Effectively Realizing the Inferred Message of an Information Graphic

Seniz Demir and Sandra Carberry
Dept. of Computer Science
University of Delaware
Newark, DE 19716
{*demir,carberry*}@*cis.udel.edu*

Stephanie Elzer
Dept. of Computer Science
Millersville University
Millersville, PA 17551
elzer@*cs.millersville.edu*

Abstract

Information graphics, such as bar charts and line graphs, that appear in popular media generally have a message that they are intended to convey. We have developed a Bayesian network that analyzes the communicative signals in an information graphic and produces a logical representation of the graphic's intended message. However, the output produced by the Bayesian network is deficient for producing natural language text. This paper presents our solution to several aspects of this problem: identifying an appropriate referent for the dependent axis, determining when to enumerate the bar labels in a message, and identifying the ontological category for the bar labels. An evaluation study shows that our methodology produces reasonable text that is much better than several baseline strategies.

Keywords

Natural language processing, corpus analysis, graph understanding, text generation

1 Introduction

Information graphics (such as bar charts and line graphs) are non-pictorial graphics that depict attributes of entities and relations among them. Although some information graphics are only intended to display data[18], the overwhelming majority of information graphics that appear in magazines and newspapers have a communicative goal or intended message. For example, the graphic in Figure 1 ostensibly is intended to convey that the percentage of GM's net earnings produced by its finance unit increased substantially in the second quarter of 2003 in contrast with the decreasing trend from the third quarter of 2002 to the first quarter of 2003. We developed a Bayesian system that exploits the communicative signals in an information graphic to produce a logical representation of the graphic's intended message[7]. However, the logical representation produced by the Bayesian system is deficient for producing text, and additional information must be extracted from the graphic if a useful summary is to be constructed.

Clark[3] contends that language is not just text and utterances, but instead includes any deliberate signal that is intended to convey a message. Thus, under Clark's definition, information graphics are a form of language. Our work shows that methodologies typically used in processing utterances and text (such as extraction of communicative signals, corpus analysis to identify common patterns, algorithms based on insights from the corpus analysis, and modification of existing software, such as parsers, to fit the needs of the problem) can also be successful on non-stereotypical forms of language.

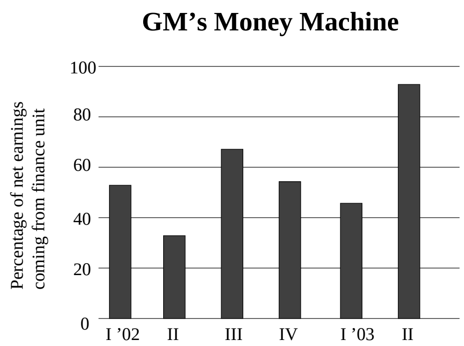

Fig. 1: *Graphic with a cute caption*

Section 2 shows the importance of recognizing the intended message of an information graphic, Section 3 discusses related research, and Section 4 gives an overview of our system that hypothesizes the intended message of an information graphic. Section 5 then discusses the problems that we had to address in order to construct text from the logical representations of the hypotheses produced by our message recognition system. Sections 6, 7 and 8 present our solutions to these problems. Section 9 discusses examples of how our system constructs text capturing the graphic's primary message, and Section 10 presents an evaluation of the resultant text and discusses future work.

2 Importance of Understanding Information Graphics

Information graphics, such as the two graphics in Figure 5, are important knowledge resources that could be used for many purposes, such as devising proposals for legislation on identity theft. To be useful, such graphics must be accessible from a digital library based on what the graphic conveys.

What about graphics in multimodal documents? We conducted a corpus study to determine the extent to which information graphics are redundant in a multimodal document[1]. We found that in over 60% of the instances in our analyzed corpus, little or none of the graphic's message was captured by the article's accompanying text. Yet the graphic's message played an important role in achieving the discourse purpose[12] of the overall document. Thus information graphics cannot be ignored, and effective summarization of a multimodal document must take into account the messages conveyed by its information graphics.

Given that information graphics cannot be ignored, it is imperative that individuals with sight-impairments be provided with a means of accessing the graphic's content. Although researchers have attempted to convey information graphics via an alternative modality (such as touch or sound), these approaches have serious limitations, such as requiring expensive equipment or requiring that the user develop a mental map of the graphic, something that is very difficult for users who are congenitally blind[14]. Our approach differs significantly from previous work: we are developing an interactive natural language system that provides the user with a brief summary of the graphic's message and then responds to followup questions requesting further detail about the graphic.

For all of the above reasons, it is important that a system be able to recognize a graphic's message. However, as shown by Corio and LaPalme[4] and our own corpus study[6], naturally occurring captions are often very general and of limited utility in identifying the graphic's message. For example, the caption on the graphic in Figure 1 captures little of what the graphic conveys — namely, a contrast between recent performance of GM's finance unit and the trend over the preceding quarters. Thus it is essential that a system be devised for recognizing the message conveyed by an information graphic.

3 Related Work

Research has addressed the problem of generating information graphics and accompanying captions[4, 15, 16]. In graphics generation, the system is given a set of data along with one or more communicative goals, and it designs a graphic that achieves these goals. Our problem is different: we are given the information graphic and must identify its communicative goal (the message that it conveys) by reasoning about the communicative signals in the graphic. Futrelle and Nikolakis[10] developed a constraint grammar for parsing vector-based visual displays and producing structured representations of the elements comprising the display, but Futrelle's goal is to produce a graphic that serves as a simpler representation of one or more graphics in a document[9]. Our work is the first to address the problem of *understanding* an information graphic — ie., recognizing the message that it conveys.

4 Graph Understanding System

We have developed a Bayesian system[7] that treats information graphics as a form of language and hypothesizes a graphic's intended message. The system takes as input an xml representation of the visual image (produced by a visual extraction module) that specifies the graphic's axes, the bars, their heights, colors, labels, any special annotations, the caption, etc. We have identified three kinds of communicative signals that appear in bar charts:

- the relative effort required for different perceptual tasks; for example, it is easier to determine the rank of an entity in a bar chart if the bars are sorted according to height than if they appear in alphabetical order of their labels. AutoBrief[15] contended that graphic designers construct graphics so that *important* perceptual tasks (the ones necessary for achieving the graphic's communicative goal) are as easy as possible. Thus the relative effort required for different perceptual tasks serves as evidence about which tasks the viewer is intended to perform in deciphering the graphic's message.
- the salience of entities in the graphic; for example, coloring a bar differently from other bars in a bar chart makes the bar salient, as does mentioning its label in the caption. Our hypothesis is that salient entities play a significant role in a graphic's message.
- the presence of suggestive verbs (such as *rising*) in a graphic's caption

Our system, described in [7], extracts this evidence from a given bar chart and enters it into a Bayesian network which hypothesizes the graphic's intended message. Leave-one-out cross validation on a corpus of 110 bar charts showed that our system has a success rate of 79.1% in identifying the graphic's message. Although our current system is limited to bar charts, we believe that our methodology is extensible to other kinds of information graphics.

5 Problems in Message Realization

Our Bayesian system produces a logical representation of a graphic's message; this representation consists of the message type (such as Maximum for messages which convey that a particular entity has the largest value in a bar chart), and the parameters of the message (such as the bar with the largest value in the case of the *Maximum* message type). For example, the system produces Maximum(First_Bar) for the graphic in Figure 2. Reiter and Dale[17] argue that templates are appropriate for many natural language problems. Since the syntactic variability in our target messages is limited, we use templates for generation, with one template defined for each of our 12 message types.

For the graphic in Figure 2, the natural language output should ideally be *"Tennis has the highest number of past nominees for the Laureus World Sports Awards among the sports listed"*. However, several problems arise:

1. the appropriate referent for the dependent axis (in this case, *number of past nominees for the Laureus World Sports Awards*) is not part of the logical representation and is not explicitly given in the graphic.
2. for some message types (such as *Maximum* in the above example), a decision must be made regarding when the labels should be enumerated in the natural language text and when only the ontological category of the labels should be given.

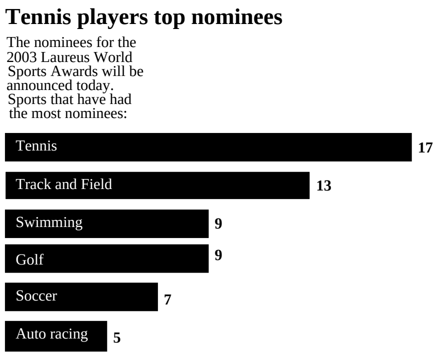

Fig. 2: *Graphic from USA Today*

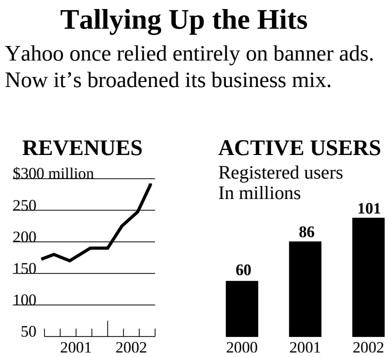

Fig. 3: *A composite graphic from Newsweek*[1]

3. the ontological category of the independent axis labels (in the above example, *sports*) must be inferred from the bar labels.

Section 6 presents our corpus study and solution to the difficult problem of realizing an appropriate referent for the dependent axis, Section 7 discusses how Gricean maxims motivated our decision about when to enumerate the bar labels, and Section 8 describes how we identify the ontological category of the labels.

6 Measurement Axis Descriptor

The first, and most serious problem, encountered in generating the message conveyed by an information graphic is the identification of an appropriate referent for the dependent axis. We will refer to this referent as the **measurement axis descriptor**. We undertook a corpus analysis in order to identify where the measurement axis descriptor appears in a graphic and to motivate heuristics for extracting it.

6.1 Corpus analysis

We analyzed 107 simple bar charts from articles in newspapers and popular magazines such as Newsweek and Business Week. We observed that graphics contain a set of component texts, in addition to the bar labels, that are visually distinguished from one another (e.g by placement, blank lines, or different fonts), which we refer to as **text levels**. Table 1 lists the various text levels, along with how often they appeared in the graphs in our corpus. In composite graphs (graphs consisting of multiple individual graphs as in Figure 3), Overall_Caption is the text that appears at the top of the overall group and serves as a caption for the whole set; Overall_Description is additional text, distinguishable from the caption, that is common to the set of graphics and often elaborates on them. Caption and Description serve the same roles for an individual graphic. For example, in Figure 3, the Overall_Caption is *"Tallying Up the Hits"* and the Overall_Description is *"Yahoo once relied entirely on banner ads. Now it's broadened its business mix"*. The Caption for the bar chart at the right of Figure 3 is *"Active Users"* and the Description is *"Registered users In millions"*. For the graphic in Figure 2, the Caption is *"Tennis players top nominees"* and the Description is *"The nominees for the 2003 Laureus World Sports Awards will be announced today. Sports that have had the most nominees:"*. Text_In_Graphic is any text residing within the borders of a graphic, such as *"U.S. Biotech Revenues, 1992-2001"* in Figure 4. Dependent_Axis_Label is the label (if any) on the dependent axis of a bar chart, such as *"Revenues(in billions)"* in Figure 4. Lastly, Text_Under_Graphic is any text under a graphic; such text generally starts with a marker symbol (such as *). As our evaluation in Section 10 shows, no single text level provides an acceptable measurement axis descriptor for all graphics. For example, Text_In_Graphic is a better measurement axis descriptor than Dependent_Axis_Label for the graphic in Figure 4, but the bar chart in Figure 3 does not have any Text_In_Graphic.

Text level	Freq. of Occurrences
Overall_Caption	31.8%
Overall_Description	17.8%
Caption	99.0%
Description	54.2%
Text_In_Graphic	39.3%
Dependent_Axis_Label	18.7%
Text_Under_Graphic	7.5%

Table 1: *Text levels in bar charts*

The goal of our corpus study was to identify how to construct a measurement axis descriptor for a graphic. Two annotators analyzed each of the 107 graphics in our corpus and determined the ideal measurement axis descriptor. We then analyzed each of the graphics to identify where the ideal measurement axis descriptor appeared. In 55.1% of the graphics, the ideal measurement axis descriptor appeared as a unit in a single text level, but in 36.5% of these instances, the text level contained additional information. For example, in the graphic at the left of Figure 5, the ideal measurement axis descriptor is *"identity-theft complaints"* which is part of the Caption *"Identity-theft complaints are skyrocketing"*. In 44.9% of the graphics, pieces of the measurement axis descriptor had to be extracted from more than one text level and melded together. In these instances, the ideal measurement axis descriptor can be viewed as consisting of a **core** or basic noun phrase from one text level that must be augmented with text from another level (or in some cases, from text in the accompanying article). For example, for the bar chart at the right of Figure 3, *"registered users"*

[1] This figure displays two of the five individual graphs comprising the composite graphic that appeared in *Newsweek*.

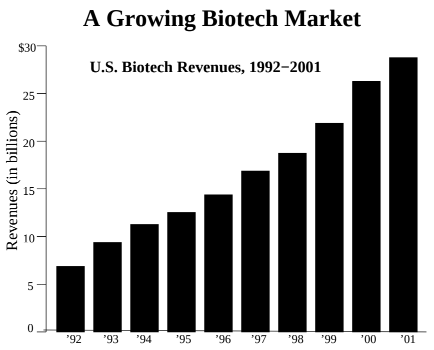

Fig. 4: *Graphic from Business Week*

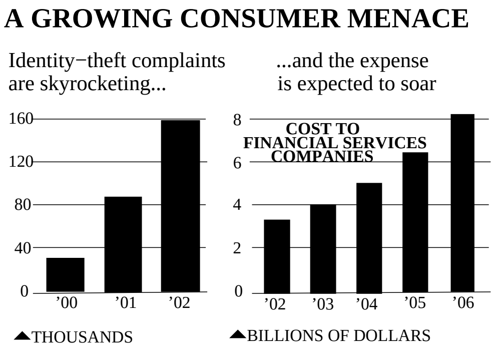

Fig. 5: *Graphic from Business Week*

is the core of the ideal measurement axis descriptor which is *"Yahoo's registered users"*.

We also found that, with the exception of Text_Under_Graphic (which typically serves as a footnote providing detail about the core), the ordering of text levels in Table 1 forms a hierarchy of textual components, with Overall_Caption and Dependent_Axis_Label respectively at the top and bottom of the hierarchy, such that the core generally appears in the lowest text level present in the graphic. For example, for every graphic in our corpus containing a Dependent_Axis_Label that was not just a scale or unit indicator (such as *millions* or *dollars*), the Dependent_Axis_Label contained the core. Where the Dependent_Axis_Label was absent from a graphic or was only a scale or unit indicator, but the graphic contained a Text_In_Graphic component, the core appeared in Text_In_Graphic in 35 of 39 graphics. Similar observations held for the text levels higher in the hierarchy. In retrospect, this is not surprising since text levels lower in the hierarchy are more specific to the graphic's content and thus more likely to contain the core of the ideal measurement axis descriptor.

We also observed the presence of cues, such as the phrase *"Here is"* or a terminating colon punctuation mark, suggesting that a text level contains the core of the measurement axis descriptor. For example, the phrase *"Here is"* realized as a contraction in the sentence "Here's the monthly construction spending", suggests that the subsequent noun phrase tells what the graphic is presenting and thus contains the core of the measurement axis descriptor.

During the corpus analysis we observed three ways in which a core from one text level was augmented to produce the ideal measurement axis descriptor:

- Expansion of the noun phrase: nouns in the core of the descriptor were replaced with a noun phrase which had the same noun as its head. For example, in the graphic in Figure 4, the Dependent_Axis_Label contains the core (*Revenues*) but the ideal measurement axis label is *"U.S. Biotech Revenues"* appearing at a higher text level; this ideal measurement axis descriptor can be viewed as an expansion of the core.
- Specialization of the noun phrase: the core was augmented with a proper noun which specialized the descriptor to a specific entity. For example, in the graphic at the right of Figure 3, the ideal measurement axis descriptor *"Yahoo's registered users"* consists of the core *"registered users"* augmented with the proper noun "Yahoo" that appears in the Overall_Description.
- Addition of detail: Text_Under_Graphic typically serves as a footnote to give specialized detail about the graphic. If the Text_Under_Graphic begins with a marker, such as an asterisk, and the core is followed by the same marker, then Text_Under_Graphic adds detail to the core. For example, in the graphic in Figure 6, "unit costs" is the core but the ideal measurement axis descriptor must also contain the information from Text_Under_Graphic, namely *"U.S. only, one available seat flown one mile, year ending June 2002"*.

6.2 Methodology

Our methodology for constructing a measurement axis descriptor is based on the insights gained from our corpus analysis. First, preprocessing extracts the scale and unit indicators from the text levels or from labels on the dependent axis (for example, the label *30%* would indicate that *percent* is the unit of measurement). Next heuristics are used to construct the core of the measurement axis descriptor by extracting a noun phrase from a text level of the graphic. Three kinds of augmentation rules, corresponding to the three kinds of augmentation observed in our corpus, are then applied to the core to produce the measurement axis descriptor. Finally, if the measurement axis descriptor does not already contain the unit of measurement (such as *percent*), the phrase indicating the unit of measurement is appended to the front of the measurement axis descriptor.

6.2.1 Heuristics

We developed 9 heuristics for identifying the core of the measurement axis descriptor. The application of the heuristics gives preference to text levels that are lower in the hierarchy, and the heuristics themselves take into account the presence of cue phrases, special characters, and the presence and position of noun phrases in a text level. The first heuristic only applies to the Dependent_Axis_Label:

- **Heuristic-1:** If the Dependent_Axis_Label contains a noun phrase that is not a scale or unit indicator, that noun phrase is the core of the measurement axis descriptor.

The second heuristic only applies to Text_In_Graphic:

- **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.

The remaining heuristics are then applied, in order, to a text level, starting with the Description text level; if a core is not identified at one text level, the heuristics are applied, in order, to the next higher text level in the hierarchy. Space limitations preclude listing all of the heuristics, but the following are two representative heuristics, in addition to Heuristic-1 and Heuristic-2 presented above. Heuristic-5 is based on observations about punctuation that suggests the presence of the core of the measurement axis descriptor, and Heuristic-8 is based on observations about the location of the core when it is part of a full sentence.

- **Heuristic-5:** If a fragment at the text level consists solely of a noun phrase followed by a colon (:), and the noun phrase is not just a proper noun, that noun phrase is the core.

- **Heuristic-8:** The core is the noun phrase preceding the verb phrase in the current sentence at the text level.

In some graphics, what is extracted as the core is conflated with a reference to the ontological category of the bar labels. If the core's head noun matches the ontological category of the bar labels, then that noun cannot be the measurement axis descriptor; thus if the noun is modified by a subsequent relative clause or a phrase beginning with *with*, then the nouns and subsequent prepositional phrases in the modifier are instead collected as the core. For example, consider Figure 2. Our heuristics would initially extract *"Sports that have had the most nominees"* as the core of the measurement axis descriptor; since *sports* is the category of the bar labels, *"nominees"* becomes the core.

6.2.2 Augmentation Rules

Augmentation rules correspond to the three kinds of augmentation observed during corpus analysis (expansion, specialization, and addition of detail), along with addition of the unit of measurement (such as *percent*). In expanding the core, the system examines text levels higher in the hierarchy than the text level from which the core was extracted; if a noun phrase appears with the same head noun as a noun in the core, and the noun phrase does not consist of just an adjective and the head noun, then the noun in the core is replaced with the larger noun phrase.

To specialize the noun phrase, the system determines whether 1) there is only one proper noun at all text levels higher in the hierarchy than the text level from which the core was extracted, or 2) there is only one proper noun in the Overall_Caption or the Caption; if one of these two criteria are satisfied and the proper noun is not a bar label in the graphic, then the possessive form of that proper noun is appended to the front of the core. (The reason for treating the Overall_Caption and Caption differently from the other text levels is that a single proper noun at these levels refers to the content of the graphic, whereas a proper noun at other levels, such as Description, may be part of an elaboration or comparison with the graphic's content and thus not necessarily specialize the core.)

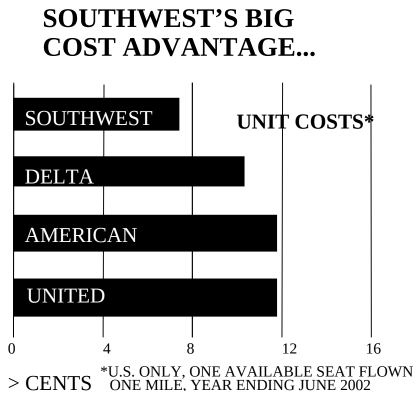

Fig. 6: *Graphic from Business Week*

To add detail to the core, the system determines whether the core was followed by a special marker, such as an asterisk, in the text level from which it was extracted; if so, the system searches for text preceded by the same marker and appends it to the core as a bracketed expression.

Finally, the measurement axis descriptor that has been constructed is checked to determine if it already gives the unit of measurement (as identified from the graphic during preprocessing); if not, a phrase indicating the unit of measurement is added to the front of the measurement axis descriptor.

6.3 Implementing the Heuristics

The heuristics must examine the parses of a graphic's text levels in order to identify the core of the measurement axis descriptor and to apply the augmentation rules. We experimented with both NP chunkers and parsers. NP chunkers are biased toward noun phrases and produced unsatisfactory results for text levels that consisted of full sentences. Thus we adopted Charniak's maximum entropy based parser[2], but had to address its bias toward imperative sentences over sentence fragments. Fragments are common as Text_In_Graphic, Caption, and Overall_Caption. When the Caption or Overall_Caption is a fragment, it may begin with a noun that can also be used as a verb. An example is *"Cost to financial services companies"* which appeared as the Caption on one of the graphics in our corpus. Unfortunately, a bias toward full sentences causes the parser to parse such fragments as imperative sentences with words such as *cost* identified as verbs. However, imperative sentences are rarely seen in the textual components of graphics. We solved this bias problem with rules such as the following:

> If WordNet[8] indicates that the first word of the input can be used as both a verb and a noun or as both a verb and an adjective, precede the input with "The" before sending it to the parser.

These rules forced the parser to prefer noun phrase fragments over imperative sentences.

7 Applying Gricean Maxims

Grice's Maxim of Quantity[11] states that one's discourse contribution should be as informative as necessary for the purposes of the exchange but not more so. Joshi, Webber, and Weischedel[13] showed that a system should not only produce correct information but should also prevent the user from drawing false inferences from the system's responses. If our system were to enumerate all entities involved in a comparison message, the response might be lengthy and the enumeration of little utility to the user. (To address instances in which the user wants the additional lengthy detail, both our system for blind individuals and our digital libraries application will include a facility for followup questions.) On the other hand, if the system never enumerates the entities (even when there are only a few), the user may make the false inference that there are too many to list. Thus we set a cut-off C, such that if the number of entities involved in a Maximum or Get-Rank message exceeds C, they are not enumerated. For the examples in this paper, C is set to 4, but we are performing human subject experiments to identify the most appropriate C value.

8 Identifying the Ontological Category

Although the dependent axis does not specify the ontological category for the bar labels, identifying the category results in better natural language than merely using a generic referent; for example, compare the phrase *"...among the sports listed"* with the phrase *"...among the entities listed"* in producing natural language text for the message conveyed by the graphic in Figure 2. We use OpenCyc ontology version v0.7.8b[5] to identify the ontological categories of bar labels.[2] For each bar label, all ontological categories it belongs to are identified. The most general and common category for at least two of the bar labels is identified as the ontological category.

9 Processing Examples

Our methodology for producing natural language text from the logical representation of a graphic's intended message has been implemented and tested on examples from many publications. Input to the natural language system is the logical representation of the graphic's message and the xml representation of the graphic's components, including the text at the various text levels.

The following examples illustrate how our system generates a graphic's message as natural language text. For the graphic in Figure 2, Heuristic-5 initially identifies the noun phrase *"sports that have had the most nominees"* as the core. However, its head noun *"sports"* matches the ontological category of the bar labels; consequently, the noun *"nominees"* in the relative clause modifying *"sports"* becomes the core. The augmentation rule for specialization finds that *"Laureus World Sports"* is the only proper noun in the text levels and forms *"Laureus World Sports's nominees"*. After adding a pre-fragment representing the unit of measurement, the measurement axis descriptor becomes *"The number of Laureus World Sports's nominees"*. Using the template for the Maximum message type, our system generates *"The bar chart titled 'Tennis players top nominees ' shows that the number of Laureus World Sports's nominees is highest for Tennis among the sports listed."* Our generated natural language for this example is imperfect and will be discussed further in Section 10.

For the graphic in Figure 4, Heuristic-1 identifies *"Revenues"* in Dependent_Axis_Label as the core. Since the core and the Text_In_Graphic, *"U.S. Biotech Revenues"*, have the same head noun, the augmentation rule for expansion produces *"U.S. Biotech Revenues"* as the augmented core. Using the template for the Increasing-Trend message type, our system renders the following natural language text: *"The bar chart titled 'A Growing Biotech Market ' shows that the dollar value of U.S. biotech revenues had a rising trend from 1992 to 2001."*

For the graphic in Figure 6, our system uses Heuristic 2, the augmentation rule for adding detail, and the template for the Minimum message type to produce the natural language text *"The bar chart titled 'Southwest's Big Cost Advantage ' shows that the cent value of unit costs (u.s. only one way available seat flown one mile, year ending june 2002) is lowest for Southwest among the entities listed: Southwest, Delta, American, and United."* Note that in this instance, our system was unable to identify the ontological category of the bar labels and therefore used the generic term *"entities"*. Note also that the bar labels were enumerated in this message since the number of bars did not exceed our cutoff of 4, whereas only the ontological category of the bar labels was given for the graphic in Figure 2.

10 Evaluation and Future Work

The quality of our generated text is largely dependent on how well we identify an appropriate measurement axis descriptor. Thus we constructed a test corpus consisting of 202 randomly selected bar charts from 19 different newspapers and magazines, along with their accompanying articles. We ran our system for each of the graphics and the resultant output was rated by two evaluators. The evaluators each assigned a rating from 1 to 5 to the system's output; if the evaluators differed in their ratings, then the lowest rating was recorded. For comparison, three baselines were computed, consisting of evaluations of the text that would be produced using each of the following three text levels as the measurement axis descriptor: Dependent_Axis_Label, Text_In_Graphic, and Caption. For the baselines, if the evaluators differed in their rating of the resultant output, the higher rating was recorded, thereby biasing our evaluation toward better scores

[2] If we could determine that the bar labels in a graphic cover all elements of a given category, then the generated message could for example say *"among airlines"* instead of *"among the airlines listed"*. However, we have not found an existing ontology that would allow us to reliably make such a determination.

for the baselines (in contrast with the scores for our system's output, where the lower score was recorded when the evaluators differed). We did not compute a baseline comparison using the text in Description as the measurement axis descriptor since that text level is most often full sentences and thus would generally produce very poor results.

5	**excellent text**
4	**very good:** very understandable but awkward
3	**good:** contains the right information but is hard to understand
2	**poor:** missing important information
1	**very bad**

The results of our evaluation are presented in Table 2. They show that our system produces natural language text that rates midway between good and very good. It is particularly noteworthy that our methodology performs much better than any of the baseline approaches.

However, further work is needed to improve our results. We need to resolve pronominal references within the text in a graphic and between texts in composite graphics, and we need to examine the internal organization/relation between graphs in composite graphics to identify the full referent of definite noun phrases. For example, for the bar chart on the right of Figure 5, the noun phrase *"the expense"* (and thus the noun phrase *"cost to financial services companies"* since *cost* and *expense* refer to the same entity) must be specialized so that it is *"the expense of identity-theft complaints"*, thereby leading to a measurement axis descriptor that is *"cost to financial services companies of identity-theft complaints"*. We also must take the tense of the text in the graphic into account in constructing a measurement axis descriptor. For example, the ideal measurement axis descriptor for the graphic in Figure 2 would indicate that the graphic is displaying the number of past nominees for the Laureus Sports Award, not current nominees. But the Description does not explicitly state *"past nominees"* and this must be inferred from the past tense in *"Sports that have had the most nominees"*.

11 Conclusion

This paper has presented our work on realizing the intended message of a simple bar chart. We have shown how Gricean maxims dictate the amount of information included in the summary and how the OpenCyc ontology is used to generate meaningful categories. We also presented our corpus analysis that explored where the ideal measurement axis descriptor appears in a graphic, discussed the insights that we gained from the corpus analysis, and presented our strategy for constructing a measurement axis descriptor by identifying a core descriptor and then augmenting it to obtain an appropriate measurement axis descriptor. Evaluation of our implemented system shows that our methodology generally produces reasonable text and that it performs far better than any of three baseline approaches. Moreover, our work illustrates how NLP methodologies can be successfully applied to non-stereotypical forms of language such as information graphics.

Approach	**Evaluation score**
Our system	3.574
Baseline-1: Dependent_Axis_Label	1.475
Baseline-2: Text_In_Graphic	1.757
Baseline-3: Caption	1.876

Table 2: *Evaluation of generated text*

Acknowledgements

We would like to thank Dr. Vijay Shanker for his valuable suggestions on addressing the problems encountered with the parser. This material is based upon work supported by the National Science Foundation under Grant No. IIS-0534948.

References

[1] S. Carberry, S. Elzer, and S. Demir. Information graphics: An untapped resource for digital libraries. In *the Proc. of SIGIR'06*, pages 581–588, 2006.

[2] E. Charniak. A maximum-entropy-inspired parser. In *the Proc. of NAACL'02*, pages 132–139, 2002.

[3] H. Clark. *Using Language*. Cambridge University Press, 1996.

[4] M. Corio and G. Lapalme. Generation of texts for information graphics. In *the Proc. of the 7th European Workshop on Natural Language Generation EWNLG'99*, pages 49–58, 1999.

[5] CycL.Cycorp. *URL: http://www.cyc.com.*

[6] S. Elzer, S. Carberry, D. Chester, S. Demir, N. Green, I. Zukerman, and K. Trnka. Exploring and exploiting the limited utility of captions in recognizing intention in information graphics. In *the Proc. of ACL'05*, pages 223–230, 2005.

[7] S. Elzer, S. Carberry, I. Zukerman, D. Chester, N. Green, and S. Demir. A probabilistic framework for recognizing intention in information graphics. In *the Proc. of IJCAI'05*, pages 1042–1047, July 2005.

[8] C. Fellbaum. *WordNet: An electronic Lexical Database*. The MIT Press, 1998.

[9] R. Futrelle. Summarization of diagrams in documents. In I. Mani and M. Maybury, editors, *Advances in Automated Text Summarization*. MIT Press, 1999.

[10] R. Futrelle and N. Nikolakis. Efficient analysis of complex diagrams using constraint-based parsing. In *the Proc. of the Third International Conference on Document Analysis and Recognition*, 1995.

[11] H. P. Grice. Logic and conversation. *Speech Acts*, 3:41–58, 1975.

[12] B. Grosz and C. Sidner. Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, 12(3):175–204, 1986.

[13] A. Joshi, B. Webber, and R. Weischedel. Living up to expectations: Computing expert responses. In *the Proc. of the National Conference on Artificial Intelligence*, pages 169–175, 1984.

[14] A. Kennel. Audiograf: A diagram-reader for the blind. In *the Proc. of the Second Annual ACM Conference on Assistive Technologies*, pages 51–56, 1996.

[15] S. Kerpedjiev and S. Roth. Mapping communicative goals into conceptual tasks to generate graphics in discourse. In *the Proc. of IUI'00*, pages 60–67, 2000.

[16] V. Mittal, J. Moore, G. Carenini, and S. Roth. Describing complex charts in natural language; a caption generation system. *Computational Linguistics*, 34:431–468, 1998.

[17] E. Reiter and R. Dale. Building applied natural language generation systems. *Natural Engineering*, 3(1):57–87, March 1997.

[18] J. Yu, J. Hunter, E. Reiter, and S. Sripada. Recognising visual patterns to communicate gas turbine time-series data. In *the Proc. of the Twenty-second SGAI International Conference on Knowledge Based Systems and Applied Artificial Intelligence*, pages 105–118, 2002.