading of S03 and S04 (see Figure 9.7 and Figure 9.8 respectively) rather than to the fixation map that is obtained after participants' reading of S10 (see Figure 9.14).

In the last step of the test of the predictions that were specified in 9.2, the use of the PP 'between' is investigated. The stimulus sentence S12 is repeated as (14) below.

(14) *Kuş sayısı 1970 ve 2005 yılları arasında yükselmiştir.*
The number of birds 1970 and 2005 years between increased.
'The number of birds increased between the years 1970 and 2005.'

Figure 9.16 shows the fixation map obtained from the eye movement recordings of the participants, after their reading of the stimulus sentence S12.

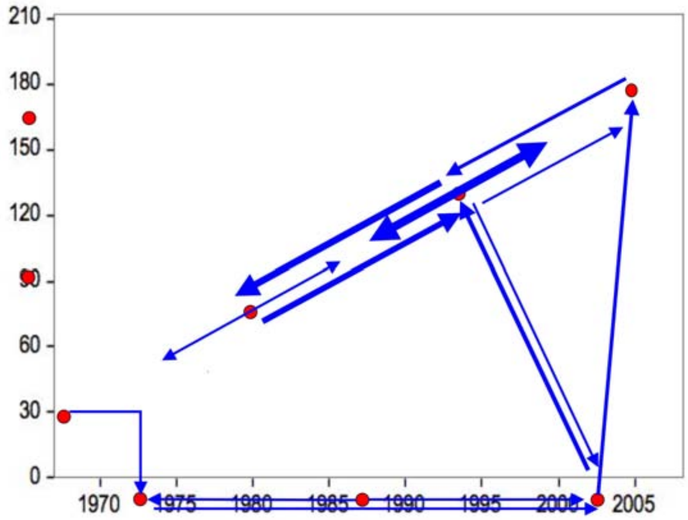


Figure 9.16. The fixation map after the stimulus sentence S12: 'The number of birds increased between the years 1970 and 2005.'

The obtained gaze patterns for S12 are close to the combination of the gaze patterns obtained for S09 and S10. The transition matrix, presented in Appendix D.2, shows all the gaze patterns including the ones that are not depicted in the fixation map.

In summary, the analyses of the participants' graph inspections after their reading of the stimuli sentences S09, S10, S11, and S12 show that the task analysis method presented in 9.2.3 can be used for predicting readers' graph inspections. In particular, the analysis of the stimulus sentences S09, S10, and S12 revealed the combinatory characteristic of decompositions of the task specified by the stimulus sentences. The PPs 'since' and 'until' exhibited different gaze patterns than 'to' and 'from': a closer inspection of the graph line was obtained after participants' reading of the stimulus sentences that involved 'since' and 'until' as PPs, compared to the reading of the sentences that involved 'from' and 'to'. In addition, the results showed that the obtained gaze patterns from the 'since' and the 'until' cases could be used to predict the gaze patterns after participants' reading of the sentences with a 'between' PP.

In the next part of the chapter, the model presented in this chapter is discussed within the framework of the previous relevant models in HCI and cognitive science.

9.4 EVALUATION OF THE MODEL WITHIN THE FRAMEWORK OF THE PREVIOUS MODELS IN HCI AND COGNITIVE SCIENCE

Several graph-based reasoning models have been proposed by researchers in the domain of HCI and cognitive science. Three of them have particular importance due to their relevance to the model presented in this chapter: Lohse's (1993) UCIE (Understanding Cognitive Information Engineering) model; the GBR (Graph-Based

Reasoning) model proposed by Peebles, Cheng, and Shadbolt (1999), and Peebles and Cheng (2001, 2002, 2003); and the GOMS-like model of perceptual task effort for bar charts proposed by Elzer, Green, and Carberry (2006).

Lohse (1993) proposed a simulation model that simulated how humans answer certain questions using different graph types, in particular line graphs and bar graphs, and tables. Lohse's model predicted response times of the users, based on the assumption that the users performed an optimal eye movement sequence, which minimizes the effort to reach a target location. Lohse's GOMS-based model used a set of cognitive engineering parameters for perceptual and cognitive operations for predicting the response times. The model of perceptual task effort for bar charts proposed by Elzer, Green, and Carberry (2006) analyzes the perceptual task effort of participants in terms a set of rules for estimating effort for specific perceptual tasks, which provides the basis for comparison between task completion times of users.

Peebles, Cheng, and Shadbolt (1999) and Peebles and Cheng (2001, 2002, 2003) reported experiments in which participants solved various problems with two different graph types: function graphs and parametric graphs. Peebles, Cheng, and Shadbolt (1999) explained the differences in response times and rated and patterns of errors between the two graph types in terms of the differences between the optimal visual scan paths the participants follow through the graph. As stated by Peebles and Cheng (2001) these optimality assumptions serve as an approximation to the experimental data. Peebles and Cheng (2001, 2002, 2003) showed that within the framework of the ACT-R/PM cognitive architecture (Byrne & Anderson, 1998) the optimality assumptions, which comprised the basis of their task analysis, approximated the detailed sequences of saccades made by the users.

The models summarized above made predictions about response times of the users during their problem solving with different types of graphs. On the other hand, a systematic analysis of the task specification, which is based on the use of particular linguistic entities, was not in the focus of these studies. Therefore, the model presented in this chapter is unique for the detailed analysis of the structural aspects of the presented sentences, which specified the tasks of graph comprehension.

The model presented in this chapter focused on the prediction of eye movement patterns rather than response times since the variability in response times of the participants was higher compared to the variability in gaze patterns. A GOMS-based cognitive model was developed for the prediction of response times by the author of the dissertation, based on a set of basic components tasks for perceptual, cognitive, and motor processes which assume fixed processing times for specific tasks, such as 1200 ms for comparing to units in memory, 70 ms for executing a mental step, and so on. Nevertheless, the high variability in the response time values resulted in inefficient models with low predictive power. Therefore, the model presented in this chapter was based on the gaze patterns of the participants rather than the response times and the analysis of the response times is left for a further study (see Appendix D.3 for the mean response times obtained in the experimental investigation).

9.5 SUMMARY

In Chapter 4, a cognitive architecture for multimodal comprehension of graph-text constellations was introduced. The main purpose of the cognitive architecture was to present the structural properties of the basic components in the modality-specific modules and their relationship to form integrated conceptual representations. This chapter presented a cognitive model that investigated a specific set of information processing tasks, as a specific instance of the architecture introduced in Chapter 4. In particular, the model showed that the gaze patterns of users on graphs could be predicted by decomposition of the tasks specified by the sentences into subtasks. The model revealed the combinatory characteristic of multimodal integration by showing that gaze patterns on graphs after the users' reading of complex sentences could be predicted by combining the gaze patterns already obtained after their reading of simpler sentences. The findings suggest that readers' gaze patterns on graphs can also be used for predicting the sentences inspected by the readers before the inspection of the graphs. The findings also showed that certain types of prepositional phrases, such as 'since', 'until', and 'between' in sentences result in different gaze pattern characteristics on the graphs, compared to other types of PPs such as 'from' and 'to'.

10 Design Principles and Guidelines for Graph-Text Constellations

The cognitive architecture and the cognitive model for multimodal comprehension of graph-text constellations and the experimental investigations reported in the previous chapters contribute to the study of the interaction between graphical and linguistic entities in graph-text constellations by revealing systematic patterns for structural and processing aspects of multimodal comprehension. The research presented in those chapters can be described as *basic research in applied situations* on graph-text constellations. The term "basic research in applied situations", inspired by Stokes' (1997, p. 73) "use-inspired basic research", has been used by Mayer (2009, p. 25; 2008) within the context of multimodal learning (i.e., *multimedia learning* in Mayer's terminology).

As stated by Mayer (2008, 2009), the goal of basic research is to contribute to theory, whereas the goal of applied research is to contribute to practice. In the domain of multimodal learning, learning theory is the hallmark for basic research, and the science of instruction is the hallmark of applied research. The basic research in applied situations then contributes to both learning theory and the science of instructional design. In particular, the science of instruction aims to contribute to "evidence-based practice – instructional practices that are consistent with research evidence" (Mayer, 2009, p. 29). The term *evidence-based practice* describes the overlapping goals of basic research and applied research. In the domain of multimodal learning, the development of guidelines and principles for designing effective multimodal instruction exemplifies the contribution of instruction science to evidence-based practice.

From the perspective of HCI and cognitive science, the term *evidence-based practice* is the blueprint of the relationship between the experimental investigations on and the practical issues concerning the design of graph-text constellations. Poorly designed graphs may lead to conceptualizations that are not compatible with conceptualizations of textual entities in graph-text constellations, which in turn result in processing difficulties during the course of multimodal comprehension. Based on the findings of the experimental investigations reported in the previous chapters, the initial goal of this chapter is to devise a set of principles and guidelines for designing graph-text constellations. The second goal is to present the implications of the guidelines for the development of intelligent artificial systems that have the capacity to support users during their course of designing and producing graph-text constellations. Therefore, the approach presented in this chapter can be characterized as evidence-based practice about the applied problems of intelligent artificial systems that present graph-text constellations to users. The design principles and guidelines presented in this chapter present a complementary contribution to the basic research in applied situations on graph-text constellations presented in the previous chapters of the dissertation.

In 10.1, the origins of design principles and guidelines for the design of human-computer interfaces are introduced. In 10.2, the guidelines for designing graph-text constellations, which were distributively presented under the discussions of the experimental findings in Section 2 of the dissertation, are brought together under the framework of two fundamental principles of comprehension of graph-text constellations. In 10.3, the implications of the guidelines are presented for the development of software tools that support users during their course designing and producing graph-text constellations.

10.1 PRINCIPLES AND GUIDELINES FOR USER INTERFACE DESIGN

As stated in Chapter 3, where the research methods in HCI and cognitive science were introduced, various usability inspection methods were proposed as low-cost methods for testing usability of human-computer interfaces in the 1980s. Design guidelines and principles facilitate usability inspection by providing the basis for usability evaluation.

The term *design guideline* is used for recommendations of good practice (Stewart & Travis, 2003). Design guidelines "prescribe good practices and caution against dangers" (Shneiderman & Plaisant, 2010, p. 74). The term *design principle* is generally used for abstract frameworks that specific design guidelines can be generated (Dix, Finlay, Abowd, & Beale, 2004). Design principles are used to analyze and compare design alternatives. Both design guidelines and design principles are derived from theoretical frameworks and from the findings of empirical investigations in relevant disciplines (e.g., cognitive and educational psychology). However, compared to design principles, design guidelines have a more suggestive, practical, and specific nature.

Among many others, Shneiderman's (1987) user interface guidelines, Nielsen's (1994) usability heuristics, and Norman's (1990) principles for transforming difficult tasks into simple ones have been the most influential design guidelines and principles for designing interactive user software in the domain of HCI. In addition to those broad-level guidelines for designing user software interface, comprehensive guidelines and principles were proposed for the more specific domain of graph design (Washburne, 1927; Tufte, 1983, 1990; Kosslyn, 1989, 1994b, 2006; Renshaw, Finlay, Tyfa, & Ward, 2004, among many others), and multimedia learning (Mayer, 2001/2009; 2005b).[1]

As an example of the concept of design guideline, consider one in a set of "general rules for the appropriate use of the various graphic, tabular, and textual forms" proposed by Washburne (1927, p. 375), which has been previously stated in Chapter 2 of the dissertation. Washburne suggests that "never present numerical data in textual (paragraph) form if there are more than one or two items to be presented" (p. 376). These guidelines suggested by Washburne are also used today. For instance, the APA Publication Manual (5th ed., 2001) offers the following as a method of presentation of data: "if you have 3 or fewer numbers, use a sentence; if you have from 4 to 20 numbers, use a table; and if you have more than 20 numbers, consider using a graph or figure instead of a table" (p.137). The suggestions made by Washburne and by the APA manual exemplify basic guidelines for presenting data in alternative representational modalities.

As an example of the concept of design principle, Mayer's (2009) twelve principles of multimedia design are presented below. These principles are based on the findings of Mayer's (2001/2009) theoretical framework for multimedia learning (viz., cognitive theory of multimedia learning) and the findings obtained in experimental investigations.

Principles for reducing extraneous processing

1. Coherence principle: People learn better when extraneous words, pictures, and sounds are excluded rather than included.
2. Signaling principle: People learn better when cues that highlight the organization of the essential material are added.
3. Redundancy principle: People learn better from graphics and narration than from graphics, narration, and on-screen text.
4. Spatial contiguity principle: People learn better when corresponding words and pictures are presented near rather than far from each other on the page or screen.
5. Temporal contiguity principle: People learn better when corresponding words and pictures are presented simultaneously rather than successively.

Principles of managing essential processing

6. Segmenting principle: People learn better when a multimedia lesson is presented in user-paced segments rather than as a continuous unit.
7. Pre-training principle: People learn better from a multimedia lesson when they know the names and characteristics of the main concepts.

[1] Mayer and colleagues use the term *multimedia* for constellations of pictorial illustrations and language, without specific emphasis on modality-specific differences between types of depictive representations (see Chapter 2 for the types of depictive representations).

8. Modality principle: People learn better from graphics and narration than from animation and on-screen text.

Principles for fostering generative processing

9. Multimedia principle: People learn better from words and pictures than from words alone.
10. Personalization principle: People learn better from multimedia lessons when words are in conversational style rather than formal style.
11. Voice principle: People learn better when the narration in multimedia lessons is spoken in a friendly human voice rather than a machine voice.
12. Image principle: People do not necessarily learn better from a multimedia lesson when the speaker's image is added to the screen. (Mayer, 2009, p. 267–268)

A closer look to Mayer's (2009) twelve principles of multimedia learning reveals that they are descriptive and open-ended statements. Guidelines, on the other hand, often are presented in instructive mood rather than descriptive statements.[2] Despite the difference between the two, guidelines and principles share the characteristic of practical flexibility. For instance, Washburne (1927) and the APA Publication Manual make contradictory statements about the appropriate number of numerals in text. The producer of the text may select one or the other, or he/she may decide to use not two or three but four numerals in text depending on the context (e.g., the requirements of the design task).

Mayer (2009) states that there is a set of conditions under which the guidelines and principles apply (namely, "boundary conditions", p. 269) and proposes a set of specific boundary conditions for the principles of multimedia learning, such as individual characteristics of the learners and the complexity and pace of the presentation. The specification of boundary conditions is a necessary step to analyze the applicability of guidelines and principles. On the other hand, the specification of boundary conditions should also have a flexible nature, since the diversity of individual characteristics and the complexity of a representation cannot be easily investigated in terms of such higher-level theoretical constructs.

This flexibility of design guidelines, principles, and boundary conditions is a peculiar characteristic of user interface software design in HCI. The tasks and requirements of software interfaces and the individual characteristics of users are highly variable and context dependent. Therefore, the experimental evidence and theories in user interface design are organized into suggestive design guidelines and principles, and in

[2] This difference between principles and guidelines reflects a general observation rather than a clear distinction between the two; the instructive mood is used in design principles as well as guidelines. In addition, the terms *guidelines*, *principles*, and *rules* are used interchangeably in the HCI literature.

some cases into predictive cognitive models (Shneiderman & Plaisant, 2010), rather than authoritative design standards.[3]

Critics often criticize principles for being a "failure to replicate" (Mayer, 2009, p. 273) and guidelines for being "too specific, incomplete, hard to apply, and sometimes wrong" (Shneiderman & Plaisant, 2010, p. 75). Nevertheless, on the positive side, proponents state that design guidelines and principles contribute to steady improvements (Shneiderman & Plaisant, 2010) and maturing in the specific fields of research will clarify the conditions under which the guidelines and principles can be applied (Mayer, 2009).

In 10.2, a set of design principles and guidelines is presented for graph-text constellations, in connection to the relevant principles and guidelines suggested in the relevant research.

10.2 DESIGN PRINCIPLES AND GUIDELINES FOR GRAPH-TEXT CONSTELLATIONS

Graph-text constellations are widely used in communicative settings of daily life such as in newspaper articles, as well as in scientific and educational settings, such as in scientific articles, textbooks, and classroom presentations, both in printed form and in electronic media. On the other hand, research on multimodal comprehension of graph-text constellations is scant. Therefore, a prescription of good practices and caution against dangers, in other words, design guidelines, and design principles for graph-text constellations are missing.

The experimental investigations reported in Section 2 of the dissertation revealed a set of findings out of which design guidelines were devised and distributively reported in the corresponding chapters. These design guidelines are presented under the more general design principles for graph-text constellations below. In 10.2.1, the fundamental principle of multimodal graph-text comprehension is presented. In 10.2.2, two guidelines are presented for the design of graph-text constellations that have a separate graph-layer – text-layer layout. In 10.2.3, the fundamental principle of comprehension of verbally annotated graphs in graph-text constellations is presented. In 10.2.4, three guidelines are presented for the design of verbally annotated graphs for graph-text constellations.

---

[3] The term *design standard* is used for formal documents that are published by institutions that are committed to prepare and publish standards, such as Deutsches Institüt für Normung (DIN) and International Organization for Standardization (ISO). Design standards are usually used for computer hardware ergonomics rather than user interface software design. In industrial ergonomics, for instance, it is possible and necessary to produce design standards, such as the application-specific standards for aircraft avionics.

10.2.1 The Fundamental Principle of Multimodal Comprehension of Graph-Text Constellations

The fundamental principle of multimodal comprehension of graph-text constellations makes the following statement.

Principle 1. Users recall better from graph-text constellations rather than from text alone.

Two experimental investigations made a comparison of participants' recall in graph-text constellations and text-only stimuli: the Study 1 in Chapter 7 and the Study 1 in Chapter 8. In Chapter 7, an overall statistically significant effect was obtained in favor of the principle; in Chapter 8, partial support was obtained.

The fundamental principle of multimodal comprehension of graph-text constellations is compatible with Mayer's (2001/2009) multimedia principle, which states that "People learn better from words and pictures than from words alone" (Mayer, 2009, p. 223). It is also compatible with the principles of Universal Design proposed by The Center for Universal Design (1997), reported in Story, Mueller, and Mace (1998). In particular, it is compatible with the principle of perceptible information, which states that "The design communicates necessary information effectively to the user, regardless of ambient conditions or the user's sensory abilities" (Story, et al., 1998, p. 61) and the guideline "Use different modes (pictorial, verbal, tactile) for redundant presentation of essential information" (p. 62).[4] In 10.2.2, two specific guidelines are presented for designing graph-text constellations with a separate graph-layer – text-layer layout.

10.2.2 Design Guidelines for Graph-Text Constellations with a Separate Graph-Layer—Text-Layer Layout

The findings of the experimental studies reported in Chapter 6 show that users spend high cognitive effort for integrating information induced by textual and graphical entities in graph-text constellations if the graph-related sentences in the text layer of the constellation have a complex structure, such as coordinate constructions that involve a series of conjoined phrases or clauses. This suggests that using simpler sentences in the text-layer of a graph-text constellation could facilitate the integration.

[4] Although the principle of perceptual information focuses on user's sensory abilities, the specific guideline that is presented within the context of the principle in Story et al. (1998) uses the term *mode* and exemplifies it with the terms *pictorial* and *verbal*, which are considered representational modalities in this dissertation.

These findings lead to the following guideline for designing graph-text constellations.

Guideline 1. Transform complex graph-related sentences into simple ones.

In Chapter 9, where a GOMS-style processing model of comprehension of graph-text constellations is presented, it was shown that comprehension processes during integration can be investigated in terms of the tasks and subtasks induced by the sentences in graph-text constellations. Therefore, Guideline 1 is compatible with Norman's (1990) seven principles for transforming difficult tasks into simple ones, in particular the specific principle which states that "simplify the structure of tasks" to minimize the users' effort for planning or problem-solving.

A second guideline for designing graph-text constellations with a separate text-layer – graph-layer layout is based on the experimental findings reported in Chapter 8. Briefly, the findings show that the concept of scale can accessed by entities in different modalities. Therefore, incongruency between the representation of scale by linguistic entities (e.g., a gradable adjective or a gradable adverb) and the representation of scale by graphical entities, i.e. the spatial configurations of graphical entities, results in high cognitive effort of participants during the course of integration of the information induced by the two modalities. The findings obtained in Chapter 8 lead to the following design guideline.

Guideline 2. Check the congruency between scalar entities in the text and spatial configuration of graphical entities in the graph.

In particular, the interpretation of scalar adjectives may interact with the position of the graphical entities in the Cartesian graph space and the interpretation of scalar adverbs may interact with the steepness of graph line. As an example, consider the graph-text constellation in Figure 10.1, which was used as experimental stimulus in the Study 1 of Chapter 8. The text layer of the graph-text constellation involves the sentence 'Auckland has rainy climate', which is called a *target sentence* in Chapter 8, and the graph depicts the annual amount of rainfall in Auckland.

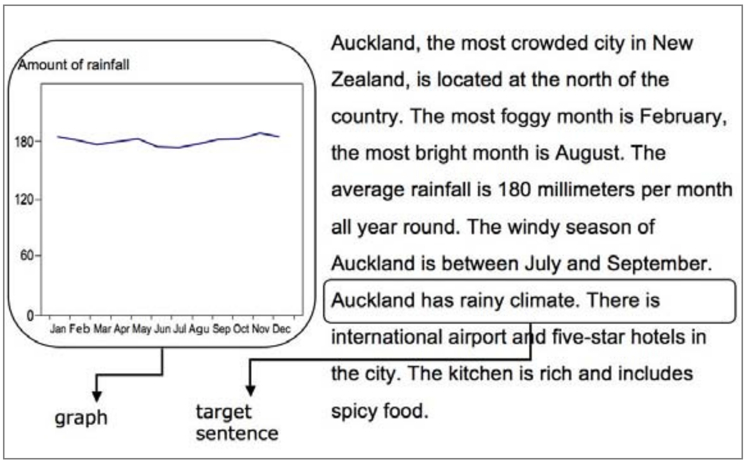


Figure 10.1. Sample stimuli from the Study 1 of Chapter 8. The rectangular regions show the specification of the two Areas of Interest (AOIs) for the analysis of eye movement recordings, which is considered an indication of cognitive effort; the rectangles were not shown to the participants.

The two graphs in the graph-text constellations presented in Figure 10.2 represent the same numerical information, i.e. the average rainfall amount of 180 millimeters per month, though with different locations with respect to the graph frame.

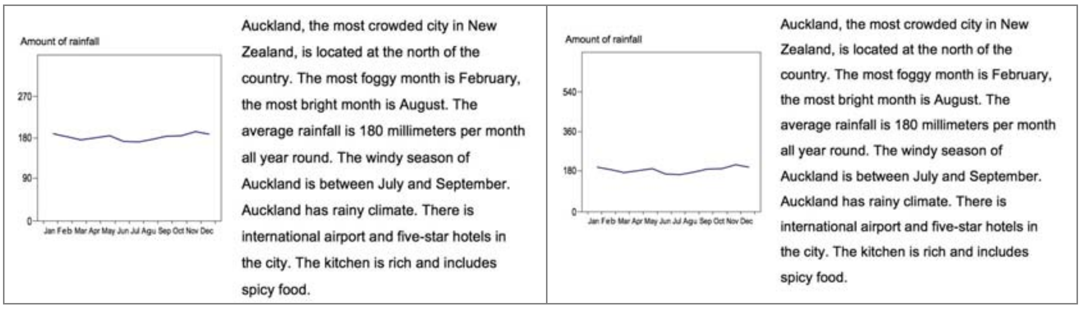


Figure 10.2. Two alternative presentations of the same numerical information about rainfall by the graphs.

The results of the experimental studies that are reported in Chapter 8 show that users spend higher cognitive effort to interpret the relevant sentences in the graph-text constellations if the scalar adjectives such as *rainy* are used together with a graph line that is located in the lower portion of the graph space in a graph-text constellation. The Guideline 2 is applicable to scalar adverbs in the sense that the interpretations of scalar adverbs interact with the steepness of the graph line segments.

Those poorly designed graph-text constellations, which cause processing difficulties in multimodal comprehension of graphical and linguistic entities, are often used in printed or online publications that are designed and produced by non-expert users. In the recent state of the art, the available software tools do not have the capacity to provide users with support during their design and production of graph-text constellations. The principles and guidelines that are presented in this chapter provide computer scientists with implications for the design of user-support systems for graph-text constellations, as discussed in 8.3. Before the discussion of the implications, design principles and guidelines for verbally annotated graphs are presented below. In 10.2.3, fundamental principle of comprehension of verbally annotated graphs in graph-text constellations is introduced and three specific guidelines for designing annotated graphs are presented (cf. graph-internal multimodality, see Chapter 1).

10.2.3 The Fundamental Principle of Comprehension of Verbally Annotated Graphs in Graph-Text Constellations

As stated in Chapter 7, written verbal annotations in graph-text constellations have a similar function to annotating graphical entities by gestures during speaking. Annotations have a similar role to those referential entities used in communication such as gestures and expressions with deictic reference. Comprehension of purely referential deictic linguistic expressions requires “going out of the language faculty and making use of the visual system” (Jackendoff, 2002, p. 307). Within the framework of the cognitive architecture presented in Chapter 4, the common underlying correspondence to conceptual entities allows gestures to accompany linguistic expressions with deictic reference, since those conceptual entities that are referred to by linguistic expressions can also be referred to by gestures. In particular, a *visual percept*, which is linked to the deictic expression via the SpS-CS interface, is the necessary construction for the act of annotating in spoken language settings. In verbally annotated graphs, graphical entities comprise a specific subcase of a visual percept. The processes and events represented by graphical entities interact with processes and events represented by verbal annotations, thus leading to a coherent interpretation of verbally annotated graphs.

This type of annotating graphical entities in graph-text constellations is frequently used in asynchronous communication settings, such as newspaper articles in print media and electronic media (Kong & Agrawala, 2009). In Chapter 7, three experimental studies that investigated verbally annotated graphs were presented. In particular, the Study 1 investigated the role of verbal annotations in graph-text constellations in multimodal comprehension of graph-text constellations. The findings of the study showed that the participants spent less effort, i.e. they spent less time for inspecting the graph, to integrate the information induced by the textual entities and graphical entities when the graphs were annotated compared to the cases when the graphs were not annotated. This finding indicates that annotations may play a bridging role between the textual entities in the text layer of the graph-text constellation and the relevant parts of graphical entities in the graph layer of the graph-text constellation, which leads to the following fundamental principle of comprehension of verbally annotated graphs in graph-text constellations.

Principle 2. Verbal annotations in graphs facilitate multimodal comprehension of graph-text constellations.

This principle is compatible with Mayer's (2001/2009) spatial contiguity principle, which states that "people learn better when corresponding words and pictures are presented near rather than far from each other on the page or screen" (Mayer, 2009, p. 267). Mayer's spatial contiguity principle has been devised from the findings from experimental studies with picture-text constellations in the domain of learning and instructional science. However, the findings presented in Chapter 7 show that the principle can be extended to graph-text constellations with annotated graphs.

This principle is also compatible with the eight golden rules of interface design suggested by Shneiderman and Plaisant (2010), particularly the principle "help preventing errors" (p. 88) in the sense that annotations reduce search effort of users. In addition, the guidelines for "getting the user's attention" (Shneiderman & Plaisant, 2010, p. 78) and the signaling principle proposed by Mayer (2009) which states that "People learn better when cues that highlight the organization of the essential material are added" (p. 109).[5] Three specific design guidelines, based on the experimental findings reported in Chapter 7, are presented below.

10.2.4 Design Guidelines for Verbally Annotated Graphs in Graph-Text Constellations

A line graph that shows the changes in a domain value over time is one of the rare means to convey dynamical information (i.e., change of a value in time) statically on paper or on screen. The Study 2 and the Study 3 reported in Chapter 7 investigated the interaction between events and processes represented by verbal annotations (shortly, event annotations) and the events and processes represented by graph lines, in annotated graphs. The results in those studies showed that comprehension of graphical entities leads to process conceptualizations and event conceptualizations

[5] A further finding in the Study 1 reported in Chapter 7 was that the users had reduced recall of the text when they were presented annotated graphs in graph-text constellations, compared to non-annotated graphs. This finding shows that annotations may attract users' attention (as evidenced by more frequent paragraphs-to-graph gaze shifts in annotated graphs compared to non-annotated graphs), which, in turn, results in a split attention effect (Chandler & Sweller, 1992). This indicates the need for further research for the specification of *boundary conditions* in which the principle applies, such as the role of graph complexity, individual characteristics of users, and the conditions under which the cognitive load introduced by annotations leads to *extraneous*, *essential*, and *generative processing* (see Mayer, 2009, for the terms in italics and the research on the specification of boundary conditions for the principles of multimedia learning).

that interact with comprehension of verbal annotations. This is exemplified by the sample stimuli used in Study 3, which are shown in Figure 10.3.

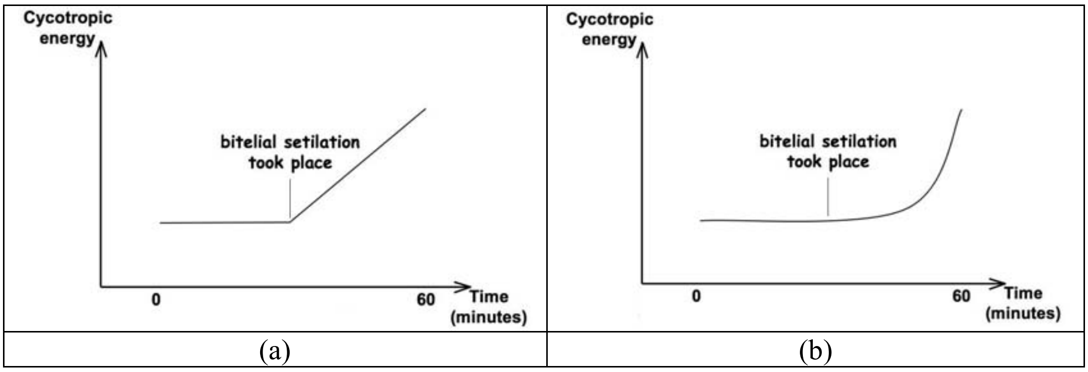
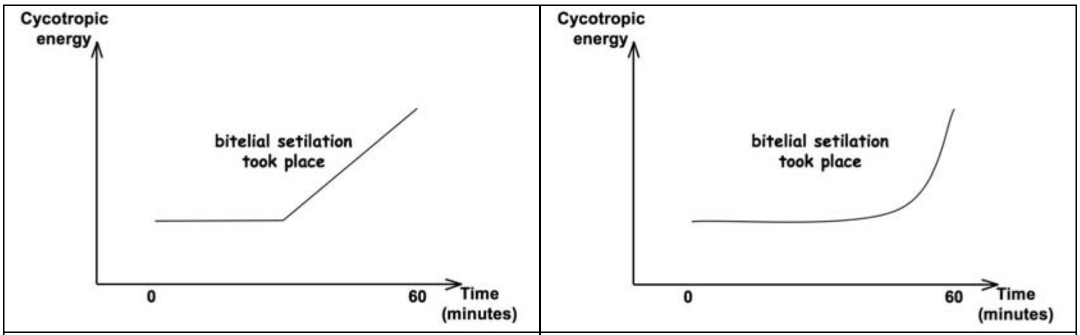


Figure 10.3. Sample stimuli from the Study 3 reported in Chapter 7 (translated into English by the author of the dissertation; the font sizes in the figure were enlarged for better visibility).

The participants were asked to report their judgments about temporal aspects of the annotation events, which were represented by the fictitious phrases such as 'bitelial setilation took place'. The results of the study showed that, when they reported their judgments for the stimuli as exemplified in Figure 10.3a, the participants stated that the annotation event was a point-like event in time. On the other hand, when they reported their judgments for the stimuli as exemplified in Figure 10.3b, most of the participants reported that the annotation event was a durative event. These findings reveal that comprehension of annotation events and the events and processes represented by graphical entities are interpreted to reach a coherent whole.

The results of the Study 3 revealed a further finding: The presence of an annotation icon, i.e. the graphical entities between the annotation text and the relevant part of the graph such as the thin vertical line shown in Figure 10.4, did not reveal changes in the subjective judgments of the participants about the temporal aspects of the annotation events compared to the cases where an annotation icon was absent. However, the participants spent higher cognitive effort to interpret the relationship between the annotation event and visually/informationally salient parts of the graph when an annotation icon was absent compared to the cases where an annotation icon was present. In particular, in the graph presented in Figure 10.4b, the two visually salient regions were (i) the region where the curvature was high, and (ii) the inflection point, which was below the middle of the annotation text.[6]

[6] The similarity of the participants' judgments about temporal aspects of the annotation events between the annotated and non-annotated graphs points to the need for the development of graphical representations for annotation icons (as alternatives to the standard straight-line annotation icon) that reduce the ambiguities about temporal aspect of the annotation events.



(a) (b)

Figure 10.4. Sample annotated graphs without annotation icons, from the Study 3 reported in Chapter 7 (translated into English by the author of the dissertation; the font sizes in the figure were enlarged for better visibility).

This finding suggests the following guideline for designing verbally annotated graphs in graph-text constellations.

Guideline 3. Use annotation icon to facilitate the construction of the relationship between annotation and the relevant part of the graphical entities.

In the Study 2 that is reported in Chapter 7, the participants were asked to report their judgments for possible causal relationships between annotation events and the events and processes represented by the graph lines. In particular, the Study 2 investigated the role of annotation position and the shape of graph lines on causal attributions concerning the relation between annotation events (i.e. events presented by the annotation text) and the processes and states represented by process-lines (increases / decreases) or state-lines (no-change). Briefly, the results of the study show that different positions of verbal annotations on graph lines result in systematically different patterns in judgments for causal attributions. For instance, the use of annotations as shown in Figure 10.5 resulted in ambiguous interpretations about causal relation between the annotation event the increase or the no-change represented by the increase process-line and the no-change state-line.

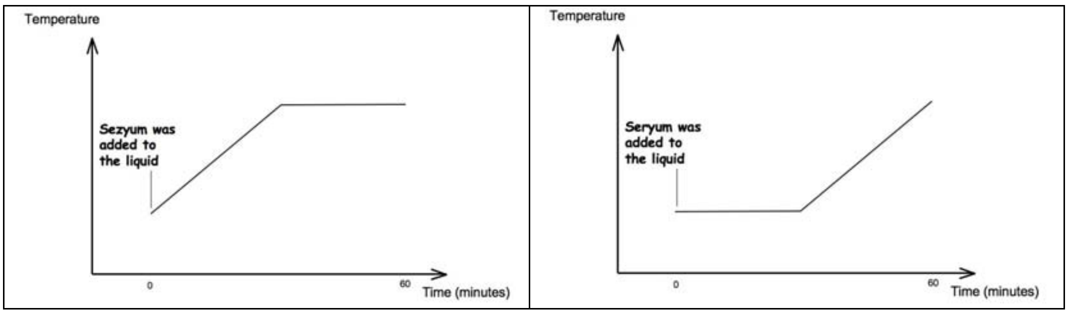


Figure 10.5. Sample annotated graphs from the Study 2 reported in Chapter 7 (translated into English by the author of the dissertation; the font sizes in the figure were enlarged for better visibility).

On the other hand, adding a second annotation resolved those ambiguities about the causal relation.

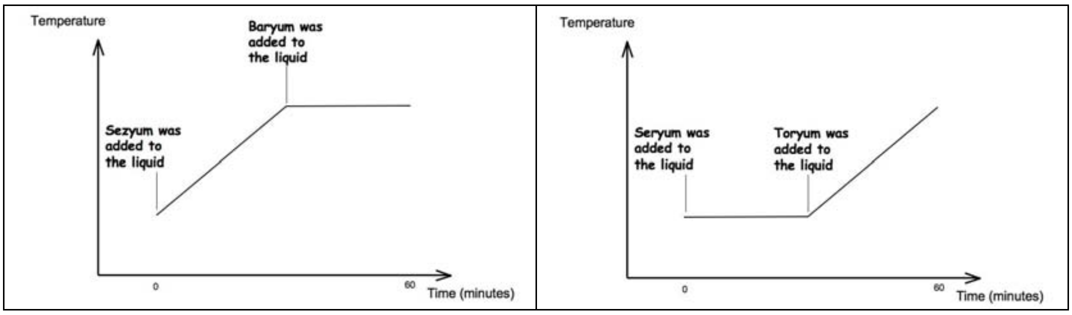


Figure 10.6. Sample annotated graphs from the Study 2 reported in Chapter 7 (translated into English by the author of the dissertation; the font sizes in the figure were enlarged for better visibility).

The findings of the Study 2 lead to the following guideline for designing verbally annotated graphs in graph-text constellations.

Guideline 4. Select the location of the annotation appropriately; use more than one annotation where necessary.

The verbally annotated graphs shown in Figure 10.5 and 10.6 involve graph lines with corners. In the Study 2 and the Study 3 of Chapter 7, graphs with smooth graph lines were used as well as graphs the graphs with graph lines that have corners (as exemplified in Figure 10.4b). As briefly presented above, the findings of the Study 3 show that the smoothness of the graph line interact with the interpretation of the temporally ambiguous annotation events. A further finding obtained in the Study 2 was that the graphs with smooth graph lines result in different causal attributions between annotated events and events and processes represented by graph lines, compared to the causal attributions in the graphs that have graph lines with a corner. These findings lead to the following guideline for designing verbally annotated graphs in graph-text constellations.

Guideline 5. Use graph-line smoothing cautiously in annotated graphs.

Most of the available software tools for graph design allow the options for smoothing the graph lines. On the other hand, the smoothing options should be cautiously used in annotated graphs, particularly in annotated graphs with event annotations.

The following part of the chapter introduces implications of the presented guidelines for the development of software tools for graph-text constellations.

10.3 IMPLICATIONS OF THE GUIDELINES FOR THE DEVELOPMENT OF SOFTWARE TOOLS FOR GRAPH-TEXT CONSTELLATIONS

The presented principles and guidelines for designing graph-text constellations provide a basis for appropriate design and good practice for multimodal graph-text comprehension. Based on the experimental findings that were reported in Section 2 of the dissertation, they provide an initial set of clues for designing graph-text constellations with a separate graph-layer–text-layer layout and graph-text constellations with annotated graphs (cf. graph internal multimodality), both in printed form and in online media (see Chapter 1 for graph-text constellations with a separate layout of text and graph and for graph-internal multimodality).

From the perspective of the user, the recent developments in user interface technology enabled people to interactively use software programs to design and generate graphs. The recent user interface technology provides users with the opportunity to design and produce graphs and graph-text constellations in myriad of formats.

From the perspective of computer science, the development of hypertext has provided opportunities for the development of assistance systems for users. For instance, most of the available word processing software tools now have spell checking and grammar checking capabilities that have been showing a steady progress over time. Software tools for annotating documents in collaborative writing have been developed and tested for usability (e.g., Bargeron & Moscovich, 2003; Brush, Bargeron, Gupta, & Cadiz, 2001; Zheng, Booth, & McGrenere, 2006). Nevertheless, the development of assistance systems for design, production, and checking of multimodal constellations has not yet fully progressed to reach the designer (i.e., producer) as the end-user. Despite the significant progress in information visuzalition, the support for visualization-based communication has been limited (Viégas and Wattenberg, 2006).

As discussed in Section 1 of the dissertation, graphs comprise a distinct class of depictive representation. The grammatical nature of graph-text constellations presents advantages over other types of multimodal constellations, such as picture-text constellations, for the development of assistance systems for users. These advantages are recently recognized in the domain of HCI. For instance, Kong and Agrawala (2009) present a set of techniques for interpreting freeform ink annotations in line graphs, based on the identification of perceptual parts (e.g., peaks, humps, cf. Hoffman & Singh, 1997) corresponding to each annotation.[7] As stated by Heer and Agrawala (2008), determining the *appropriate design decisions* and *technical*

[7] The term *freeform ink annotation* is used for graphical means for pointing by annotations (Heer, Viégas, & Wattenberg, 2009).

mechanisms is a challenge for enabling effective asynchronous collaboration in visual analysis environments. Accordingly, the guidelines for designing graph-text constellations presented in 10.2 can provide a primary basis for the development of assistance systems that are similar to spell checking and grammar checking facilities in the available word processing software tools. The implications of the presented guidelines for the development of graph-text software tools are presented below.

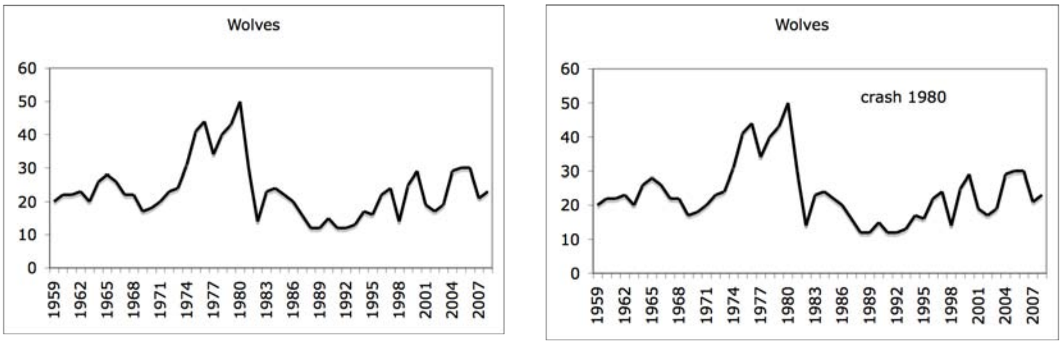
Consider the Guideline 2, repeated below, for the development of an assistance facility for graph-text constellations with a separate text-layer–graph-layer layout.

Guideline 2. Check the congruency between scalar entities in the text and spatial configuration of graphical entities in the graph.

This guideline can be used for the development of an assistance facility, which checks the congruence between the textual entities in the text layer and the graphical entities in the graph layer. Technically, such an assistance facility for congruence checking for scalar adjective basically does the following: check the text layer of a graph-text constellation for the availability of scalar adjectives and classify the available ones into *positive* and *negative* scalar adjectives (e.g., rainy vs. non-rainy); check the location of the graph line with respect to the horizontal axis of the graph, and classify the graph line into *high* and *low* values (i.e., the graph line in the upper part of the graph space vs. the graph line in the lower part of the graph frame); check the congruency between the text and the graph by comparing the obtained pair with the appropriate ones (i.e., a positive adjective – high graph pair or a negative adjective – low graph pair).

As stated in 10.1, graph-text constellations often involve verbally annotated graphs, based on the fact that the act of pointing is a characteristic behavior of humans in multimodal communication settings. Verbal annotations are widespread in both printed form and in hypertext. Nevertheless, the recently available software tools have limited capacity for annotation design. It was stated in Chapter 2 that popular software programs such as spreadsheet tools as well as scientific software tools for expert use, such as numerical and technical computing software have graph producing tools with a wide range of options for graph types and formatting. On the other hand, most of these software tools offer limited capabilities for design of verbal annotations for graphs.

For instance, one of the most popular spreadsheet software tool that is used to produce graphs by non-expert people, Microsoft Excel® does not have an explicit annotation generation tool in its recent version (Microsoft® Excel® 2004 for Mac, version 11.5.4). On the other hand, graphs that are generated by Excel can be indirectly annotated by the 'data label' tool; once the data label is created, the content of the label can be changed by the user, thus producing an indirect solution for the annotation problem. These two steps are exemplified in the graphs shown in Figure 10.7. The graphs show the wolf population in the Isle Royale between the years 1959 and 2008.[8] The graph producer might want to annotate the sudden decline in the 1980s by the annotation text 'crash 1980'. Accordingly, he/she produces the annotation text by using the above-described method.



(a) (b)

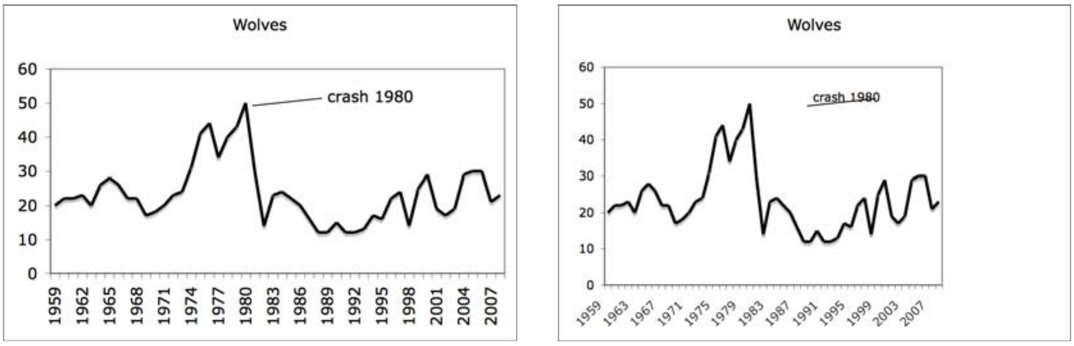
Figure 10.7. (a) Sample wolf population progression graph, (b) an annotation produced to show the wolf crash in 1980.

Moreover, following the Guideline 3, which is repeated below, the graph producer may want to create an explicit pointer that relates the text to the data points represented by the graph (viz., annotation icon).

Guideline 3. Use annotation icon to facilitate the construction of the relationship between annotation and the relevant part of the graphical entities.

However, there is no explicit way provided by the tool; therefore, an annotation icon has to be drawn separately in the exemplified software package. The graph producer uses the drawing tool to draw a straight line between the annotation text and the graph line. Then he/she decides to add some more textual information in the graph frame and resizes the frame in which the graph line is placed. However, the resulting annotation looses the appropriate annotating function, as shown in Figure 10.8.

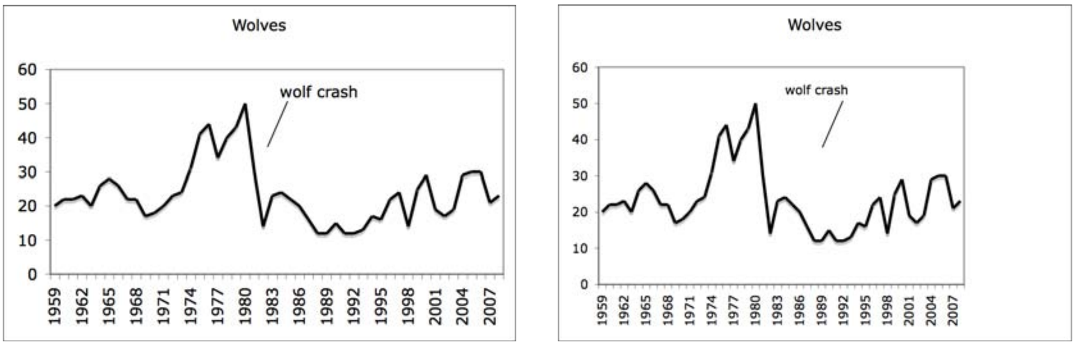
[8] The graphs are based on the source data received from the Isle Royal Project web page, http://www.isleroyalewolf.org/data/data/womoabund.html, retrieved on October 17, 2009.



(a) (b)

Figure 10.8. (a) The annotation text on the graph, accompanied by an annotation icon, (b) the obtained constellation after resizing the graph space.

The graph producer may easily recognize the inappropriate location and orientation of the annotation in Figure 10.8b. However, by using the same method, the graph producer's resizing may result in inappropriate relocation and reorientation in annotation icons that are not easily recognized by eye. Such a case is exemplified in Figure 10.9. The annotated graph in Figure 10.9b is a resized version of the annotated graph in Figure 10.9a.



(a) (b)

Figure 10.9. (a) The annotation text on the graph, accompanied by an annotation icon, (b) the constellation obtained after resizing the graph space.

The presented cases exemplify the problem of *reflow*. The term *reflow* is used for changing the visual properties of a text in the screen so that the space it takes is rearranged. It can be exemplified by changing the font size, which in turn results in the change in the location of the textual entities on the screen. For instance, one word in the third row may be in the second row after the reflow. The problem of reflow occurs when the accompanying content in the document, such as a mark (e.g., a star) does not follow the new pattern specified after the reflow. The reflow problems and technical solutions to it have been discussed for freeform ink annotations in text-only documents (e.g., Brush et al., 2001; Bargeron & Moscovich, 2003).

As stated by the Guidelines 4, repeated below, such reflow problems in annotation text and annotation icon may result in misconceptualizations of the relationship between the events and processes represented by annotations and events and processes represented by graph lines.

Guideline 4. Select the location of the annotation appropriately; use more than one annotation where necessary.

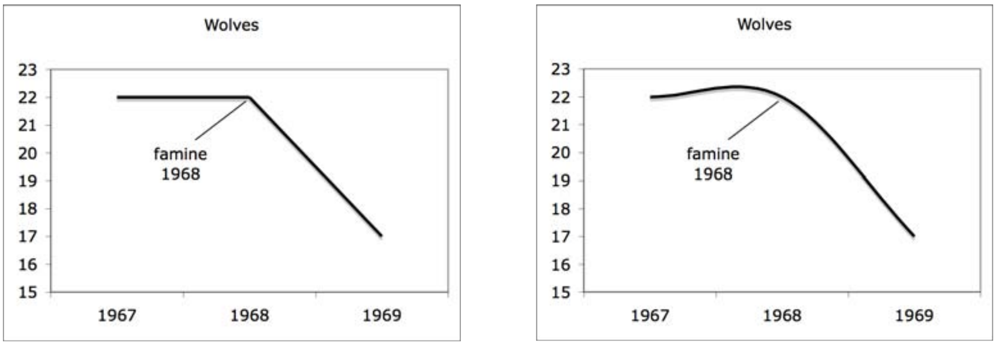
One technical solution to remedy the reflow problem in annotation icons is to define a set of anchor points in the graph space. One end of the annotation icon should be anchored to the annotation text, which is often placed in a textbox. Therefore, the anchor point can be placed on the textbox, rather than the text itself. At the other end of the annotation icon, a tiny amount of gap is necessary between the end of the annotation and the graph line, since a direct contact of the annotation icon to the graph line may result in a branching-appearance of the graph line. This requires setting an invisible anchor point between the end of the annotation icon and the graph line. Once the anchor point is specified, the end point of the annotation icon and the graph line can be affixed to the anchor point (see Bargeron & Moscovich, 2003; Brush et al., 2001, for technical mechanisms for the reflow problems in documents with freeform ink annotations).

To sum up, the reflowing of a verbally annotated graph introduces a set of problems that arise due to the contradictions to the good practice of multimodal graph-text comprehension, as exemplified above. The design of technical mechanisms for solving these problems is a recent task of application oriented computer science research.

A further implication of the presented design guidelines for graph-text constellations is derived from the Guideline 5, repeated below, which was devised from the experimental findings presented in Chapter 7.

Guideline 5. Use graph-line smoothing cautiously in annotated graphs.

Smoothing options for line graphs are useful tools to produce graphs not only for a better aesthetic appearance but also for a better visibility of the trends. Smoothing reduces the cluttered appearance of a noisy graph line thus producing a more easily comprehensible graph. On the other hand, smoothing is not always a better way of representing data graphically. A problematic case is exemplified in Figure 10.10. The graphs show the change in the wolf population for three years. The graph in 10.10b is produced by applying the option “smoothed line” of the graph drawing facility to the graph in 10.10a.



(a) (b)

Figure 10.10. (a) An annotated graph with a graph line that has a straight corner, (b) the same graph after applying the "smoothed line" option provided by the graph producing software.

The two graphs shown in Figure 10.10 lead to different conceptualization of the events and processes represented by the graph line and those represented by the annotations. One problem with the exemplified smoothing option is the artifact of the increase representation on the left segment of the graph line. More importantly, the causal attribution between the 'famine' event and the decrease in the wolf population is weaker on the graph in 10.10b compared to the graph in 10.10a since the graph in 10.10b shows that the population decrease literally started before the event of famine; this is a contradiction to the temporal precedence of causes compared to effects (see Chapter 7, Study 2 for experimental investigations on causal attribution).

In summary, the available software tools for producing annotated graphs and graph-text constellations offer limited capabilities to users as producers of graph-text constellations. Clearly, the exemplified cases reflect situations that are specific to a version of a widely used general-purpose software tool, and these tools have been under steady progress. The purpose of the exemplified cases is to specify implications of the presented guidelines for the development of software tools for designing graph-text constellations.

10.4 SUMMARY

The common conceptual basis for linguistic and graphical entities, as stated in the theoretical framework presented in Section 1 of the dissertation, suggests that multimodal comprehension of graph-text constellations is based on the construction of inter-modal coreference links and coherence links during the course of comprehension of linguistic and graphical entities. The empirical studies that are presented in Section 2 of the dissertation show how changes in one modality influence comprehension of the other modality in graph-text constellations, thus providing support for readers' effort for the construction of integrated conceptual representations of the representations with different modalities.

The empirical findings of the dissertation indicate that poorly designed combinations of graphs and text in graph-text constellations and poorly designed combinations of graphical entities and verbal annotations in verbally annotated graphs result in processing difficulties during the course of multimodal comprehension. In the recent state of the art, the design and production of graph-text constellations are based on the experience and practice of the designers rather than theory, guidelines or systematic empirical research. The available software tools have not yet capacity to provide support to designers as end-users during their course of designing and producing graph-text constellations.

This chapter presented a set of principles and guidelines for appropriate design and good practice for multimodal comprehension of graph-text constellations, based on the empirical findings reported in Section 2 of the dissertation. The following table summarizes the principles and guidelines for graph-text constellations presented in 10.2.

- The fundamental principle of multimodal comprehension of graph-text constellations makes the following statement.
    - ***Principle 1.*** Users recall better from text-graph constellations rather than from text alone.
        - ***Guideline 1.*** Transform complex graph-related sentences into simple ones.
        - ***Guideline 2.*** Check the congruency between scalar entities in the text and spatial configuration of graphical entities in the graph.

- The fundamental principle of comprehension of verbally annotated graphs in graph-text constellations.
    - ***Principle 2.*** Verbal annotations in graphs facilitate multimodal comprehension of graph-text constellations.
        - ***Guideline 3.*** Use annotation icon to facilitate the construction of the relationship between annotation and the relevant part of the graphical entities.
        - ***Guideline 4.*** Select the location of the annotation appropriately; use more than one annotation where necessary.
        - ***Guideline 5.*** Use graph-line smoothing cautiously in annotated graphs.

In 10.3, the principles and guidelines for the design of graph-text constellations were further discussed with respect to their implications for the design and development of software tools that have the capacity to provide end-users during their course of designing and producing graph-text constellations.

Section 4

Conclusion

11 Conclusions, Open Questions, and Future Research

Graph-text constellations are one of those cognitive means by which we accomplish high-level problem solving and communication tasks in daily life. Graphs have been widely used since the eighteenth century, and it has been recognized that a competent use of graphs and diagrams facilitates scientific reasoning (Tufte, 1983; Larkin & Simon, 1987; Cheng & Simon, 1995; Catley, 2006). Graph literacy has recently been an essential part of educational curricula (NCTM, 2000; Catley, 2006). Not only in scientific domains but also in non-scientific domains graph-text constellations are widely used both in printed media and in electronic media, as in newspapers articles, magazine articles, and in online user blogs on the WWW. The development of graph-producing software tools has enabled both non-expert people and knowledge visualization experts to design and produce graphs.

Human beings' effective use of external representations in communication and in problem solving is subject to a successful integration of information induced by external representations in different representational modalities. In multimodal comprehension of graph-text constellations, humans have to construct the correspondences between linguistic entities, graphical entities, and domain entities during the course of comprehension (i.e., the coreference relations between the text layer, the graph layer, and the layer of domain entities, see Chapter 1). Moreover, humans, in principle, achieve a coherent interpretation by employing conceptual relations, such as constructing causal relationships between the events and processes represented by graphical and linguistic entities. On the other hand, graphs have the capacity to represent quantitative information in perceptually different ways. The representation of the same information in perceptually different ways may lead non-expert people to have different and inconsistent conceptualizations of the represented information. Those misconceptualizations in graph comprehension may risk the integration of the information induced by different representational modalities, which, in turn, may hamper the effective communication and problem solving by means of graph-text constellations.

The available computer technologies provide non-expert users with software tools for designing and producing myriad types of graphs. Nevertheless, they offer limited capacity to augment language by graphs: They do not possess the capacity to inspect graph-text constellations that may lead non-experts to perceptually induced misconceptualizations. In particular, there is a need for software tools to check the consistency between graphical and linguistic entities, which can have similar functions to spell checkers and grammar checkers in the available word processing software, are missing. An essential step for the development of such artificial systems that have the capability to make suggestions for appropriately augmenting language and graphs is the study of human multimodal comprehension of graph-text constellations, since multimodal communication and problem solving by humans have peculiar characteristics that can recently be approximated by artificial systems only to a limited extent. This dissertation investigated multimodal comprehension of graph-text constellations by humans—as natural cognitive systems—for providing the basic knowledge for the design and development of artificial cognitive systems that have the capacity to support users during their design and production of graph-text constellations.

This chapter is composed of two parts. In 11.1, the conclusions of the dissertation are presented, under the framework of the multidisciplinary approach of cognitive science and HCI. In 11.2, open questions for further research are presented.

11.1 CONCLUSIONS

HCI and cognitive science are two domains of research where the study of natural cognitive systems and the study of artificial cognitive systems overlap. In the past several decades, depictive external representations have been a topic of experimental research in both cognitive science (e.g., Monaghan & Stenning, 1998) and HCI (e.g., Scaife & Rogers, 1996). On the other hand, the investigation of the interaction between individuals and external representations of particular types has been limited. In particular, the research on external representations of specific types has been scarce. This dissertation focused on a specific type of external representation, namely line graphs that represent change of a quantitative measurement over time. Interpretation of line graphs is an essential part of problem solving and communication tasks in various domains, such as identification of trends in economy, in medical diagnosis and treatment, and the interpretation of statistical data in experimental research, among many others.

Graphs comprise a distinct class of quasi-symbolic depictive representation compared to symbolic representations (e.g., natural and artificial languages) and depictive representations (e.g., pictorial illustrations and diagrammatic illustrations). Graphs are cognitive human artifacts that are used for externalizing our mental conceptualization of the world. They have peculiar *structural* characteristics that overlap between symbolic representations and depictive representations: Like language, graphs are grammatical means of problem solving and communication; like pictorial illustrations, graphs are depictions of our knowledge of domain entities. On the other hand, the previous research on graphs has often focused on *processing aspects* of graph perception and comprehension, rather than the *structural aspects* of graphs, except for the seminal study of Kosslyn (1989). After Kosslyn, Pinker (1990) presented a close view that considered graphs as means of communicating quantitative information with grammatical structure. Kosslyn (1989) and Pinker (1990) presented their approach to graph comprehension from a general psychological perspective. This dissertation extends their approach to graph comprehension research in several respects. First, a complementary theoretical framework to graph comprehension was developed by extending its reach beyond the comprehension of graphical entities, making it possible to embody comprehension of linguistic entities, thus leading to integrated conceptual representations of graphical and linguistic entities. In particular, the construction of referential links between linguistic and graphical entities in graph-text constellations was investigated under the theoretical framework of the representational modularity hypothesis (Jackendoff, 1996, 2002). Second, through experiments with human participants, various aspects of the interaction between language comprehension and graph comprehension were empirically investigated. Third, based on the findings of the empirical investigations, design principles and guidelines were devised for an appropriate design of graph-text constellations and for the development of software tools that would provide the end-users with support for appropriate designs during their course of designing and producing graph-text constellations. These aspects are further discussed below.

The study of the construction of referential links (i.e., coreference and coherence links) between linguistic and graphical entities in graph-text constellations is connected to the study of the interaction between language and visual percepts during the course of comprehension. In the dissertation, the representational modularity hypothesis (Jackendoff, 1996, 2002) was adopted as the basis for multimodal comprehension of graph-text constellations, where graphs were conceived as visual percepts with grammatical structure. The representational modality hypothesis analyzes the interaction between language and visual percepts in terms of the two interface components, namely conceptual and spatial structures. Conceptual and spatial structures are essential parts of the interface between linguistic systems and the sensory systems. In particular, the conceptual structure (CS) encodes aspects of human understanding such as predicate-argument structure and category membership, and the Spatial Structure (SpS) encodes spatial properties such as object shape (also see Pinker & Jackendoff, 2005; Jackendoff & Pinker, 2005). The overlap between CS and SpS is constituted by the notions that have reflexes in both systems, such as PATH and PLACE in natural language. In particular, shape nouns, spatial verbs (i.e., verbs of change in space), spatial prepositional phrases, and adverbial modifiers exemplify the class of linguistic entities that involve spatial concepts in their lexical structure.

In this dissertation, taking the concepts that have reflexes both in CS and SpS (i.e., the *common conceptual basis* for the representations in different representational modalities) as the working assumption, the correspondence between textual entities and graphical entities (e.g., the term 'peak' in the text layer and the peak as a graphical entity in the graph layer in a graph-text constellation) was investigated in terms of the concepts that are accessible by both textual and graphical entities. For instance, the frequently used spatial verbs in graph-text constellations, such as *decrease*, *decline*, and *fall* involve the process concept DECREASE_OF_VALUE($_{TEMP}$, $_{VALUE}$), where the *TEMP* argument stands for the temporal interval during which the process is occurring, and the *VALUE* argument stands for the amount of decrease. The process concept DECREASE_OF_VALUE is a specification of a mapping from the temporal domain to the value domain. In the terminology of topology, this mapping corresponds to a *path* in the value space. Moreover, the process concepts such as INCREASE and DECREASE have spatial properties, which constitute their geometric specification. For instance, the process concept INCREASE holds the necessary condition *VALUE(BEGIN(DECREASE)) < VALUE(END(DECREASE))*. The *SOURCE* and the *GOAL* arguments of the path, when specified—for instance, by prepositional phrases—determine the *VALUE(BEGIN(DECREASE))* and the *VALUE(END(DECREASE))* correspondingly. Those spatial properties are accessed during the course of comprehension of graphical entities, which is based on a set of processes that transform visual information into spatial and conceptual representations. In particular, the graph schema, i.e. the long-term knowledge structure that involves the necessary information for interpretation of graphs, provides a syntactic and combinatory system at the interface between CS and SpS, thus providing the basis for the construction of coreference relations in graph-text constellations.

Based on the theoretical framework of the common conceptual basis for representations in different modalities, a coarse-grained cognitive architecture for multimodal comprehension of graph-text constellations was developed. The processes of multimodal comprehension of graph-text constellations were described in terms of the interaction between the processes of the two major comprehension modules, namely the language comprehension module and the graph comprehension module. The integrated conceptual and spatial representations were constructed as an outcome of this interaction. The architecture described the information flow at a coarse level, and its purpose was to provide the basic framework for the empirical investigations of various aspects of the interaction between comprehension of linguistic entities and graphical entities, rather than to investigate the inter-modal processing details. Various aspects of the construction of inter-modal coreference and coherence relations were analyzed by empirical investigations, as presented below.

The empirical investigations reported in the dissertation were conceived as the experimental studies of the interaction between users (i.e., participants), display (i.e., stimuli), and methods (i.e., the methods of interaction and analysis). The interaction was investigated in terms of the analysis of participants' eye movements during the course of their reading of the presented graph-text constellations, and in terms of a set of complementary methods such as the analysis of participants' measurement of recall and the analysis of subjective judgment reports.

Each empirical investigation revealed how the readers integrated the information induced by the representations to reach a coherent interpretation and how the changes in one modality led to substantial influences in comprehension of the other modality. The findings of the empirical investigations that were reported in Chapter 6 revealed that the changes at syntactic level of the sentences in a graph-text constellation influenced comprehension of graphical entities. In particular, the results showed that the readers integrated the information induced by textual and graphical entities by executing reciprocal gaze fixations (i.e., gaze shifts) between the text layer and the graph layer of the graph-text constellations with a separate text-layer–graph-layer layout. The syntactic complexity of textual entities in the text layer of the graph-text constellations influenced the inspection of graphical entities, which, in turn, influenced the integration of the information contributed by the two representational modalities, as shown by the higher number of gaze shifts between the text layer and the graph layer under the complex-syntax conditions compared to simple-syntax conditions. The reciprocal nature of the input increments (i.e., gaze fixations) between the two representational modalities and the influence of the syntactic complexity of textual entities on the inspection of graphical entities indicate the moment-by-moment integration of the information contributed by the two representational modalities and the *incremental characteristic* of the integration processes. The further analysis of the consecutively inspected locations in the text layer and in the graph layer of the graph-text constellations showed that the readers often accessed corresponding textual and graphical entities, providing empirical support for the working assumption of the common conceptual basis for the representations with the two representational modalities and for the instantiation of the graph schema during the course of the construction of coreference relations.

The two empirical investigations that were reported in Chapter 7 and Chapter 8 showed how the changes in spatial properties of graphical entities influenced comprehension of linguistic entities. In particular, causal attribution concerning the relationship between processes and events represented by verbal annotations and processes and events represented by graphical entities was investigated, as reported in Chapter 7. The findings of the experimental studies showed that the position of the verbal annotations and the shape properties (in particular, the smoothness) of the graph lines influenced the readers' judgments about the causal relationship between the events and processes represented by the verbal annotations and the graphical entities. Moreover, multimodal comprehension of verbal annotations and graphical entities led not only to coherence judgments concerning causal attribution but also to coherence judgments concerning temporal aspects of the events and processes represented by verbal annotations and graphical entities. The smooth graph lines resulted in a durative attribution of the temporal aspects of the related events that were represented by verbal annotations, whereas graph lines that had sharp corners produced a punctual attribution. To sum up, the studies reported in Chapter 7 revealed that the comprehension of verbal event annotations and the comprehension of graphical entities interact to construct a coherent whole, which, in turn, provide empirical support for the assumption of the common conceptual basis for multiple representational modalities.

The experimental investigations that were reported in Chapter 8 analyzed the construction of the coherence relations between graphical and textual entities in graph-text constellations in terms of the consistency between the scalar properties, i.e. the concept of scale. The concept of scale is another conceptual entity at the interface between CS and SpS, which is accessed by scalar adjectives/adverbs (e.g. *tall*, *slowly* etc.) and by the spatial configuration of the graphical entities in the graph space, i.e. the location of the slope of a graph segment. The findings of the experimental studies showed that readers have difficulties in comprehension of the graph-text constellations in case of an inconsistency between the scale information induced by the textual and graphical entities in a graph-text constellation. Consequently, this shows the readers' effort to reach a coherent whole during the course of multimodal comprehension of graph-text constellations.

In Chapter 9, in order to develop a GOMS-style cognitive model for multimodal comprehension of spatial verbs and spatial prepositional phrases in graph-text constellations, the participants' eye movement patterns in a sentence-graph verification task were analyzed. The cognitive model was developed as a specific instance of the cognitive architecture for multimodal comprehension of graph-text constellations, which was developed in Chapter 4. The model showed that a GOMS-style decomposition of the linguistically specified tasks into subtasks could be used for predicting graph-inspection gaze patterns of readers. This finding also indicates that the available gaze patterns on graphs could be used for predicting the sentences that were comprehended by the readers. Moreover, the model revealed the *combinatory characteristic* of multimodal integration processes. In particular, the model showed that after their reading of complex sentences, the gaze patterns of the readers on graphs could be predicted by combining the available gaze patterns after their reading of simpler sentences.

In summary, the experimental studies that are reported in the dissertation provided empirical support for the cognitive plausibility of the analysis of multimodal graph-text comprehension in terms of the basic assumptions of the developed architecture, i.e. the common conceptual basis for multiple representational modalities and the conceptual entities that have reflexes at the CS-SpS interface. Therefore, the dissertation not only adopted the representational modularity hypothesis (Jackendoff, 1996, 2002) as the theoretical framework, but also provided additional empirical support for Jackendoff's conceptual analysis of the interaction between language and visual percepts in terms of the CS-SpS interface.

Beyond extending Kosslyn's (1989) and Pinker's (1990) graph comprehension approaches to multimodal comprehension, this dissertation additionally supplements Kosslyn (1989), who suggested a set of *acceptability principles* for graph design, by developing a set of principles and guidelines for designing graph-text constellations and for designing software tools that have the capacity to help non-expert graph producers during their design and production of graph-text constellations. Based on the experimental findings reported in Section 2 of the dissertation, two fundamental principles and five guidelines for designing graph-text constellations were devised, as presented in Chapter 10. These principles and guidelines are summarized below.

A. The fundamental principle of multimodal comprehension of graph-text constellations makes the following statement.

Principle 1. Users recall better from text-graph constellations rather than from text alone.

Guideline 1. Transform complex graph-related sentences into simple ones.

Guideline 2. Check the congruency between scalar entities in the text and spatial configuration of graphical entities in the graph.

B. The fundamental principle of comprehension of verbally annotated graphs in graph-text constellations.

Principle 2. Verbal annotations in graphs facilitate multimodal comprehension of graph-text constellations.

Guideline 3. Use annotation icon to facilitate the construction of the relationship between annotation and the relevant part of the graphical entities.

Guideline 4. Select the location of the annotation appropriately; use more than one annotation where necessary.

Guideline 5. Use graph-line smoothing cautiously in annotated graphs.

The principles and guidelines for designing graph-text constellations were further analyzed to provide a primary basis for the development of assistance systems in the design of graph-text constellations. In particular, most of the available tools for producing verbal annotations in graphs are not adapted to a reflow of the graph. For instance, changing the dimensions of a graph or smoothing a graph line may lead to misconceptualizations of the events and processes represented by verbal annotations due to the incorrect repositioning or mispositioning of the verbal annotation. Moreover, a reflow of the graph may lead to a change in conceptualization of a scalar entity, such as *slowly*, which is influenced by the slope of the graph line in the graph space. Those problematic cases and possible technical solutions were discussed in Chapter 10. In the rest of this chapter, open questions for further research are presented.

11.2 OPEN QUESTIONS AND FUTURE RESEARCH

Human multimodal comprehension of graph-text constellations is a complex information processing task, guided by almost automatically performed perceptual and cognitive information processing subtasks. Accordingly, an indispensable part of the study of human multimodal comprehension for the purpose of developing intelligent artificial systems is a set of theoretical and methodological assumptions that make possible to conduct repeatable empirical investigations and to reach statistically valid results. Specifying the scope and limitations of the study and devising open research questions, those assumptions pave the way for future research. In this part of the chapter, based on the findings reported in the dissertation, open research questions and the claims for future research are presented.

The first open research question concerns the cognitive architecture for multimodal comprehension of graph-text constellations (Chapter 4). The architecture was presented in the form of an information flow architecture with a focus on the structural properties of the representations in the two representational modalities. The architecture was based on the working assumption of the common conceptual basis for multiple representational modalities. The interaction between the two major modality-specific modules, namely the language comprehension module and the graph comprehension module, was analyzed in terms of the shared concepts between the representations in the two modalities (e.g., the path concept), and empirical investigations were conducted to seek for the evidence for the construction of coreference relations between the representational modalities, which, in turn, presented partial evidence for the integrated conceptual and spatial representations. The presented architecture did not focus on the information processing within the particular modules, i.e. the language comprehension module and the graph comprehension module. Accordingly, the specific direction of information flow and feedback loops between the submodules of the modality-specific modules were not in the focus of the dissertation; therefore, they were not discussed under the framework of the architecture. In addition, the architecture is undetermined with respect to the internal structure of the integration module and with respect to the details of the interaction processes. Consequently, the architecture is *coarse-grained* in its recent form. These assumptions show the aspects of the architecture that need development in future research. A more comprehensive architecture should specify the processing characteristics of the module-specific information processes, the processing characteristics of the integration processes, and the nature of input increments and conceptual increments. In particular, two open research sub-questions for future investigation are (1) which level of incremental entities are involved in integration? (2) how do integrated representations influence modality-specific comprehension?

The second open research question concerns the properties of the stimuli that were used in the empirical investigations. The majority of the stimuli were expository graph-text constellations that were based on the newspaper articles (Chapter 7, Study 1), travel guides (Chapter 7), and public reports (Chapter 6, Chapter 9). The stimuli in the two experiments were designed by the experimenter (Chapter 7, Study 2 & Study 3). The design of the stimuli, which was based on the publicly available expository graph-text constellations (i.e., naturalistic stimuli, see Chapter 5) in the majority of the experiments, brought the advantage of ecological validity of the empirical investigations, but at the expense of the loss of experimental control over the stimulus material, which resulted in variability in the results. In future research, more experiments would enable further control over the experimental factors, thus providing a better generalizability of the results.

A relevant open research question about the stimuli is the role of shape properties of graphical entities (e.g., complexity of the graph line in terms of smoothness) in multimodal comprehension of graph-text constellations. This dissertation focused on the influence of comprehension of linguistic entities on inspection of graphical entities in graph-text constellations. On the other hand, as discussed in Chapter 2, visual salience of graphical entities influences the processing of perceptually driven information, which, in turn, might influence multimodal comprehension of graph-text constellations. The visual salience is often specified by the critical points such as maxima, minima, and point of inflection (Attneave, 1954; Freeman, 1978; Feldman & Singh, 1997; De Winter & Wagemans, 2004; 2006). In the reported experiments, homogenizing the material across the experimental conditions reduced the influence of visual salience. In particular, at least three different graphs were used for the specification of a graph-related experimental condition. For instance, an "increase" graph is represented by at least three different graphs, each having represented the process of increase with a different shape of the graph line. On the other hand, a systematic investigation of the role of shape properties in multimodal comprehension of graph-text constellations should involve the analysis of shape parameters, such as smoothness, the specification of particular elementary visual operators in comprehension of line graphs (cf. visual routines, Ullman, 1984), and the analysis of the granularity of the correspondence between linguistic entities and graphical entities, such as the granularity of the correspondence between the term 'increase' and the graph line that represents a process of increase. Moreover, the study of graph-text constellations that involve line graphs with non-temporal variables would also reveal how particular graph shapes influence multimodal comprehension of linguistic and graphical entities.

Another open research question is the high variability in participants' skills and abilities. A distinction between *visualizers* (i.e., the learners with high visual abilities and low verbal abilities) and *verbalizers* (i.e., the learners with high verbal abilities and low visual abilities) has generally been recognized by researchers in cognitive and educational psychology (cf. cognitive and learning styles, Kozhevnikov, Hegarty, & Mayer, 2002; Kozhevnikov, Kosslyn, & Shephard, 2005; Shah & Miyake, 1996, among many others). Further individual differences among students, such as differences in mechanical and spatial abilities, are measured by Differential Aptitude Tests (Bennett, Seashore, & Wesman, 1947). The findings in multimedia learning showed that picture-text constellations increased the performance of students with low verbal abilities compared to the performance of students with high verbal abilities (Winn, 1987). Moreover, students with high spatial abilities learned better with picture-text constellations compared to students with low spatial abilities (cf. the individual differences principle, Mayer, 2001/2009). The study of individual differences is connected to the study of the role of human working memory system and the role of memory components during the course of multimodal comprehension. The research on human memory and the particular working memory components, namely the verbal working memory and the visuospatial working memory, has been systematically developed in the domain of cognitive and experimental psychology in the last several decades (e.g., Baddeley 1986, 1997, 2000; Jonides, Lewis, Nee, Lustig, Berman, & Moore, 2008). The more specific research on the role of particular working memory components in multimodal comprehension is currently a developing domain of research (Gyselinck et al., 2008; Cowan et al., 2006; Seufert et al., 2009).

The characterization and consideration of cognitive styles and the role of memory components are important factors not only in empirical investigations but also in user interface design. However, more research is needed to prove cognitive styles to be robust foundations to explain the divergent findings on in the educational and cognitive psychology literature, including the domain of multimodal learning with picture-text constellations (Clark & Feldon, 2005; Ainsworth, 1999; 2006) and the domain of multimodal graph-text comprehension. In one of the experimental investigations reported in this dissertation (the Study 1 in Chapter 7), the spatial and verbal abilities of the participants were measured by the verbal span test and the arrow span test (adapted from Shah & Miyake, 1996); however, the participants' reactions to the experimental stimuli revealed results that were independent of the span test scores. Those non-significant differences obtained in the reported experimental investigations, such as the span test results, and the participants' recall of the text, as reported in Chapter 6, would be prevented by conducting more experimental investigations with higher number of participants and by applying alternative empirical techniques for the analysis of individual differences.

Finally, the framework presented in this dissertation can be used for the development of *accessible* graph-text constellations for visually impaired people. The design of graphical user interfaces (GUIs) has been one of the most salient developments in the past several decades of computer technologies. The GUIs have been often based on visual metaphors such as windows and visual icons. Although the use of visual metaphors in GUIs provides sighted persons with the opportunity to design and produce myriad visualizations, it limits the accessibility of information for individuals, who have visual impairment (Dix et al., 2004), thus contributing to *digital divide* in the society (Jacko, Leonard, & Scott, 2009).

Recently, most of the available graph-producing software tools have not been able to offer accessible user interfaces except for a number of software tools that provide the opportunity for employing alternative sensory modalities for comprehension of graphical entities (viz., sensory substitution for graph comprehension). In particular, tactile graphs with a physical form, such as embossed graphs or thermographic printers, computer-generated graphs working through virtual touch (e.g., by force feedback devices, Yu, Ramloll, Brewster, & Riedel, 2001), sonified graphs (see Ratanasit & Moore, 2005, for a review), and a combination of haptic and audio representations (Yu & Brewster, 2003) have been developed for accessible use by visually impaired people. With the help of the common conceptual basis for multiple representational modalities, the available graph comprehension tools for sensory substitution can be further developed to comprise representational substitution, i.e. language support. The language support for accessible graph-text comprehension can be achieved by incorporating verbal annotations into tactile and auditory graphs, either in printed form (i.e., braille) or narrative form. Verbal annotations would help visually impaired persons' comprehension of graphs by facilitating their access to a coherent interpretation of the information represented by the tactile and auditory graphical entities. In particular, comprehension of tactile graph-text constellations with verbal annotations by blind users can be empirically investigated during their course of interaction with tactile desktop devices, such as force feedback and tactile navigation devices (e.g., the PHANToM® (SensAble) haptic interface), and tactile tablets (e.g., the T3 (RNC) Talking Tactile Tablet). The development of language support in sonified and tactile graphs is recently an unexplored domain of research at present.

Appendix

Appendix A

A.1 SAMPLE GRAPHS FROM PLAYFAIR, LAMBERT, AND WATT

In the three figures below, sample graphs from W. Playfair (Figure A.1), from J. H. Lambert (Figure A.2), and from J. Watt (Figure A.3) are shown.

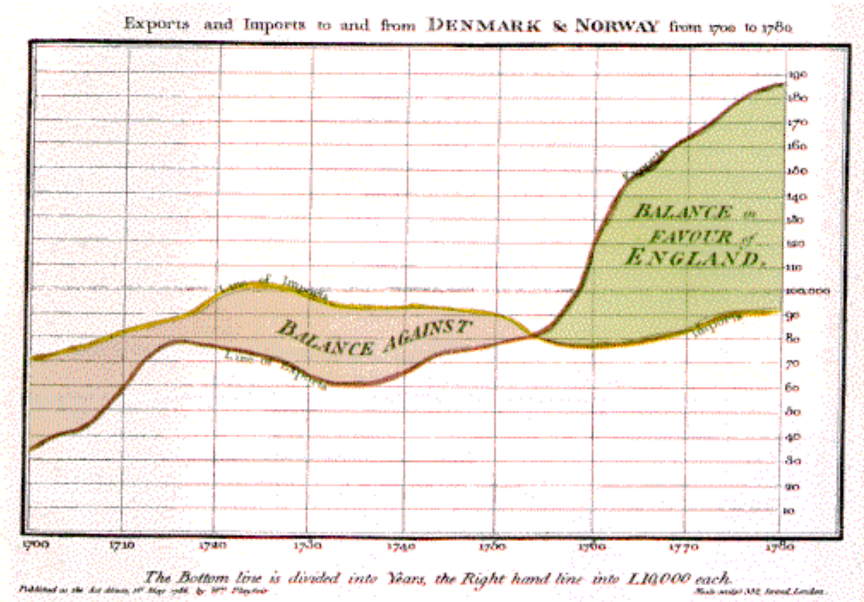


Figure A.1. William Playfair's (1759-1823) trade balance time-series graph. Excerpted from Wikipedia, http://en.wikipedia.org/wiki/William_Playfair, retrieved on 18 March 2009. Original source: Playfair, W. (1786), *The Commercial and Political Atlas*, London: Corry.

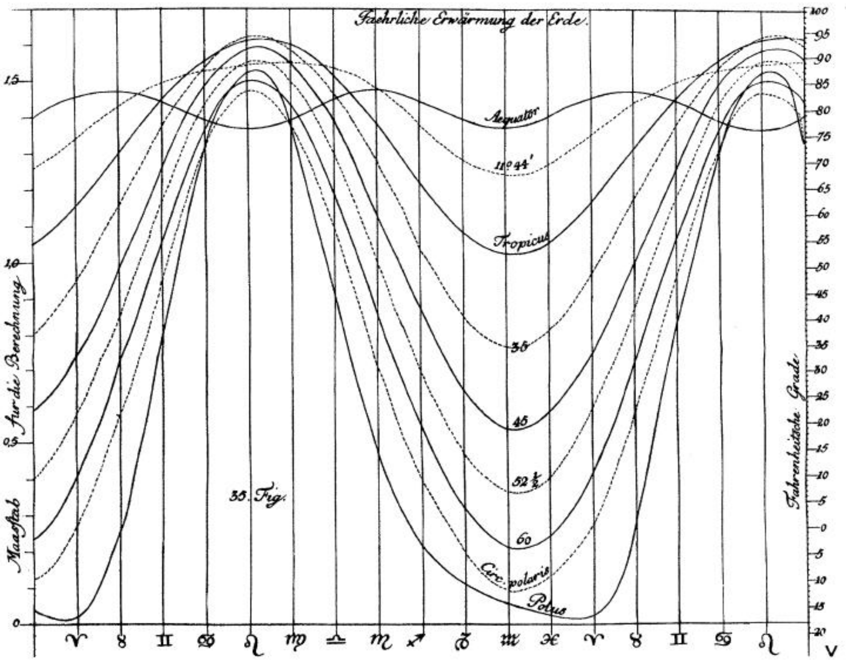


Figure A.2. Johann Heinrich Lambert's (1728-1777) graphical analysis of periodic variation in soil temperature. Excerpted from Friendly (2001) Gallery of Data Visualization, http://www.math.yorku.ca/SCS/Gallery/, retrieved on 18 March 2009. Original source: Lambert, Johann Heinrich (1779). *Pyrometrie; oder, vom maasse des feuers und der wärme mit acht kupfertafeln.* Berlin. (see *Service de la documentation University of Strasbourg - Digital old books* for electronic copy, http://num-scd-ulp.u-strasbg.fr:8080/, accessed on 18 March 2009).

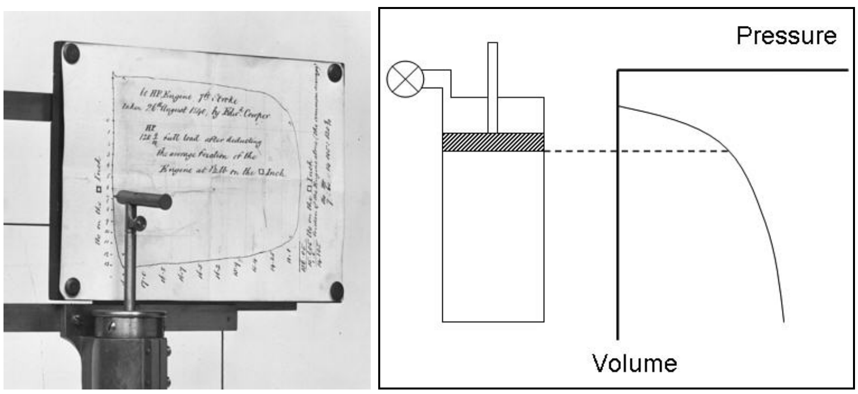
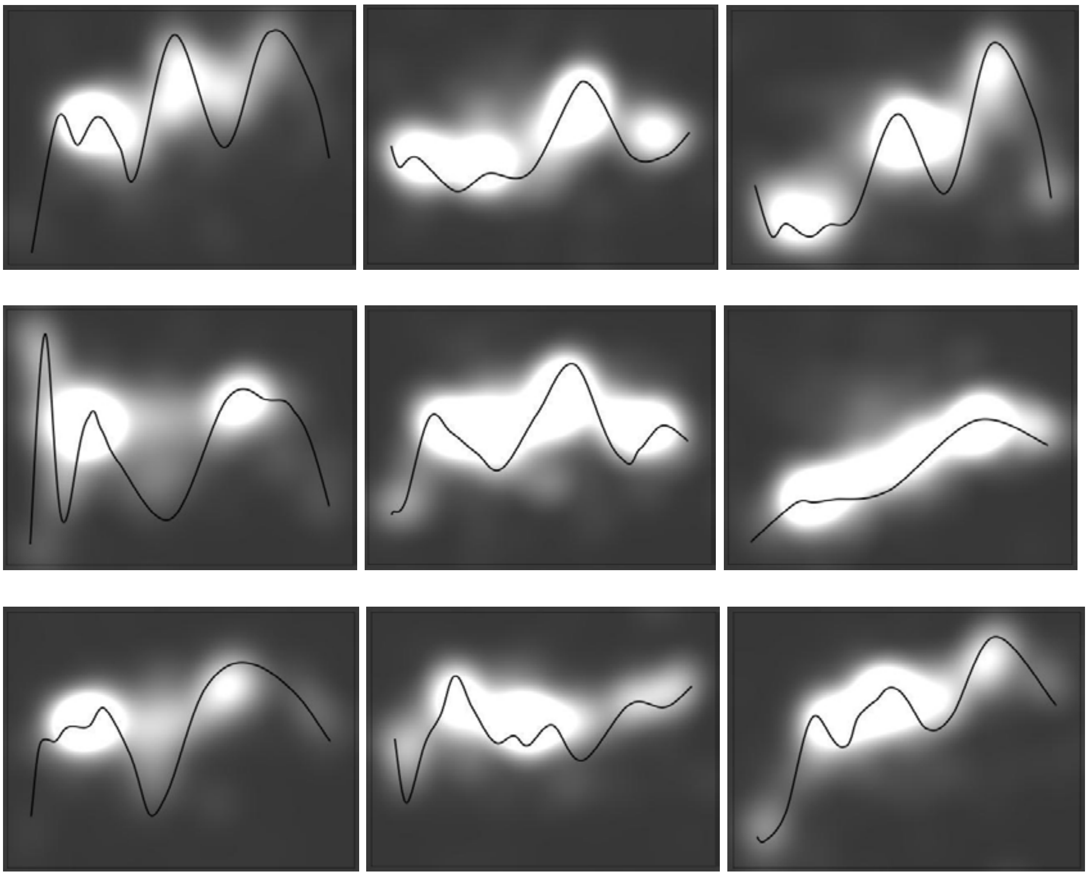


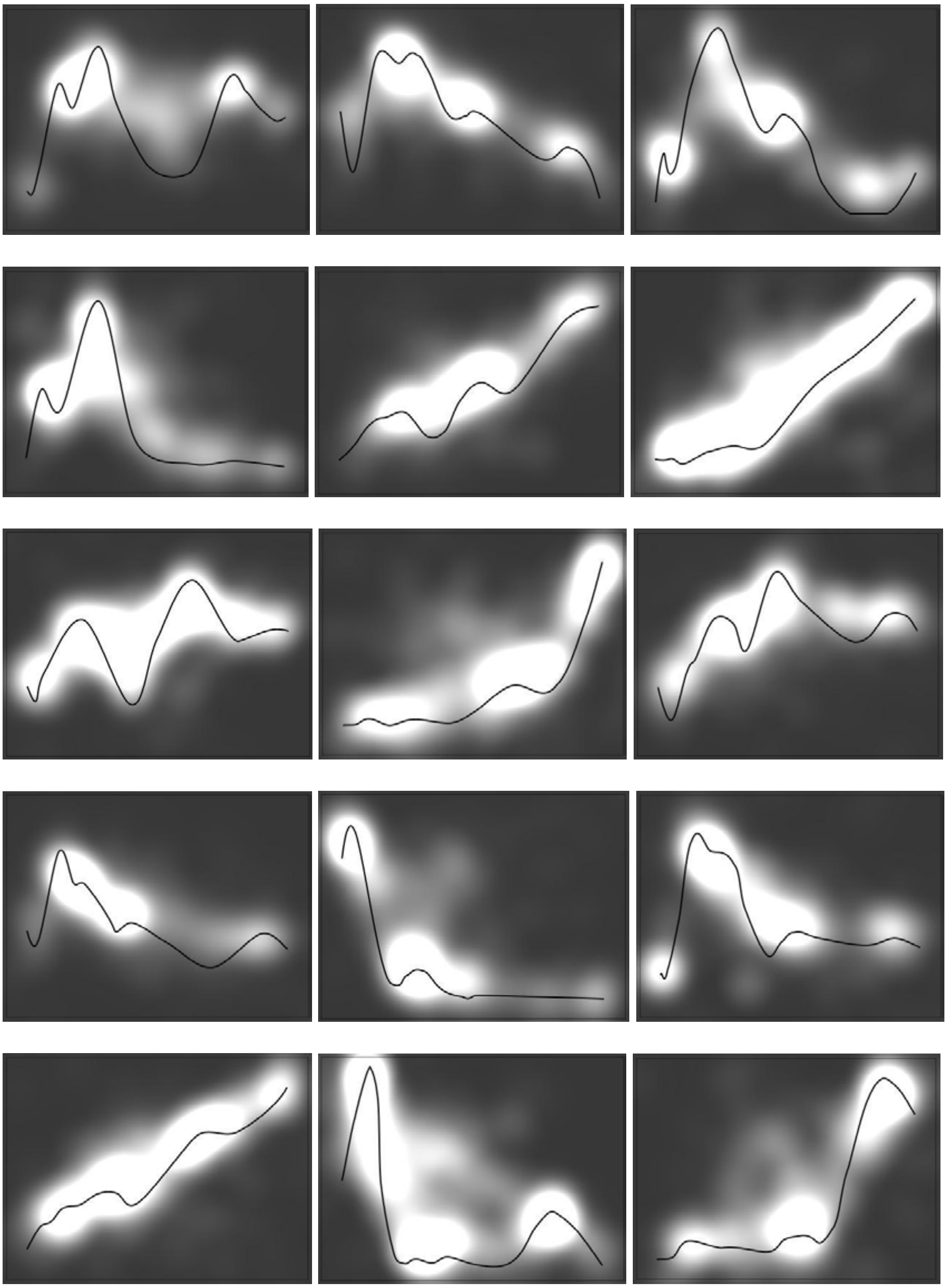
Figure A.3. James Watt's indicator for automatic recording of pressure vs. volume in steam engine. The photograph on the left shows the indicator device. Excerpted from Friendly (2001) Gallery of Data Visualization, http://www.math.yorku.ca/SCS/Gallery/, retrieved on 18 March 2009. The illustration on the right shows a schematic diagram of the indicator's producing a pressure-volume graph. Excerpted from Wikipedia, http://en.wikipedia.org/wiki/Indicator_diagram, retrieved on 18 March 2009.

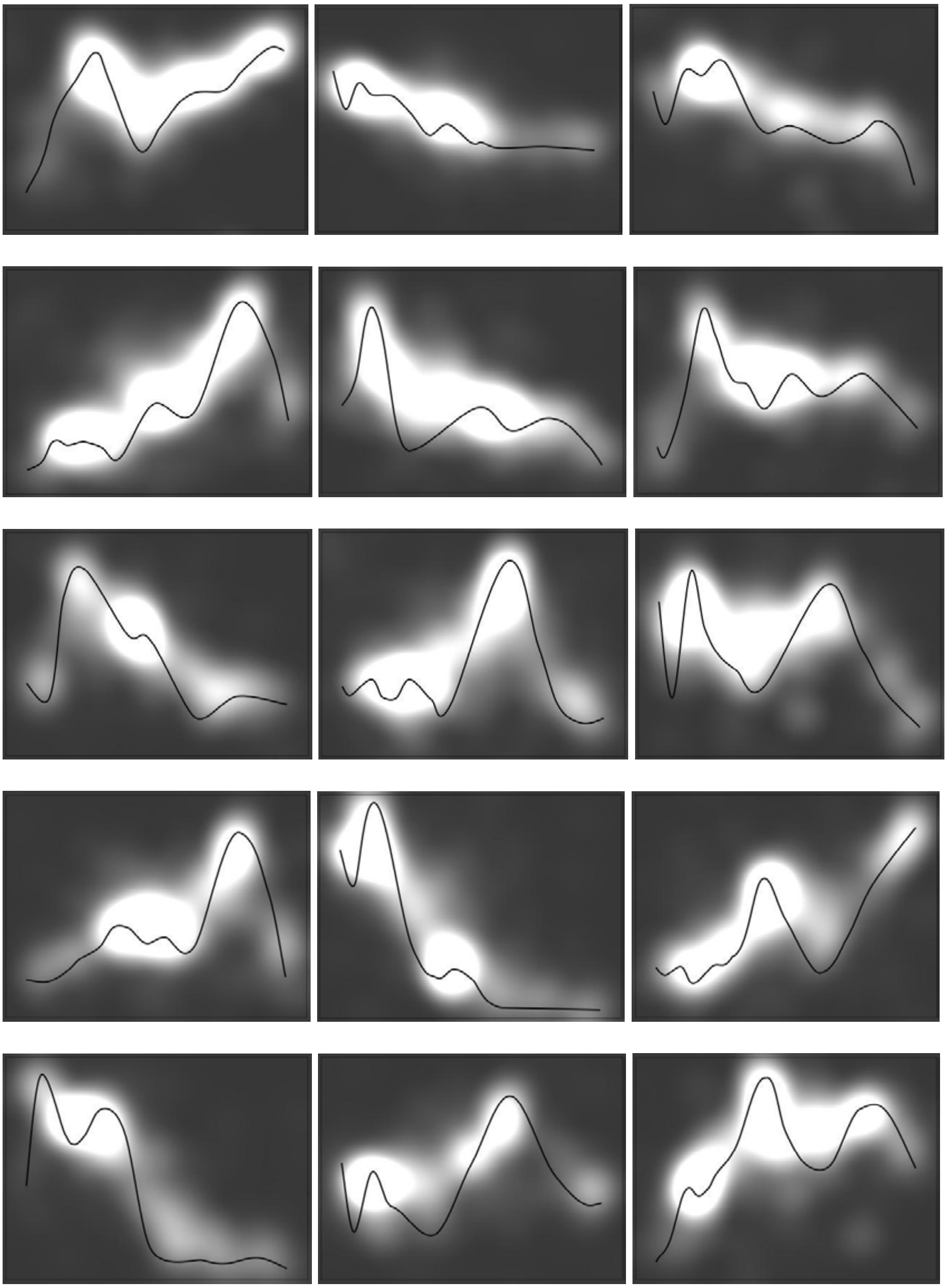
A.2 VISUALLY SALIENT REGIONS IN LINE GRAPHS: A SMALL-SCALE STUDY

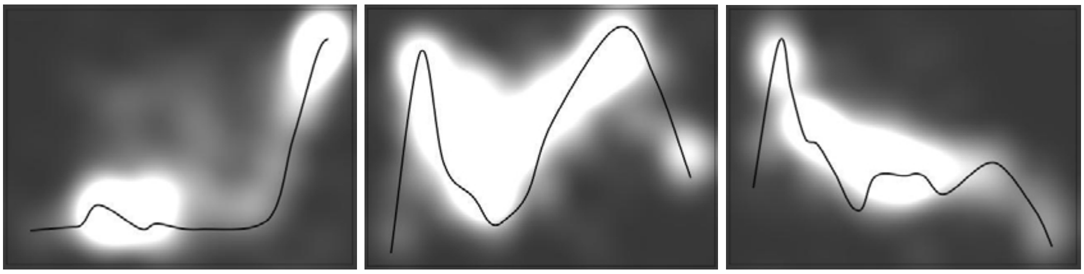
The results of a small-scale study that investigated participants' inspection of graph lines are presented below. Ninety-one graduate or undergraduate students were presented 42 line graphs, without any labels or numbers. The graphs were redrawn based on the "Waterbird Census at Bolinas Lagoon, Marin County, CA" by Wetlands Ecology Division, Point Reyes Bird Observatory (PRBO) Conservation Science, http://www.prbo.org/cms/366, retrieved on September 25, 2009.

The participants were informed that they were expected to inspect about forty graphs, each presented for three seconds. The eye movements of the participants were recorded by a 50 Hz eye tracker. The figures below show the results in a qualitative manner. Each figure is a map of gaze time of the participants' inspections on a stimulus graph. The brighter a region on the graph the longer the mean gaze time of the participants. Since the graphs were not accompanied by text , the resulting gaze time maps reflect the visually salient regions. In other words, these patterns present not-linguistically-guided gaze time patterns in a set of selected line graphs.









Appendix B

This appendix is composed of two parts. In the first part B.1, samples from the experimental stimuli for Experiment 1 to Experiment 5 are presented. The second part B.2 is an appendix for the results of the experiments.

B.1 EXPERIMENTAL STIMULI

B.1.1 Sample Pre-target Sentences, Target Sentences, and Post-target Sentences

Sample pre-target sentences and post-target sentences used in Experiments 1 to 5 are presented below.

Sample pre-target sentences. Sanderlings are small and plump species. They have a stout black bill, black legs and feet. Summer adults have a rufous head and breast, and a white belly. By winter, rufous areas have been replaced by pale gray, and birds look almost white and paler than any other shorebird.
Sample post-target sentences. Sanderlings have one of the most extensive winter ranges of all shorebirds in the world. They can be found worldwide on almost every sandy, coastal beach during the winter.

The target-sentence conditions are exemplified below.

Condition 1 (TC-1). The number of birds in the lagoon peaked twice.

Condition 2 (TC-2). The number of birds in the lagoon peaked twice. The peaks occurred at 1985 and 2000.

Condition 3 (TC-3). The number of birds in the lagoon peaked twice. The peaks occurred at 1985 and 2000. The number of birds fell to about 200 in 1990.

Condition 4 (TC-4). The number of birds in the lagoon peaked twice. The peaks occurred at 1985 and 2000. The number of birds fell to about 200 in 1990. As of 2005, there are about 400 birds in the lagoon.

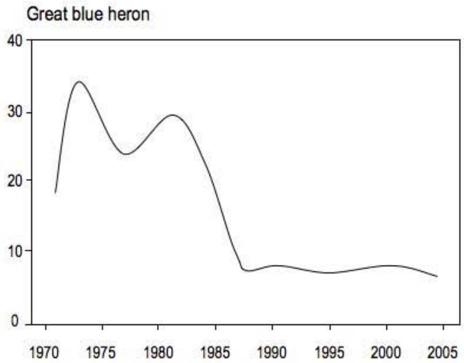
B.1.2 The Instructions

The instructions for the experiment are presented below (translated into English by the author of the dissertation).

Wintering birds in Manyas lagoon

In this study, you will see information about 12 bird species that visit Manyas lagoon for wintering, from a database record of Manyas wintering birds, which has been kept since 1970. We expect you to investigate the presented information and answer a single question for each bird species.

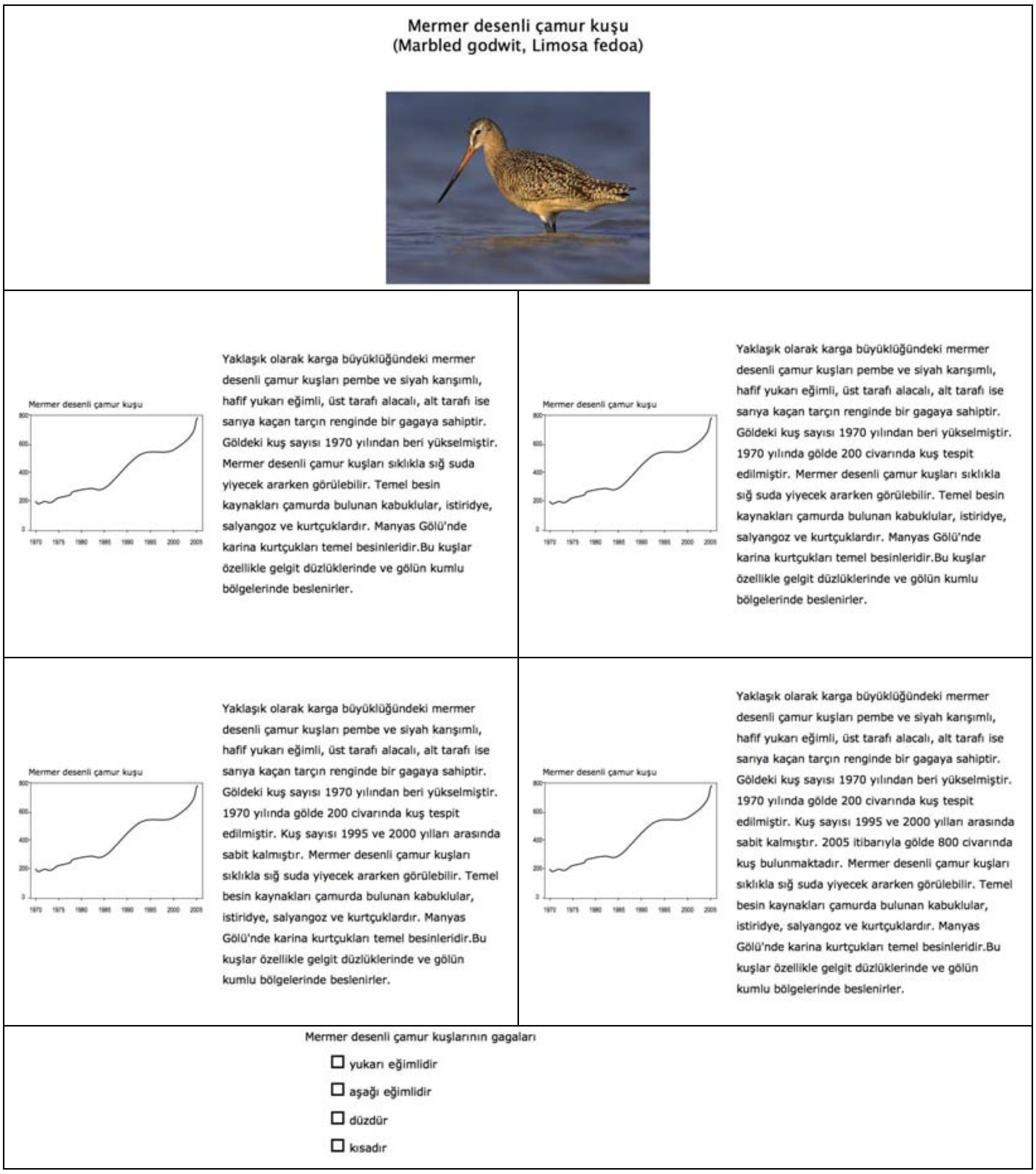
You will first see a screen that includes the picture of the bird (photograph as exemplified in left below). In the next screen, you will see information about wintering habits, about diet characteristics, and about changes in the number of birds in the lagoon. In this screen, you will see a graph that shows the changes in the number of birds in the lagoon (as shown in below middle). The sample graph below shows the change in the number of great blue herons in the lagoon between 1970 and 2005. If you think that you did not understand this graph, or if you have not seen such a graph before, please inform the experimenter. In the next screen, you will see a multiple-choice question that is about the presented information. You will click to submit your answer and move to the next bird species. We expect you to apply this procedure for the 12 bird species.



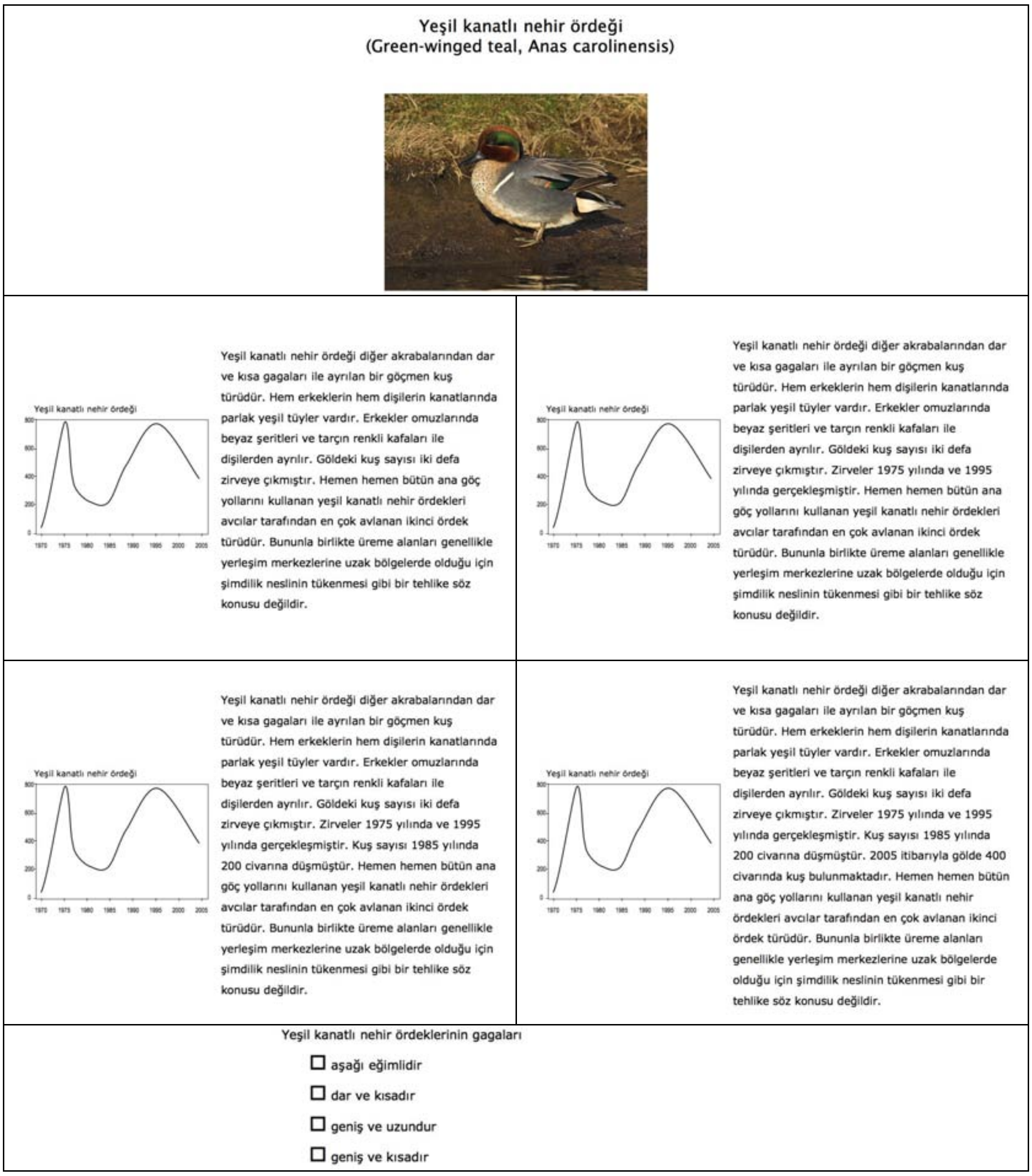
It is not possible to go to the previous screen during the study. Except for the question and answer screen, you will proceed to the next screen by pressing a key on the keyboard. To proceed to the next screen and to start the study, please press a key on the keyboard.

B.1.3 Sample Stimuli from Experiment 1

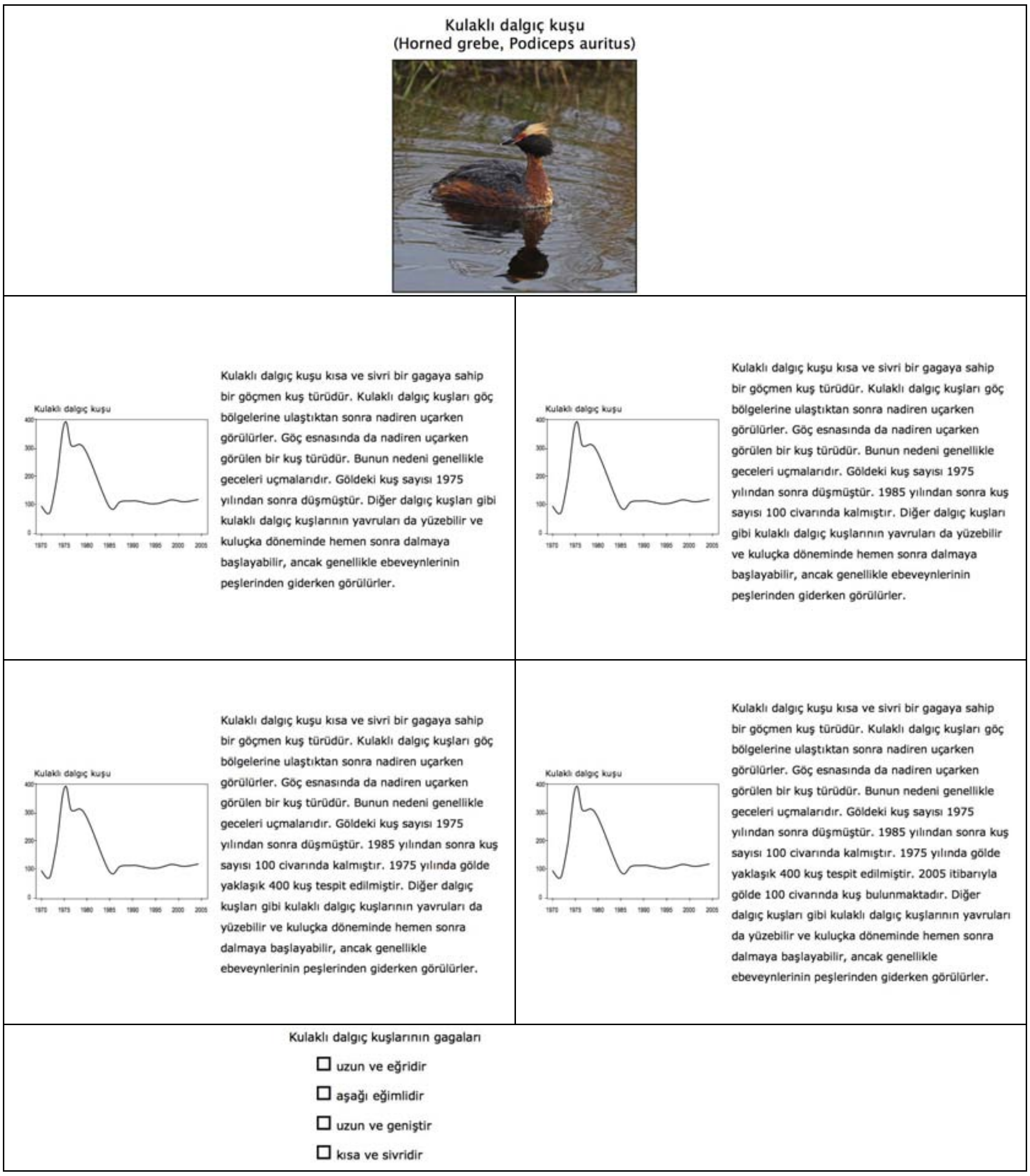
The participants were presented 12 graph-text constellations. Each presentation was composed of three screens. In the first screen, a photo of the bird is presented. In the second, one of the four the graph-text constellation is presented. In the figure below, four graph-text constellations (in Turkish) are shown with four target sentence conditions (TC-1 at the top-left, TC-2 at the top-right, TC-3 at the bottom-left, TC-4 at the bottom-right). Finally, a multiple-choice posttest question is presented. The following figure exemplifies material from the graph condition GC-1.



The figure below exemplifies the stimuli from the graph condition GC-2.

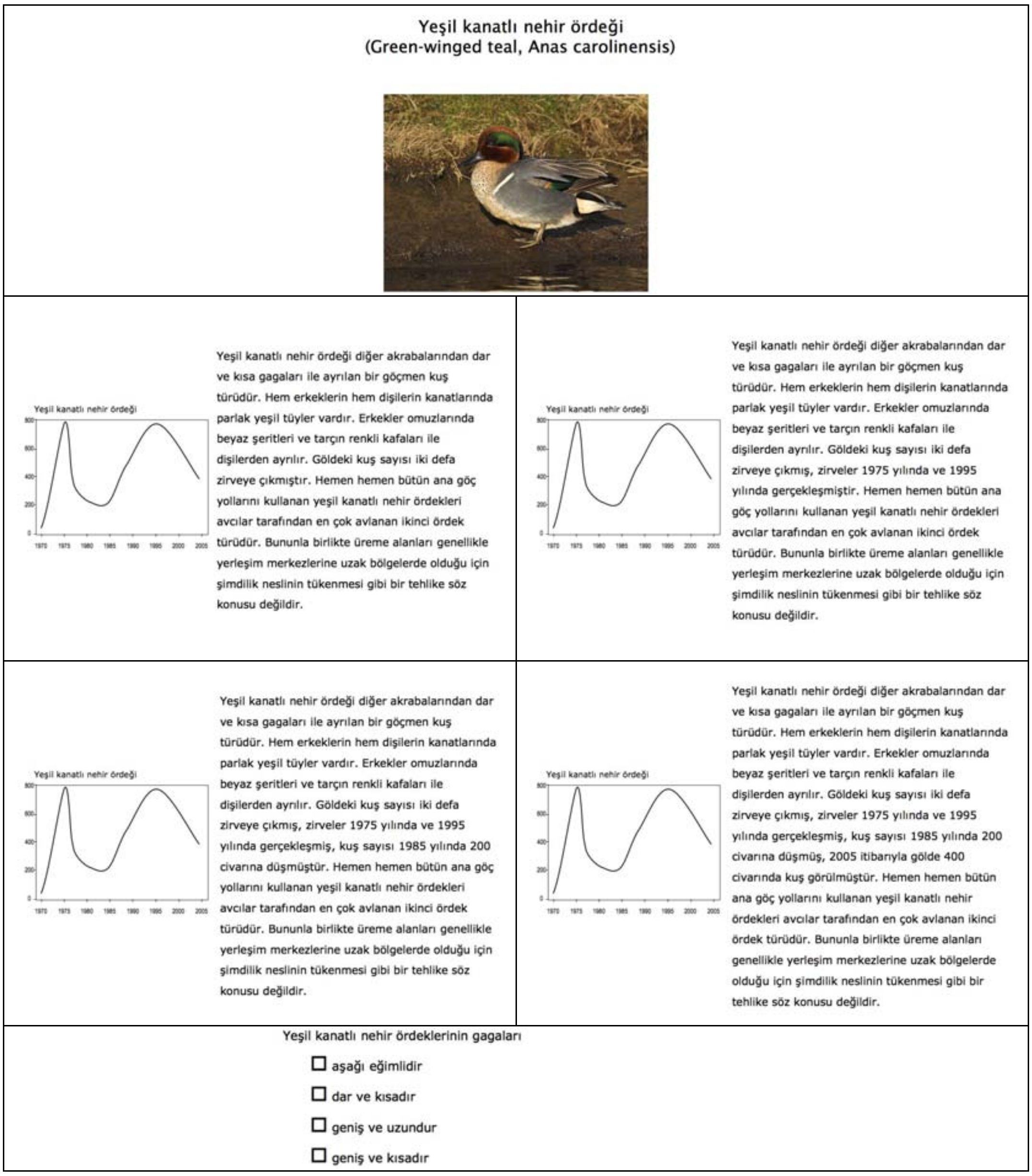


The figure below exemplifies the stimuli from the graph condition GC-3.



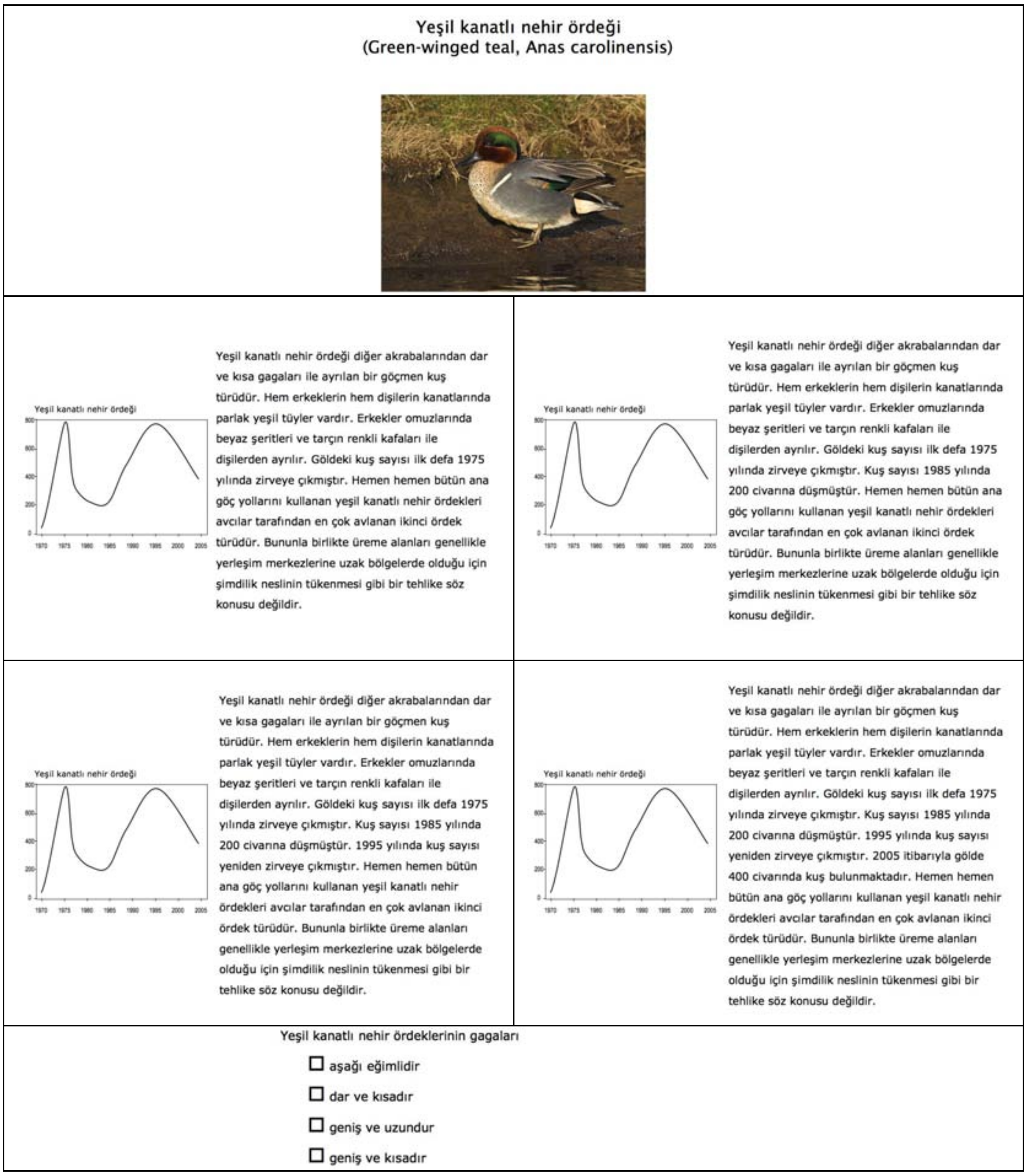
B.1.4 Sample Stimuli from Experiment 2

The figure below exemplifies the stimuli from the graph condition GC-2 in Experiment 2. The difference between Experiment 2 and Experiment 1 was that the target sentences involved suspended affixation as coordinate structure.



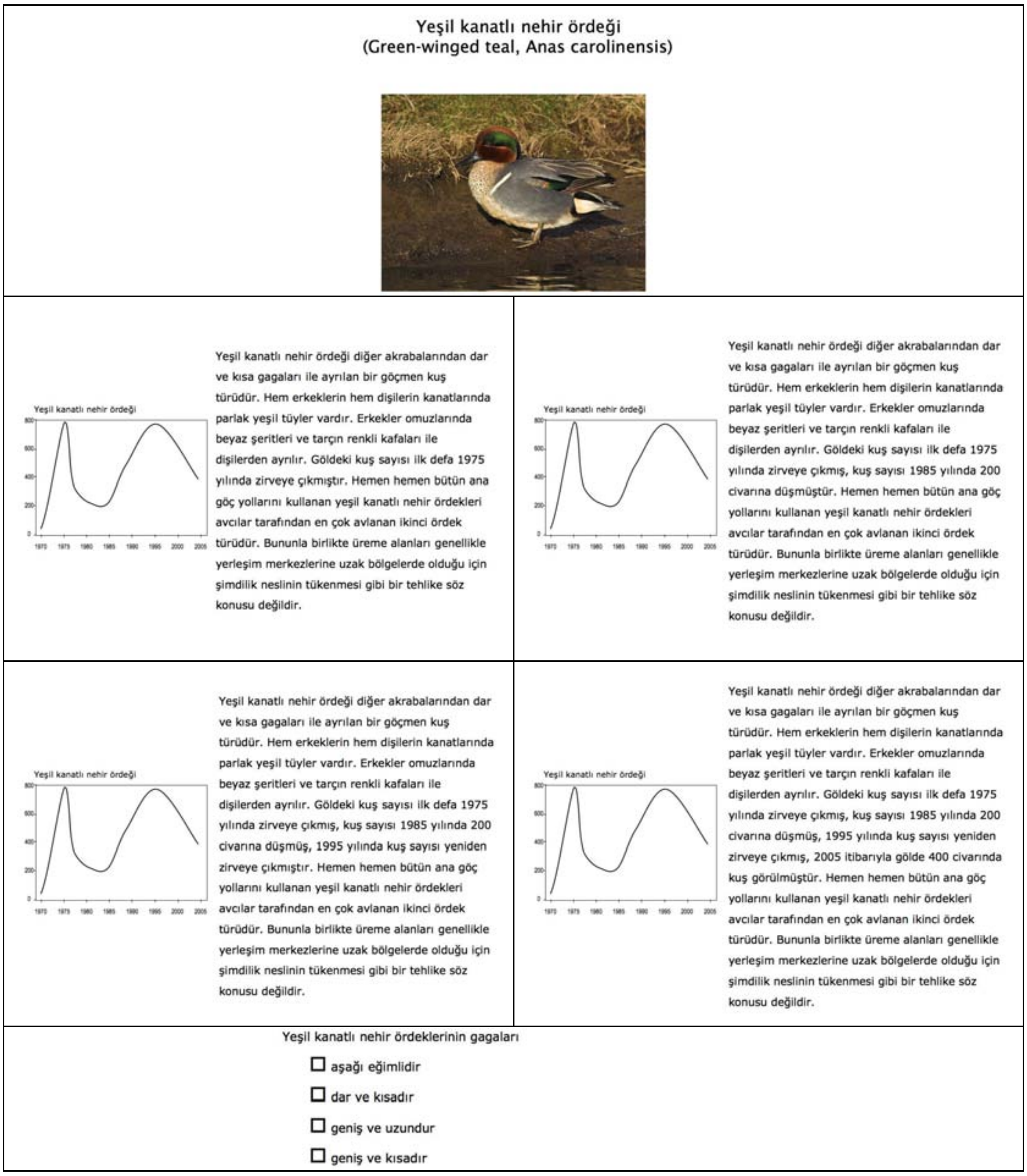
B.1.5 Sample Stimuli from Experiment 3

The figure below exemplifies the stimuli from the graph condition GC-2 in Experiment 3. In Experiment 3, the target sentences were rearranged such that the temporal order of the events and processes was consistent with the order of the target sentences in the text.



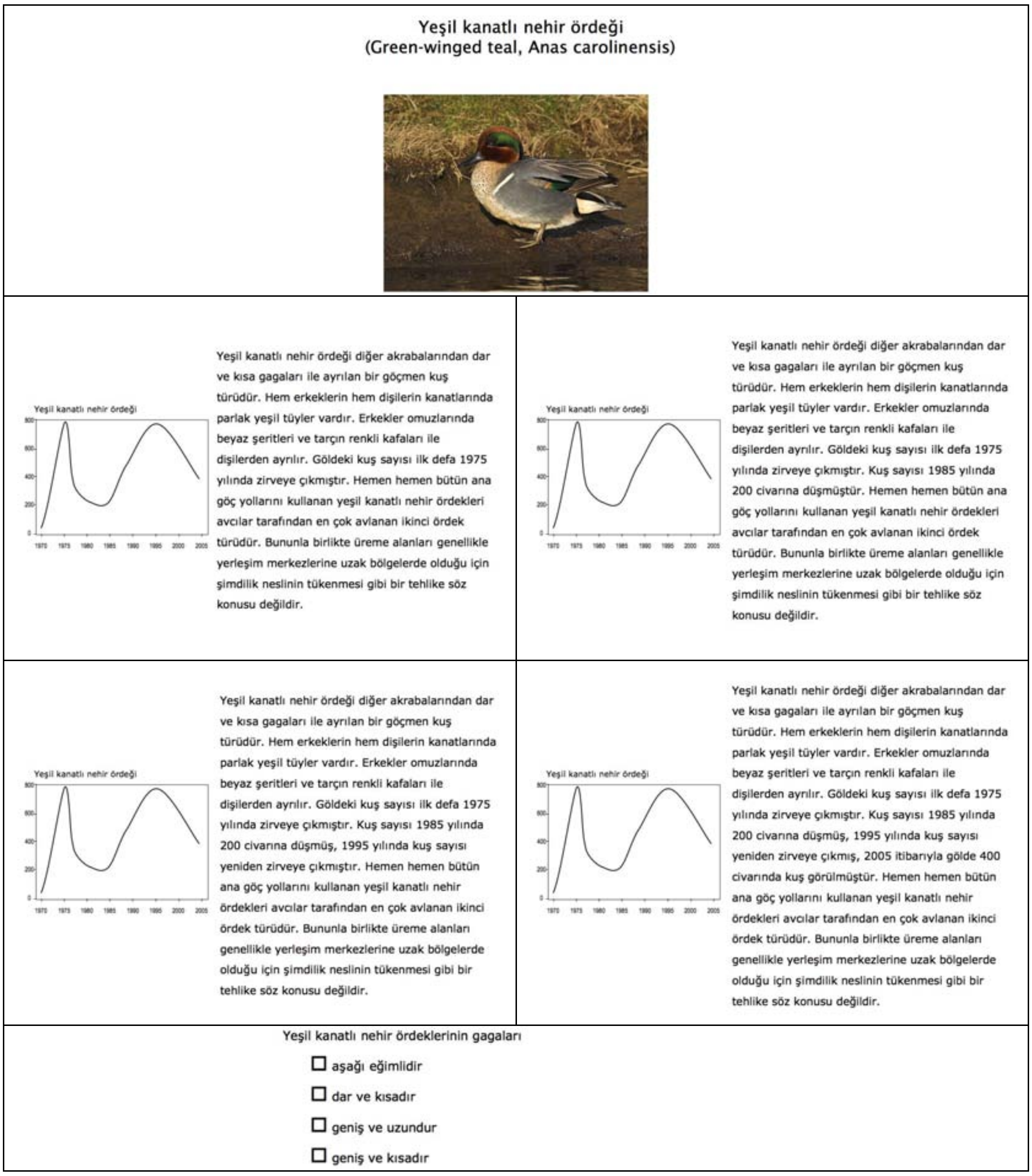
B.1.6 Sample Stimuli from Experiment 4

The figure below exemplifies the stimuli from the graph condition GC-2 in Experiment 4. The difference between Experiment 4 and Experiment 3 was that the target sentences involved suspended affixation as coordinate structure.



B.1.7 Sample Stimuli from Experiment 5

The figure below exemplifies the stimuli from the graph condition GC-2 in Experiment 5. The difference between Experiment 5 and Experiment 4 was that the target sentences involved a partial coordinate structure in Experiment 5 instead of a full coordinate structure.



B.2 APPENDIX TO THE ANALYSIS OF THE RESULTS

The following parts from B.2.1 to B.2.5 show the numerical details of the Experiments 1 to Experiment 5 respectively.

B.2.1 Experiment 1

Table B.1. The mean number of gaze shifts into the graph layer in the four target sentence conditions and the three graph conditions in Experiment 1. TC is the abbreviation for 'target sentence condition'. The numbers in parentheses show standard deviations.

	TC-1	TC-2	TC-3	TC-4
Graph Condition 1	0.27 (0.45)	0.75 (0.90)	1.41 (1.16)	1.40 (1.34)
Graph Condition 2	0.43 (0.48)	0.52 (0.69)	1.52 (1.48)	2.00 (1.39)
Graph Condition 3	0.52 (0.51)	1.07 (1.00)	1.63 (1.54)	1.90 (1.59)

The numbers given in Table B.1 are depicted in the graphs in Figure B.1.

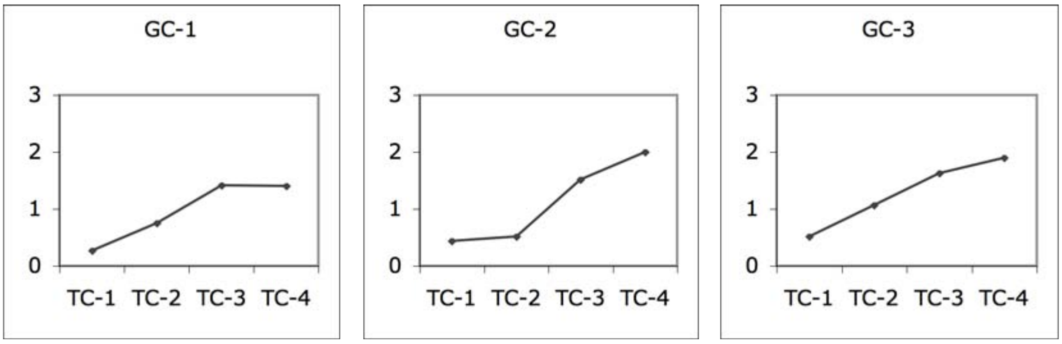


Figure B.1. The mean number of gaze shifts into the graph layer in Experiment 1. The graphs show the results for the target sentence conditions (TC) and the graph conditions (GC).

Pairwise comparisons revealed that in graph condition GC-1, the difference between TC-1 and TC-2, and the difference between TC-2 and TC-3 were significant. However, the difference between TC-3 and TC-4 was not significant. In graph condition GC-2, the difference between TC-2 and TC-3 was significant, however the difference between TC-1 and TC-2 and the difference between TC-3 and TC-4 were not significant. In GC-3, the difference between TC-1 and TC-2 was significant, however, the difference between TC-2 and TC-3 and the difference between TC-3 and TC-4 were not significant.

Table B.2 shows the mean number of gaze shifts after the target sentences for all experimental conditions.

Table B.2. The mean number of gaze shifts into the graph layer during participants' reading of the target sentences in the four target sentence conditions and the three graph conditions in Experiment 1. TC is the abbreviation for 'target sentence condition', GC for 'graph condition', and S for 'target sentence'. The numbers in parentheses show standard deviations.

	TC-1	TC-2		TC-3			TC-4			
	S1	S1	S2	S1	S2	S3	S1	S2	S3	S4
GC-1	0.33 (0.60)	0.31 (0.47)	0.44 (0.62)	0.47 (0.57)	0.24 (0.44)	0.70 (0.70)	0.27 (0.45)	0.27 (0.45)	0.67 (0.66)	0.20 (0.41)
GC-2	0.47 (0.57)	0.31 (0.47)	0.21 (0.49)	0.45 (0.51)	0.55 (0.68)	0.52 (0.68)	0.38 (0.49)	0.45 (0.57)	0.66 (0.72)	0.52 (0.63)
GC-3	0.64 (0.70)	0.50 (0.57)	0.57 (0.73)	0.53 (0.62)	0.63 (0.75)	0.47 (0.51)	0.52 (0.57)	0.66 (0.77)	0.38 (0.62)	0.34 (0.55)

Table B.3 shows the results for the answers to the posttest questions.

Table B.3. The results for the posttest scores in Experiment 1. The numbers in parentheses show standard deviations.

	Target Sentence Conditions TC-1 and TC-2	Target Sentence Conditions TC-3 and TC-4
Appearance	0.41 (0.92)	0.42 (0.91)
Feeding	0.30 (0.96)	0.74 (0.68)
Graph	0.85 (0.53)	0.96 (0.27)

B.2.2 Experiment 2

Table B.4. The mean number of gaze shifts into the graph layer in the four target sentence conditions and the three graph conditions in Experiment 2. TC is the abbreviation for 'target sentence condition'. The numbers in parentheses show standard deviations.

	TC-1	TC-2	TC-3	TC-4
Graph Condition 1	0.25 (0.44)	0.71 (0.91)	1.54 (1.32)	1.58 (1.59)
Graph Condition 2	0.38 (0.49)	0.96 (1.00)	1.52 (1.66)	1.96 (2.10)
Graph Condition 3	0.50 (0.51)	1.67 (1.40)	1.71 (1.46)	2.23 (2.28)

The numbers given in Table B.4 are depicted in the graphs in Figure B.2.

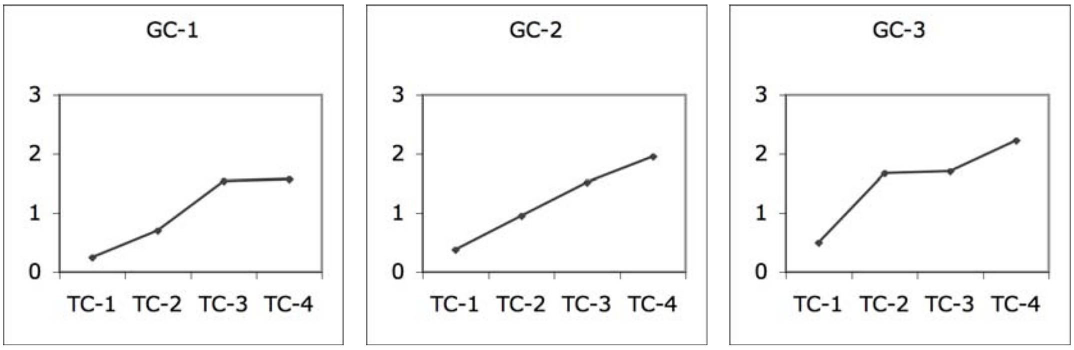


Figure B.2. The mean number of gaze shifts into the graph layer in Experiment 2. The graphs show the results for the target sentence conditions (TC) and the graph conditions (GC).

Pairwise comparisons revealed that in graph condition GC-1, the difference between TC-1 and TC-2, and the difference between TC-2 and TC-3 were significant. However, the difference between TC-3 and TC-4 was not significant. In graph condition GC-2, the difference between TC-1 and TC-2 was significant, however the difference between TC-2 and TC-3 and the difference between TC-3 and TC-4 were not significant. In GC-3, the difference between TC-1 and TC-2 and the difference between TC-3 and TC-4 were significant, however, the difference between TC-2 and TC-3 was not significant.

Table B.5 shows the mean number of gaze shifts after the target sentences for all experimental conditions.

Table B.5. The mean number of gaze shifts into the graph layer layer during participants' reading of the target sentences in the four target sentence conditions and the three graph conditions in Experiment 2. TC is the abbreviation for 'target sentence condition', GC for 'graph condition', and S for 'target sentence'. The numbers in parentheses show standard deviations.

	TC-1	TC-2		TC-3			TC-4			
	S1	S1	S2	S1	S2	S3	S1	S2	S3	S4
GC-1	0.33 (0.64)	0.13 (0.45)	0.58 (0.83)	0.29 (0.46)	0.50 (0.59)	0.75 (0.85)	0.33 (0.64)	0.33 (0.56)	0.33 (0.48)	0.58 (0.58)
GC-2	0.46 (0.66)	0.30 (0.47)	0.65 (0.83)	0.30 (0.47)	0.57 (0.90)	0.65 (1.11)	0.22 (0.42)	0.39 (0.66)	0.61 (0.66)	0.74 (1.29)
GC-3	0.67 (0.82)	0.75 (0.74)	0.92 (1.06)	0.33 (0.48)	0.67 (0.76)	0.71 (0.86)	0.41 (0.59)	0.77 (1.02)	0.36 (0.58)	0.68 (0.78)

Table B.6 shows the results for the answers to the posttest questions.

Table B.6. The results for the posttest scores in Experiment 2. The numbers in parentheses show standard deviations.

	Target Sentence Conditions TC-1 and TC-2	Target Sentence Conditions TC-3 and TC-4
Appearance	0.36 (0.94)	0.28 (0.97)
Feeding	0.44 (0.91)	0.44 (0.91)
Graph	0.83 (0.56)	0.74 (0.68)

B.2.3 Experiment 3

Table B.7. The mean number of gaze shifts into the graph layer in the four target sentence conditions and the three graph conditions in Experiment 3. TC is the abbreviation for 'target sentence condition'. The numbers in parentheses show standard deviations.

	TC-1	TC-2	TC-3	TC-4
Graph Condition 1	0.48 (0.67)	0.65 (0.83)	1.39 (1.20)	1.43 (1.70)
Graph Condition 2	0.74 (0.54)	1.23 (1.13)	1.83 (1.19)	2.26 (1.54)
Graph Condition 3	0.65 (0.88)	1.09 (1.12)	1.87 (1.52)	2.32 (1.92)

The numbers given in Table B.7 are depicted in the graphs in Figure B.3.

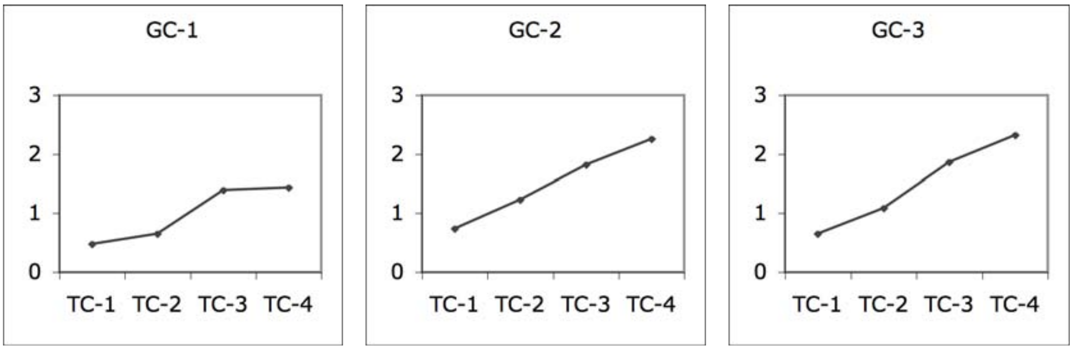


Figure B.3. The mean number of gaze shifts into the graph layer in Experiment 3. The graphs show the results for the target sentence conditions (TC) and the graph conditions (GC).

Pairwise comparisons revealed that in graph condition GC-1, the difference between TC-2 and TC-3 was significant. However, the difference between TC-1 and TC-2 and the difference between TC-3 and TC-4 were not significant. In graph condition GC-2, the difference between TC-2 and TC-3 was significant, however the difference between TC-1 and TC-2 and the difference between TC-3 and TC-4 were not significant. In GC-3, the difference between TC-1 and TC-2 and the difference between TC-2 and TC-3 were significant, however, the difference between TC-3 and TC-4 was not significant.

Table B.8 shows the mean number of gaze shifts after the target sentences for all experimental conditions.

Table B.8. The mean number of gaze shifts into the graph layer layer during participants' reading of the target sentences in the four target sentence conditions and the three graph conditions in Experiment 3. TC is the abbreviation for 'target sentence condition', GC for 'graph condition', and S for 'target sentence'. The numbers in parentheses show standard deviations.

	TC-1	TC-2		TC-3			TC-4			
	S1	S1	S2	S1	S2	S3	S1	S2	S3	S4
GC-1	0.48 (0.67)	0.35 (0.49)	0.30 (0.56)	0.43 (0.59)	0.26 (0.45)	0.70 (0.63)	0.26 (0.54)	0.13 (0.34)	0.57 (0.73)	0.48 (0.67)
GC-2	0.74 (0.54)	0.73 (0.88)	0.50 (0.51)	0.78 (0.60)	0.70 (0.93)	0.35 (0.57)	0.65 (0.65)	0.74 (0.81)	0.30 (0.47)	0.57 (0.66)
GC-3	0.65 (0.88)	0.70 (0.93)	0.39 (0.58)	0.78 (0.74)	0.43 (0.59)	0.65 (0.71)	0.73 (0.94)	0.41 (0.50)	0.82 (0.85)	0.36 (0.49)

Table B.9 shows the results for the answers to the posttest questions.

Table B.9. The results for the posttest scores in Experiment 3. The numbers in parentheses show standard deviations.

	Target Sentence Conditions TC-1 and TC-2	Target Sentence Conditions TC-3 and TC-4
Appearance	0.46 (0.90)	0.41 (0.92)
Feeding	0.51 (0.87)	0.54 (0.85)
Graph	0.78 (0.63)	0.83 (0.57)

B.2.4 Experiment 4

Table B.10. The mean number of gaze shifts into the graph layer in the four target sentence conditions and the three graph conditions in Experiment 4. TC is the abbreviation for 'target sentence condition'. The numbers in parentheses show standard deviations.

	TC-1	TC-2	TC-3	TC-4
Graph Condition 1	0.60 (0.58)	0.68 (0.90)	1.52 (1.64)	2.20 (2.18)
Graph Condition 2	0.72 (0.68)	1.44 (1.47)	1.88 (1.59)	3.08 (2.50)
Graph Condition 3	0.80 (0.76)	1.32 (1.57)	1.72 (1.67)	2.52 (2.69)

The numbers given in Table B.10 are depicted in the graphs in Figure B.4.

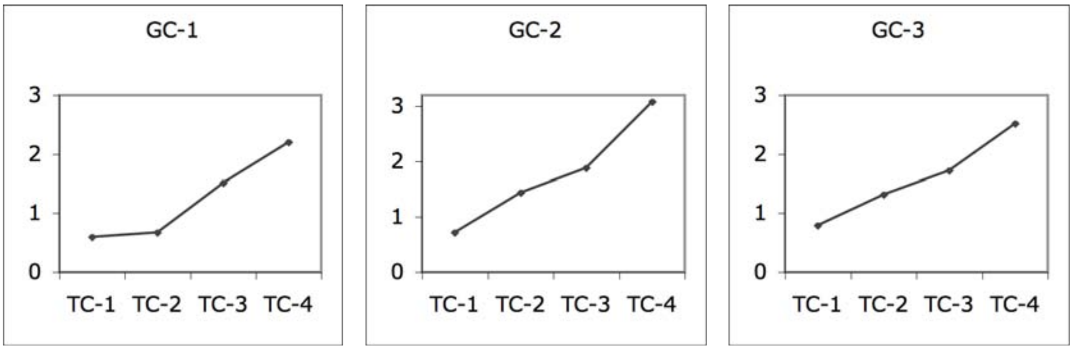


Figure B.4. The mean number of gaze shifts into the graph layer in Experiment 4. The graphs show the results for the target sentence conditions (TC) and the graph conditions (GC).

Pairwise comparisons revealed that in graph condition GC-1, the difference between TC-2 and TC-3 was significant. However, the difference between TC-1 and TC-2 and the difference between TC-3 and TC-4 were not significant. In graph condition GC-2, the difference between TC-1 and TC-2 and the difference between TC-3 and TC-4 was significant, however the difference between TC-2 and TC-3 was not significant. In GC-3, there were no significant differences among the means.

Table B.11 shows the mean number of gaze shifts after the target sentences for all experimental conditions.

Table B.11. The mean number of gaze shifts into the graph layer layer during participants' reading of the target sentences in the four target sentence conditions and the three graph conditions in Experiment 4. TC is the abbreviation for 'target sentence condition', GC for 'graph condition', and S for 'target sentence'. The numbers in parentheses show standard deviations.

	TC-1	TC-2		TC-3			TC-4			
	S1	S1	S2	S1	S2	S3	S1	S2	S3	S4
GC-1	0.60 (0.58)	0.52 (0.71)	0.16 (0.37)	0.44 (0.65)	0.24 (0.52)	0.84 (1.03)	0.28 (0.46)	0.40 (0.65)	0.92 (1.35)	0.60 (1.29)
GC-2	0.72 (0.68)	0.60 (0.76)	0.84 (0.94)	0.36 (0.57)	0.96 (0.98)	0.56 (0.71)	0.68 (0.63)	0.68 (0.85)	0.84 (1.21)	0.88 (0.67)
GC-3	0.80 (0.76)	0.64 (0.91)	0.68 (0.85)	0.52 (0.77)	0.48 (0.65)	0.72 (1.02)	0.72 (0.84)	0.44 (0.92)	0.48 (0.59)	0.88 (1.01)

Table B.12 shows the results for the answers to the posttest questions.

Table B.12. The results for the posttest scores in Experiment 4. The numbers in parentheses show standard deviations.

	Target Sentence Conditions TC-1 and TC-2	Target Sentence Conditions TC-3 and TC-4
Appearance	0.52 (0.86)	0.15 (1.00)
Feeding	0.51 (0.87)	0.58 (0.83)
Graph	0.74 (0.69)	0.95 (0.33)

B.2.5 Experiment 5

Table B.13. The mean number of gaze shifts into the graph layer in the four target sentence conditions and the three graph conditions in Experiment 5. TC is the abbreviation for 'target sentence condition'. The numbers in parentheses show standard deviations.

	TC-1	TC-2	TC-3	TC-4
Graph Condition 1	0.36 (0.67)	0.64 (0.81)	1.64 (1.43)	1.91 (2.12)
Graph Condition 2	0.73 (0.79)	1.27 (1.10)	1.60 (1.11)	2.18 (2.48)
Graph Condition 3	1.00 (0.89)	1.09 (0.94)	2.27 (1.27)	1.91 (1.22)

The numbers given in Table B.13 are depicted in the graphs in Figure B.5.

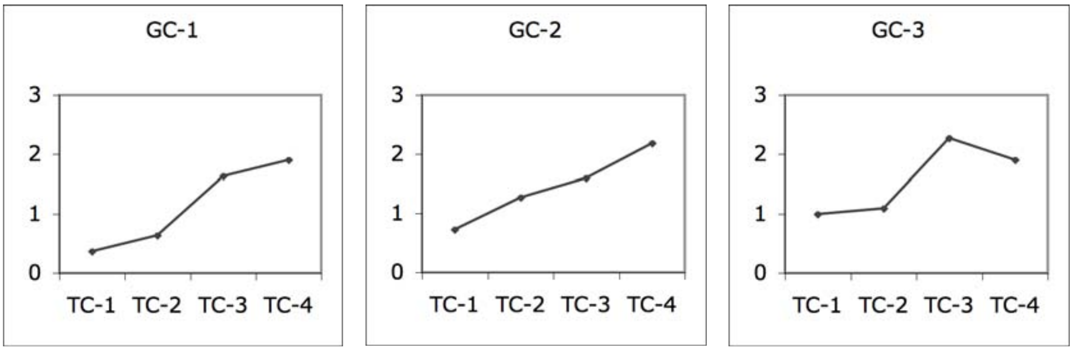


Figure B.5. The mean number of gaze shifts into the graph layer in Experiment 5. The graphs show the results for the target sentence conditions (TC) and the graph conditions (GC).

Pairwise comparisons revealed that in the graph type conditions GC-1 and GC-3, the difference between TC-2 and TC-3 was significant; however, the difference between TC-1 and TC-2 and the difference between TC-3 and TC-4 were not significant. In GC-2, the difference between TC-1 and TC-2, the difference between TC-2 and TC-3, and the difference between TC-3 and TC-4 were not significant.

Table B.14 shows the mean number of gaze shifts after the target sentences for all experimental conditions.

Table B.14. The mean number of gaze shifts into the graph layer during participants' reading of the target sentences in the four target sentence conditions and the three graph conditions in Experiment 5. TC is the abbreviation for 'target sentence condition', GC for 'graph condition', and S for 'target sentence'. The numbers in parentheses show standard deviations.

	TC-1	TC-2		TC-3			TC-4			
	S1	S1	S2	S1	S2	S3	S1	S2	S3	S4
GC-1	0.36 (0.67)	0.45 (0.69)	0.18 (0.40)	0.45 (0.52)	0.36 (0.50)	0.82 (0.98)	0.36 (0.50)	0.18 (0.60)	0.73 (1.01)	0.64 (1.03)
GC-2	0.73 (0.79)	0.64 (0.50)	0.64 (0.67)	0.70 (0.48)	0.40 (0.52)	0.50 (0.53)	0.73 (0.79)	0.55 (0.82)	0.55 (0.82)	0.36 (0.50)
GC-3	1.00 (0.89)	0.91 (0.94)	0.18 (0.40)	0.64 (0.81)	0.64 (0.50)	1.00 (1.10)	0.73 (0.90)	0.27 (0.47)	0.64 (0.67)	0.27 (0.47)

Table B.15 shows the results for the answers to the posttest questions.

Table B.15. The results for the posttest scores in Experiment 5. The numbers in parentheses show standard deviations.

	Target Sentence Conditions TC-1 and TC-2	Target Sentence Conditions TC-3 and TC-4
Appearance	0.39 (0.93)	0.28 (0.97)
Feeding	0.33 (0.97)	0.44 (0.92)
Graph	0.78 (0.65)	0.56 (0.86)

Appendix C

C.1 A SAMPLE STORY PRESENTED TO THE PARTICIPANTS IN STUDY 2

A sample story presented to the participants as the context information in Study 2 is shown below (translated into English by the author of the dissertation):

> In this part of the study, you play the role of a chemist. You are making research on a recently discovered fluid. You decide to investigate the relation between the liquid and different elements by studying temperature change.
>
> For this purpose, you prepare tubes filled with the liquid, and by adding small amount of substances into the tubes; you record the change in temperature with devices which measure the temperature continuously. The devices plot the temperature change graphs on the screen.
>
> Please investigate the graph which shows the 60 minute change in temperature and interpret what happened. After the interpretation of the graph, bring the question to the screen and make your evaluation using the scale.

C.2 APPENDIX TO THE ANALYSIS OF THE RESULTS IN STUDY 2

C.2.1 Mean Gaze Times

The mean gaze times of the participants on the graph AOI and the Q-VAS AOI, and the gaze patterns on the graph line are presented below.

C.2.1.1 Left-Annotation Condition

Table C.1. Mean gaze times of the participants in the left-annotation condition, GC-1. The numbers are in milliseconds. SD is used for standard deviation.

	Graph AOI				Q-VAS AOI			
	Group A		Group B		Group A		Group B	
	M	SD	M	SD	M	SD	M	SD
Q-1	1603	1525	4144	6317	4527	2763	6486	5235
Q-2	2445	3697	4245	4504	6935	4593	6878	3146
Q-3	1237	1659	2280	1871	4461	2226	5448	2450

Table C.2. Mean gaze times of the participants in the left-annotation condition, GC-2. The numbers are in milliseconds.

	Graph AOI				Q-VAS AOI			
	Group A		Group B		Group A		Group B	
	M	SD	M	SD	M	SD	M	SD
Q-1	3116	3435	3829	2884	5289	2260	5466	3843
Q-2	1329	1152	2195	3540	4921	2700	5549	2517
Q-3	1079	1206	1828	1677	3456	1702	4376	2455

Table C.3. Mean gaze times of the participants in the left-annotation condition, GC-3. The numbers are in milliseconds.

	Graph AOI				Q-VAS AOI			
	Group A		Group B		Group A		Group B	
	M	SD	M	SD	M	SD	M	SD
Q-1	1558	1732	2435	1891	5640	4059	4682	2783
Q-2	1829	2046	2681	1586	5723	4132	4842	2429
Q-3	1765	1626	2862	2554	5347	2818	4398	4011

C.2.1.2 Middle-Annotation Condition

Table C.4. Mean gaze times of the participants in the middle-annotation condition, GC-1. The numbers are in milliseconds.

	Graph AOI				Q-VAS AOI			
	Group A		Group B		Group A		Group B	
	M	SD	M	SD	M	SD	M	SD
Q-1	1204	1318	1563	1504	4083	2309	3370	2035
Q-2	816	872	1799	1426	3710	2009	3350	1761
Q-3	1014	978	1562	1043	3656	2432	3028	1496

Table C.5. Mean gaze times of the participants in the middle-annotation condition, GC-2. The numbers are in milliseconds.

	Graph AOI				Q-VAS AOI			
	Group A		Group B		Group A		Group B	
	M	SD	M	SD	M	SD	M	SD
Q-1	1111	1152	1847	1843	4230	2883	3559	2103
Q-2	1019	772	1728	2268	2519	1370	3375	2330
Q-3	1098	666	1561	1132	2584	1715	3177	1308

Table C.6. Mean gaze times of the participants in the middle-annotation condition, GC-3. The numbers are in milliseconds.

	Graph AOI				Q-VAS AOI			
	Group A		Group B		Group A		Group B	
	M	SD	M	SD	M	SD	M	SD
Q-1	1208	1460	1779	1695	4380	2872	3872	1514
Q-2	803	861	2286	1395	4849	2789	3528	1601
Q-3	737	647	962	592	3142	1712	2623	1353

C.2.1.3 Double-Annotation Condition

Table C.7. Mean gaze times of the participants in the double-annotation condition, GC-1. The numbers are in milliseconds.

	Graph AOI				Q-VAS AOI			
	Group A		Group B		Group A		Group B	
	M	SD	M	SD	M	SD	M	SD
Q-1	1399	1158	2460	2156	3121	1357	2808	1748
Q-2	1396	992	3329	3571	5385	2474	3777	2626
Q-3	736	974	2987	3312	3217	1684	3002	1202
Q-4	749	722	2543	1989	3392	1353	3167	1778

Table C.8. Mean gaze times of the participants in the double-annotation condition, GC-2. The numbers are in milliseconds.

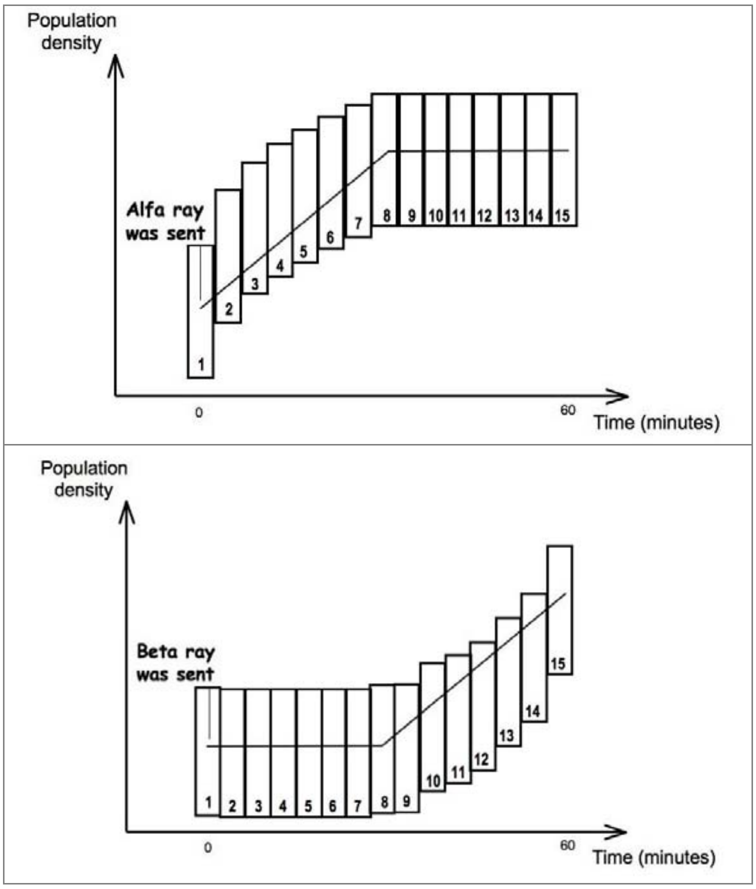
	Graph AOI				Q-VAS AOI			
	Group A		Group B		Group A		Group B	
	M	SD	M	SD	M	SD	M	SD
Q-1	1523	1362	2526	2066	3349	1640	3559	1617
Q-2	1383	2026	1821	1169	3588	1145	3457	1950
Q-3	742	706	1721	1354	3065	1613	2517	1478
Q-4	684	532	1488	820	3642	2031	3764	1895

Table C.9. Mean gaze times of the participants in the double-annotation condition, GC-3. The numbers are in milliseconds.

	Graph AOI				Q-VAS AOI			
	Group A		Group B		Group A		Group B	
	M	SD	M	SD	M	SD	M	SD
Q-1	1143	951	2445	2150	2861	1378	2906	1849
Q-2	1615	1361	2140	1444	3715	1372	3552	1612
Q-3	980	642	2873	2998	3732	1892	3083	2302
Q-4	598	382	1620	1607	3586	1775	3325	2115

C.2.2 Gaze Distributions on the Graph Lines

The purpose of the analysis of gaze distributions is to investigate how the specific questions influence inspection of the graph lines. For the analysis of gaze distributions, three AOI templates were prepared, one for each graph type condition GC-1, GC-2, and GC-3. Each AOI template consisted of 15 equal-size rectangles, numbered from left to right, as exemplified in Figure C.1. The same template was used for both Group A and Group B stimuli, in all the three annotation conditions.



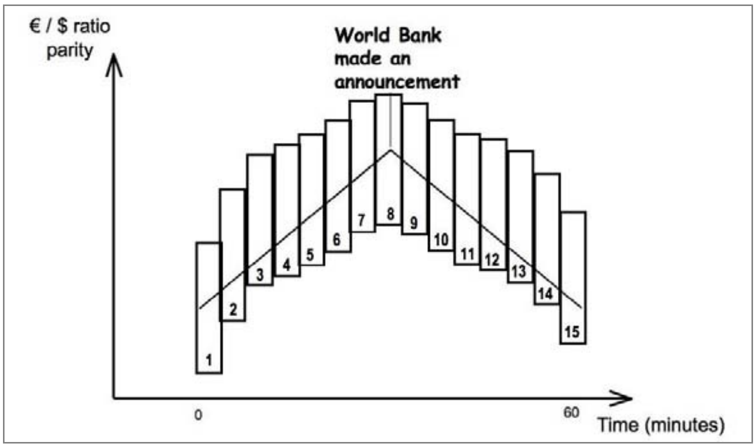
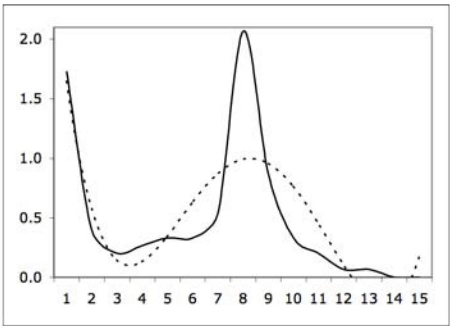
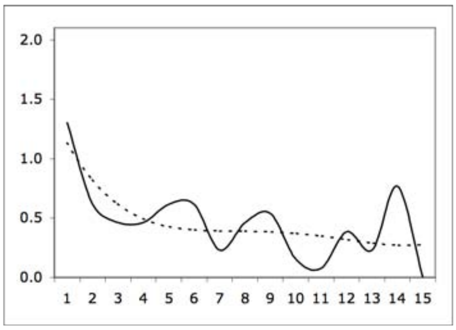
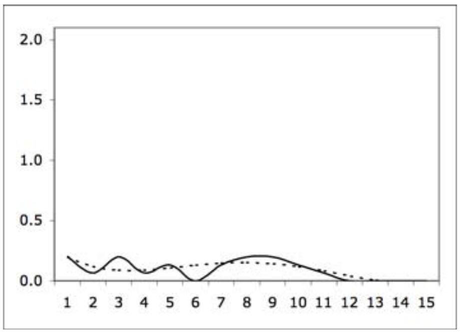
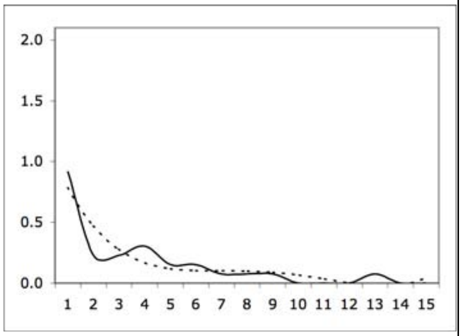


Figure C.1. The AOI templates for GC-1, GC-2, and GC-3.

C.2.2.1 Left-Annotation Condition

Figure C.2 shows the distribution of mean fixation counts among the 15 AOIs in the left-annotation condition, in GC-1. The participants were presented the annotated graphs before the questions. The first row in the figure (namely, No-Q) shows the distribution of fixation counts in the absence of a question. The second row, labeled Q-1 shows the distribution on the AOIs after the first question appeared on the screen. Accordingly, the next two rows show the distribution for Q-2 and Q-3. The dotted lines show the fourth-order polynomial estimation of the curve.

	Group A	Group B
No-Q		
Q-1		

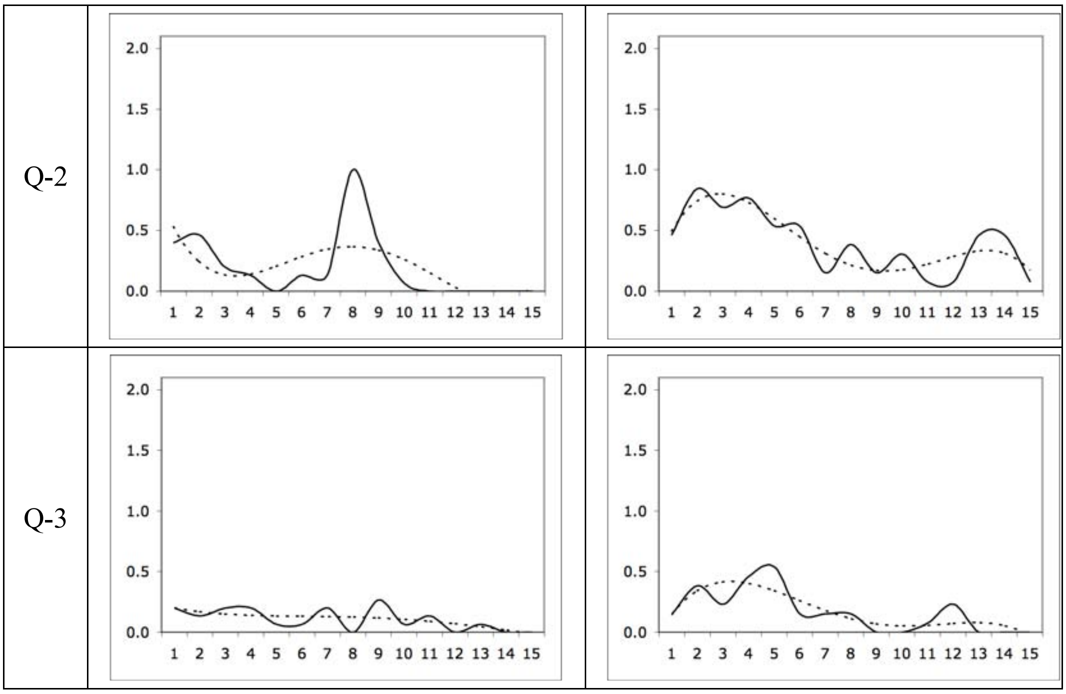


Figure C.2. Distribution of mean fixation counts on the AOIs in GC-1 in the left-annotation condition. The dotted lines show the fourth-order polynomial curve estimation.

The distributions of the mean fixation counts show how the questions changed the distributions of the fixations on the graphical entities. In the first presentation of the annotated graphs without a question (the No-Q row), the mean fixation counts were high for the left end and the middle of the graph line in Group A, whereas the mean fixation counts exhibited a decreasing trend from left to right in Group B. After the presentation of the first question, the mean fixation counts were low in Group A. This shows that few participants in Group A inspected the graph after the presentation of Q-1. In Group B, the fixations were shifted to the left of the graph, with a similar decreasing trend of distribution to the No-Q case. After the presentation of the next question, the participants in Group A inspected the left end and the middle part of the graph line, with more fixations at the middle. In Group B, the participants made more inspections on the left half of the graph than the right half. After the presentation of Q-3, the mean fixations decreased in Group A, compared to Q-2. The Q-3 distribution in Group B was similar to Q-2 distribution with reduced means.

The fixation distributions for the other conditions are presented below.

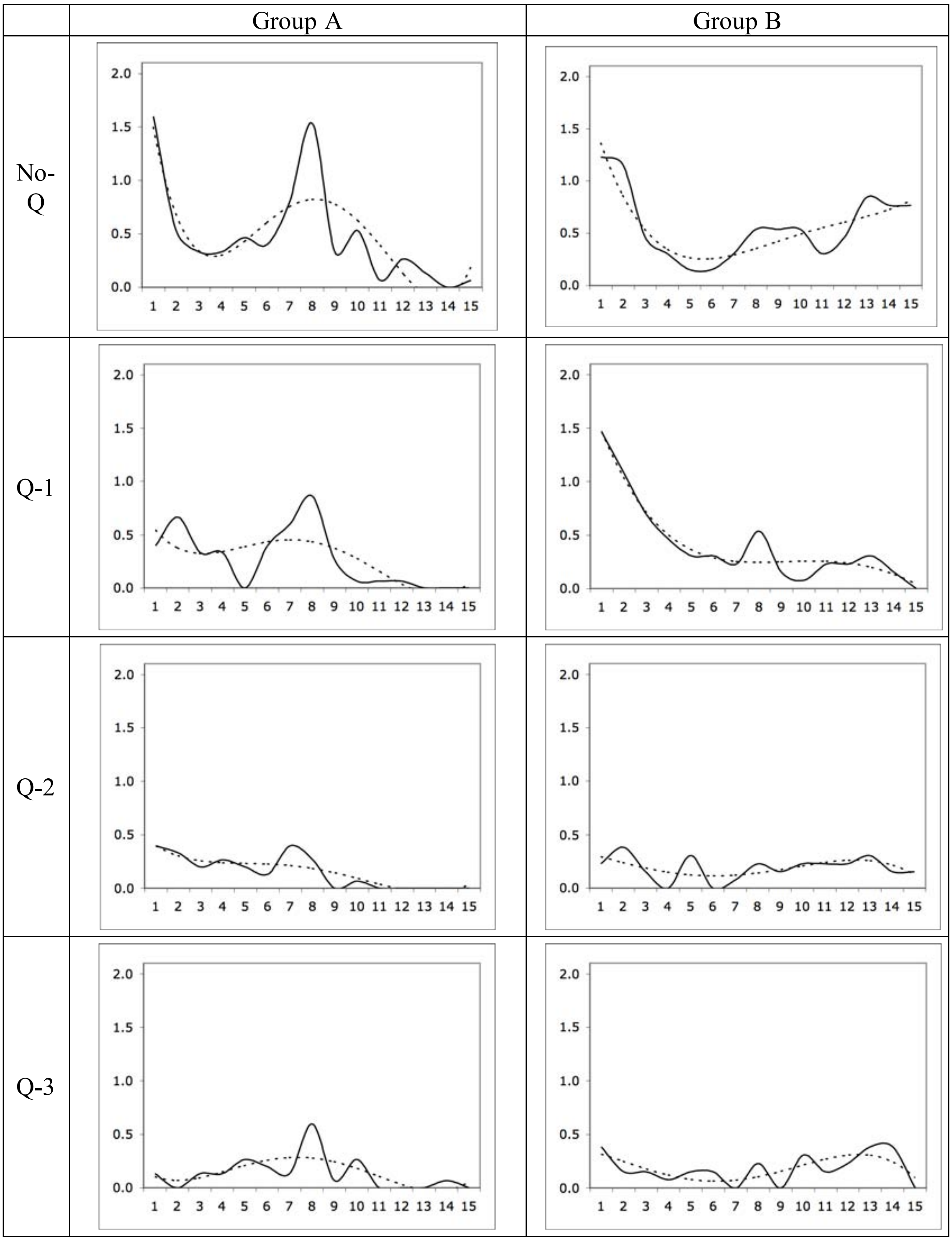


Figure C.3. Distribution of mean fixation counts on the AOIs in GC-2 in the left-annotation condition. The dotted lines show the fourth-order polynomial curve estimation.

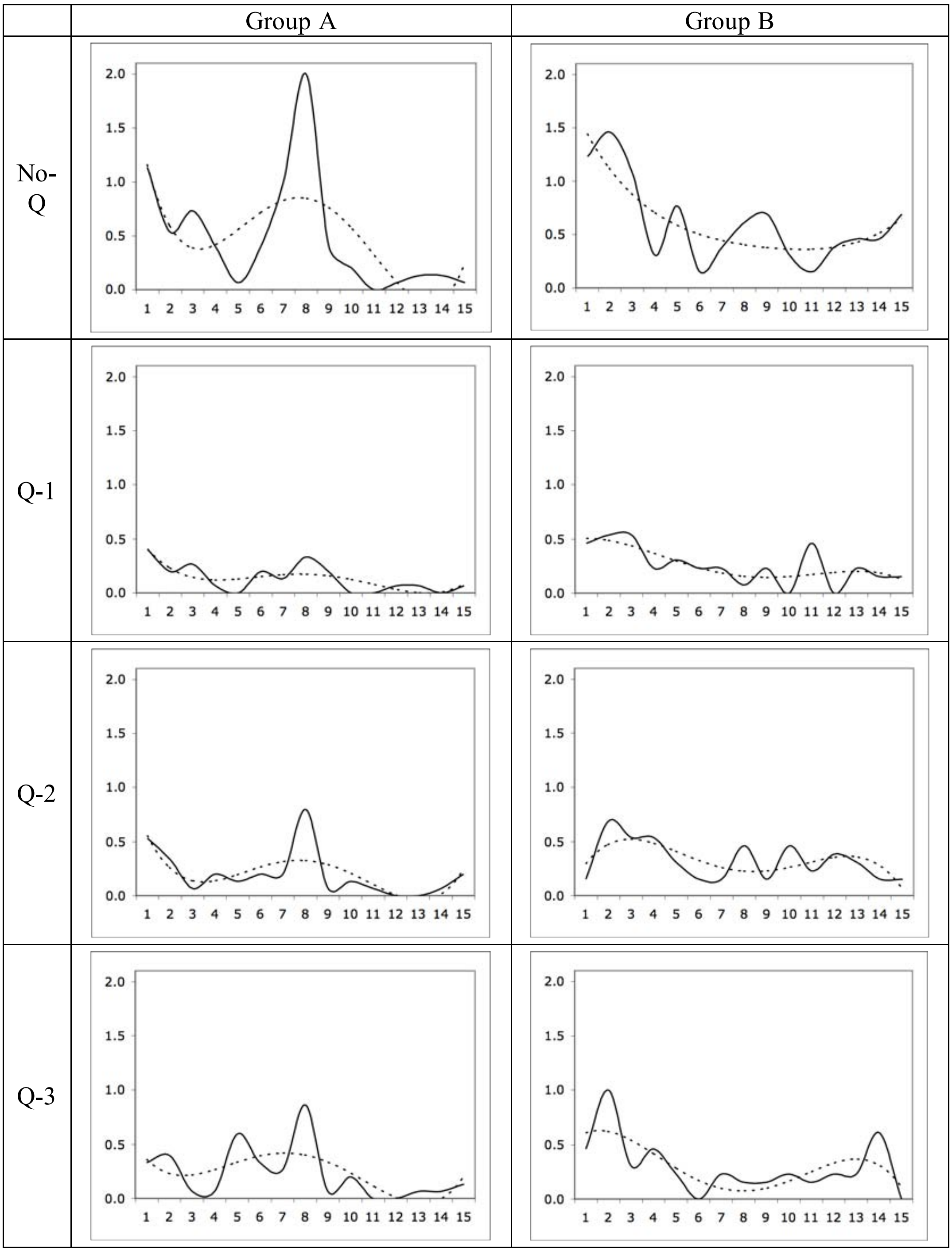


Figure C.4. Distribution of mean fixation counts on the AOIs in GC-3 in the left-annotation condition. The dotted lines show the fourth-order polynomial curve estimation.

C.2.2.2 Middle-Annotation Condition

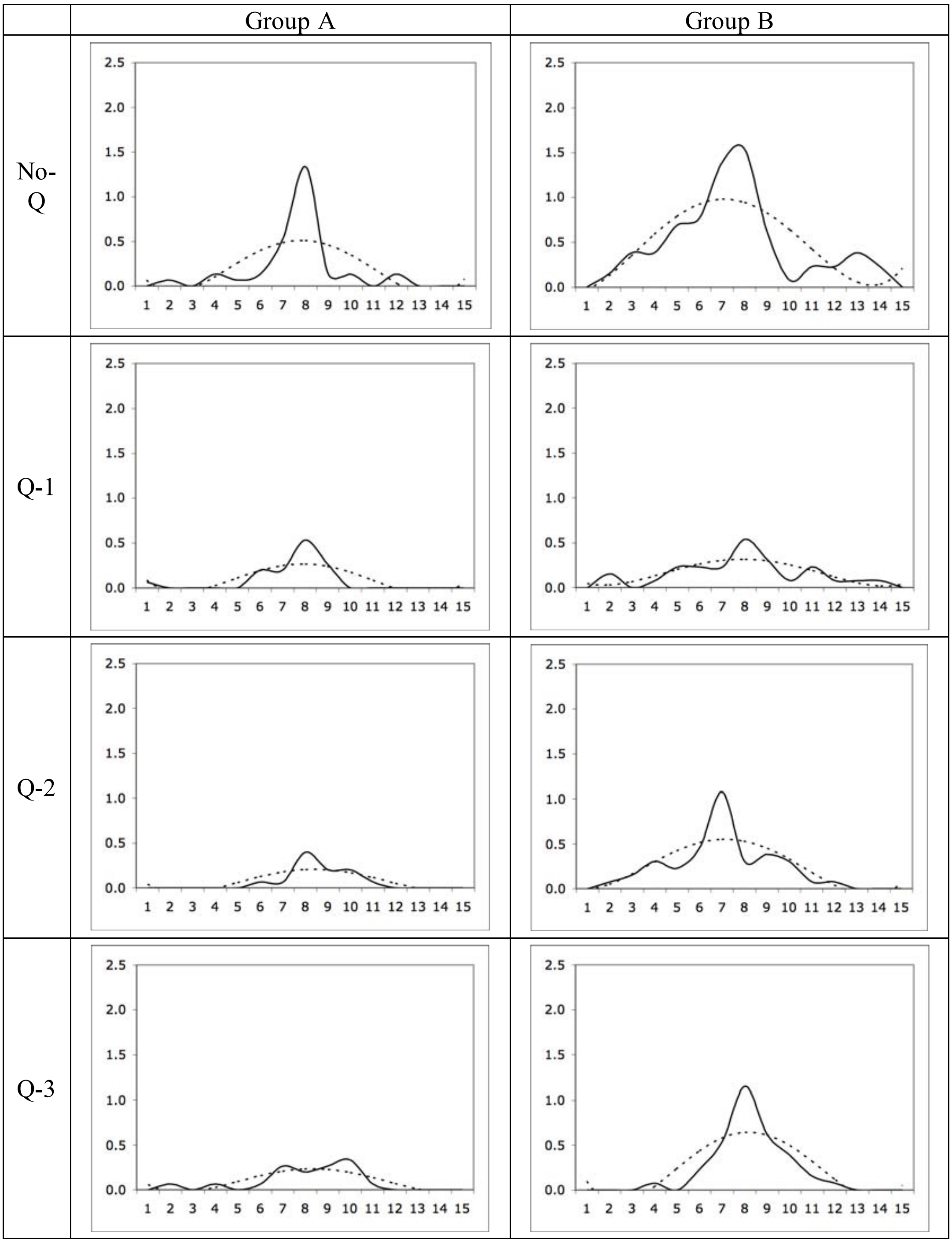


Figure C.5. Distribution of mean fixation counts on the AOIs in GC-1 in the middle-annotation condition. The dotted lines show the fourth-order polynomial curve estimation.

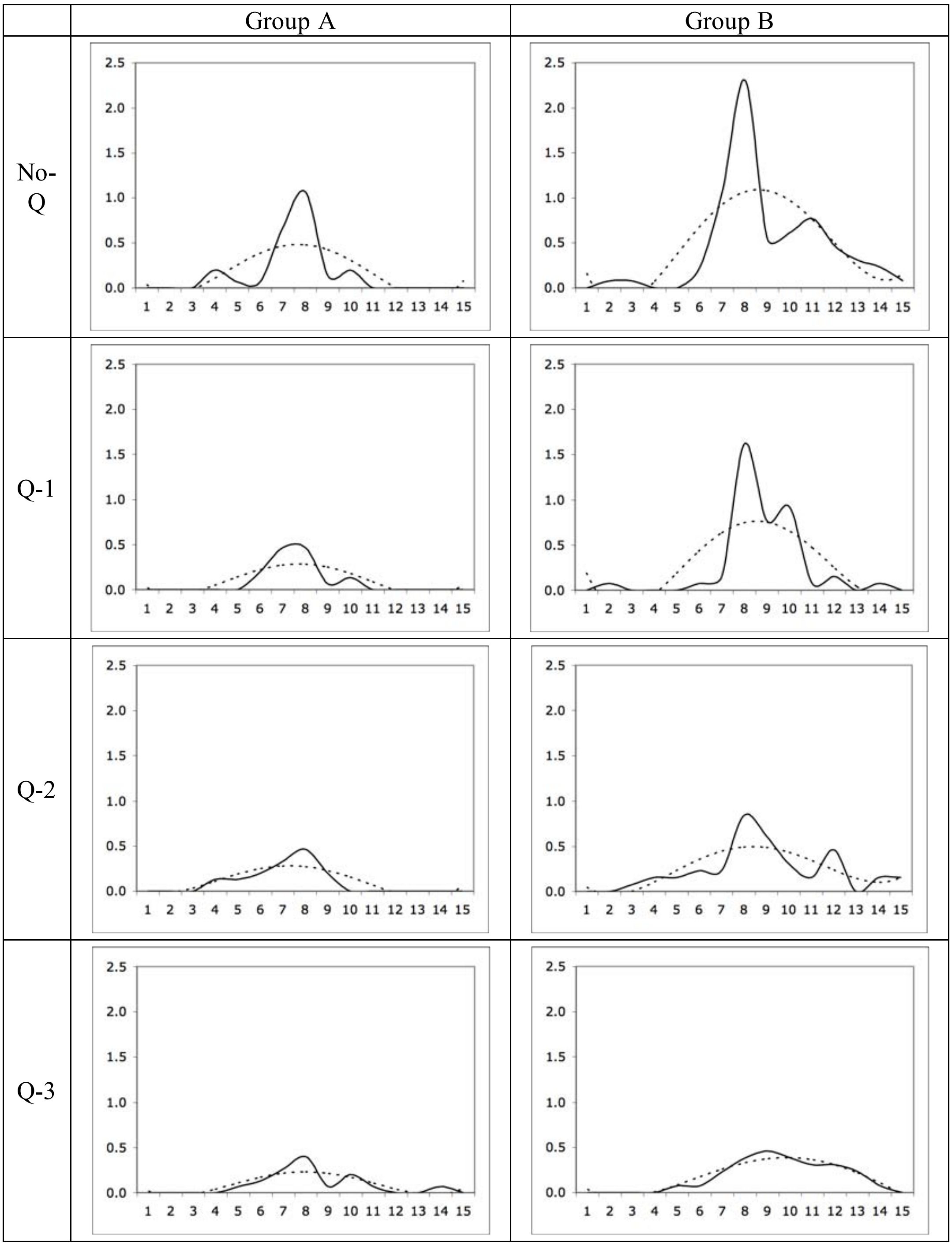


Figure C.6. Distribution of mean fixation counts on the AOIs in GC-2 in the middle-annotation condition. The dotted lines show the fourth-order polynomial curve estimation.

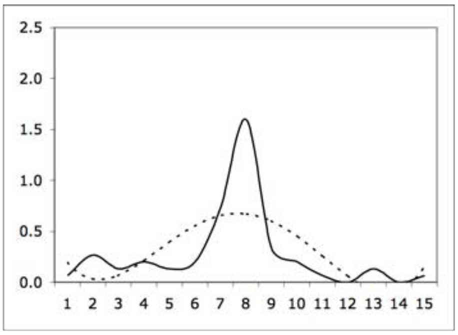
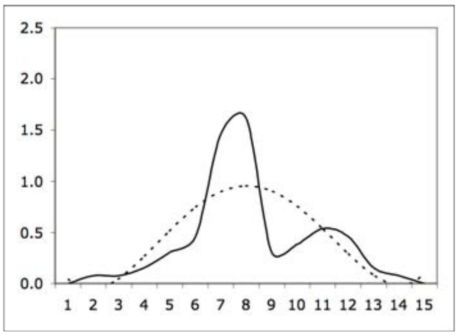
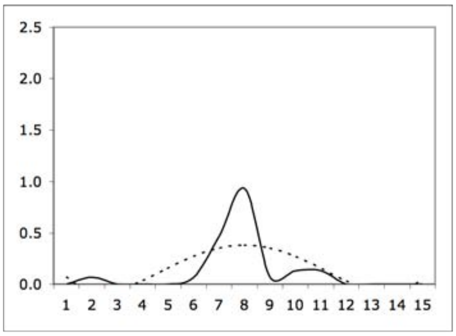
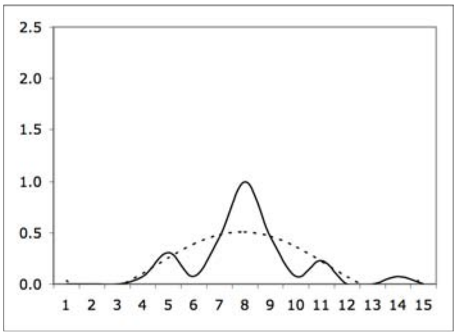
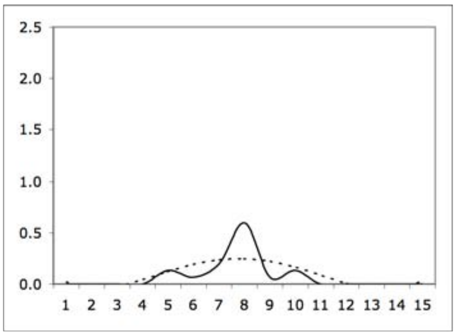
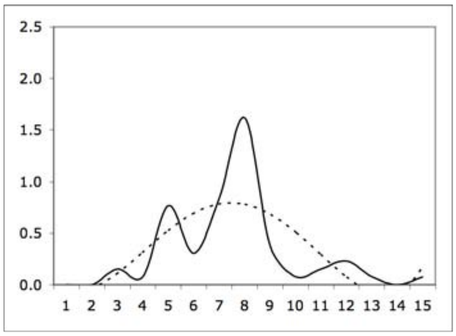
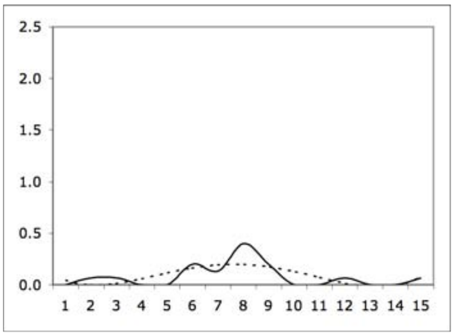
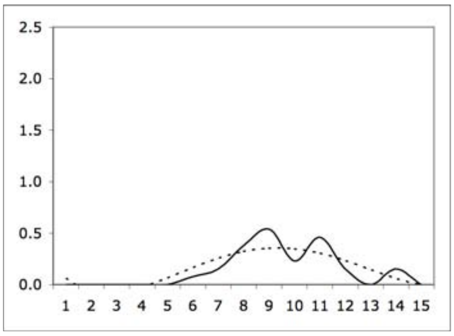
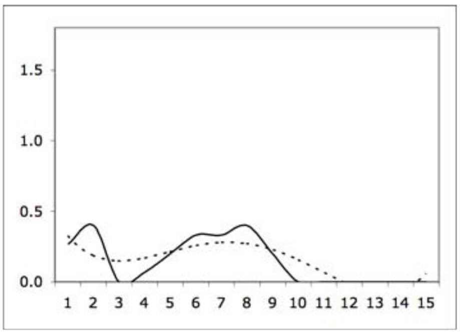
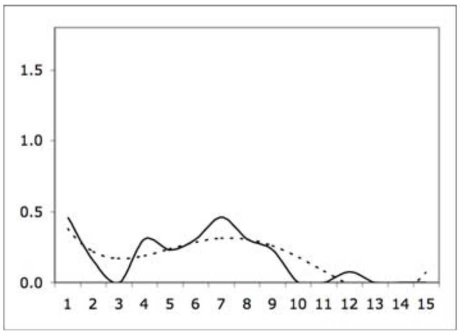
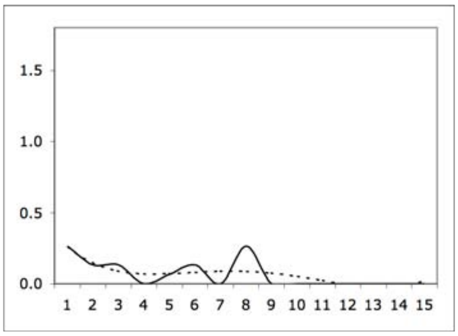
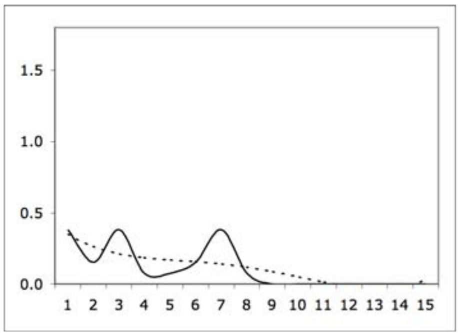
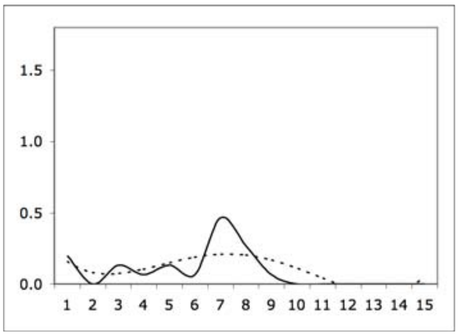
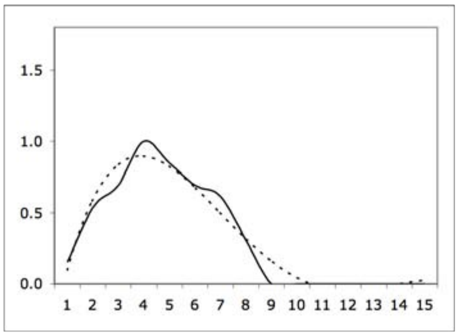
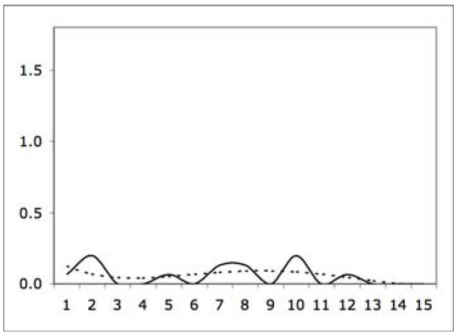
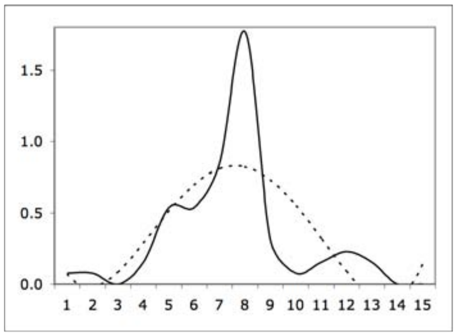
	Group A	Group B
No-Q		
Q-1		
Q-2		
Q-3		

Figure C.7. Distribution of mean fixation counts on the AOIs in GC-3 in the middle-annotation condition. The dotted lines show the fourth-order polynomial curve estimation.

C.2.2.3 Double-Annotation Condition

	Group A	Group B
No-Q		
Q-1		
Q-2		
Q-3		

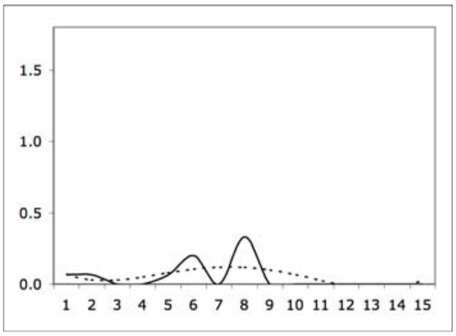
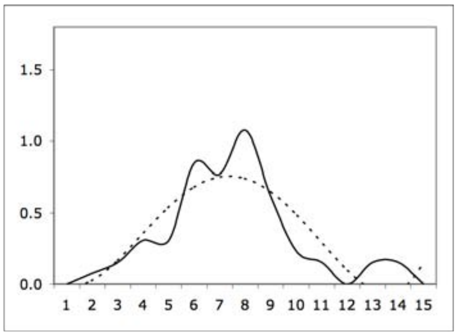
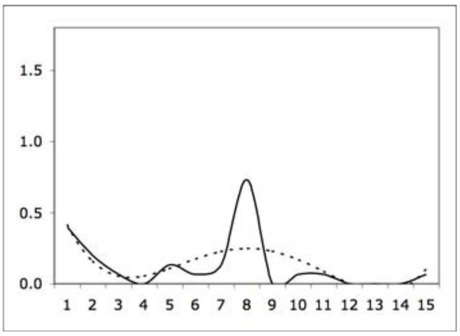
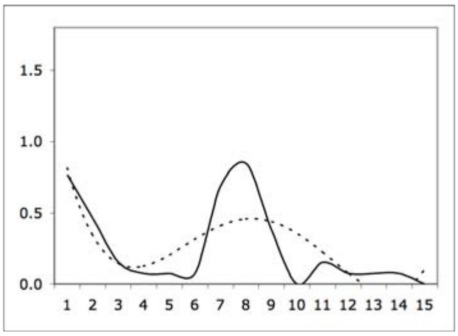
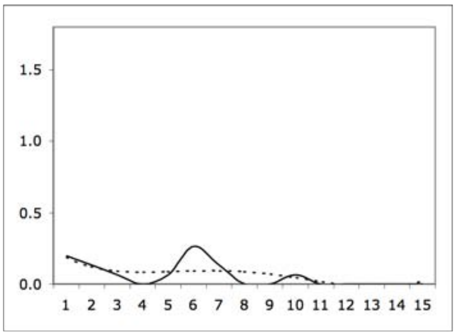
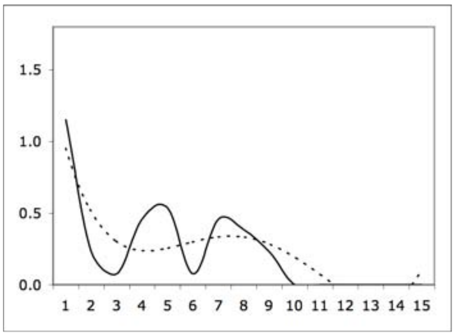
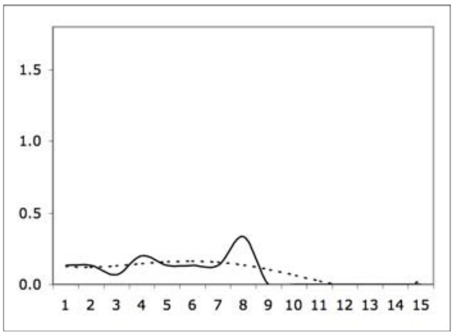
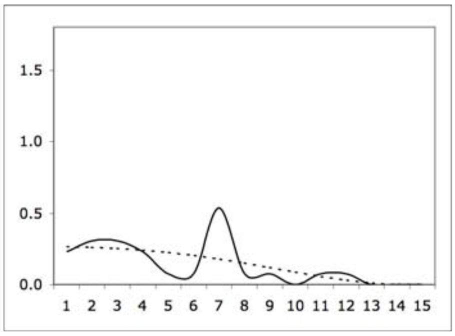
Q-4		

Figure C.8. Distribution of mean fixation counts on the AOIs in GC-1 in the double-annotation condition. The dotted lines show the fourth-order polynomial curve estimation.

	Group A	Group B
No-Q		
Q-1		
Q-2		

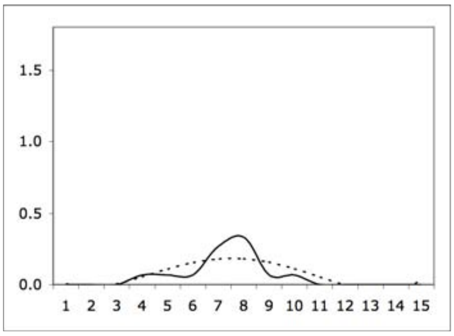
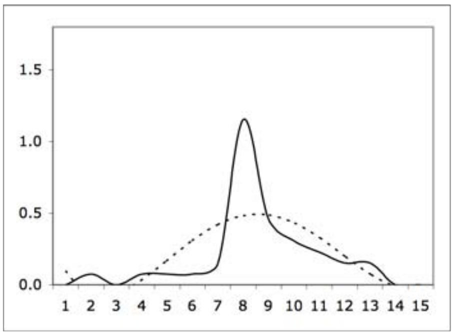
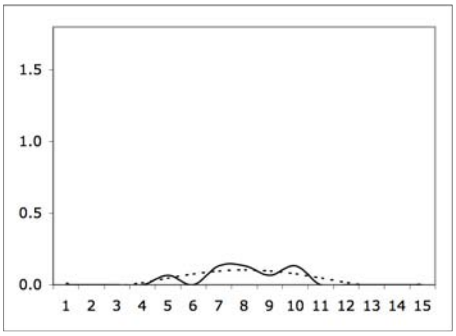
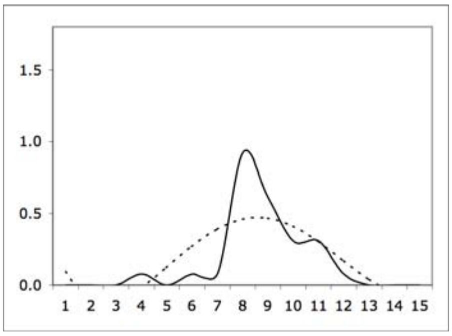
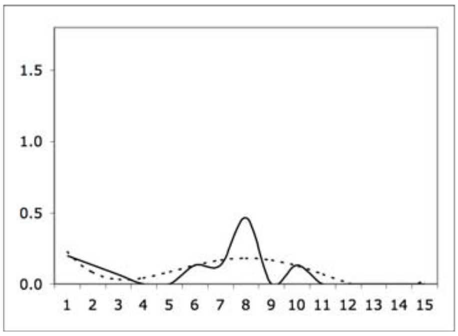
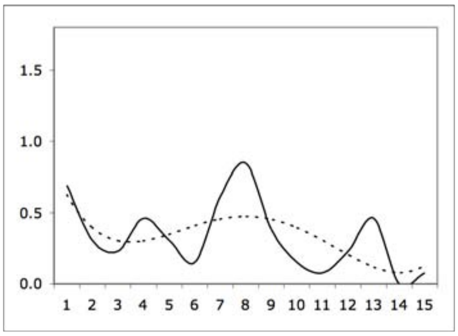
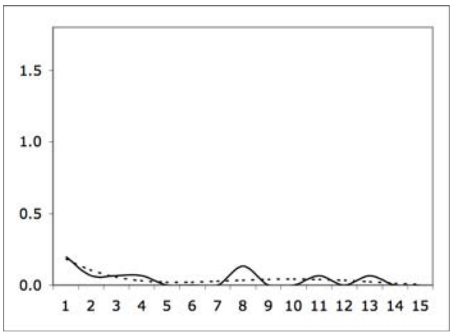
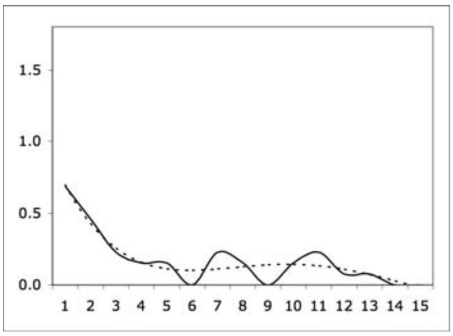
Q-3		
Q-4		

Figure C.9. Distribution of mean fixation counts on the AOIs in GC-2 in the double-annotation condition. The dotted lines show the fourth-order polynomial curve estimation.

	Group A	Group B
No-Q		
Q-1		

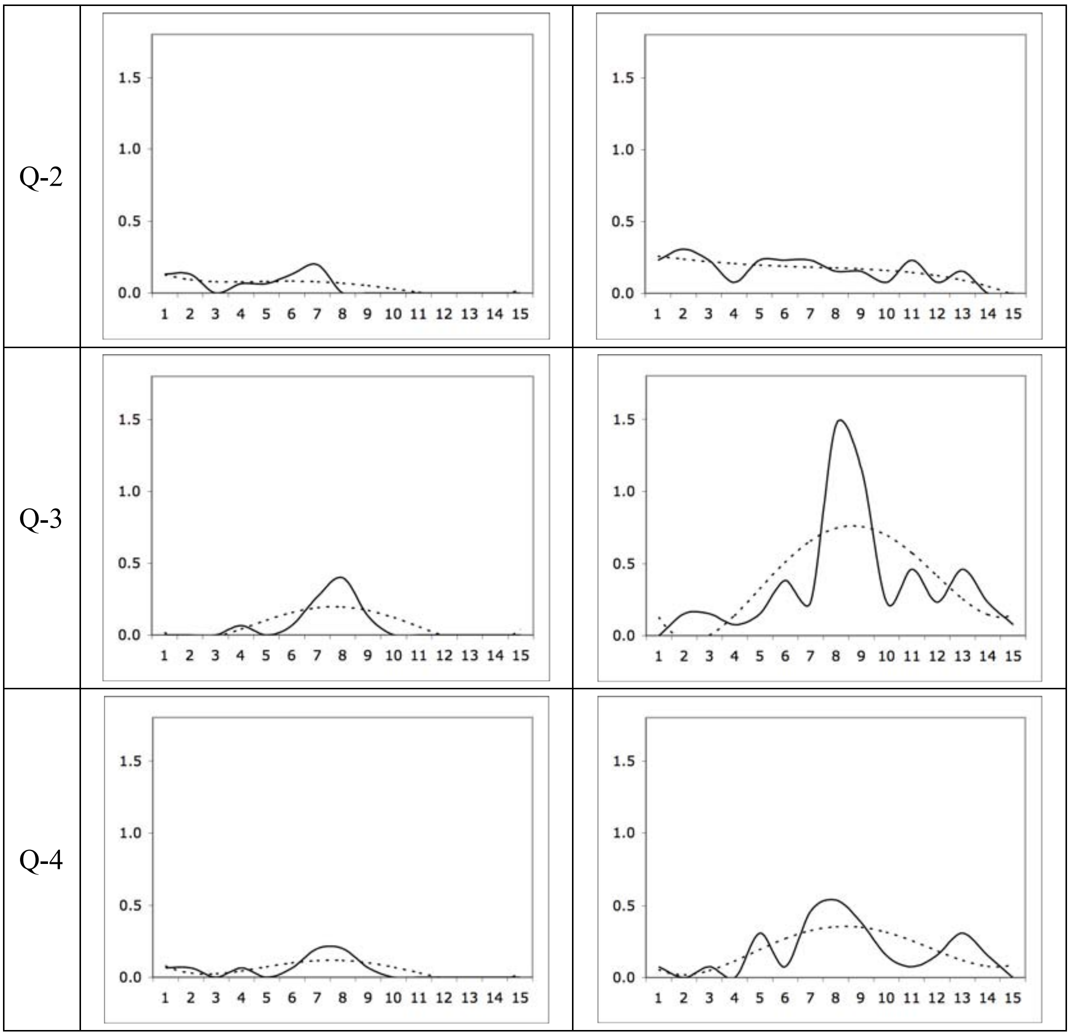
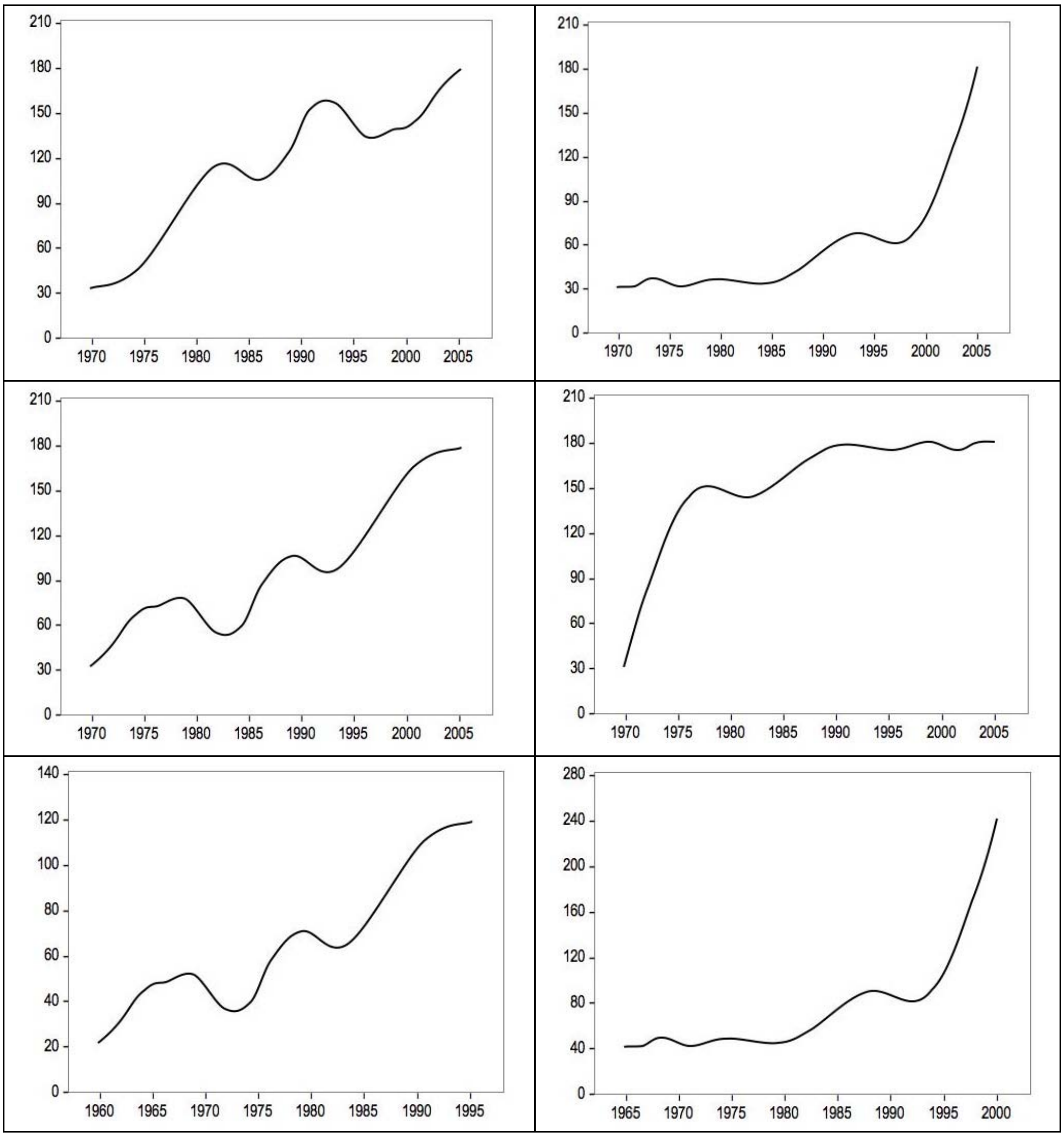


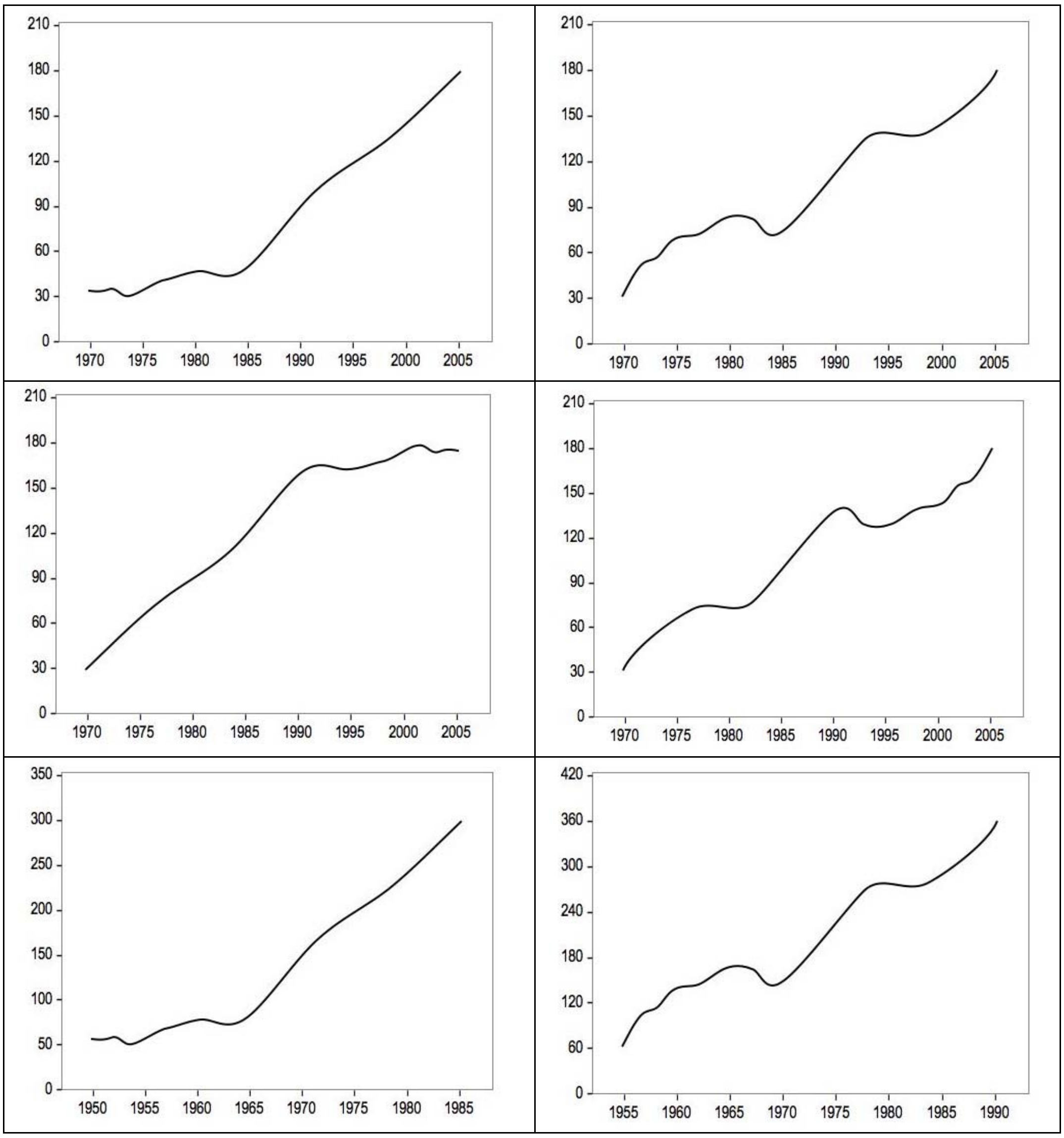
Figure C.10. Distribution of mean fixation counts on the AOIs in GC-3 in the double-annotation condition. The dotted lines show the fourth-order polynomial curve estimation.

Appendix D

D.1 GRAPH STIMULI OF THE EXPERIMENT

The figures below show the 12 graph stimulus that were used in the experiment.





210
180
150
120
90
60
30
0
1970 1975 1980 1985 1990 1995 2000 2005
210
180
150
120
90
60
30
0
1970 1975 1980 1985 1990 1995 2000 2005
210
180
150
120
90
60
30
0
1970 1975 1980 1985 1990 1995 2000 2005
210
180
150
120
90
60
30
0
1970 1975 1980 1985 1990 1995 2000 2005
350
300
250
200
150
100
50
0
1950 1955 1960 1965 1970 1975 1980 1985
420
360
300
240
180
120
60
0
1955 1960 1965 1970 1975 1980 1985 1990

D.2 TRANSITION MATRICES

The transition matrices that are generated from eye movement patterns are presented below.

The transition matrix from the specified AOIs after participants' reading of the stimuli sentence S02 is presented below.

to / from	X1	X	X2	Y	Y2	GY1	G2	GL	GR	*SUM (from)*
X1	***.14***	.02	.01	.00	.00	***.18***	.00	.01	.00	.36
X	***.06***	***.06***	.02	.00	.00	.01	.00	.01	.00	.14
X2	.01	.01	.00	.00	.00	.01	.00	.01	.01	.05
Y	.02	.00	.01	.00	.00	.01	.00	.00	.00	.04
Y2	.01	.00	.00	.00	.00	.00	.00	.00	.00	.01
GY1	***.09***	.02	.00	***.03***	.00	***.07***	.01	.01	.01	.24
G2	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
GL	***.06***	.02	.00	.00	.00	.02	.00	.00	.00	.10
GR	.01	.00	.01	.01	.01	.01	.00	.01	.01	.06
SUM (to)	.39	.12	.05	.03	.01	.31	.01	.05	.03	

The transition matrix from the specified AOIs after participants' reading of the stimuli sentence S03 is presented below.

to / from	X1	X	X2	Y	Y2	GY1	G2	GL	GR	*SUM (from)*
X1	.00	.00	.02	.00	.00	.00	.00	.00	.00	.02
X	.00	.01	.02	.00	.00	.00	.00	.00	.01	.06
X2	.00	.00	***.06***	.00	.00	.00	***.11***	.00	***.03***	.22
Y	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Y2	.00	.02	.01	.00	***.03***	.01	***.11***	.00	***.03***	.21
GY1	.01	.00	.00	.00	.00	.00	.00	.00	.00	.02
G2	.00	.00	***.06***	.00	***.14***	.00	***.03***	.00	***.03***	.27
GL	.00	.00	.01	.00	.00	.00	.00	.00	.00	.02
GR	.00	.01	***.04***	.00	***.04***	.00	***.04***	.01	.02	.18
SUM (to)	.02	.06	.24	.00	.22	.01	.29	.01	.14	

The transition matrix from the specified AOIs after participants' reading of the stimuli sentence S04 is presented below.

to / from	X1	X	X2	Y	Y2	GY1	G2	GL	GR	*SUM (from)*
X1	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01
X	.00	.02	***.04***	.00	.00	.00	.00	.00	.01	.09
X2	.00	.01	***.05***	.01	.01	.00	***.08***	.00	.02	.21
Y	.00	.00	.00	.00	.02	.00	.00	.00	.00	.03
Y2	.00	.00	.02	.00	***.03***	.00	***.11***	.00	.02	.20
GY1	.00	.00	.00	.00	.00	.00	.00	.00	.00	.02
G2	.00	.00	***.07***	.00	***.10***	.00	***.04***	.00	.02	.24
GL	.00	.00	.00	.00	.00	.00	.00	.00	.01	.02
GR	.00	.02	***.04***	.00	***.03***	.00	***.03***	.00	***.04***	.17
SUM (to)	.00	.08	.23	.03	.21	.02	.27	.02	.12	

The comparison of the transitions matrices between S03 and S04 shows that the total distribution of gaze fixations (cf. the values in the *SUM (to)* rows and the *SUM (from)* columns) is similar. On the other hand, the weight of the vertical projections are almost equal in both directions in the S04 matrix shown below (cf. the weight of the gaze pattern from the X2-AOI to the G2-AOI is .08 and the weight of the gaze pattern from the G2-AOI to the X2-AOI is .07), whereas the weight of the vertical projection from the X2-AOI to the G2-AOI in S03 (.11) was higher than the weight of the vertical projection from the G2-AOI to the X2-AOI (.06). In other words, the directionality of the mapping, which is specified by the syntactic order of lexical entities, is reflected in the weights of the projection gaze patterns.

The transition matrix from the specified AOIs after participants' reading of the stimuli sentence S05 is presented below.

to / from	X1	X	X2	Y	Y2	GY1	G2	GL	GR	*SUM (from)*
X1	***.04***	.00	.02	.00	.00	***.05***	.00	.01	.01	.13
X	.02	.02	.02	.00	.00	.00	.00	.00	.00	.05
X2	.00	.01	***.04***	.00	.01	.00	***.05***	.00	.00	.13
Y	.00	.00	.00	.01	.01	.00	.00	.00	.00	.02
Y2	.00	.00	.01	.00	.02	.01	***.07***	.01	.01	.12
GY1	***.04***	.02	.01	.01	.02	***.03***	.01	.00	.01	.16
G2	.00	.00	***.03***	.01	***.04***	.00	.02	.01	***.04***	.16
GL	.00	.01	.01	.00	.00	.02	.00	.00	.01	.06
GR	.00	.01	.01	.01	.02	.00	***.04***	.01	***.05***	.16
SUM (to)	.11	.06	.14	.03	.13	.14	.19	.05	.15	

The transition matrix from the specified AOIs after participants' reading of the stimuli sentence S06 is presented below.

from \ to	X1	X	X2	Y	Y2	GY1	G2	GL	GR	*SUM (from)*
X1	***.07***	.02	***.03***	.00	.00	***.04***	.00	.02	.01	.17
X	.01	.01	.02	.00	.00	.00	.00	.01	.00	.05
X2	.01	.01	***.04***	.00	.00	.00	***.05***	.00	.02	.14
Y	.00	.00	.00	.00	.00	.00	.00	.00	.00	.02
Y2	.00	.00	.01	.00	.02	.02	***.06***	.00	.02	.13
GY1	***.04***	.00	.01	.01	.02	***.03***	.00	.01	.02	.14
G2	.00	.00	.01	.00	***.05***	.00	***.03***	.01	.02	.13
GL	.02	.01	.01	.01	.00	.02	.00	.02	.02	.10
GR	.00	.00	.02	.00	***.03***	.01	.02	.02	.02	.13
SUM (to)	.15	.05	.14	.02	.13	.13	.16	.08	.12	

The transition matrix from the specified AOIs after participants' reading of the stimuli sentence S07 is presented below.

from \ to	X1	X	X2	Y	Y2	GY1	G2	GL	GR	*SUM (from)*
X1	***.05***	***.04***	.02	.01	.00	***.03***	.00	.01	.00	.15
X	***.04***	.02	**.05**	.01	.00	.00	.00	.00	.00	.13
X2	.02	.02	***.04***	.00	.00	.00	***.04***	.00	***.03***	.15
Y	.00	.00	.00	.00	.01	***.03***	.00	.00	.00	.05
Y2	.01	.01	.01	.00	.00	.01	***.04***	.00	.01	.10
GY1	.02	.01	.01	.00	***.03***	***.03***	.01	.00	.01	.13
G2	.00	.00	.01	.00	***.05***	.00	.01	.01	.02	.11
GL	.01	.02	.00	.00	.00	.01	.00	.01	.00	.05
GR	.00	.01	.01	.01	.00	.01	.02	.01	***.05***	.13
SUM (to)	.15	.13	.16	.04	.11	.12	.13	.04	.13	

The transition matrix from the specified AOIs after participants' reading of the stimuli sentence S08 is presented below.

to / from	X1	X	X2	Y	Y2	GY1	G2	GL	GR	*SUM (from)*
X1	.01	.00	.00	.01	.00	***.05***	.00	.00	.01	.09
X	.01	.00	.01	.00	.00	.01	.00	.00	.00	.05
X2	.01	.00	***.03***	.00	.01	.00	.01	.00	.01	.08
Y	.00	.00	.00	.00	***.03***	.01	.00	.00	.00	.05
Y2	.00	.01	.00	.00	.02	.01	***.09***	.01	.02	.17
GY1	.01	.00	.00	.02	***.04***	***.04***	.02	.01	***.03***	.17
G2	.00	.01	***.04***	.00	***.03***	.01	.01	.01	.02	.14
GL	.02	.01	.00	.00	.00	***.03***	.00	.00	.01	.09
GR	.01	.00	.02	.01	***.04***	.00	***.03***	.01	***.04***	.17
SUM (to)	.08	.04	.11	.05	.17	.16	.16	.06	.16	

The transition matrix from the specified AOIs after participants' reading of the stimuli sentence S09 is presented below.

to / from	X1	X	X2	Y	Y2	GY1	G2	GL	GR	*SUM (from)*
X1	***.05***	.01	.01	.00	.00	***.03***	.00	***.05***	***.03***	.17
X	***.05***	.02	.02	.00	.00	.00	.00	.02	.00	.11
X2	.01	.01	.01	.00	.00	.00	.02	.01	.01	.06
Y	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Y2	.00	.00	.00	.00	.00	.00	.00	.00	.00	.02
GY1	.01	.00	.00	.00	.00	.01	.00	***.03***	.00	.07
G2	.00	.00	.01	.00	.01	.00	.00	.00	***.04***	.07
GL	.02	***.04***	.00	.00	.00	.02	.00	***.08***	***.09***	.27
GR	.01	.01	.02	.00	.01	.01	***.05***	***.04***	***.10***	.24
SUM (to)	.15	.09	.06	.00	.02	.08	.08	.23	.28	

The transition matrix from the specified AOIs after participants' reading of the stimuli sentence S10 is presented below.

to / from	X1	X	X2	Y	Y2	GY1	G2	GL	GR	*SUM (from)*
X1	.02	.01	.02	.01	.00	.00	.00	.01	.00	.07
X	.01	.01	.02	.00	.00	.00	.01	.00	.01	.07
X2	.00	.00	***.03***	.00	.00	.00	***.03***	.01	***.08***	.17
Y	.01	.00	.00	.01	.00	.00	.00	.00	.00	.02
Y2	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01
GY1	.02	.00	.00	.00	.00	.01	.00	.02	.00	.05
G2	.00	.01	.02	.00	.01	.00	.01	.02	***.03***	.10
GL	.01	.02	.02	.00	.00	.01	.01	***.04***	***.07***	.18
GR	.01	.01	***.06***	.00	.00	.01	***.03***	***.08***	***.13***	.33
SUM (to)	.07	.06	.17	.02	.01	.05	.10	.18	.33	

The transition matrix from the specified AOIs after participants' reading of the stimuli sentence S11 is presented below.

to / from	X1	X	X2	Y	Y2	GY1	G2	GL	GR	*SUM (from)*
X1	.00	.00	.01	.01	.01	.01	.00	.00	.00	.03
X	.00	.01	.02	.00	.00	.00	.00	.01	.01	.05
X2	***.03***	.01	.02	.00	***.03***	.00	***.04***	.01	.01	.13
Y	.00	.00	.01	.00	.01	.00	.00	.00	.00	.01
Y2	.00	.01	.00	.01	***.04***	.01	***.16***	.01	***.03***	.26
GY1	.01	.01	.00	.00	.00	.00	.00	.01	.01	.03
G2	.00	.00	***.04***	.00	***.14***	.00	***.06***	.00	.02	.27
GL	.01	.01	.00	.00	.00	.01	.01	.01	.01	.06
GR	.00	.02	***.03***	.00	***.04***	.01	***.04***	.01	.01	.16
SUM (to)	.04	.06	.12	.01	.28	.03	.31	.04	.10	

The transition matrix from the specified AOIs after participants' reading of the stimuli sentence S12 is presented below.

to / from	X1	X	X2	Y	Y2	GY1	G2	GL	GR	*SUM (from)*
X1	.02	.02	***.03***	.00	.00	.01	.02	.02	.01	.13
X	***.03***	.02	***.03***	.00	.00	.00	.00	.00	.00	.09
X2	.01	.00	.02	.00	.00	.00	***.04***	.00	***.05***	.13
Y	.00	.00	.00	.00	.00	.01	.00	.00	.00	.02
Y2	.00	.00	.00	.00	.00	.00	.00	.01	.00	.03
GY1	***.03***	.00	.00	.00	.00	.00	.00	.00	.02	.06
G2	.00	.00	.02	.00	.02	.00	.02	.01	***.04***	.10
GL	.02	.01	.01	.00	.00	.02	.00	***.03***	***.07***	.16
GR	.02	.02	***.03***	.00	.00	.00	***.03***	***.08***	***.10***	.29
SUM (to)	.14	.08	.14	.01	.03	.05	.11	.15	.29	

D.3 RESPONSE TIMES

The table below shows the twelve stimulus sentences of the experiment and participants' response time after their reading of each sentence, which is also equivalent to participants' mean gaze times on the graphs.

	Sentences	*Gaze Time (ms)*
S01	The number of birds is about 180 in the year 2005.	2313 (531)
S02	The number of birds is about 30 in the year 1970.	2478 (993)
S03	The number of birds in the year 2005 increased to about 180.	3729 (1353)
S04	The number of birds increased to about 180 in the year 2005.	3546 (1626)
S05	The number of birds increased from about 30 in the year 1970 to about 180 in the year 2005.	4440 (2564)
S06	The number of birds increased between the years 1970 and 2005 from about 30 to about 180.	4779 (2311)
S07	The number of birds increased from about 30 to about 180 between the years 1970 and 2005.	4284 (1626)
S08	The number of birds increased from about 30 to about 180.	3211 (1970)
S09	The number of birds increased since year 1970.	4878 (2696)
S10	The number of birds increased until the year 2005.	4482 (2652)
S11	The number of birds increased until about 180.	2605 (936)
S12	The number of birds increased between the years 1970 and 2005.	4168 (1736)

BIBLIOGRAPHY

Acartürk, C., & Habel, C. (2009). Attributing causality in multimodal language-graphics comprehension [abstract]. In *ESLP 2009, the Annual Meeting of the Embodied & Situated Language Processing*. Rotterdam, Netherlands.

Acartürk, C., Habel, C., & Çağıltay, K. (2008). Multimodal comprehension of graphics with textual annotations: The role of graphical means relating annotations and graph lines. In J. Howse, J. Lee & G. Stapleton (Eds.), *Diagrammatic representation and inference: Lecture notes in computer science* (Vol. 5223, pp. 335-343). Berlin/Heidelberg: Springer.

Acartürk, C., Habel, C., Çağıltay, K., & Alaçam, O. (2008). Multimodal comprehension of language and graphics: Graphs with and without annotations. *Journal of Eye Movement Research, 1(3:*2), 1-15.

Adams, A., & Cox, A. L. (2008). Questionnaires, in-depth interviews, and focus groups. In P. Cairns & A. L. Cox (Eds.), *Research methods for human-computer interaction* (pp. 17-34). New York: Cambridge University Press.

Ahn, W.-k., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54*(3), 299-352.

Ainley, J., Nardi, E., & Pratt, D. (1998). Graphing as a computer mediated tool. In A. Olivier & K. Newstead (Eds.), *Proceedings of the 22nd annual conference of the International Group for the Psychology of Mathematics Education* (pp. 243–258). Stellenbosch, South Africa: University of Stellenbosch.

Ainsworth, S. E. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*, 183-198.

Alesandrini, K. L. (1984). Pictures and adult learning. *Instructional Science, 13*, 63-77.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum.

Anderson, R. E., & Helstrup, T. (1993). Visual discovery in mind and on paper. *Memory & Cognition, 21*, 283-293.

Annett, J. (2003). Hierarchical Task Analysis. In E. Holnagel (Ed.), *Handbook of cognitive task design* (pp. 17-35). Mawhah, NJ: Lawrence Erlbaum.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61*, 183-193.

Avrahami, J., & Kareev, Y. (1994). The emergence of events. *Cognition, 53*, 239-261.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A. D. (1997). *Human memory: Theory & practice* (Revised ed.). Hove: Psychology Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Bargeron, D., & Moscovich, T. (2003). Reflowing digital ink annotations. In *CHI '03: Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 385-393). New York: ACM Press.

Barkowsky, T., Latecki, L. J., & Richter, K.-F. (2000). Schematizing maps: Simplification of geographic shape by discrete curve evolution. In C. Freksa, W. Brauer, C. Habel & K. F. Wender (Eds.), *Spatial cognition II* (pp. 41-53). Berlin: Springer.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577-660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617-645.

Beniger, J. R., & Robyn, D. L. (1978). Quantitative graphics in statistics: A brief history. *The American Statistician, 32*(1), 1-9.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1947). *Differential Aptitude Tests*. New York: Psychological Corp.

Bernard, R. M. (1990). Using extended captions to improve learning from instructional illustrations. *British Journal of Educational Technology, 21*, 212-225.

Bernsen, N. O. (1994). Foundations of multimodal representations: A taxonomy of representational modalities. *Interacting with Computers, 6*(4), 347-371.

Best, L. A., Smith, L. D., & Stubbs, D. A. (2001). Graph use in psychology and other sciences. *Behavioural Processes, 54*, 155-165.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-147.

Bierwisch, M. (1989). The semantics of gradation. In M. Bierwisch & E. Lang (Eds.), *Dimensional adjectives* (pp. 71-261). Berlin: Springer-Verlag.

Blandford, A., Cox, A. L., & Cairns, P. (2008). Controlled experiments. In P. Cairns & A. L. Cox (Eds.), *Research methods for human-computer interaction* (pp. 1-16). New York: Cambridge University Press.

Bodemer, D., & Faust, U. (2006). External and mental referencing of multiple representations. *Computers in Human Behavior, 22*, 27-42.

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience, 9*(1), 27-38.

Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 95-129). Cambridge, MA: MIT Press.

Brush, A. J., Bargeron, D., Gupta, A., & Cadiz, J. J. (2001). Robust annotation positioning in digital documents. In *CHI '01: Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 285-292). New York: ACM Press.

Buehner, M. J., & May, J. (2003). Rethinking temporal contiguity and the judgement of causality: Effects of prior knowledge, experience, and reinforcement procedure. *The Quarterly Journal of Experimental Psychology, 56A*(5), 865-890.

Butler, D. L. (1993). Graphics in psychology: Pictures, data, and especially concepts. *Behavior Research Methods, Instruments, and Computers, 25*, 81-92.

Byrne, M. D., & Anderson, J. R. (1998). Perception and action. In J. R. Anderson & C. Lebiere (Eds.), *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum.

Cairns, P., & Cox, A. L. (2008). *Research methods for human-computer interaction*. New York: Cambridge University Press.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum.

Carlson, L. A. (1999). Selecting reference frame. *Spatial Cognition and Computation, 1*, 365-379.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*, 75-100.

Carroll, P. J., Young, R. J., & Guertin, M. S. (1992). Visual analysis of cartoons: A view from the far side. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 444-461). New York: Springer.

Carswell, C. M., Emery, C., & Lonon, A. M. (1993). Stimulus complexity and information integration in the spontaneous interpretation of line graphs. *Applied Cognitive Psychology, 7*, 341-357.

Catley, K. M. (2006). Darwin's missing link - a novel paradigm for evolution education. *Science Education, 90*, 767-783.

Chabris, C. F., & Kosslyn, S. M. (2005). Representational correspondence as a basic principle of diagram design. In S.-O. Tergan & T. Keller (Eds.), *Knowledge and information visualization: Searching for synergies. Lecture notes in computer science* (Vol. 3426, pp. 36-57). Berlin: Springer.

Chambers, D., & Reisberg, D. (1985). Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception & Performance, 11*(3), 317-328.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*, 233-246.

Chandrasekaran, B., & Narayanan, N. H. (1992). Perceptual representation and reasoning. In B. Chandrasekaran & H. Simon (Eds.), *AAAI Spring Symposium technical report SS-92-02: Reasoning with diagrammatic representations*. Manlo Park, CA: AAAI Press.

Chatterjee, A. (2001). Language and space: Some interactions. *Trends in Cognitive Sciences, 5*(2), 55-61.

Cheng, P. C., & Simon, H. A. (1995). Scientific discovery and creative reasoning with diagrams. In S. M. Smith, T. B. Ward & R. A. Finke (Eds.), *The creative cognition approach* (pp. 205-228). Cambridge, MA: MIT Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation in physics problems by experts and novices. *Cognitive Science, 5*(2), 121-152.

Chomsky, N. (1959). A review of B. F. Skinner's Verbal Behavior. *Language, 35*, 26-58.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Clark, H. H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing: where language, culture, and cognition meet* (pp. 243-268). London: Erlbaum.

Cleveland, W. S. (1984). Graphs in scientific publications. *The American Statistician, 38*(4), 261-269.

Cleveland, W. S. (1993). A model for studying display methods of statistical graphics. *Journal of Computational and Graphical Statistics, 2*, 323-364.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association, 77*, 541–547.

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science, 229*, 828–833.

Cohen, E. H., & Singh, M. (2007). Geometric determinants of shape segmentation: Tests using segment identification. *Vision Research, 47*(22), 2825-2840.

Cook, M. P. (2006). Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Science Education, 90*(6), 1073-1091.

Coradeschi, S., & Saffiotti, A. (2003). An introduction to the anchoring problem. *Robotics and Autonomous Systems, 43*(2-3), 85-96.

Costigan-Eaves, P., & Macdonald-Ross, M. (1990). William Playfair (1759-1823). *Statistical Science, 5*(3), 318-326.

Cowan, N., Saults, J. S., & Morey, C. C. (2006). Development of working memory for verbal–spatial associations. *Journal of Memory and Language, 55*, 274-289.

Cowen, L., Ball, L. J., & Delin, J. (2002). An eye movement analysis of web page usability. In X. Faulkner, J. E. Finlay & F. Detienee (Eds.), *People and computers XVI: Memorable yet invisible* (pp. 317-335). Berlin: Springer.

Cox, A. L., & Peebles, D. J. (2008). Cognitive modelling in HCI research. In P. Cairns & A L.. Cox (Eds.), *Research methods for human-computer interaction* (pp. 70-87). New York: Cambridge University Press.

Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction, 9*, 343-363.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of Processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(671-684).

Cresswell, M. (1977). The semantics of degree. In B. Partee (Ed.), *Montague grammar* (pp. 261-292). New York: Academic Press.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge, MA: Cambrigde University Press.

Culbertson, H. M., & Powers, R. D. (1959). A study of graph comprehension difficulties. *Educational Technology Research and Development, 7*(3), 97-110.

d'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 415-427). New York: Springer Verlag.

De Winter, J., & Wagemans, J. (2004). Contour-based object identification and segmentation: Stimuli, norms and data, and software tools. *Behavior Research Methods, Instruments, & Computers, 36*(4), 604-624.

De Winter, J., & Wagemans, J. (2006). Segmentation of object outlines into parts: A large-scale integrative study. *Cognition, 99*, 275-325.

Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2004). *Human-Computer Interaction* (3rd ed.). Harlow: Pearson Education Limited.

Do, E. Y.-L., Gross, M. D., Neiman, B., & Zimring, C. (2000). Intentions in and relations among design drawings. *Design Studies, 21*, 483-503.

Duchowski, A. T. (2002). A breadth-first survey of eye tracking applications. *Behavior Research Methods, Instruments, & Computers, 34*(4), 455-470.

Duchowski, A. T. (2007). *Eye tracking methodology: Theory & practice* (2nd ed.). London: Springer-Verlag.

Duffy, S. A. (1992). Eye movements and complex comprehension processes. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 462-471). New York: Springer Verlag.

Eells, W. C. (1926). The relative merits of circles and bars for representing component parts. *Journal of the American Statistical Association, 21*(154), 119-132.

Eilam, B., & Poyas, Y. (2008). Learning with multiple representations: Extending multimedia learning beyond the lab. *Learning and Instruction, 18*(4), 368-378.

Elby, A. (2000). What students' learning of representations tells us about constructivism. *Journal of Mathematical Behavior, 19*(4), 481-502.

Elzer, S., Green, N., Carberry, S., & Hoffman, J. (2006). A model of perceptual task effort for bar charts and its role in recognizing intention. *User Modeling and User-Adapted Interaction, 16*, 1-30.

Engelhardt, Y. (2006). Objects and spaces: The visual language of graphics. In D. Barker-Plummer, R. Cox & N. Swoboda (Eds.), *Diagrammatic representation and* *inference: Lecture notes in computer science* (Vol. 4045, pp. 104-108). Berlin: Springer-Verlag.

Engelhardt, Y. (2007). Syntactic structures in graphics. *IMAGE, 5*, 23-35.

Ericsson, K. A., & Simon, H. A. (1980). Verbal report as data. *Psychological Review, 87*(3), 215-251.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Boston: MIT Press.

Eschenbach, C. (1995). *Zählangaben - maßangaben. Bedeutung und konzeptuelle interpretation von numeralia*. Wiesbaden: DUV, Dt. Univ.-Verl.

Eschenbach, C. (1999). Geometric structures of frames of reference and natural language semantics. *Spatial Cognition and Computation, 1*, 329-348.

Eschenbach, C., Habel, C., Kulik, L., & Leßmöllmann, A. (1998). Shape nouns and shape concepts: A geometry for 'corner'. In C. Freksa, C. Habel & K. F. Wender (Eds.), *Spatial cognition* (pp. 177-201). Berlin: Springer.

Eschenbach, C., Tschander, L., Habel, C., & Kulik, L. (2000). Lexical specifications of paths. In C. Freksa, W. Brauer, C. Habel & K. F. Wender (Eds.), *Spatial cognition II* (pp. 127-144). Berlin: Springer.

Feldman, J., & Singh, M. (2005). Information along contours and object boundaries. *Psychological Review, 112*(1), 243-252.

Feliciano, G. D., Powers, R. D., & Kearl, B. E. (1963). The presentation of statistical information. *Educational Technology Research and Development, 11*(3), 32-39.

Ferreira, F., & Tanenhaus, M. K. (2007). Introduction to special issue on language-vision interactions. *Journal of Memory and Language, 57*, 455-459.

Fienberg, S. E. (1979). Graphical methods in statistics. *American Statistician, 33*(4), 165-178.

Findlay, J. M., & Gilchrist, I. D. (1998). Eye guidance and visual search. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 295-312). Oxford, UK: Elsevier.

Fodor, J. (1975). *The language of thought*. Cambidge, MA: Hardvard University Press.

Fodor, J. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.

Forbus, K. D. (1995). Qualitative spatial reasoning: Framework and frontiers. In J. Glasgow, H. Narayanan & B. Chandrasekaran (Eds.), *Diagrammatic reasoning: Cognitive and computational perspectives* (pp. 183-202). Cambridge, MA: MIT Press.

Freeman, H. (1978). Shape description via the use of critical points. *Pattern Recognition, 10*(3), 159-166.

Friendly, M., & Denis, D. (2005). The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences, 41*(2), 103-130.

Funkhouser, H. G. (1937). Historical development of the graphical representation of statistical data. *Osiris, 3*, 269-404.

Futrelle, R. P. (1990). Strategies for diagram understanding: Object/spatial data structures, animate vision and generalized equivalence. In *Proceedings of the 10th international conference on pattern recognition* (pp. 403-408). Los Alamitos, CA: IEEE CS Press.

Futrelle, R. P. (1999). Summarization of diagrams in documents. In I. Mani & M. Maybury (Eds.), *Advances in automated text summarization* (pp. 403-421). Cambridge, MA: MIT Press.

Garcia, M. R., & Stark, P. A. (1991). *Eyes on the news*. St. Petersburg, FL: The Poynter Institute.

Gibson, E. (1998). Syntactic complexity: Locality of syntactic dependencies. *Cognition,* *68*(1), 1-76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. P. Marantz & W. O'Neil (Eds.), *Image, language, brain* (pp. 95-126). Cambridge, MA: MIT Press.

Gibson, E., & Pearlmutter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences, 2*, 262-268.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language, 31*, 129-151.

Glenberg, A. M., & McDaniel, M. A. (1992). Mental models, pictures, and text: Integration of spatial and verbal information. *Memory & Cognition, 20*, 458-460.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foreground during text comprehension. *Journal of Memory and Language, 26*, 69-83.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language, 43*, 379-401.

Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. New York: Routledge.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics, 24*, 631-645.

Goldschmidt, G. (1991). The dialectics of sketching. *Creativity Research Journal, 4*, 123-143.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 1411-1423.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review Psychology, 48*, 163-189.

Graham, J. L. (1937). Illusory trends in the observation of bar graphs. *Journal of Experimental Psychology, 6*(20), 597-608.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: A validation of GOMS for prediction and explanation of real-world task performance. *Human-Computer Interaction, 8*(3), 237-309.

Gray, W. D., Young, R. M., & Kirschenbaum, S. S. (1997). Introduction to special issue on cognitive architectures and human-computer interaction. *Human Computer Interaction, 12*(4), 301-309.

Green, N. (2005). Design of information graphics for causal arguments. In *Proceedings of the international joint conference on artificial intelligence, workshop on computer models of natural argument*. Edinburgh, UK.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science, 29*, 261-290.

Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006). The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1253-1262). New York: ACM Press.

Guhe, M., & Habel, C. (2001). The Influence of resource parameters on incremental conceptualization. In E. M. Altmann, A. Cleeremans, C. D. Schunn & W. D. Gray (Eds.), *Proceedings of the Fourth International Conference on Cognitive Modeling.*

July 26–28, 2001, George Mason University, Fairfax, VA. Mahwah, NJ: Lawrence Erlbaum (with minor corrections).

Guhe, M., Habel, C., & Tschander, L. (2004). Incremental generation of interconnected preverbal messages. In T. Pechmann & C. Habel (Eds.), *Multidisciplinary approaches to language production.* (pp. 7-52). Berlin: Mouton de Gruyter.

Gyselinck, V., Jamet, E., & Dubois, V. (2008). The role of working memory components in multimedia comprehension. *Applied Cognitive Psychology, 22*, 353-374.

Habel, C. (1990). Propositional and depictorial representations of spatial knowledge: The case of path concepts. In R. Studer (Ed.), *Natural language and logic. Lecture notes in computer science* (Vol. 459, pp. 94-117). Berlin: Springer.

Habel, C. (2005). Verbs and directions. In L. A. Carlson & E. v. d. Zee (Eds.), *Functional features in language and space* (pp. 93–112). Oxford: Oxford University Press

Habel, C., & Acartürk, C. (2007). On reciprocal improvement in multimodal generation: Co-reference by text and information graphics. In I. van der Sluis, M. Theune, E. Reiter & E. Krahmer (Eds.), *Proceedings of MOG 2007: The workshop on multimodal output generation* (pp. 69-80). University of Aberdeen, UK.

Habel, C., & Acartürk, C. (2009). Eye-tracking evidence for multimodal language-graphics comprehension: The role of integrated conceptual representations. In C. Navarretta, P. Paggio, J. Allwood, E. Alsén & Y. Katagiri (Eds.), *Proceedings of NODALIDA 2009: The workshop on multimodal communication - from human behaviour to computational models* (Vol. 6, pp. 9-14). Odense, Denmark: Northern European Association for Language Technology (NEALT).

Habel, C., & Eschenbach, C. (1997). Abstract structures in spatial cognition. In C. Freksa, M. Jantzen & R. Valk (Eds.), *Foundations of computer science – potential – theory – cognition* (pp. 369-378). Berlin: Springer.

Haberlandt, K. (1994). Methods in reading research. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1-30). San Diego, CA: Academic Press.

Hagmayer, Y., & Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. *Memory & Cognition, 30*(7), 1128-1137.

Hankins, T. L. (1999). Blood, dirt, and nomograms: A particular history of graphs. *Isis, 90*, 50-80.

Hardy, I., Schneider, M., Jonen, A., Stern, E., & Möller, K. (2005). Fostering diagrammatic reasoning in science education. *Swiss Journal of Psychology, 64*(3), 207-217.

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335-346.

Harrison, M., Campos, J. C., & Loer, K. (2008). Formal analysis of interactive systems: Opportunities and weaknesses. In P. Cairns & A. L. Cox (Eds.), *Research methods for human-computer interaction* (pp. 88-111). New York: Cambridge University Press.

Hauser, M. D., Chomsky, N., & Fitch, W. (2002). The faculty of language: What Is it, who has it, and how did it evolve? *Science, 298*(5598), 1569-1579.

Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research, 38*(1), 125-137.

Heer, J., & Agrawala, M. (2008). Design considerations for collaborative visual analytics. *Information Visualization, 7*, 49-62.

Heer, J., Viégas, F., & Wattenberg, M. (2009). Voyagers and voyeurs: Supporting asynchronous collaborative information visualization. *Communications of ACM, 52*(1), 87-97.

Hegarty, M., Carpenter, P. A., & Just, M. A. (1991). Diagrams in the comprehension of scientific texts. In R. Barr, M. L. Kamil, P. Mosenthal & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 641-688). Mahwah, NJ: Erlbaum.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*, 717-742.

Henderson, J. M. (2003). Human gaze control in real-world scene perception. *Trends in Cognitive Sciences, 7*, 498-504.

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In *The interface of language, vision, and action: Eye movements and the visual world* (pp. 1 - 58). New York: Psychology Press.

Hewett, T. T. (Ed.). (1992). *ACM SIGCHI curricula for human-computer interaction*. New York: ACM Press.

Hill, W. C., & Hollan, J. D. (1991). Deixis and the future of visualization excellence. In *Proceedings of the 2nd conference on visualization* (pp. 314-320).

Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin, 107*(1), 65-81.

Hoffman, D. D., & Singh, M. (1997). Salience of visual parts. *Cognition, 63*, 29-78.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction, 7*(2), 174-196.

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

Horst, S. (2005). The computational theory of mind. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy (Spring 2009 edition), URL = <http://plato.stanford.edu/archives/spr2009/entries/computational-mind/>*.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

Jackendoff, R. (1996). The architecture of the linguistic-spatial interface. In P. Bloom, M. Peterson, L. Nadel & M. Garrett (Eds.), *Language and space* (pp. 1-30). Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition, 97*(2), 211-225.

Jacko, J. A., Leonard, V. K., & Scott, I. U. (2009). Perceptual impairments: New advancements promoting technological access. In A. Sears & J. A. Jacko (Eds.), *Human-Computer Interaction: Designing for diverse users and domains* (pp. 93-110). Boca Raton, London, & New York: CRC Press.

James, W. (1890). *The principles of psychology*. New York: Holt.

Jastrow, J. (1890). Minor contributions. Studies from the Laboratory of Experimental Psychology of the University of Wisconsin. *American Journal of Psychology, 3*, 43-58.

John, B. E., & Kieras, D. E. (1996a). Using GOMS for user interface design and evaluation:

Which technique? *ACM Transactions on Computer-Human Interaction, 3*(4), 287-319.

John, B. E., & Kieras, D. E. (1996b). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction, 3*(4), 320-351.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59*, 193-224.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329-354.

Kabak, B. (2007). Turkish suspended affixation. *Linguistics, 45*(2), 311-347.

Kelter, S., Kaup, B., & Claus, B. (2004). Representing a described sequence of events: A Dynamic view of narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 451-464.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science, 11*, 201-258.

Kennedy, C. (1999). *Projecting the adjective: The syntax and semantics of gradability and comparison*. New York: Garland [Santa Cruz: University of California, Santa Cruz dissertation, 1997].

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language, 81*(2), 345-381.

Kieras, D. E. (1997). A guide to GOMS model usability evaluation using NGOMSL. In M. Helander & T. Landauer (Eds.), *The handbook of human-computer interaction* (2nd ed., pp. 733-766). Amsterdam: North-Holland.

Kieras, D. E. (2004). Task analysis and the design of functionality. In A. Tucker (Ed.), *The computer science and engineering handbook* (2nd ed., pp. 46-41 -- 46-25). Boca Raton: CRC.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction, 4*(12), 391-438.

Kintsch, W. (1988). The role of knowledge in discourse comprehension. A construction integration model. *Psychological Review, 95*, 163-182.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.

Klein, E. (1980). A semantic for positive and comparative adjectives. *Linguistics and Philosophy, 4*, 1-45.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition, 95*(1), 95-127.

Koerner, C. (2004). Sequential processing in comprehension of hierarchical graphs. *Applied Cognitive Psychology, 18*, 467-480.

Koerner, C., & Gilchrist, I. D. (2004). Eye movements in a simple spatial reasoning task. *Perception, 33*(4), 485-494.

Kong, N., & Agrawala, M. (2009). Perceptual interpretation of ink annotations on line charts. In *UIST 2009: Proceedings of the 22nd ACM symposium on user interface*

software and technology (pp. 233-236). New York: ACM.
Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research, 29*(6), 627-645.
Kornfilt, J. (1996). On some copular clitics in Turkish. In A. Alexiadou, N. Fuhrhop, P. Law & S. Löhken (Eds.), *In ZAS Papers in Linguistics 6* (pp. 96-114). Berlin: Zentrum für Allgemeine Sprachwissenschaft.
Kornfilt, J. (1997). *Turkish*. London and New York: Routledge.
Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA.: Harvard University Press.
Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology, 3*(3), 185-226.
Kosslyn, S. M. (1994a). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.
Kosslyn, S. M. (1994b). *Elements of graph design*. New York: W. H. Freeman.
Kosslyn, S. M. (2006). *Graph design for the eye and mind*. New York: Oxford University Press.
Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2003). Mental imagery: Against the nihilistic hypothesis. *Trends in Cognitive Sciences, 7*, 109-111.
Kowler, E. (1999). Eye movements and visual attention. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 306-308). Cambridge, MA: MIT Press.
Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cognition and Instruction, 20*, 47-77.
Kozhevnikov, M., Kosslyn, S. M., & Shephard, J. (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory & Cognition, 33*, 710-726.
Köhler, W. (1929). *Gestalt psychology*. New York: Liveright.
Lagnado, D., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 451-460.
Laird, J. E., Rosenbloom, P. S., & Newell, A. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence, 33*, 1-64.
Lambert, J. H. (1779). *Pyrometrie; oder, vom maasse des feuers und der wärme mit acht kupfertafeln*. Berlin.
Land, M. F., & Hayhoe, M. (2001). In what ways do eyemovements contribute to everyday activities? *Vision Research, 41*(25-26), 3559–3565.
Landau, B., & Jackendoff, R. (1993). What and where in spatial language and spatial cognition. *Behavioral and Brain Sciences, 16*(2), 217-265.
Landauer, T. (1995). *The trouble with computers: Usefulness, usability, and productivity.* Cambridge, MA: MIT Press.
Langley, P., Laird, J. E., & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research, 10*(2), 141-160.
Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.
Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research, 60*, 1-64.
Levelt, W. J. M. (1989). *Speaking, from intention to articulation*. Cambridge, MA: MIT Press.
Levie, W. H. (1987). Research on pictures: A guide to the literature. In D. M. Willows & H.

A. Houghton (Eds.), *The psychology of illustration* (Vol. 1, pp. 1-50). New York: Springer-Verlag.

Levinson, S. C. (1996). Frames of reference and Molyneux's question: Crosslinguistic evidence. In P. Bloom, M. A. Peterson, L. Nadel & M. F. Garrett (Eds.), *Language and space* (pp. 109-169). Cambridge, MA: MIT Press.

Levitin, D. J. (2002). Experimental design in psychological research. In *Experimental design in psychoacoustic research* (pp. 299-328). Cambridge, MA: MIT Press.

Lewis, C., Poison, P., Wharton, C., & Rieman, J. (1990). Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In *CHI '90: Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 235-242). New York: ACM Press.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research, 25*, 93-115.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10*, 447-454.

Liversedge, S. P., Rayner, K., White, S. J., Vergilino-Perez, D., Findlay, J. M., & Kentridge, R. W. (2004). Eye movements when reading disappearing text: Is there a gap effect in reading? *Vision Research, 44*, 1013–1024.

Loftus, G. R. (1981). Tachistoscopic simulations of eye fixations on pictures. *Journal of Experimental Psychology: Human Learning and Memory, 7*(5), 369-376.

Loftus, G. R. (1993). A picture is worth a thousand p-values: On the irrelevance of hypothesis testing in the computer age. *Behavior Research Methods, Instrumentation and Computers, 25*, 250-256.

Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human-Computer Interaction, 8*(4), 353-388.

Lohse, G. L., Biolsi, K., Walker, N., & Rueter, H. (1994). A classification of visual representations. *Communications of ACM 37*(12), 36-49.

Lohse, G. L., Rueter, H., Biolsi, K., & Walker, N. (1990). Classifying visual knowledge representations: A foundation for visualization research. In *Proceedings of the 1st conference on visualization* (pp. 131-138). San Francisco, CA: IEEE Computer Society Press.

Lovett, A., & Forbus, K. (2009a). Using a visual routine to model the computational of positional relations. In N. Taatgen, H. van Rijn, J. Nerbonne & L. Schomaker (Eds.), *Proceedings of the 21st annual conference of the Cognitive Science Society*. Amsterdam, The Netherlands.

Lovett, A., & Forbus, K. (2009b). Computing human-like qualitative topological relations via visual routines. In I. Bratko & J. Žabkar (Eds.), *Proceedings of the 23rd international qualitative reasoning workshop*. Ljubljana, Slovenia.

Ludlow, P. (1989). Implicit comparison classes. *Linguistics and Philosophy, 12*, 519-533.

Majid, A., Sanford, A. J., & Pickering, M. J. (2006). Covariation and quantifier polarity: What determines casual attribution in vignettes *Cognition, 99*, 35-51.

Mandler, J. M. (1986). On the comprehension of temporal order. *Linguistic and Cognitive Processes, 1*, 309-320.

Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.

Marr, D., & Poggio, T. (1977). From understanding computation to understanding neural circuitry. *Neuroscience Research Progress Bulletin, 15*, 470-488.

Marslen-Wilson, W., & Tyler, L. K. (1980). Central processes in speech understanding.

Philosophical Transactions of the Royal Society, Series B, 295, 317–332.

Mautone, P. D., & Mayer, R. E. (2007). Cognitive aids for guiding graph comprehension. *Journal of Educational Psychology, 99*(3), 640-652.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge, MA: Cambridge University Press.

Mayer, R. E. (Ed.). (2005a). *The Cambridge handbook of multimedia learning*. Cambridge, MA: Cambridge University Press.

Mayer, R. E. (2005b). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). Cambridge, MA: Cambridge University Press.

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist, 63*(8), 760-769.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

Mevarech, Z. R., & Kramarsky, B. (1997). From verbal descriptions to graphic representations: Stability and change in students' alternative conceptions. *Educational Studies in Mathematics, 32*(3), 229-263.

Meyer, D. E., & Kieras, D. E. (1997a). A computational theory of executive cognitive processes and multiple-task performance 1. Basic mechanisms. *Psychological Review, 104*(1), 3-65.

Meyer, D. E., & Kieras, D. E. (1997b). A computational theory of executive cognitive processes and multiple-task performance 2. Accounts of psychological refractory-period phenomena. *Psychological Review, 104*(4), 749-791.

Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*(3), 398-407.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In D. R. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. II, pp. 419-491). New York: John Wiley.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

Millis, K. K., Graesser, A. C., & Haberlandt, K. (1993). The impact of connectives on the memory for expository text. *Applied Cognitive Psychology, 7*, 317-339.

Milroy, R., & Poulton, E. C. (1978). Labelling graphs for improved reading speed. *Ergonomics, 21*(1), 55-61.

Miyake, N. (1997). Making internal processes external for constructive collaboration. In J. Marsh, C. Nehaniv & B. Gorayska (Eds.), *CT '97: Proceedings of the 2nd international conference on cognitive technology: Humanizing the information age* (pp. 119-123): IEEE Computer Society Press.

Monaghan, P., & Stenning, K. (1998). Effects of representational modality and thinking style on learning to solve reasoning problems. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th annual conference of the Cognitive Science Society* (pp. 716-721). Mahwah, NJ: Lawrence Erlbaum.

Montello, D. R. (2002). Cognitive map-design research in the twentieth century: Theoretical and empirical approaches. *Cartography and Geographic Information Science, 29*(3), 283-304.

Motter, B. (1999). Attention in the animal brain. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 41-43). Cambridge, MA: MIT Press.

Myers, G. (1988). Every picture tells a story: Illustrations in E.O. Wilson's Sociobiology. *Human Studies, 11*, 235-269.

Narens, L., & Luce, R. D. (1986). Measurement: The theory of numerical assignments. *Psychological Bulletin, 99*(2), 166-180.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *ACM 1975 Turing Award Lecture*, 287-317.

Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 25-64). New York: John Wiley & Sons.

Nielsen, J., & Mack, R. L. (Eds.). (1994). *Usability inspection methods*. New York: John Wiley & Sons.

Norman, D. A. (1990). *The design of everyday things*. New York: Doubleday.

Noton, D., & Stark, L. W. (1971). Scan paths in eye movements during pattern perception. *Science, 171*, 308-311.

Novick, L. R. (2006). The importance of both diagrammatic conventions and domain-specific knowledge for diagram literacy in science: The hierarchy as an illustrative case. In D. Barker-Plummer, R. Cox & N. Swoboda (Eds.), *Diagrammatic representation and inference: Lecture notes in computer science* (Vol. 4045, pp. 1-11). Berlin: Springer.

Oestermeier, U., & Hesse, F. W. (2000). Verbal and visual causal arguments. *Cognition, 75*, 65-104.

Ohtsuka, K., & Brewer, W. F. (1992). Discourse organization in the comprehension of temporal order in narrative texts. *Discourse Processes, 15*, 317-336.

Olson, J. R., & Olson, G. M. (1990). The growth of cognitive modeling in human-computer interaction. *Human Computer Interaction, 1990*(5), 221-265.

Osherson, D. N., Kosslyn, S. M., & Hollerbach, J. M. (1990). *Visual cognition and action: An invitation to cognitive science* (Vol. 1). Cambridge, MA: MIT Press.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 259-303). Hillsdale, NJ: Erlbaum.

Paoletti, G. (2004). Writing-to-learn and graph-drawing as aids of the integration of text and graphs. In G. Rijlaarsdam, H. van den Bergh, M. Couzijn & H. V. D. Bergh (Eds.), *Effective learning and teaching of writing: A handbook of writing in education* (2nd ed., Vol. 14, pp. 587-598). New York: Kluwer Academic Publishers.

Pashler, H. E. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.

Peden, B. F., & Hausmann, S. E. (2000). Data graphs in introductory and upper level psychology textbooks: A content analysis. *Teaching of Psychology, 27*(2), 93-97.

Peebles, D. J., & Cheng, P. C.-H. (2001). Graph-based reasoning: From task analysis to cognitive explanation. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the twentythird annual conference of the Cognitive Science Society* (pp. 762-767). Mahwah, NJ: Lawrence Erbaum.

Peebles, D. J., & Cheng, P. C.-H. (2002). Extending task analytic models of graph-based reasoning: A cognitive model of problem solving with Cartesian graphs in ACT-R/PM. *Cognitive Systems Research, 3*, 77-86.

Peebles, D. J., & Cheng, P. C.-H. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors, 45*(1),

28-35.
Peebles, D. J., Cheng, P. C.-H., & Shadbolt, N. (1999). Multiple processes in graph-based reasoning. In M. Hahn & S. S. Stoness (Eds.), *Proceedings of the 21st annual conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence, Erlbaum.
Peeck, J. (1987). The role of illustrations in processing and remembering illustrated text. In D. M. Willows & H. A. Houghton (Eds.), *The psychology of illustration* (Vol. 1, pp. 115-151). New York: Springer-Verlag.
Pickering, M. J., McElree, B., & Garrod, S. (2004). Interactions of language and vision restrict "visual world" interpretations. Manuscript.
Pinker, S. (1981). A theory of graph comprehension. Occasional Paper 10. Massachusetts Institute of Technology, Center for Cognitive Science.
Pinker, S. (1989). *Learnability and cognition*. Cambridge, MA: MIT Press.
Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73-126). Hillsdale, NJ: Erlbaum.
Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition, 95*(2), 201-236.
Playfair, W. (1786). *The commercial and political atlas*. London: Corry.
Ponsoda, V., Scott, D., & Findlay, J. M. (1995). A probability vector and transition matrix analysis of eye movements during visual search. *Acta Psycholgica, 88*, 167-185.
Poole, A., & Ball, L. J. (2005). Eye tracking in HCI and usability research. In C. Ghaoui (Ed.), *Encyclopedia of human computer interaction* (pp. 211-219). Hershey: Idea Group Reference.
Posner, M. I., & Fernandez-Duque, D. (1999). Attention in the human brain. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 43-46). Cambridge, MA: MIT Press.
Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*(2), 160-174.
Preim, B., Michel, R., Hartmann, K., & Strothotte, T. (1998). Figure captions in visual interfaces. In *Proceedings of the working conference on advanced visual interfaces* (pp. 235-246). L'Aquila, Italy.
Publication Manual of the American Psychological Association (5th ed.). (2001). Washington, DC: American Psychological Association.
Purcell, A. T., & Gero, J. S. (1998). Drawings and the design process. *Design Studies, 19*, 389-430.
Pustejovsky, J. (1995). *The generative lexicon*. Cambridge: MIT Press.
Putnam, H. (1961). Brains and behavior, originally read as part of the program of the American Association for the Advancement of Science, Section L (History and Philosophy of Science), December 27, 1961. Reprinted in N. Block (ed.), *Readings in the philosophy of psychology*, (Vol. 1, pp. 24-36), Cambridge, MA: Harvard University Press, 1980.
Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin, 80*, 1-24.
Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.
Pylyshyn, Z. W. (2003). *Seeing and visualizing: It's not what you think*. Cambridge, MA: MIT Press.
Ratanasit, D., & Moore, M. M. (2005). Representing graphical user interfaces with sound: A

review of approaches. *Journal of Visual Impairment and Blindness, 99*(2), 69-84.

Ratwani, R. M., Trafton, J. G., & Boehm-Davis, D. A. (2008). Thinking graphically: Connecting vision and cognition during graph comprehension. *Journal of Experimental Psychology: Applied, 14*(1), 36-49.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rayner, K., & Pollatsek, A. (1992). Eye movements and scene perception. *Canadian Journal of Psychology, 46*(3), 342-476.

Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied, 7*, 219-226.

Rayner, K., & Sereno, S. C. (1994). Eye movements in reading. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 57-81). San Diego, CA: Academic Press.

Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science, 20*(1), 6-10.

Reichle, E. D. (2006). Computational models of eye-movement control during reading: Theories of the ‘‘eye–mind’’ link. *Cognitive Systems Research, 7*, 2-3.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E–Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research, 7*, 4-22.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*(4), 445-476.

Reisberg, D. (1987). External representations and the advantages of externalizing one's thoughts. In E. Hunt (Ed.), *Proceedings of the 9th annual conference of the Cognitive Science Society* (pp. 281-293). Seattle, WA/Hillsdale, NJ: Erlbaum.

Renshaw, J. A., Finlay, J. E., Tyfa, D., & Ward, R. D. (2003). Designing for visual influence: An eye tracking study of the usability of graphical management information. In M. Rauterberg et al. (Ed.), *Proceedings of the 9th IFIP TC13 international conference on human-computer interaction* (pp. 144-151): IOS Press.

Renshaw, J. A., Finlay, J. E., Tyfa, D., & Ward, R. D. (2004). Understanding visual influence in graph design through temporal and spatial eye movement characteristics. *Interacting with Computers, 16*, 557-578.

Richardson , D. C., Dale, R., & Spivey, M. J. (2007). Eye movements in language and cognition: A brief introduction. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson & M. J. Spivey (Eds.), *Methods in Cognitive Linguistics* (pp. 325-346).

Richardson, D. C., & Spivey, M. J. (2004a). Eye tracking: Characteristics and methods. In G. Wnek & G. Bowlin (Eds.), *Encyclopedia of biomaterials and biomedical engineering*. New York: Marcel Dekker.

Richardson, D. C., & Spivey, M. J. (2004b). Eye tracking: Research areas and applications. In G. Wnek & G. Bowlin (Eds.), *Encyclopedia of biomaterials and biomedical engineering*. New York: Marcel Dekker.

Richter, E. M., Engbert, R., & Kliegl, R. (2006). Current advances in SWIFT. *Cognitive Systems Research, 7*, 23-33.

Rinck, M., Gámez, E., Día, J. M., & de Vega, M. (2003). Processing of temporal information: Evidence from eye movements. *Memory & Cognition, 31*(1), 77-86.

Rotstein, C., & Winter, Y. (2004). Total adjectives vs. partial adjectives: Scale structure and higher order modifers. *Natural Language Semantics, 12*, 259-288.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition, 17*(6), 759-769.

Sapir, E. (1944). Grading: A study in semantics. *Philosophy of Science, 11*, 93-116.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies, 45*, 185-213.

Schiano, D., & Tversky, B. (1992). Structure and strategy in encoding simplified graphs. *Memory & Cognition, 20*(1), 12-20.

Schnotz, W. (2002). Commentary - Towards an integrated view of learning from text and visual displays. *Educational Psychology Review, 14*(1), 101-120.

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49-69). Cambridge, MA: Cambridge University Press.

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations *Learning and Instruction, 13*, 227-237.

Seufert, T., Schütze, M., & Brünken, R. (2009). Memory characteristics and modality in multimedia learning: An aptitude-treatment-interaction study. *Learning and Instruction, 19*, 28-42.

Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General, 124*, 43-61.

Shah, P., Freedman, E., & Vekiri, I. (2005). The comprehension of quantitative information in graphical displays. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 426-476). New York: Cambridge University Press.

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review, 14*(1), 47-69.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 25*, 4-27.

Shneiderman, B. (1987). *Designing the user interface: Strategies for effective human-computer interaction*. Reading, MA: Addison-Wesley.

Shneiderman, B., & Plaisant, C. (2010). *Designing the user interface: Strategies for effective human-computer interaction* (5th ed.). Upper Saddle River, NJ: Addison-Wesley/Pearson.

Simkin, D., & Hastie, R. (1987). An information processing analysis of graph perception. *Journal of the American Statistical Association, 82*(398), 454-465.

Skinner, B. F. (1957). *Verbal behavior*. Acton, MA: Copley.

Spence, I. (2000). The invention and use of statistical charts. *Journal de la Société Francaise de Statistique, 141*, 77-81.

Spence, I. (2005). No humble pie: The origins and usage of a statistical chart. *Journal of Educational and Behavioral Statistics, 30*, 353-368.

Spence, I., & Lewandowsky, S. (1991). Displaying proportions and percentages. *Applied Cognitive Psychology, 5*, 61-77.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science, 19*, 97-140.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677-680.

Stevens, S. S. (1975/1986). *Psychophysics. Introduction to its perceptual, neural, and social prospects*. New Brunswick (USA) Oxford: Transaction Books (1st ed., Wiley, New

York, 1975).

Stewart, T., & Travis, D. (2003). Guidelines, standards, and style guides. In J. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (pp. 991-1005). Mahwah, NJ: Lawrence Erlbaum.

Stokes, D. E. (1997). *Pasteur's quadrant – Basic science and technological innovation*. Washington, DC: Brookings Institution Press.

Story, M. F., Mueller, J. L., & Mace, R. L. (1998). *The universal design file: Designing for people of all ages and abilities*. Relaigh: North Carolina State University.

Sun, R. (2009). Theoretical status of computational cognitive modeling. *Cognitive Systems Research, 10*(2), 124-140.

Suwa, M., & Tversky, B. (1997). What do architects and students perceive in their design sketches? A protocol analysis. *Design Studies, 18*, 385-403.

Suwa, M., & Tversky, B. (2002). External representations contribute to the dynamic construction of ideas. In M. Hegarty, B. Meyer & N. H. Narayanan (Eds.), *Diagrammatic representation and inference* (pp. 341-343). New York: Springer-Verlag.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Tabachneck-Schijf, H. J. M., Leonardo, A. M., & Simon, H. A. (1997). CaMeRa: A computational model of multiple representations. *Cognitive Science, 21*, 305-350.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632.

Tappe, H., & Habel, C. (1998). Verbalization of dynamic sketch maps: Layers of representation and their interaction. [full version of one page abstract / Poster presented at CogSci '98: The annual conference of the Cognitive Science Society]. Madison WI, *URL= <http://www.informatik.uni-hamburg.de/WSV/sprachproduktion/CogSci98.ps.gz>*.

The Center for Universal Design. (1997). *The principles of Universal Design, Version 2.0*. Raleigh, NC: North Carolina State University.

Thomas, N. J. T. (2009). Mental imagery. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy (Winter 2009 edition), URL = <http://plato.stanford.edu/archives/win2009/entries/mental-imagery/>*.

Tschander, L., Schmidtke, H., Habel, C., Eschenbach, C., & Kulik, L. (2003). A geometric agent following route instructions. In F. Freksa, W. Brauer, C. Habel & K. F.Wender (Eds.), *Spatial cognition III* (pp. 89-111). Heidelberg: Springer.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphic Press.

Tufte, E. R. (1990). *Envisioning information*: Graphics Press.

Tversky, B. (2004). Semantics, syntax, and pragmatics of graphics. In K. Holmqvist & Y. Ericsson (Eds.), *Language and visualisation* (pp. 141-158). Lund: Lund University Press.

Ullman, S. (1984). Visual routines. *Cognition, 18*, 97-106.

Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *The Quarterly Journal of Experimental Psychology, 57A*, 165-182.

van der Meer, E., Beyer, R., Heinze, B., & Badel, I. (2002). Temporal order relations in language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 770-779.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language, 82*(4), 767-794.

Velleman, P. F., & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *The American Statistician, 47*(1), 65-72.

Vendler, Z. (1957). Verbs and times. *Philosophical Review, 56*, 143-160.

Verkuyl, H. (1972). On the compositional nature of the aspects. *Foundations of Language Supplement Series, 15*, Dordrecht: Reidel.

Verkuyl, H. (1993). *A theory of aspectuality. The interaction between temporal and atemporal structure*. Cambridge: Cambridge University Press.

Vernon, M. D. (1953). Presenting information in diagrams. *Educational Technology Research and Development, 1*(3), 147-158.

Viégas, F. B., & Wattenberg, M. (2006). Communication-minded visualization: A call to action. *IBM Systems Journal, 45*(4), 801-812.

Visser, W. (2006). Designing as construction of representations: A dynamic viewpoint in cognitive design research. *Human Computer Interaction, 21*, 103-152.

von Helmholtz, H. (1925). *Handbuch der physiologischen optik (Treatise on physiological optics) Vol. III, (Translated from the Third German ed.)*. Rochester, NY: The Optical Society of America.

von Stechow, A. (1984). Comparing semantic theories of comparison. *Journal of Semantics, 3*, 1-77.

Wahlster, W., André, E., Finkler, W., Profitlich, H.-J., & Rist, T. (1993). Plan-based integration of natural language and graphics generation. *Artificial Intelligence Review, 63*, 387-427.

Wainer, H., & Velleman, P. F. (2001). Statistical graphics: Mapping the pathways of science. *Annual Review of Psychology, 52*, 305-335.

Wallen, E., Plass, J. L., & Brünken, R. (2005). The function of annotations in the comprehension of scientific texts: Cognitive load effects and the impact of verbal ability. *Educational Technology Research and Development, 53*(3), 59-71.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition, 85*, 79-112.

Washburne, J. N. (1927). An experimental study of various graphic, tabular and textual methods of presenting quantitative material. *Journal of Educational Psychology, 18*, 361-376.

Weaver, C. A., III, & Kintsch, W. (1991). Expository text. In R. Barr, M. L. Kamil, P. Mosenthal & P. D. Pearson (Eds.), *Handbook of reading research* (pp. 230-244). White Plaines, NY: Longman.

Webb, N., & Renshaw, T. (2008). Eyetracking in HCI. In P. Cairns & A. L. Cox (Eds.), *Research methods for human-computer interaction* (pp. 35-69). New York: Cambridge University Press.

Wertheimer, M. (1923/1938). Untersuchungen zur Lehre von der Gestalt II. Psychologische Forschung, 4, 301-350. (Excerpts translated into English as 'Laws of organization in perceptual forms' in W. D. Ellis (Ed.), A source book of Gestalt psychology (pp. 71-

88). New York: Hartcourt, Brace and Co., and as 'Principle of perceptual organization' in D.C. Beardslee & Michael Wertheimer (Eds.), Readings in Perception, Princeton, NJ: D. Van Nostrand Co.).

Wewers, M. E., & Lowe, N. K. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing & Health, 13*(4), 227-236.

Wharton, C., Rieman, J., Lewis, C., & Poison, P. (1994). The cognitive walkthrough method: A practitioner's guide. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 105-140). New York: John Wiley & Sons.

Winn, B. (1987). Charts, graphs, and diagrams in educational materials. In D. M. Willows & H. A. Houghton (Eds.), *The psychology of illustration* (Vol. 1, pp. 152-198). New York: Springer-Verlag.

Wooding, D. S. (2002). Eye movements of large populations: II. Deriving regions of interest, coverage, and similarity using fixation maps. *Behavior Research Methods, Instruments and Computers, 34*(4), 518-528.

Wooding, D. S., Mugglestone, M. D., Purdy, K. J., & Gale, A. G. (2002). Eye movements of large populations: I. Implementation and performance of an autonomous public eye tracker. *Behavior Research Methods, Instruments and Computers, 34*(4), 509-517.

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

Yu, W., & Brewster, S. A. (2003). Evaluation of multimodal graphs for blind people. *Universal Access in the Information Society, 2*(2), 105-124.

Yu, W., Ramloll, R., Brewster, S. A., & Riedel, B. (2001). Exploring computer-generated line graphs through virtual touch. In *ISSPA 2001: Proceedings of the 6th international symposium on signal processing and its applications, special session on enhanced user interfaces for non-visual image display* (pp. 72-75). Kuala-Lumpur, Malaysia.

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions, A, 226*, 267-298.

Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & Cognition, 27*, 1073-1079.

Zhang, J. (1996). A representational analysis of relational information displays. *International Journal of Human-Computer Studies, 45*, 59-74.

Zhang, J. (1997). Distributed representation as a principle for the analysis of cockpit information displays. *International Journal of Aviation Psychology, 7*(2), 105-121.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*, 87-122.

Zhang, J., & Norman, D. A. (1995). A representational analysis of numeration systems. *Cognition, 57*, 271-295.

Zhang, J., & Wang, H. (2009). An exploration of the relations between external representations and working memory. *PLoS ONE, 4*(8), e6513.

Zheng, Q., Booth, K., & McGrenere, J. (2006). Co-authoring with structured annotations. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 131-140). New York: ACM Press.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1196-1207.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological*

Science, 6, 292-297.

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 224-245). Cambridge, MA: Cambridge University Press.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162-185.