What is being Measured in an Information Graphic?

Seniz Demir[1], Stephanie Elzer Schwartz[2], Richard Burns[3], and Sandra Carberry[4]

[1] TUBITAK-BILGEM, Turkey
seniz.demir@tubitak.gov.tr
[2] Dept. of Computer Science, Millersville University, USA
stephanie.schwartz@millersville.edu
[3] Dept. of Computer Science, West Chester University, USA
rburns@wcupa.edu
[4] Dept. of Computer and Information Sciences, University of Delaware, USA
carberry@cis.udel.edu

Abstract. Information graphics (such as bar charts and line graphs) are widely used in popular media. The majority of such non-pictorial graphics have the purpose of communicating a high-level message which is often not repeated in the text of the article. Thus, information graphics together with the textual segments contribute to the overall purpose of an article and cannot be ignored. Unfortunately, information graphics often do not label the dependent axis with a full descriptor of what is being measured. In order to realize the high-level message of an information graphic in natural language, a referring expression for the dependent axis must be generated. This task is complex in that the required referring expression often must be constructed by extracting and melding pieces of information from the textual content of the graphic. Our heuristic-based solution to this problem has been shown to produce reasonable text for simple bar charts. This paper presents the extensibility of that approach to other kinds of graphics, in particular to grouped bar charts and line graphs. We discuss the set of component texts contained in these two kinds of graphics, how the methodology for simple bar charts can be extended to these kinds, and the evaluation of the enhanced approach.

1 Introduction

Information graphics (such as simple bar charts and line graphs) are non-pictorial graphics that depict entities and the relations between these entities. Some information graphics are constructed only to visualize raw data. However when such graphics appear in newspapers and magazines, they generally have a communicative goal or message that is intended to be conveyed to the graph viewer. The graphic designer makes deliberate choices in order to bring that message out such as highlighting certain aspects of the graphic via annotations or the use of different colors. For example, the bar chart in Figure 1 ostensibly conveys the high-level message that "The mortgage program assets of the Chicago Federal

A. Gelbukh (Ed.): CICLing 2013, Part I, LNCS 7816, pp. 501–512, 2013.

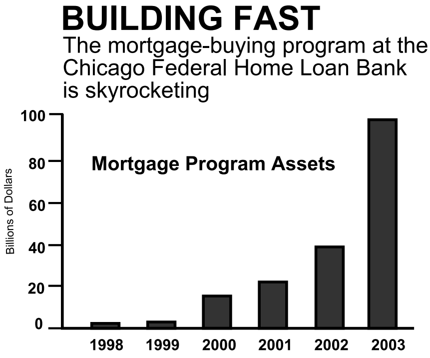

Fig. 1. Graphic from Business Week

Home Loan Bank had a rising trend from 1998 to 2003". Clark [4] contends that language is not just text and utterances, but instead includes any deliberate signal (such as facial expressions) that is intended to convey a message. Thus under Clark's definition, an information graphic is a form of language.

Information graphics often appear as part of a multimodal document in popular media. Carberry et al. [3] contend that the high-level message of an information graphic is the primary contribution of the graphic to the article in which it appears. However, very often the intended message of a graphic is not repeated in the article's text. This is in contrast with scientific documents which generally describe what is graphed in their graphics. Moreover, the high-level message of a graphic often cannot be gleaned from the graphic's caption [6], as can be seen in Figure 1. Recognizing the intention of information graphics is an important step in fully comprehending a multimodal document. Thus, information graphics which contribute to the overall purpose of an article [9] cannot be ignored.

Several applications can greatly benefit from the realization of the graphic's high-level message such as the summarization of multimodal documents and the retrieval of graphics in a digital library. However, additional information should be extracted from the graphic if the high-level message is to be realized. One such piece of information is a descriptor for what is being measured in the graphic. Unfortunately, information graphics often do not label the dependent axis with a descriptor and in such cases the descriptor must be extracted from the text of the graphic. For example, the dependent axis label of the graphic in Figure 1 does not convey what is being measured in that graphic (i.e., "the dollar value of the Chicago Federal Home Loan Bank's mortgage program assets").

In our earlier work [5], we explored a corpus of simple bar charts to identify where to look for the pieces of a descriptor (which is referred to as the "measurement axis descriptor"). The insights gained from the study were used to develop a set of heuristics and augmentation rules that can be applied to a simple bar chart in order produce its measurement axis descriptor. The evaluation study showed that the approach generally produces reasonable descriptors as compared to several baselines. Although the heuristic-based approach is unique in that it showed that methodologies generally used in processing texts (such as finding patterns via corpus analysis) can successfully be applied to non-stereotypical forms of language, its extensibility to other kinds of graphics has not been previously validated.

Different kinds of graphics have different characteristics (such as the number of high-level messages, the type of underlying data – discrete versus continuous, and the distribution of component texts). Thus, one cannot assume that the heuristic-based solution would generate appropriate measurement axis descriptors for all kinds of graphics. This paper presents how a reasonable measurement axis descriptor for grouped bar charts and line graphs can be produced. The characteristics of these two kinds of graphics represent sufficient variability and hence the task of identifying descriptors for these kinds should be treated differently. We greatly drew from our earlier work on simple bar charts as we developed our generation approach. We first explored the component texts contained in these kinds of graphics and identified their similarities and differences from those of simple bar charts. We then investigated how the heuristic-based solution could be extended to grouped bar charts and line graphs and evaluated the generalizability of the overall methodology to more complex graphics.

The rest of this paper is organized as follows. Section 2 discusses related research and Section 3 describes the textual components of grouped bar charts and line graphs. Section 4 gives an overview of the heuristic-based approach and describes how this approach is extended to these kinds of graphics. Finally, Sections 5 presents the evaluation of the enhanced approach and Section 6 concludes the paper.

2 Related Work

The intention of information graphics has been studied by previous research. Kerpedjiev et al. [10] developed a method for automatically realizing intentions in graphics. The PostGraphe system [8] generates a multimodal report containing graphics with accompanying text once a data set is given along with the communicative intentions of the author. Elzer et al. [7] developed a Bayesian system that automatically recognizes the high-level message of simple bar charts. The intention recognition systems for more complex graphics were later developed. Wu et al. [13] implemented a system which recognizes the most likely high-level message in line graphs. Burns et al. [2] investigated the high-level messages that grouped bar charts can convey and developed a methodology for recognizing these messages.

Producing the measurement axis descriptor for a graphic is indeed a referring expression generation problem. Research has been studying the generation of referring expressions (determining a set of properties that would single out the target entity among other entities) for some years [11]. In recent years, it is considered as postprocessing in extractive multidocument summarization where the goal is to improve text coherence [1]. One prominent work in this area improves the coherence of a multidocument summary of newswire texts by regenerating referring expressions for the first and subsequent mentions of people names [12]. The referring expressions are extracted from input documents via a set of rewrite rules. Since our task is extracting a reasonable referring expression for the dependent axis from the graphic's text, it is similar to but more complex than this body of research. First, the referring expression often does not appear as a single unit and thus must be produced by extracting and combining pieces of information from the graphic's text. Second, even if it appears as a label on the dependent axis, it still needs to be augmented.

3 Textual Components of Information Graphics

Graphic designers often use text within and around an information graphic in order to present related information about the graphic. The set of component texts contained in graphics, individual or composite[1], are visually distinguished from one another by different fonts or styles, by blank lines, or by different directions and positions in the graphic. Despite the differences between the number of component texts present in a graph, an alignment or leveling of text contained in a graphic ("text levels") is observed. In this section, we describe the similarities and differences between the textual content of the three kinds of graphics that are considered in this work.

3.1 Simple Bar Charts

In our earlier study [5], a corpus of 107 bar charts (which we refer to as the "bar_chart_corpus") was collected from 11 different magazines (such as Business Week and Newsweek) and newspapers. These graphics were either individual bar charts or charts that appear along with other graphics (some of these graphs were of different kinds) in a composite. Seven text levels were observed in the collected corpus, but every text level did not appear in all graphics. Two of these text levels (Overall Caption and Overall Description) applied to composite graphs and referred to the entire collection of graphics in the composite whereas the remaining text levels applied to each individual graph:

- **Overall Caption:** The text that appears at the top of a composite graph (i.e., the lead caption on the composite) and serves as a caption for the whole set of individual graphs.
- **Overall Description:** The text that often appears under the Overall Caption in a composite but distinguished from it by a different font or a blank line. This text component is pertinent to all graphs in a composite and elaborates on them.
- **Caption:** The text that comprises the lead caption on a graphic.
- **Description:** The text that appears under the lead caption of a graphic.
- **Text In Graphic:** The only text that appears within the borders of a graphic.

[1] Composite graphics consist of multiple individual graphs.

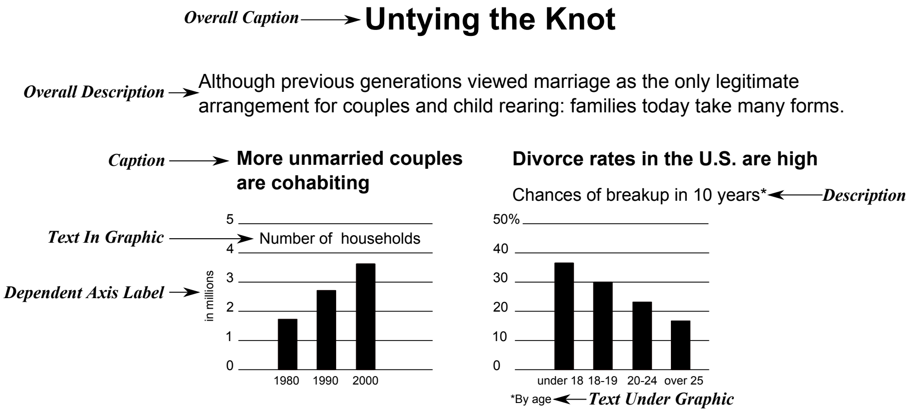

Fig. 2. A composite graphic

– **Dependent Axis Label:** The text that explicitly labels the dependent axis of a graphic.
– **Text Under Graphic:** The text under a graphic that usually starts with a marker symbol (such as * or **) and is essentially a footnote[2].

For example, Figure 2 presents a composite graphic which contains all text levels recognized for simple bar charts. In the study, it was observed that almost all graphics have a Caption (99%) and the Description appeared in slightly more than half of the graphics (54.2%). The Text Under Graphic was the least observed text level (7.5%). Only 18.7% of the graphics had an explicitly labelled Dependent Axis Label. With the exception of one graphic, the Dependent Axis Label (if present) consisted solely of a phrase used to give the unit and/or scale of the values presented in the graphic (such as "in millions" in the left bar chart in Figure 2). These phrases, referred to as unit (e.g., "dollars") and scale (e.g., "billions") indicators, were also observed as part of other text levels.

3.2 Grouped Bar Charts

Grouped bar charts are a kind of information graphic which visually display quantifiable relationships of values with an additional grouping dimension. The grouping dimension makes these graphics more complex to comprehend than simple bar charts. We undertook a corpus analysis in order to explore the set of component texts contained in grouped bar charts. We collected a corpus of 120 grouped bar charts from popular media such as USA Today, The Economist, NewsWeek, and Time. The charts in the corpus (which we refer to as the "grouped_chart_corpus") are diverse in several ways, including the presentation

[2] Each Text Under Graphic has a referrer elsewhere that ends with the same marker and that referrer could be at any text level of the graphic. In the case of multiple such texts within the same graphic, each Text Under Graphic is differentiated with a different marker.

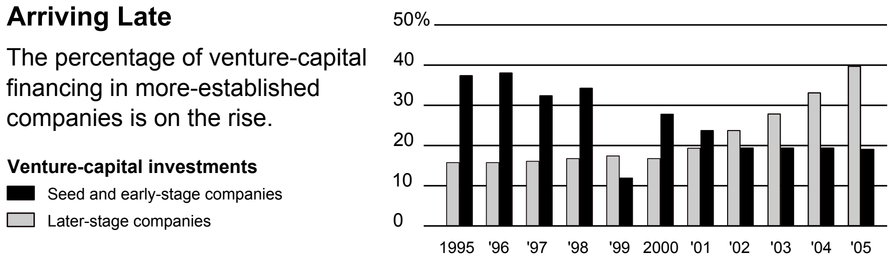

Fig. 3. Graphic from The Wall Street Journal

of the underlying data, the use of visual patterns, and the positioning of textual content. However, we recognized a leveling of component texts in grouped bar charts which is similar to that of simple bar charts. For instance, the text levels of grouped bar charts are also visually distinguishable from one another and not every text level presents in all grouped bar charts.

The seven text levels observed in simple bar charts were also recognized in grouped bar charts. In addition to these levels, another text level which we call the "Legend Descriptor" was observed in the corpus. The Legend Descriptor is the text that appears at the top of the labels of groupings in the chart. The Legend Descriptor is distinguished from the group labels and the preceding text (if exists) by a different font or a blank line. The Legend Descriptor and the group labels, surrounded with a box of solid lines in some cases, appear at different positions in grouped bar charts, including at the top of the graph and within the graph borders. For example, in the graph shown in Figure 3, the Legend Descriptor (i.e., "Venture-capital investments") and the group labels (i.e., "Seed and early-stage companies" and "Later-stage companies") appear at the left side of the dependent axis. The second column of Table 1 presents how often the recognized text levels appear in the grouped_chart_corpus. The Caption and the Text In Graphic are the most frequent and least frequent text levels in the corpus respectively.

Table 1. Text levels in grouped bar charts and line graphs

Text level	**Frequency** grouped bar charts	**Frequency** line graphs
Overall Caption	42 ∼ 35%	24 ∼ 20%
Overall Description	35 ∼ 29.2%	20 ∼ 16.7%
Caption	116 ∼ 96.7%	115 ∼ 95.8%
Description	94 ∼ 78.3%	90 ∼ 75%
Text In Graphic	5 ∼ 4.17%	13 ∼ 10.8%
Legend Descriptor	17 ∼ 14.2%	
Dependent Axis Label	21 ∼ 7.5%	12 ∼ 10%
Text Under Graphic	8 ∼ 6.7%	1 ∼ 0.83%

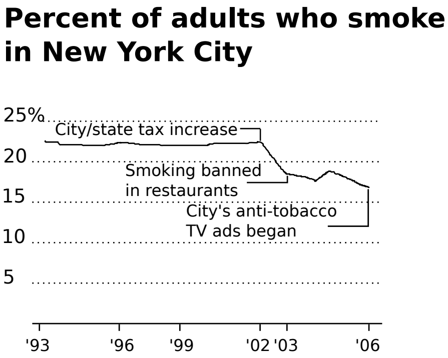

Fig. 4. Graphic from USA Today

3.3 Line Graphs

Line graphs are the preferred medium for conveying trends in quantitative data over an ordinal independent axis. For line graphs, we followed the same methodology for examining their component texts and how these texts are distributed around each graph. We first collected a corpus of 120 line graphs (which we refer to as the "line_graph_corpus") from newspapers (such as USA Today and Los Angeles Times) and magazines (such as BusinessWeek and Forbes). The seven text levels described in Section 3.1 were also recognized in this corpus. In addition, line graphs were observed to share a characteristic that was not seen in bar charts and grouped bar charts. In simple and grouped bar charts, the annotations are mainly for presenting values of specific bars. Although value annotations are also present in line graphs, the annotations at sample points of line graphs diverge substantially in nature. For example, Figure 4 presents a line graph from the corpus where the annotations are used for describing events occured within the time period depicted in the graph. As shown in the third column of Table 1, the Caption appears the most in the line_graph_corpus whereas the Text Under Graphic is the least observed text level.

4 Generating a Referring Expression for the Dependent Axis

We use the main aspects and overall methodology developed for simple bar charts in order to identify the measurement axis descriptor for grouped bar charts and line graphs. The methodology is based on the results of an analysis of a large corpus of simple bar charts (i.e., the bar_chart_corpus) and the identified measurement axis descriptors for these charts by two human annotators. The human annotators not only used the information residing within the component texts of a bar chart but also the graphic's article and their commonsense knowledge in order identify its measurement axis descriptor. There are a number of observations gained from this study, the most important of which are:

- The measurement axis descriptors often cannot be extracted as a whole from a single text level but rather formed by putting together pieces from component texts of the same graph or other graphs in the composite, or from the article's text.
- The measurement axis descriptor consists of a core noun phrase or wh-phrase that appears in one text level of the graphic, possibly augmented with words appearing elsewhere. For example, for the left bar chart in Figure 2, "number of households" is the core of the measurement axis descriptor "U.S.'s number of households".
- Even in cases where the dependent axis is explicitly labelled with a descriptor, it still has to be augmented with pieces of information extracted from other text levels.
- The textual components of a graphic form a hierarchy according to their placements and the core of the measurement axis descriptor generally appears in the lowest text level present in the graph. The ordering of the text levels in Table 1 (except the Legend Descriptor[3] and the Text Under Graphic) forms that hierarchy where the Dependent Axis Label and Overall Caption are at the bottom and top of it.
- If multiple sentences or sentence fragments are contained in a text level, the core of the measurement axis descriptor typically appears near the end of that level.
- Some cues are strong indicators of where to look for the core such as "Here is" and "Here are". The core generally follows such phrases and appears as a noun phrase in the same text level.

The methodology for generating a measurement axis descriptor consists of four steps. First, preprocessing deletes the unit/scale indicators and the ontological category of the bar labels (in cases where explicitly marked by the prepositions "by") from the text levels. Then, a set of heuristics is applied to the text levels in order to extract the core of the descriptor. Next, three kinds of augmentation rules are applied to the core in order to produce the measurement axis descriptor[4]. None of these augmentations might be applicable and in such cases the core forms the whole descriptor. Finally, postprocessing appends a phrase to indicate the unit of measurement to the front of the descriptor if it does not already contain the unit of measurement.

9 heuristics are developed for identifying the core. The heuristics are dependent on the parses of text levels and give preference to the text levels that are lower in the hierarchy. Heuristics are applied in order to the text levels (starting from the lowest text level) until the core is extracted from a level. The following are three sample heuristics:

- **Heuristic 2:** If Text In Graphic consists solely of a noun phrase, then that noun phrase is the core; otherwise, if Text In Graphic is a sentence, the noun phrase that is the subject of the sentence is the core.

[3] This text level is not observed in simple bar charts.
[4] In detail information about all heuristics and augmentation rules can be found in [5].

- **Heuristic 4:** If the current sentence at the text level consists solely of a wh-phrase followed by a colon (:) or a question mark (?), that wh-phrase is the core.
- **Heuristic 6:** If a fragment at the text level consists solely of a noun phrase, and the noun phrase is not a proper noun, that noun phrase is the core.

A sample augmentation rule used to fill out the descriptor is as follows:

- **Specialization of the noun phrase:** The core is augmented with a proper noun which specializes the descriptor to a specific entity; that proper noun is either the only proper noun present at any text level higher in the hierarchy than the level from which the core is extracted, or the only proper noun in the Overall Caption or the Caption.

For the graphic in Figure 1, Heuristic 2 identifies "mortgage program assets" in the Text In Graphic as the core. Since there is only one proper noun in the Description, the augmentation rule for specialization produces "Chicago Federal Home Loan Bank's mortgage program assets" as the augmented core. After adding a pre-fragment, the measurement axis descriptor becomes "The dollar value of Chicago Federal Home Loan Bank's mortgage program assets".

4.1 Measurement Axis Descriptor for Grouped Bar Charts

In grouped bar charts, we recognized all of the text levels observed in simple bar charts. We argue that the heuristics applicable to these text levels in simple bar charts can also be considered for the corresponding textual components of grouped bar charts. However, another text level, the Legend Descriptor, may appear in grouped bar charts. We therefore determined a new hierarchy of text levels for grouped bar charts which is the ordering shown Table 1 (except the Text Under Graphic). As in simple bar charts, the Overall Caption and the Dependent Axis Label are at the top and bottom of the hierarchy respectively. We also developed a new heuristic that is restricted to the Legend Descriptor:

- **Heuristic 10:** If Legend Descriptor consists solely of a noun phrase, then that noun phrase is the core; otherwise, if Legend Descriptor is a sentence, the noun phrase that is the subject of the sentence is the core.

Since the Legend Descriptor is the second lowest text level in the hierarchy, the heuristic specialized to this level is applied in between the heuristics that are restricted to the Dependent Axis Label (Heuristic 1) and the Text In Graphic (Heuristic 2). For example, for the graphic in Figure 3, our enhanced approach uses Heuristic 10 and a pre-fragment to construct "the percentage of venture-capital investments" as the referring expression for the dependent axis.

4.2 Measurement Axis Descriptor for Line Graphs

All text levels observed in line graphs are already covered by the heuristics developed for simple bar charts. Thus, rather than developing new heuristics specific to line graphs, we decided to use the heuristics of simple bar charts in order to produce the measurement axis descriptor for line graphs. For example, for the graphic in Figure 4, our approach runs Heuristic 6 to produce "the percent of adults who smoke in New York City" as the descriptor of that graph.

5 Evaluation

To evaluate our enhanced methodology on grouped bar charts and line graphs, we followed the same evaluation strategy that was carried out to evaluate the generated descriptors for simple bar charts [5]. First, a test corpus of graphics was collected and the produced measurement axis descriptors for these graphics were rated by two human evaluators. Each evaluator assigned a rating from 1 to 5 to each measurement axis descriptor: 1(very bad), 2 (poor, missing important information), 3 (good, contains the right information but is hard to understand), 4 (very good, interpreted as understandable but awkward), and 5 (excellent text). In cases where the evaluators assigned different ratings to the same graph, the lowest rating was recorded. Using the same scale, the evaluators also evaluated a number of baselines which used texts appearing at different text levels. For the baselines, if the evaluators differed in their ratings, the higher rating was recorded. This biased the evaluation toward better scores for the baselines in contrast to the identified measurement axis descriptors.

The evaluation of simple bar charts was performed on a corpus of 205 graphics [5]. The Dependent Axis Label, Text In Graphic, and Caption were used as baselines in the evaluation[5]. The evaluation score for the produced measurement axis descriptors was midway between good and very good and far better than the scores for the three baselines as shown in the second column of Table 2. For the evaluation of grouped bar charts, we collected a corpus of 120 graphics from popular media[6]. We used the Dependent Axis Label, Legend Descriptor, Text In Graphic, and Caption as baselines. The evaluation scores collected for the grouped bar charts are shown in the third column of Table 2. Similarly, we collected 120 line graphs from different newspapers and magazines in order to evaluate the generated descriptors for these graphs[7]. The Dependent Axis Label, Text In Graphic, and Caption were the baselines of the evaluation. The evaluation scores assigned by the same evaluators, who also rated the texts produced for grouped bar charts, are shown in the fourth column of Table 2.

The scores given in Table 2 show that our enhanced approach generates measurement axis descriptors which are midway between good and very good for grouped bar charts and line graphs. It is noteworthy to mention that the average score of the produced descriptors for grouped bar charts is higher than that of simple bar charts. Moreover, the average score assigned to the descriptors of line graphs is close to the score that simple bar charts received. In both grouped

[5] The Description was not used as a baseline in the evaluation since that text level is most often full sentences and thus would generally produce very poor results.

[6] These graphics are different from those contained in the grouped_chart_corpus.

[7] None of these graphics are contained in the line_graph_corpus.

Table 2. Evaluation results of simple bar charts, grouped bar charts, and line graphs

Evaluated Text	**Avg. Score** simple bar charts	**Avg. Score** grouped bar charts	**Avg. Score** line graphs
Measurement Axis Descriptor	**3.574**	**3.867**	**3.467**
Dependent Axis Label	1.475	1.367	1.075
Legend Descriptor		1.383	
Text In Graphic	1.757	1.142	1.25
Caption	1.876	2.683	2.733

bar charts and line graphs, the evaluation score of the produced descriptors is far better than the scores for the baselines. Our analysis also showed that the text level from where the core of the measurement axis descriptor is extracted varies among grouped bar charts and line graphs (shown in Table 3) and that the three kinds of augmentation are used in both kinds of graphics. Although the average score that the Caption received in grouped bar charts and line graphs is significantly higher than the scores for other baselines, there are a number of cases where this text is not enough on its own. For example, for the graphic in Figure 3, the Caption "Arriving Late" does not convey what is being measured in the graphic. Thus, both evaluators assigned the lowest rating to this text level in the evaluation. On the other hand, the produced descriptor for the same graph "The percentage of venture-capital investments" received the highest score of 5 from both evaluators.

Table 3. The use of heuristics in grouped bar charts and line graphs

Corpus	**H_1**	**H_2**	**H_3**	**H_4**	**H_5**	**H_6**	**H_7**	**H_8**	**H_9**	**H_10**
Grouped bar charts	11	5	1	4	13	64	4	5	0	13
Line graphs	5	9	0	1	16	66	7	8	8	0

If a text level presents in a graphic, our enhanced generation approach uses the same heuristics applicable to that level no matter in which kind of graphic it appears. However, it was unclear that this approach would work well in different kinds of graphics before the evaluations. But, the collected scores demonstrated that the heuristics applicable to a particular text level work reasonably well in different kinds of graphics. This can be seen from the fact that the average scores of the measurement axis descriptor for different kinds of graphics are close to each other in the rating scale and higher than the scores for the baselines.

6 Conclusion

This paper presents how a heuristic-based approach for generating a referring expression for the dependent axis (i.e., what is being measured in the graphic) of simple bar charts can be extended to different kinds of graphics, in particular to grouped bar charts and line graphs. We identified the textual components of grouped bar charts and line graphs and described the similarities and differences of these components from the texts of simple bar charts. We presented how our existing approach can be enhanced to cover the observed text levels in these kinds of graphics. An evaluation study showed that our approach generally produces reasonable referents for the dependent axis of grouped bar charts and line graphs. In the future, we will address the task of generating such referents for the dependent axis of all kinds of information graphics.

References

1. Belz, A., Kow, E., Viethen, J., Gatt, A.: The grec challenge 2008: Overview and evaluation results. In: The Proceedings of the 5th International Natural Language Generation Conference (2008)
2. Burns, R., Carberry, S., Elzer, S., Chester, D.: Automatically Recognizing Intended Messages in Grouped Bar Charts. In: Cox, P., Plimmer, B., Rodgers, P. (eds.) Diagrams 2012. LNCS, vol. 7352, pp. 8–22. Springer, Heidelberg (2012)
3. Carberry, S., Elzer, S., Demir, S.: Information graphics: An untapped resource for digital libraries. In: The Proceedings of the ACM Special Interest Group on Information Retrieval Conference, pp. 581–588 (2006)
4. Clark, H.: Using Language. Cambridge University Press (1996)
5. Demir, S., Carberry, S., Elzer, S.: Issues in realizing the overall message of a bar chart. In: Recent Advances in Natural Language Processing, vol. 5, pp. 311–320. John Benjamins (2007)
6. Elzer, S., Carberry, S., Chester, D., Demir, S., Green, N., Zukerman, I., Trnka, K.: Exploring and exploiting the limited utility of captions in recognizing intention in information graphics. In: The Proceedings of the Annual Meeting on Association for Computational Linguistics, pp. 223–230 (2005)
7. Elzer, S., Carberry, S., Zukerman, I., Chester, D., Green, N., Demir, S.: A probabilistic framework for recognizing intention in information graphics. In: The Proceedings of the International Joint Conference on Artificial Intelligence, pp. 1042–1047 (2005)
8. Fasciano, M., Lapalme, G.: Intentions in the coordinated generation of graphics and text from tabular data. Knowledge and Information Systems 2(3), 310–339 (2000)
9. Grosz, B., Sidner, C.: Attention, intentions, and the structure of discourse. Computational Linguistics 12(3), 175–204 (1986)
10. Kerpedjiev, S., Green, N., Moore, J., Roth, S.: Saying it in graphics: from intentions to visualizations. In: The Proceedings of the Symposium on Information Visualization, pp. 97–101 (1998)
11. Krahmer, E., van Erk, S., Verleg, A.: Graph-based generation of referring expressions. Computational Linguistics 29(1), 53–72 (2003)
12. Nenkova, A., McKeown, K.: References to named entities: a corpus study. In: The Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, pp. 70–72 (2003)
13. Wu, P., Carberry, S., Elzer, S., Chester, D.: Recognizing the Intended Message of Line Graphs. In: Goel, A.K., Jamnik, M., Narayanan, N.H. (eds.) Diagrams 2010. LNCS, vol. 6170, pp. 220–234. Springer, Heidelberg (2010)