

Generating Textual Summaries of Bar Charts

Seniz Demir Sandra Carberry Kathleen F. McCoy

Department of Computer Science

University of Delaware

Newark, DE 19716

{demir, carberry, mccoy}@cis.udel.edu

Abstract

Information graphics, such as bar charts and line graphs, play an important role in multimodal documents. This paper presents a novel approach to producing a brief textual summary of a simple bar chart. It outlines our approach to augmenting the core message of the graphic to produce a brief summary. Our method simultaneously constructs both the discourse and sentence structures of the textual summary using a bottom-up approach. The result is then realized in natural language. An evaluation study validates our generation methodology.

1 Introduction

Information graphics, such as bar charts and line graphs, are an important component of a multimodal document. However, summarization has focused primarily on the text of a document. But as shown in (Carberry et al., 2006), information graphics in magazine and newspaper articles often convey a message that is not repeated in the article’s text. Thus information graphics cannot be ignored.

Individuals with sight impairments can access the text of an electronic document via text to speech systems. But such individuals are stymied when they encounter information graphics. The SIGHT system (Elzer et al., 2007) has the goal of providing the user with the message and knowledge that one would gain from viewing the graphic, rather than providing alternative access to what the graphic looks like. In the current Bayesian network implementation, the system uses the communicative sig-

nals present in the graphic to recognize one of the twelve message categories that can be conveyed by a bar chart and produces a logical representation of what we will refer to as the **core message** conveyed by the graphic; this representation is translated into natural language via templates. For example, the system determines that the core message of the graphic in Figure 1 is that the bar for the United States has the maximum value of the entities listed, and the system produces Maximum(First.Bar) as the logical representation of that message. However, this is insufficient as a summary of the graphic since it doesn’t convey the particularly significant features of the graphic such as the fact that the number of hacker attacks in the United States is far greater than in the other countries listed.

In this paper, we explore the generation of an effective initial summary of a bar chart within the SIGHT system. Input to our system is a logical representation of the graphic’s core message (as produced by SIGHT) and the XML representation of the graphic which specifies the components of the graphic such as the number of bars and the heights of each bar. Our goal is to generate a succinct coherent summary of a graphic that captures its core message and the most important and significant features. At the same time, the summary need not include all information that could be extracted from the graphic since future work on SIGHT includes a mechanism for responding to follow-up questions from the user.

Our work is unique in that it generates a summary of the content of a bar chart with no domain restriction and constructs a high-quality but brief summary by incorporating the graphic’s core mes-

Countries With the Most Hacker Attacks, 2002

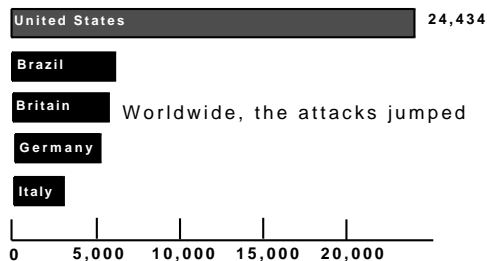


Figure 1: Graphic conveying a maximum bar.

sage and the most important and significant features of the graphic. Our system produces a coherent organization of the content of the summary that we hypothesize lends itself to follow-up questions, applies the notion of syntactic complexity in choosing how to aggregate information into sentence-sized pieces, utilizes a sentence realizer to convey the output in natural language, and applies heuristics for constructing referents for graphical elements. Thus, our work addresses several generation problems.

Section 2 discusses related work in the area of graph summarization. Section 3 presents our methodology for identifying the content of initial summaries. Section 4 describes our approach for organizing the summaries. Section 5 presents some issues that we addressed in realizing the summaries and a few example summaries generated by our system. Section 6 discusses the results of an evaluation study that validates our methodology.

2 Related Work

Graph summarization has received some attention. The SumTime project uses pattern recognition techniques to generate textual summaries of automatically generated time-series data; different systems have been designed in three domains, such as SumTime-Turbine (Yu et al., 2007) for data from gas turbine engines. However, each of these systems is domain-dependent. We view this as a very different problem from the one that we address in this paper, since we are working on domain independent graphical representations. The iGRAPH-Lite system (Ferres et al., 2007), whose main objective is to make the information in a graphic accessible to blind users via keyboard commands, uses tem-

plates to provide a short textual summary of what the graphic looks like, but their summary is not concerned with the higher level knowledge conveyed by the graphic. The goal of Futrelle’s project (Futrelle, 1999) is to produce a summary graphic that captures the content of one or more graphics. However, the end result is itself a graphic, not text.

3 Informational Content of the Summaries

In (McCoy et al., 2001), we reported an informal experiment in which human subjects were asked to write a brief summary of a series of line graphs with the same high-level intention. This experiment led to three observations:

- the intended message of the graphic was contained in all summaries.
- summaries of the same graphic by different subjects were similar.
- summaries of different graphics in the same message category (such as *Rising_Trend*) varied in the information provided.

Subjects included more than the core message of the graphic in their summaries. The extra information could be explained as capturing features of the graphic that were visually salient. It was hypothesized that what is taken as visually salient in a graphic relates to the overall message of the graphic. For example, in the line graphs analyzed, if the overall message of the graphic is an increasing trend and the variance in that trend is large, then the variance is salient. The fact that the summaries only included the core message and the visually salient features, correlates with Grice’s Maxim of Quantity (Grice, 1975) which states that one’s discourse contribution should be as informative as necessary for the purposes of the current exchange but not more so.

To extend these observations to other kinds of information graphics, particularly to simple bar charts, we needed to identify the kinds of features that were salient with respect to the graphic’s overall message. For this purpose, we conducted a set of formal experiments with human subjects. Our goal was a set of rules that took into account the message category (such as *Rising_Trend* or *Rank_of_Entity*) and the visual features of the graphic to specify what propositions should be included in the initial summary.

3.1 Experimental Setup

Based on our experience with summarizing graphics, we first identified the set of all propositions (PROP_ALL) that reflect information that we envisioned someone might ask about a simple bar chart. Twenty graduate students from a variety of departments participated in the experiments. Each subject was given an information graphic along with a sentence conveying the intended message of the graphic (as identified by the SIGHT system) and a subset of PROP_ALL, and was asked to classify these additional propositions into one of three classes according to how important they felt it was to include that proposition in the initial summary: **essential**, **possible**, and **not important**. Subjects were told to assume that the graphic was part of an article that the user is reading and that he would be able to ask follow-up questions after receiving the summary.

Twenty-one graphics were used in the experiments and each participant was given six graphics to view. The graphics were either selected from articles in popular media or constructed from scratch to present a number of different salient features within the same graphic. Eight different message categories were tested in the experiments¹ and the graphics from the same message category reflected different salient features that had been observed in a corpus of collected graphics.

3.2 Computationalizing Proposition Selection

To analyze the experiment’s results, we assigned a numeric score to each category (essential=3, possible=1, not important=0) indicating the level of importance assigned by the subjects. We computed an importance level (**IL**) for each proposition that might be included in a summary of a graphic, estimating how important it is to be included in the initial summary. The **IL** of a proposition is computed based on the average of the importance assigned by each subject. We classify a proposition as a highly-rated proposition in a graphic (and therefore worthy of being included in the initial summary) if its importance level is equal to or above the importance level that would be obtained if at least half of the

subjects had classified the proposition as essential. The particular assignments of values to categories shown above was used in order to emphasize instances in which subjects placed a proposition in the essential category.

For each message category, we analyzed the sets of highly-rated propositions identified for the graphics associated with that message category and hand constructed a set of content identification rules specifying whether a proposition should be included in the initial summary. If a proposition was classified as highly-rated for all graphics in a particular message category, the content identification rule stated that the proposition should be included in the initial summary for every graphic whose core message fell into that message category. For the other highly-rated propositions for a message category, we identified a feature that was visually salient only in the graphics where the proposition was marked as essential, and our content identification rule used the presence of this feature in the graphic as a condition for the proposition to be included in the initial summary. Thirty content identification rules were defined in the system.

Besides the propositions capturing salient features of the graphic, we observed that the subjects selected propositions whose absence might lead the user to draw false conclusions by default (for example, the propositions indicating that the trend starting from 1997 does not cover the full range of bar labels in the graphic in Figure 2). We constructed rules to add such content. This correlates with the maxim in (Joshi et al., 1984) which states that a system should not only produce correct information but should also prevent the user from drawing false inferences.

The following are glosses of two representative content identification rules applicable to a graphic whose core message is an increasing trend:

- If (message_category **equals** ‘increasing_trend’) then **include**(propositions conveying the rate of increase of the trend)
- If (message_category **equals** ‘increasing_trend’) and (coverage(graphic) **not equal** coverage(trend)) then **include**(propositions indicating that the trend does not cover the full range of bar labels)

To see how our rules might affect the generated

¹We did not test message categories that are the opposite of categories used in the experiments, such as *Minimum* which is the opposite of *Maximum* and thus can be modeled similarly.

summary, consider the graphic in Figure 2. The SIGHT system recognizes the core message as an increasing trend from 1997 to 2002. The content identification rules defined for increasing trend select the following pieces of additional information to include in the initial summary of the graphic:

- The overall rate of increase of the trend, which is moderate
- The range of the bar values in the trend: \$480,000 to \$1,230,000
- The fact that there is an unusually steep rise between 2000 and 2001
- The period that the graphic covers, which is from 1996 to 2002

4 Organizing Coherent Summaries

At this point in the processing, the system has identified the informational content to be conveyed in the initial summary, which consists of the core message and the set of propositions identified by the content identification rules. We need to represent and organize the information to be communicated, determine how the content should be aggregated into sentence-sized pieces, and make decisions about referring expressions (Reiter and Dale, 1997).

To represent and help organize the propositions that should be included in the initial summary, we use two kinds of predicates. Relative predicates, such as *differs*, are used to express relations between the graphical elements. For example, **differs(bar(A),maximum_bar,50%)** shows that the percentage difference between the values of bar(A) and the maximum bar is 50%. Attributive predicates, such as *has_attribute*, are used to elaborate the graphical elements. For example, **has_attribute(trend,“type”,“increasing”)** shows that the trend observed in the graphic is an increasing trend. We refer to the first argument of a predicate as its **main entity** and the others as **secondary entities**.

Since a top-down planning approach does not guarantee that all propositions will be covered by the final text plan (Marcu, 1997), we use a bottom-up planner for structuring the initial summaries. The propositions selected for inclusion in the initial summaries can be classified as **message_related**, **specific**, or **computational** based on the type of in-

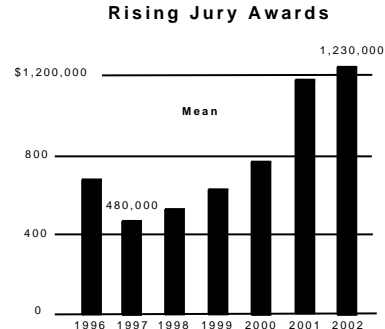


Figure 2: Graphic with an increasing trend.

formation they convey. The core message of the graphic is captured by the message-related propositions. Specific propositions focus on specific pieces of information in the graphic such as the propositions conveying the unusually large rise between 2000 and 2001 in the graphic in Figure 2. On the other hand, computational propositions require computations or abstractions over the whole graphic, such as the propositions conveying the rate of increase in the graphic in Figure 2. Once the propositions to be included in the summary are identified, we assign them to one of these three classes.

We hypothesize that the message-related class of propositions should be presented first since this places emphasis on the core message of the graphic. We anticipate that the user will ask follow-up questions after receiving the initial summary. Therefore, it is appropriate to close the initial summary with propositions from the computational class so that the whole graphic is in the user’s focus of attention (Grosz and Sidner, 1986). Thus we hypothesize that a good ordering of propositions in the initial summary is the message-related class, the specific class, and finally the computational class. This produces a partial ordering of the propositions to be included in the summary.

Each proposition can be realized as a single sentence. For example, **shows(graphic,trend)** can be realized as “The graphic shows a trend” or “There is a trend in the graphic”. Consequently, a set of propositions can be viewed as a set of single sentences. Figure 3 shows the propositions in the message-related class for the graphic in Figure 2, along with a possible realization for each.

Although each proposition could be conveyed as a single sentence, the result is unnatural and not very

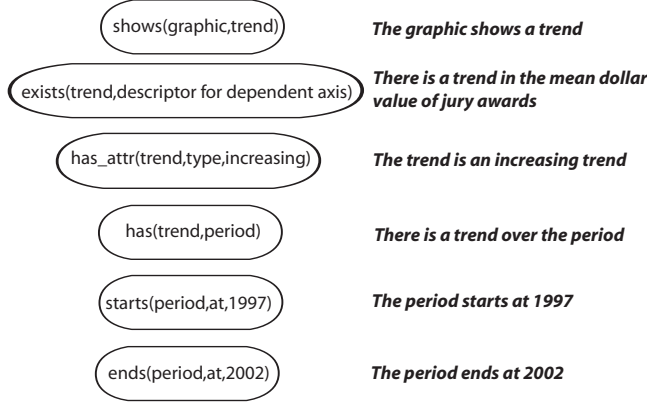


Figure 3: Propositions in Message-related Class.

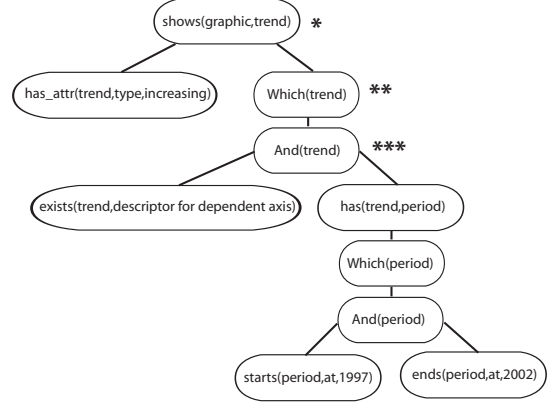


Figure 4: Best Structure of Message-related Class.

coherent. Thus, we define operators to relate propositions and explore aggregating them. The operators work on trees; initially, each proposition is the root of a single node tree. Each tree represents one or more propositions that can be realized as a single sentence, and operators combine individual trees in a class into more complex trees. The following are two such operators which work on relative predicates:

- **And_Operator:** This operator combines two trees if their root propositions share the same main entity. An And_predicate with the same main entity forms the root of the new tree, and the trees that are combined form the descendents of this root.
- **Which_Operator:** This operator attaches one tree as a descendent of another tree, connected by a Which_predicate, if the main entity of the proposition at the root of the first tree is a secondary entity in the proposition at the root of the second tree. That particular entity forms the main entity of the Which_predicate.

One possible result of applying our operators to the set of propositions (single node trees) in Figure 3 produces the single but more complex tree shown in Figure 4. The And_predicate (***) in Figure 4 is produced by the And_Operator and the resultant subtree is attached to the predicate **shows** by the Which_Operator.

4.1 Evaluating Structures

Different combinations of operators produce different sets of trees, each of which represents a different text structure and consequently leads to differ-

ent realized text. These structures must be evaluated to determine which one is the best. We don't want a structure where each proposition is realized as a single sentence nor a structure where groups of propositions are realized with sentences that are too complex. Our objective is to find a structure which stands at a mediatory point between these extremes.

Each tree in a structure represents a set of propositions that can be realized as a single sentence. The most straightforward way of realizing a tree would be conjoining the realizations of subtrees rooted by an And_predicate, embedding the realization of a subtree rooted by a Which_predicate as a relative clause, and realizing a subtree that consists solely of an attributive predicate as an adjective or a prepositional phrase. However, care must be taken that the sentence realization of a tree is not too complex.

Research has used a number of different measures to assess syntactic complexity of written text and spoken language samples (Roark et al., 2007). We apply the notion of syntactic complexity to evaluate the semantic units (predicates) that will be realized. The revised D-level sentence complexity scale (Covington et al., 2006) forms the core of our syntactic complexity measure. The D-Level scale measures the complexity of a sentence according to the sequence in which children acquire the ability to use different types of sentences. The sentence types with the lowest score are those that children acquire first and therefore are the simplest types. Among the seven levels defined in the revised D-Level scale, the levels of interest in our work are (in order of increasing complexity): simple sentences, conjoined sentences, sentences with a rela-

<p>Message_Related: <i>The graphic shows an increasing trend in the mean dollar value of jury awards over the period from 1997 to 2002.</i></p> <p>Specific: <i>The value of these awards ranges from 480,000 dollars to 1,230,000 dollars. The graphic covers a period from 1996 to 2002. Between 2000 and 2001, an unusually steep rise is observed in the value of these awards.</i></p> <p>Computational: <i>Moderate increases are observed every year during the period from 1997 to 2002 in the value of these awards.</i></p>	<p>Message_Related: <i>Although the graphic covers a period from 1996 to 2002, the graphic shows an increasing trend in the mean dollar value of jury awards over the period from 1997 to 2002.</i></p> <p>Specific: <i>The value of these awards ranges from 480,000 dollars to 1,230,000 dollars.</i></p> <p>Computational: <i>Except for a steep rise between 2000 and 2001, moderate increases are observed every year in the value of these awards.</i></p>
---	--

Table 1: Realization of the Initial Overall Structure.

tive clause, and sentences with more than one level of embedding. However, the definition of sentence types at each level is too general. Therefore, we use a finer distinction between sentence types within each D-level, such as a simple sentence with more than one preposition has a higher complexity than a simple sentence with a single preposition.

To measure the complexity of sentences that will realize a structure, we define a number of complexity estimators. A tree consisting of a single node is identified as having the lowest syntactic complexity. We use And_predicate and Which_predicate estimators to estimate the complexity of the sentences used to realize more complex trees. To do this, estimators look for realization opportunities that would produce lower complexity values than what the most straightforward realization would produce. For instance, the And_predicate estimators check whether or not the realizations of two subtrees rooted by an And_predicate can be combined into a single sentence which is not a compound sentence consisting of two independent sentences. These estimators have similar considerations to the clause-combining operations used by Walker et.al (2002) in the SPoT sentence planner. For example, one of the And_predicate estimators can successfully determine that the realizations “*There is a trend in the mean dollar value of jury awards*” and “*There is a trend over the period from 1997 to 2002*”² can be combined into “*There is a trend in the mean dollar value of jury awards over the period from 1997 to*

²Assume that the subtree in Figure 4 rooted by Which(period) can be realized as “from 1997 to 2002”.

Table 2: Realization of the Final Overall Structure.

2002” for the tree in Figure 4. This results in a complexity score which is lower than the score of a compound sentence consisting of two conjoined independent sentences. The Which_predicate estimators check whether a tree rooted by a Which_predicate can be realized as an adjunct to the modified entity rather than as a more complex relative clause. For example, one of the Which_predicate estimators can identify that * and the subtree rooted at ** in Figure 4 can be realized as “*The graphic shows a trend in the mean dollar value of jury awards over the period from 1997 to 2002*”.

Center-embedded relative clauses are more difficult to comprehend than corresponding right-branching clauses (Kidd and Bavin, 2002). Our complexity metric for evaluating the complexity of a tree penalizes Which_predicates that will be realized as a relative clause in the middle of a sentence more than one that appears at the end. Once the complexity metric has evaluated the complexity of each tree in a candidate structure, the complexities of the trees are summed to get the complexity of the structure. Our metric for selecting the best structure balances the number of sentences and their complexity.

The initial overall structure of the summary consists of the best structures for the message_related, specific, and computational classes. As a final step, we check whether we can improve the evaluation of the overall structure of the summary by moving trees or subtrees between the best structures for the three classes. For example, the best structure for the specific class might contain a tree that conveys information about an entity introduced by a proposition

in the message_related class. Moving this tree to the message-related class and using the Which operator to combine it with the tree introducing the entity (thereby realizing it as a relative clause) might improve the evaluation of the overall structure of the summary. To be consistent with the motivation behind the initial groupings of the propositions, we do not allow movements from the message_related class or any movement that will empty the computational class. Table 1 presents the summary that our system generates for the graphic in Figure 2 before the movements between classes, and Table 2 presents the summary after the movements.

5 Realizing Summaries

To realize the summaries in natural language, we use the FUF/SURGE surface realizer (Elhadad and Robin, 1996) with some changes made to address a few problems encountered with respect to the use of conjunctions and subject-ellipses. Different strategies are defined in the system for aggregating the realizations of trees that are linked with operators. The strategy selected by the system is based on the relation (such as concession) that holds between the propositions at the root of the trees and the syntactic forms of their realization opportunities. For example, the system uses different strategies for aggregating the trees rooted by the And_predicates in Figure 4, where the tree rooted by And(period) is realized as a combination of prepositional phrases and the tree rooted by And(trend) is realized as a full sentence containing a set of prepositional phrases.

5.1 A Descriptor for the Dependent Axis

The dependent axis of an information graphic is often not labelled with a full descriptor of what is being measured and therefore a mechanism for extracting an appropriate descriptor had to be developed. We undertook a corpus analysis and implemented a system to realize the descriptors (Demir et al., 2007). Our corpus analysis found that the full descriptor often must be built from pieces of text extracted from different places in the graphic. We identified seven text levels (text components) which form a hierarchy (the top being the Overall_Caption and the bottom being the Dependent_Axis_Label), and we observed that the texts at the lower levels are more

Annual percent change in global output

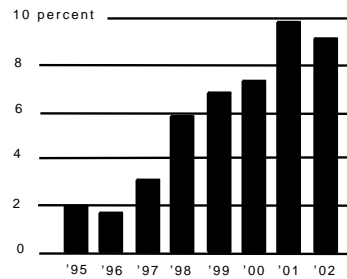


Figure 5: Graphic conveying a contrast change.

likely to contribute to the descriptor than the texts at the higher levels. We developed a set of heuristics and augmentation rules for constructing the descriptor for the dependent axis and validated them on a previously unseen corpus of graphics.

5.2 Referent Generation

The descriptor that our system constructs is always a noun phrase, but it may be quite long. Our referring expression generator uses the full descriptor when the dependent axis is first referenced in the text, but only the head noun for subsequent references. To relate the head noun to the descriptor in the text, the demonstratives “this” or “these” is added to the front of the head noun unless it follows a comparison such as “more”. This simple mechanism appears to work well for our initial summaries.

5.3 Example Summaries

For the graphic in Figure 1, the SIGHT system posits that the graphic’s core message is that the bar for the United States has the maximum value among the bars listed. Our system adds additional propositions and produces the following summary:

“The graphic shows that United States with 24,434 has the highest number of hacker attacks among the countries³ Brazil, Britain, Germany, Italy, and United States. United States has 4.9 times more attacks than the average of other countries.”

For the graphic in Figure 5, the SIGHT system posits that the core message is that the change at the last bar is in contrast with the previous increasing trend. Our system generates the following summary:

³Our system has a module which identifies the ontological category of the bar labels (Demir et al., 2007).

“The graphic shows a small drop in 2002 in contrast with the increasing trend in annual percent change in global output over the period from 1995 to 2001. Except for a small drop in 1996, varying increases are observed every year during the period from 1995 to 2001 in this change, which shows an overall increase of 485.4 percent and shows a decrease of 8.7 percent between 2001 and 2002.”

6 Evaluation

The work presented in this paper consists of several different components, all of which we claim contribute to the quality of the graphical summary. Our evaluation focused on three of these and evaluated whether or not our decisions with respect to these components contributed to the perceived quality of the summary: the organization and ordering of the informational content (**O**), the aggregation of the information into more complex structures (**A**), and the metric used to evaluate the structures that represent different possible aggregations of the informational content (**E**). We conducted an experiment with fifteen participants where they were presented with four different summaries of twelve graphics from a variety of domains. We focused on increasing and decreasing trend graphics since they have the greatest variety of possible summaries. For each of the graphics, the participants were asked to rank randomly ordered summaries in terms of their quality in conveying the informational content. The summaries varied according to the test parameters:

- **S_O+A+E+**: A summary that uses the ordering rules, the aggregation rules, and is rated highest by the evaluation metric. This is the summary produced by our system.
- **S_O+A+E-**: A summary that uses the ordering and aggregation rules, but was not rated highest by the evaluation metric.
- **S_O-A+E+**: A summary where the propositions are randomly ordered, but aggregation takes place, and the aggregation is rated best by the evaluation metric.
- **S_O-A-E-**: A summary consisting of single sentences that are randomly ordered.

The results of the experiment are presented in Table 3. It is particularly noteworthy that the summary generated by our system was most often (65.6% of

Summary_type	Best	2 nd	3 rd	4 th
S_O+A+E+	65.6%	26.6%	6.7%	1.1%
S_O+A+E-	16.7%	32.2%	33.3%	17.8%
S_O-A+E+	16.7%	30%	40%	13.3%
S_O-A-E-	1%	11.2%	20%	67.8%

Table 3: Ranking of Summary Types.

the time) rated as the best summary and overwhelmingly (92.2% of the time) rated as one of the top two summaries. The table shows that omitting the evaluation metric (S_O+A+E-) or omitting ordering of propositions (S_O-A+E+) results in summaries that are substantially less preferred by the subjects. We gained insights into improving the system’s performance by looking at the comments made by the subjects when our summary was not selected as the best. For example, it appears that the subjects did not like summaries where exceptions were fronted on the core message with an “although” clause rather than following the core message (S_O+A+E-). We hypothesize that fronting the exception detracted from the core message. This is easily remedied in our evaluation metric. Overall, the results shown in Table 3 support our methodology for generating summaries. In the future, we will test the summaries with blind individuals to determine their effectiveness in providing alternative access to graphics.

7 Conclusion

This paper has presented our methodology for generating brief and coherent summaries of simple bar charts. Our work is the first to address the problem of summarizing the content of bar charts. We have presented our approach for identifying the appropriate content of an initial summary, ordering and aggregating the included propositions, and evaluating the resultant summary structures to select the best one. Overall, the results of an evaluation study validate our ordering, aggregation, and evaluation methodology.

Acknowledgements

We would like to thank Dr. Stephanie Elzer for her advice, help, and implementation of the SIGHT system upon which this work is built and Dr. Charles Callaway for his valuable help in addressing the problems encountered with the realizer.

References

- Sandra Carberry, Stephanie Elzer, and Seniz Demir. 2006. Information graphics: an untapped resource for digital libraries. In *Proc. of SIGIR'2006*.
- Michael A. Covington, Congzhou He, Cati Brown, Lorina Naci, and John Brown. 2006. How complex is that sentence? A proposed revision of the Rosenberg and Abbeduto D-Level scale. *Research Report, AI Center, University of Georgia*.
- Seniz Demir, Sandra Carberry, and Stephanie Elzer. 2007. Effectively realizing the inferred message of an information graphic. In *Proc. of RANLP'2007*.
- Michael Elhadad and Jacques Robin. 1996. An overview of SURGE: a re-usable comprehensive syntactic realization component. In *Proc. of INLG'1996*.
- Stephanie Elzer, Edward Schwartz, Sandra Carberry, Daniel Chester, Seniz Demir, and Peng Wu. 2007. A browser extension for providing visually impaired users access to the content of bar charts on the web. In *Proc. of WEBIST'2007*.
- Leo Ferres, Petro Verkhogliad, Gitte Lindgaard, Louis Boucher, Antoine Chretien, and Martin Lachance. 2007. Improving accessibility to statistical graphs: the iGraph-Lite system. In *Proc. of ASSETS'2007*.
- Robert Futrelle. 1999. Summarization of diagrams in documents. In *Advances in Automated Text Summarization*. MIT Press, pp. 403-421.
- Paul Grice. 1975. Logic and conversation. In *Syntax and Semantics, Speech Acts*, 3:41-58.
- Barbara Grosz and Candace Sidner. 1986. Attention, intentions, and the structure of discourse. In *Computational Linguistics*, 12(3):175-204.
- Aravind Joshi, Bonnie Webber, and Ralph Weischedel. 1984. Living up to expectations: computing expert responses. In *Proc. of NCAI'1984*.
- Evan Kidd and Edith Bavin. 2002. English-speaking childrens comprehension of relative clauses: evidence for general-cognitive and language-specific constraints on development. In *Journal of Psycholinguistic Research*, 31(6):599-617.
- Daniel Marcu. 1997. The rhetorical parsing, summarization, and generation of natural language texts. *PhD thesis, Department of Computer Science, University of Toronto*.
- Kathleen F. McCoy, Sandra Carberry, Tom Roper, and Nancy Green. 2001. Towards generating textual summaries of graphs. In *Proc. of HCI'2001*.
- Ehud Reiter and Robert Dale. 1997. Building applied natural language generation systems. In *Natural Language Engineering*, 3(1):57-87.
- Brian Roark, Margaret Mitchell, and Kristy Hollingshead. 2007. Syntactic complexity measures for detecting mild cognitive impairment. In *Proc. of BioNLP'2007*.
- Marilyn Walker, Owen Rambow, and Monica Rogati. 2002. Training a sentence planner for spoken dialogue using boosting. In *Computer Speech and Language: Special Issue on Spoken Language Generation*, 16(3-4):409-433.
- Jin Yu, Ehud Reiter, Jim Hunter, and Chris Mellish. 2007. Choosing the content of textual summaries of large time-series data sets. In *Natural Language Engineering*, 13(1):25-49.