Eye Tracking in Multimodal Comprehension of Graphs

Cengiz Acartürk[1], Christopher Habel[2]

[1]Middle East Technical University, Informatics Institute,
06800 Ankara, Turkey
acarturk@metu.edu.tr
[2]University of Hamburg, Computer Science, Vogt-Koelln-Str. 30,
22527 Hamburg, Germany
habel@informatik.uni-hamburg.de

Abstract. Eye tracking methodology has been a major empirical research approach for the study of online comprehension processes in reading and scene viewing. The use of eye tracking methodology for the study of diagrammatic representations, however, has been relatively limited so far. The investigation of specific types of diagrammatic representations, such as statistical graphs is even scarce. In this study, we propose eye tracking as an empirical research approach for a systematic analysis of multimodal comprehension of line graphs. Based on a framework of multimodal comprehension of graphs and texts, which focuses on the role of spatial concepts, we present an experimental investigation of linguistic guidance and eye movement control in comprehension of time-domain line graphs with a particular focus on the analysis of gaze patterns in graph inspections.

Keywords: eye tracking, statistical graphs, line graphs, multimodal comprehension.

1 Introduction

The study of diagrams has been subject to interdisciplinary research, thus involving the application of theoretical and empirical research approaches originated from different disciplines, such as formal analysis, usability inspection, cognitive modeling and controlled experiments. The research that focuses on human comprehension of diagrams usually employs cognitive modeling and/or controlled experiments with human participants. A major goal in controlled experiments is to measure certain aspects of human comprehension in terms of independent variables, such as the time between the onset of the stimuli and the response of the participants (response time), the accuracy of the judgment responses (response accuracy) or eye movements. Those measures that provide information about comprehension after the completion of a task, such as response time and response accuracy are called *offline comprehension measures*. On the other hand, more information about the course of comprehension can be obtained by employing *online comprehension measures*. Online comprehension measures are based on measuring psychophysical aspects of comprehension, such as the recording of eye movements or brain activity during the course of comprehension.

Eye movements of human participants during their course of inspection of stimuli provide information about attentional processes that take place during comprehension. Therefore eye movement recordings provide richer and robust data compared to offline measures by allowing the researcher to access information about indirect indicators of possible cognitive processes during comprehension. In particular eye movement data may reveal information about allocation of attention to different components of a diagrammatic representation during comprehension [1].

A closer look to the analysis of eye movement data reveals approximately 120 types of eye movement measures, selected according to the needs of the research. Eye movement measures can be classified into four major classes [2].

- *Movement measures* refer to properties of movement of the eyes, such as direction, amplitude, duration and velocity.
- *Position measures* refer to properties related to fixation positions such as position dispersion, similarity, duration and dilation.
- *Numerosity measures* refer to number, proportion and rate of countable eye movement entities, such as saccades, fixations and regressions.
- *Latency measures* refer to time difference between onset of the stimuli and eye movement events (e.g., fixations, saccades); *distance measures* refer to comparison of two simultaneous entities (e.g., eye position and mouse position).

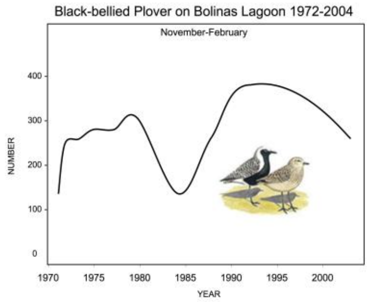
The selection of the appropriate eye movement measure depends on the specifications in experimental investigations. Eye movement measures that belong to fixations and saccades (e.g., mean duration per fixation, total gaze time on the stimuli) are relatively well investigated as indirect indicators of cognitive processes in several domains. For instance, eye movement analysis has been widely used for the study of cognitive processes in reading comprehension [3]. The major eye movement measures in reading are fixation duration and the number and percentage of regressions (backward eye movements during the course of reading). According to the classification introduced above [2], these measures belong to the class of numerosity measures. During the past two decades, research in psycholinguistics has shown that language comprehension—in particular, spoken language comprehension—interacts with visual perception by analyzing position of fixations on the entities in a visual scene [4]. Visual scene viewing is a domain in which eye movement analysis has been used widely [5]. In scene viewing, a broader range of eye movement measures has been used including fixation duration, total gaze time on the stimuli and scanpath analysis.

The use of eye tracking has been relatively limited in diagrams research compared to reading and visual scene viewing. On the other hand, the research on diagrams reveals the use of a broad range of eye movement measures: Position measures have been used alongside gaze time in the development of cognitive models for inspection of specific diagrams types. For instance, in computer simulation models of graph perception, predictions for necessary eye movement positions and gaze time are used for modeling graph inspection tasks [6], [7] and for predicting perceptual task effort in accomplishing the tasks [8]. The number of transitions between regions on the display stimuli has been analyzed (viz., scanpath analysis) for supporting models of graph comprehension processes [9]. The analyses have also covered usability investigations of different diagram types in addition to cognitive modeling, such as comparative studies for 2-D and 3-D statistical graphs [10] and evaluation of effectiveness of interactive maps [11]. Consequently, a wide range of eye movement measures have been used, selected according to the empirical approach used in the research domain and the specific experimental tasks.

In the present study, we propose the eye movement analysis as an empirical approach for the study of multimodal comprehension of statistical graphs and text. We introduce the basic concepts in multimodal comprehension in Section 2. A theoretical framework for the analysis of multimodal comprehension of line graphs and text is presented in Section 3. In Section 4, we present a case study in which we employed a sentence-graph verification paradigm for the analysis of the predictions made by the theoretical framework.

2 Multimodal Comprehension of Graphs and Texts

Statistical graphs, such as line graphs and bar graphs, are usually accompanied by language, either in written form or spoken form. When graphs are presented in company with language, readers seem to integrate information provided by both modalities. The pivot of the integration in multimodal comprehension of linguistic representations and graphical representations is the construction of reference and co-reference relations. In Figure 1, verbal expressions such as ‘peaked at 300 birds’ or ‘a drop to 120 in 1985’ correspond to entities in the domain of discourse by means of internal spatial and conceptual representations that mediate between language and the domain of discourse. And similarly, entities in the domain of discourse are depicted by graphical configurations.



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.

Fig. 1. A sample graph-text document. Excerpted from a public waterbird census report published Wetlands Ecology Division of PRBO Conservation Science, retrieved on Mar. 24, 2012, http://www.prbo.org/cms/index.php?mid=367#bbpl

The type of the (non-linguistic) representation (e.g., pictorial illustrations, maps, statistical graphs) is a key determinant in multimodal comprehension due to peculiar internal characteristics of the representation's information content. For instance, the visual resemblance between the representing entity and the represented entity is a major characteristic of pictorial illustrations [12]. In particular, pictorial illustrations and the objects that they refer to have similar spatial layouts (cf. iconic similarity). Pictorial illustrations, however, do not have an internal syntax in the sense of representational formats as discussed in [13, p.31] [14]. On the other hand, statistical graphs are representational artifacts and they possess internal syntactic structures. Their syntactic analysis provides the necessary input for further processes at semantic and pragmatic-level analyses of graph comprehension [15], [16]. Despite the extensive investigations of pictorial illustrations and text in the domain of instructional science (cf. multimedia learning) [17], [18], those characteristics of statistical graphics and their implication on comprehension processes have been virtually unnoticed.

In graph-text documents, the *inventory* of entities referred to by language or by statistical graphs is not only restricted to *objects* (in a strict sense), but also covers *classes* (e.g. referred to by 'population of birds'), as well as *states*, *processes* and *events* etc. For instance, the 'decrease of population' depicted by the right-most part of the graph line in Figure 1 refers to a *process* concept, which in the terminology of graph interpretation is often described as *trend*. To sum up, both linguistic representations (verbal expressions) and graph representations (graph elements and graph configurations) are linked among each other and they are linked to entities in the domain of discourse, via reference and co-reference relations.

In the present study, based on the theoretical framework, we employ the task analysis method for the specification of a set of linguistically-guided graph comprehension tasks in a sentence-graph verification paradigm. We then evaluate the task analysis method by analyzing the results of an experimental investigation in which the participants performed graph inspections after they read sentential expressions in sentence-graph pairs. The following sections present the theoretical framework, the task analysis method for the specification of graph inspection tasks and the experimental evaluation of the task analysis.

3 Theoretical Framework: Coreference in Multimodal Comprehension

Taking language comprehension as a starting point, we proposed a theoretical framework of multimodal comprehension of graph and text, which focuses on the role of spatial concepts in multimodal graph-text comprehension [14], [19].[1] The working assumption in this extended framework has been that spatial concepts mediate between spatial terms in language (e.g., shape nouns, verbs of change in space and spatial prepositional phrases), the entities in the domain of discourse (e.g., the peak of population at 300) and spatial configurations in a second external representation, namely a graphical entity (e.g., a peak on the graph). In other words, spatial concepts provide a common conceptual basis for concurrent processing of linguistic and graphical entities in graph-text comprehension. Based on these theoretical assumptions, we investigated the structural aspects of graph-text combinations. In a set of experimental investigations, we looked into the construction of referential links between graphical entities and linguistic entities. In a set of experimental investigations, we employed various experimental approaches, including the analysis of readers' eye movement patterns and parameters during their course of comprehension. The investigations covered graph-text documents with a separate layout of graph and text (e.g., Figure 1 above), as well as verbally and graphically annotated graphs [22], [23], [24].

In comprehension of line graph and text, graphical entities and linguistic entities have to be linked to entities in the domain. Corresponding to the different types of entities in the domain, such as *objects*, *states*, *processes* and *events*, etc., the conceptual system, that is crucial for the integration of text-based and of graph-based information, consists of concepts of different ontological type. In time-domain line graphs, the basic concepts that are accessed by linguistic expressions such as 'rise' and 'fall' are process concepts, namely INCREASE($_{\_TEMP}$, $_{\_VALUE}$) and DECREASE($_{\_TEMP}$, $_{\_VALUE}$). Correspondingly, diagonal line segments access the corresponding process concepts, e.g. INCREASE_P($_{\_PATH}$,$_{\_SRSG}$). The PATH argument specifies the graphical entities (in this case the line segments in a line graph) as directed linear entities [21]. The SRS_G argument provides the spatial reference system specified by the graph framework [14]. In particular, the mapping from the temporal domain to the value domain, for instance in INCREASE($_{\_TEMP}$, $_{\_VALUE}$), is—in topological terminology—a specification of the mapping from the temporal domain to a 'position' in the value space. A position in the value space is referred to by a point on a PATH (represented by the graph line proper) and the concept INCREASE_P($_{\_PATH}$,$_{\_SRSG}$) is a specification of this correspondence. Accordingly, (1a) involves the process concept INCREASE with explicit *GOAL TEMP* and

[1] The framework for spatial and conceptual representations in Section 3 of the current paper is based on [20], [21].

GOAL VALUE arguments whereas (1b) involves the process concept INCREASE with explicit *GOAL TEMP* and *GOAL VALUE* arguments as well as *SOURCE TEMP* and *SOURCE VALUE* arguments. The *TEMP* and *VALUE* arguments are specified by the horizontal location and vertical location respectively in the graph framework.

(1) a. The number of birds increased to about 180 in 2005.
    b. The number of birds increased from about 30 in 1970 to about 180 in 2005.

The tasks analysis method employs the structural analysis of time-domain line graphs presented above for specifying particular information processing tasks in the course of a user's interaction with a complex information display [25], [26]. Task analysis is used for breaking down the task to be accomplished by the user into a set of subtasks. The task analysis method, applied to graph comprehension suggests that humans are expected to construct the mapping between the two arguments. Readers usually construct this mapping by inspecting the relevant points with respect to the horizontal location and vertical location in the graph frame. For this, the reader executes vertical saccades and horizontal saccades (henceforth, projections) between the points on the graph line and the axes. In the present study, we assume that in multimodal comprehension of graphics and text, the tasks are specified by the sentences and the task analysis method is employed for predicting the readers' graph inspection tasks. For instance, the inspection of the line graph that follows (1a) or (1b) should consist of at least three subtasks.

- A vertical projection between the horizontal axis and the corresponding point on the path
- A horizontal projection between the point and the vertical axis
- An inspection of the spatial properties of the INCREASE process concept which holds the following 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

The vocabulary of multimodal graph-text comprehension comprises verbal expressions, such as process concepts as INCREASE together with their arguments. The arguments are specified by path functions, such as prepositional phrases (PP) 'to' and 'from'.[2] The vocabulary consists of aspectual modifier PPs such as 'since' and 'until', as exemplified in (2), as well.

(2) a. The number of birds increased since 1970.

[2] The vocabulary of multimodal graph-text comprehension in Turkish, which is the language of the experiment, has the corresponding postpositions for the English PPs. A fine-grained analysis of the differences between Turkish postpositions and English PPs is beyond the scope of the present study.

b. The number of birds increased until 2005.[3]

In (2), the PPs ‘since’, ‘until’ and ‘between’ modify the process concept INCREASE by specifying SOURCE and GOAL arguments in a similar way to the modification introduced by the PPs ‘from’ and ‘to’ respectively though with higher emphasis on the act of increase.

In summary, the task analysis approach we propose in the present study takes the structural analysis of the vocabulary of multimodal comprehension as the basis for the predictions about what humans might be expected to perform. Focusing on linguistic guidance in graph comprehension, the tasks analysis method is used for decomposing the tasks specified by sentential descriptions, and then this information is used to analyze humans’ inspections (i.e., gaze patterns) of line graphs in sentence-graph pairs by employing sentence-graph verification paradigm. In the following section, we present an empirical evaluation of the structural analysis and task specification in terms of eye movements of human participants in graph inspection tasks.

4 The Experiment

This section reports the experiment that was conducted as a case study for the evaluation of the theoretical framework presented in the previous section and the evaluation of the outcomes of eye movement analysis. Thirty-six participants (mean age 21.8, *SD* = 1.62) were asked to report if a set of displayed sentence-graph pairs were correct or incorrect by pressing a key on the keyboard. Participants’ eye movements were recorded during their course of inspection of line graphs, as described below.

4.1 Materials, Design and Procedure

The experimental stimuli consisted of 12 sentences and 12 line graphs in time domain.[4] The materials (excerpted from the original source, see Figure 1 above) were modified according to the purpose of systematic investigation aimed at the experiment. Accordingly, the stimuli sentences were designed as described in Section 3.[5] The graphs represented the population of waterbirds between 1970 and 2005. All the graphs were broad representations of states and processes in the stimuli sentences.

[3] Despite the marginal acceptability in English translations, the use of ‘since’ and ‘until’ leads to acceptable sentences in Turkish. In the present study, English translations of the stimuli sentences were prepared by the first author.

[4] In this paper, we present a comparative analysis of six sentence-graph pairs (two pairs per each stimuli type), leaving a more complete analysis to an extended study.

[5] Four of the stimuli sentences used in the experiment are presented in (1) and (2) above. The remaining two sentences were word-order alternations of them (given in Table 1 and Table 2 below). We leave the investigation of word order in task specification and its influence on graph inspections to a further analysis. At this stage, we note that word order variation is flexible in Turkish, which was the language of the experiment.

Each sentence-graph pair consisted of a stimulus sentence followed by a stimulus graph. Figure 2 shows a sample sentence-graph pair from the experiment.

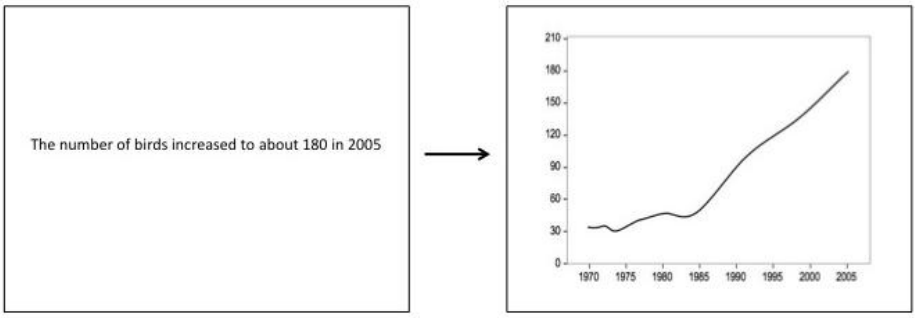


Fig. 2. A sample graph-sentence pair.

The experiment was conducted in single sessions. The task of the participant was to decide whether the displayed graph was correct or incorrect given the stimulus sentence. The presentation order of the sentences and the graphs were randomized within sentence-graph pairs so that local effects such as visual saliency in graphs were minimized. There was no time limitation set for reading the sentence and inspecting the graph given the stimuli sentence. Accordingly, the participants had control over when the display was changed from the text to the graph.

A 50 Hz. non-intrusive Tobii eyetracker recorded gaze patterns of the participants. The eye tracker was integrated into a 17” TFT monitor with a resolution of 1024x768 pixels. The spatial resolution and the accuracy of the eye tracker were 0.25° and 0.50° respectively. For the analysis of eye movement data, areas of interest (AOIs) were specified on the graph proper (one AOI at the start point, one at the endpoint and two in between) and AOIs were specified on the axis labels (one AOI covers the SOURCE values, one covers the GOAL values and one covers the values at the middle, separately for both axes). A schematic representation of AOI specification is shown in Figure 3.

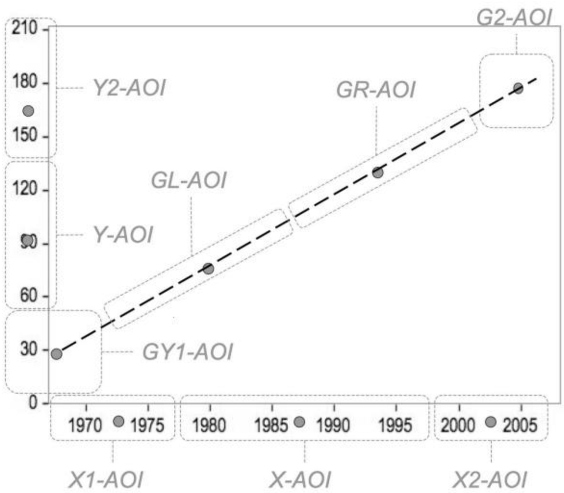
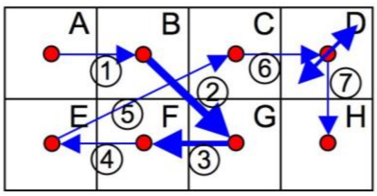


Fig. 3. A schematic representation of AOI specification. The dashed line is a schematic representation for the graph lines of different type in the experiment.

The eye movements were analyzed in terms of transitions of gaze between the AOIs. The transitions were analyzed by means of transition matrices [27], [28]. In a transition matrix, each cell shows the number of transitions from the AOI placed on a row to an AOI placed on a column. Figure 4 shows a sample gaze map on a grid of eight AOIs and the transition matrix for it.



from \ to	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	

Fig. 4. A sample gaze map and the transition matric for the illustrated gaze map.

In the gaze map shown in Figure 4, the transitions between the AOIs are represented by arrows. The thickness of an arrow shows its weight in the transition matrix. The values in a transition matrix are usually shown in terms of their weights, i.e. the number in each cell is divided by the total value of all the numbers in the matrix thus leading to a normalized transition matrix.

An eye movement measure that employs transition matrices is *transition matrix entropy*. Entropy is a measurement concept that is used for calculation of uncertainty associated with a random variable [2], [29]. The entropy in a transition matrix is calculated by the following equation, R is the normalized transition matrix, r_i values are normalized cell values and *p* shows probability of a value.

$$H(R) = - \Sigma\, p(r_i)\log_2 p(r_i) \text{ for } r_i \in R$$

When applied to a transition matrix, entropy is a measurement of the randomness of scanpath distribution across AOIs: a high value of the entropy has been proposed to align with the preference for an exploratory investigation of the display stimuli, whereas a low value indicates eye movement transitions between a few AOIs [2], [30]. Accordingly, the entropy obtained from a transition matrix is an indicator of the skewness of the transition distribution: a high value of entropy indicates an even distribution of transitions between different AOIs whereas a low value indicates a skewed distribution [30]. In the following section, the results are shown in terms of transition matrices for the sentence-graph pairs.

4.2 Results

The participants' responses for the 6 congruent sentence-graph pairs resulted in 36 x 6 = 216 eye movement protocols. Approximately 6% of the recorded data were not included into the analysis due to partial calibration problems in the eyetracker. All the remaining data were manually coded and transcribed, and the transition matrices were constructed for each graph-sentence pair by calculating the mean number of gaze transitions between the specified AOIs.

Figure 5 shows the gaze map and the transition matrix for a sample stimuli sentence 'The number of birds increased to about 180 in 2005'. In the gaze map, a thicker arrow shows a higher weight of the transition among all the transitions. Double-sided arrows show transitions within a specific AOI. Accordingly, the figure consists of vertical and horizontal arrows that show projections and diagonal arrows that show transitions on the graph line and the transitions between the graph line and the two axes. The exemplified gaze map in Figure 5 is a partial depiction of the transition matrix because the transitions below a threshold of 3% weight are not shown in the gaze map for clarity of the figure.

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	

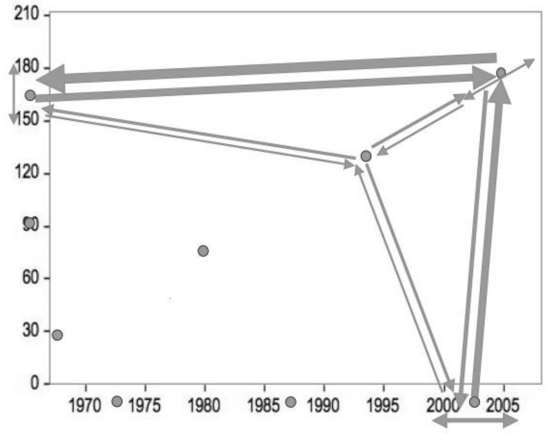


Fig. 5. The transition matrix (top) and the corresponding gaze map for the sample stimuli sentence ‘The number of birds increased in 2005 to about 180.’

Some of the values shown in the transition matrix in Figure 5 are zero, because no gaze transition was observed between the two AOIs (e.g., the transition from the G2 to X1 is zero, see Figure 3 for AOI specification). The numbers in bold show the transitions that have a higher weight than the threshold of 0.03 (i.e., 3% of the total gaze transitions) and they are shown in the gaze map.

The amount of vertical projections and horizontal projections is 42% of total gaze transitions. The inspection of the graph line for the INCREASE process, which was a required inspection task in the task specification, covers 13% of all the transitions for the stimuli sentence ‘The number of birds increased in 2005 to about 180.’ Consequently, a total of 55% of the gaze transitions on the map was already anticipated by the task analysis. Table 1 shows the results for all the analyzed stimuli sentences.

Table 1. The percentage of the transitions as anticipated by the task analysis method. The percentages show the sum of horizontal- and vertical-projections and graph-line inspections. The % column shows the percentage of the total experiment data anticipated by the task specification. S: source arguments, G: goal arguments, AM: aspectual modifiers. The sentences are literal translations from the language of the experiment to English (see fn. 3)

	Stimuli Sentence	**%**
G	The number of birds increased to about 180 in 2005	50
	The number of birds increased in 2005 to about 180 (*a word-order-alternation of the previous stimuli sentence*)	55
S + G	The number of birds increased from about 30 in 1970 to about 180 in 2005	53
	The number of birds increased between 1970 and 2005 from about 30 to about 180 (*a word-order-alternation of the previous stimuli sentence*)	50
AM	The number of birds increased since 1970	47
	The number of birds increased until 2005	47

The results suggest that the percentages of the anticipated eye movement transitions by the task analysis method is around 50% with some variability depending on the stimuli type of the sentence. Accordingly, the findings obtained by the task analysis show that it could be possible to anticipate the eye movements of the participants on the graph (after their reading of a stimuli sentence) at a certain extent.

On the other hand, the task analysis does not provide clear information about how different stimuli sentences influence eye movements in graph inspections. A fine-grained analysis of the influence of specific stimuli sentences on graph inspections can be achieved by using further eye movement measures. An appropriate measure for this is the transition matrix entropy, as described above. Table 2 shows the calculated entropy values for the transitions matrices, for each sentence stimuli.

Table 2. Entropy values for the transition matrices. S: source arguments, G: goal arguments, AM: aspectual modifiers.

	Stimuli Sentence	**H(R)**
G	The number of birds increased to about 180 in 2005	3.88
G	The number of birds increased in 2005 to about 180 (*a word-order-alternation of the previous stimuli sentence*)	3.78
S + G	The number of birds increased from about 30 in 1970 to about 180 in 2005	4.99
S + G	The number of birds increased between 1970 and 2005 from about 30 to about 180 (*a word-order-alternation of the previous stimuli sentence*)	5.20
AM	The number of birds increased since 1970	4.60
AM	The number of birds increased until 2005	4.62

A higher entropy value suggests an exploratory investigation of the graph [30], [2], whereas a lower value suggests that the participants performed a relatively small amount of gaze transitions between the areas of interest (AOIs). Therefore, the findings obtained in the experiment suggest that the inspections after different stimuli sentences lead to graph inspections with different characteristics. When the goal information is available together with the source information (i.e., when explicit values for the SOURCE VALUE argument is provided in the sentence as well as the values for the GOAL VALUE argument), an *exploratory pattern* of eye movements is observed. On the other hand, the presence of the values of only the SOURCE VALUE argument leads to a less number of gaze transitions between the AOIs. The gaze patterns in the absence of aspectual modifiers lead to neither an exploratory pattern nor a transition-pattern within a few AOIs but a gaze pattern in between.

In summary, the two methods presented in this paper revealed two different aspects of multimodal comprehension of graphs and texts. The task analysis method revealed that the gaze patterns could be partially anticipated by decomposing sentential expressions into their components. On the other hand, the analysis of eye movement patterns by the entropy calculation provided more specific information about the influence of specific types of sentential expressions and their components on graph inspections.

5 Conclusion and Future Work

Line graphs in time domain are interpreted as visualizations of states and processes, which are concepts that aggregate data points, often described as *trend* in the terminology of interpretation of graphs. The analysis of eye movement data gives hints about participants' inspections of trends, in terms of their inspection of graphical entities and alphanumerical labels such as source- and goal-argument values. Eye movement data can be analyzed by different measures, including the measurement of gaze parameters such as gaze time as well as the measurement of the gaze maps.

Based on the computational architecture for multimodal comprehension of text-graphics documents, which employs the parallelism of language comprehension and graph comprehension by use of a common representational formalism [14], [19], [22], we presented in this paper a detailed analysis with respect to the spatial concepts, in particular the PATH concept and their pivotal role for multimodal integration using the method of task analysis. The structural analysis of the linguistic entities in terms of spatial concepts allowed a systematic decomposition of the tasks into subtasks by the task analysis approach.

We presented an experimental study which tested the predictions made by the task analysis by employing the sentence-graph verification paradigm. In the present study, the sentence-graph verification paradigm is conceived as a representation of real-world settings in which the reader investigates a graph after reading a page by flipping the page in a printed newspaper or magazine article, or by clicking a link to open the pop-up window in a webpage. The results of the experimental investigation revealed that the task specification approach, in its current form, is able to anticipate approximately half of the human gaze patterns in graph inspections that follow different types of stimuli sentences. A further investigation of the gaze patterns by the analysis of transition matrix entropy revealed more findings about the influence of different stimuli sentences on graph inspections. On the other hand, the results indicate that the task analysis approach has its limitations when applied to graph interpretation situations in real-world settings.

Future research will address the study of graphs incorporating more than one line, i.e. depicting time series with respect to a set of entities. Comprehending such multiple line graphs includes the task of recognizing relations between lines or line-segments, which are not represented explicitly in the graph, in addition to the task of processing individual graph-lines. In particular, the crossing of lines (same or nearly same value) and the "quasi-parallelity" of lines (similarity of change of values, possible a common trend) are relations crucial for solving complex problems, such as comparing shares using stock market charts. Text as well as graphical cues that aim to attract the attention of the reader to certain parts of the graph [24] can trigger attention shifts among the lines during multiple-line graph-comprehension. Thus we will use the task analysis approach for predicting 'inspections' based on empirical results from eye tracking studies to identify 'optimal' locations of the cues on the graph line and to specify conditions for accompanying text or verbal annotations providing hints for discovering graphically implicit relations.

Acknowledgments. The research reported in this paper has been partially supported by DFG (German Science Foundation) in ITRG 1247 'Cross-modal Interaction in Natural and Artificial Cognitive Systems' (CINACS). We thank METU HCI Research and Application Laboratory for their technical support. We also thank two anonymous reviewers for their helpful comments and suggestions.

References

1. Koerner, C. (2004). Sequential processing in comprehension of hierarchical graphs. *Applied Cognitive Psychology, 18*, 467-480.
2. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
3. Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *The psychology of reading. Second Edition.* Psychology Press.
4. Tanenhaus, M. K., & Trueswell, J. C. (2005). Eye movements as a tool for bridging the language-as-product and language-as-actions traditions. 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. 3-37). Cambridge, MA: MIT Press.
5. Henderson, J. M. (2003). Human gaze control in real-world scene perception. *Trends in Cognitive Sciences, 7*, 498-504.
6. Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human-Computer Interaction*, 8(4), 353-388.
7. 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.
8. 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.
9. 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.
10. 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 ninth IFIP TC13 international conference on human-computer interaction* (pp. 144-151). IOS Press.
11. Coltekin, A., Heil, B., Garlandini, S., & Fabrikant, S. I. (2009). Evaluating the effectiveness of interactive map interface designs: A case study integrating usability metrics with eye-movement analysis. *Cartography and Geographic Information Science, 36*(1), 5-17.
12. Alesandrini, K. L. (1984). Pictures and adult learning. *Instructional Science*, 13, 63-77.
13. Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA.: Harvard University Press.
14. Acartürk, C. (2010). Multimodal comprehension of graph-text constellations: An information processing perspective. University of Hamburg Dissertation, Hamburg, Germany.
15. Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3(3), 185-226.
16. 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.

17. Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge, MA: Cambridge University Press.
18. Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, 20, 167-171.
19. 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.
20. 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.
21. 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.
22. 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).
23. Habel, C., & Acartürk, C. (2011). Causal inference in graph-text constellations: Designing verbally annotated graphs. *Tsinghua Science and Technology*, 16(1), 7-12.
24. Acartürk, C. (2012). Points, lines and arrows in statistical graphs. In P. Cox, P. Rodgers & B. Plimmer (Eds.), *Diagrams 2012, Lecture Notes in Artificial Intelligence* (Vol. 7352, pp. 95-101). Berlin Heidelberg: Springer-Verlag.
25. Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum.
26. 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.
27. Ponsoda, V., Scott, D., & Findlay, J. M. (1995). A probability vector and transition matrix analysis of eye movements during visual search. *Acta Psychologica*, 88, 167-185.
28. Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, 24, 631-645.
29. Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423, 623–656.
30. Shic, F., Chawarska, K., Bradshaw, J., & Scassellati, B. (2008). Autism, eye-tracking, entropy. In *Proceedings of the 7th IEEE International Conference on Development and Learning (ICDL 2008)*.