Experience with a Mixed Semantic/Syntactic Parser

Peter J. Haug, Spence Koehler, Lee Min Lau, Ping Wang, Roberto Rocha, and Stanley M. Huff
Department of Medical Informatics, LDS Hospital, Primary Children's Medical Center, and the University of Utah, Salt Lake City, Utah

The value of the computerized medical record is derived in part from the availability of medical information in a coded form accessible to manipulation by processes designed for automated decision support, medical research, and computer assistance in the management of health care delivery. To meet these needs medical reports captured and stored as natural language documents must be encoded. Below we discuss an ongoing formative process aimed at developing a natural language understanding system for chest x-ray reports. Comparative data showing the progress of this process is presented.

INTRODUCTION

In recent years the prospect of a fully computerized medical record has been the source of increasing enthusiasm. A variety of groups have formed to promote this ideal and to encourage the development of a system for encoding medical data that will first, represent a standard model supportable in any medical information system and second, allow a complete representation of relevant medical data in a form fully accessible to computers.

The latter of these two goals is particularly challenging for that portion of the record that is composed of medical reports, dictated by physicians and other health care providers, transcribe using word processing equipment, and stored in a natural language format as a part of the computerized medical record. This approach to data capture makes the information in the dictated and transcribed report readily available to users of the medical information system. However, the data in the report is unavailable to computer-mediated processes that seek to provide medical decision support or that are used to analyze grouped medical data in order to support quality assurance initiatives, to enable medical research, and to provide data for the overall management of health care organizations.

To address these problems a number of researchers are investigating natural language understanding technologies. Tools based in this discipline are designed to read and understand the content of natural language based reports and to capture salient medical facts in a coded form accessible to computerized analysis. Although a variety of different free-text, medical documents have been the focus of natural language understanding systems (NLUS)[1,2,3], a significant amount of recent work has focused on x-ray reports, particularly reports concerning chest radiographs[4,5]. This has been an area where we have also previously developed a NLUS[6,7]. This system was used for more than five years and its output contributed useful information to a decision support system used by the LDS Hospital Infection Disease Department[8]. Recently we have been developing a more general tool for parsing and understanding free-text medical documents[9]. The focus of this system is to 1) test a new theory concerning the integration of syntactic and semantic techniques and to 2) generate a coded output consistent with a new model for medical data storage proposed for the next generation of the HELP medical information system[10]. The name that we have given to this new natural language parser is SymText (for Symbolic Text Processor). Here we describe an ongoing formative evaluation of this tool and present a set of results illustrating the direction of this evaluation process.

METHODS

The underlying structure of SymText has been described previously[9] and will be mentioned only briefly here. The system combines a syntactic parser using augmented transition networks[11] and transformational grammars[12] with a model of semantics based on the Bayesian network statistical formalism[13]. One of the principal goals of this research is to explore those interactions between syntax and semantics that can be modeled using these tools. To more easily manage the semantics we have broken the concepts common to chest x-ray reports into three concept spaces, a concept space dealing with abnormalities seen on the films, a concept space dealing with medical diseases or conditions inferred from these abnormal findings, and a concept space dealing with the various tubes, lines, and other appliances which are introduced into patients' chests during the course of therapy. The results from these three concept spaces are grouped together in the analyses below.

The output of this system are tables of concepts and words. The words are terms from a sentence that the parser believes it can properly interpret by relating these terms to underlying medical concepts. The concepts are those chosen by the parser to represent the meaning of the words in the sentence.

0195–4210/95/$5.00 © 1995 AMIA, Inc.

Figure 1: The output of SymText for an Observation (Finding).

With Network : "Finding Net 1parent"
Sub Sentence : "THE MILD HAZY OPACITY SEEN IN THE RIGHT UPPER LOBE HAS DECREASED SINCE YESTERDAY"
Instantiated : (("MILD" (1010 . 5)) ("HAZY" (1006 . 34)) ("OPACITY" (1005 . 5)) ("IN" (1012 . 4)) ("RIGHT" (1016 . 1)) ("UPPER" (1017 . 4)) ("LOBE" (1015 . 6)) ("DECREASED" (1022 . 73)))
For Event :
"
Instantiated Event:

Node ID	Node Name	Node Value
1001	*Observation :	*localized infiltrate (0.998679)
1002	*State :	*present (0.932927)
1003	State Term :	null (0.921965)
1004	*Topic :	*poorly-marginated opacity (infiltrate) (1.0)
1005	Topic Term :	opacity~n (1.0)
1006	Topic Modifier :	hazy~adj (1.0)
1007	Measurement ID :	null (0.991736)
1008	Topographic location :	null (0.586777)
1009	*Severity :	*low severity (0.692044)
1010	Severity Term :	mild~adj (1.0)
1011	*Topic/Location Link :	null (0.682542)
1012	Link Term:	in (1.0)
1013	*Anatomic Information :	*right upper lobe (1.0)
1014	Anat. Location Mod :	null (0.939394)
1015	Anatomic Location :	lobe~n (1.0)
1016	Anatomic Location Mod1 :	right (1.0)
1017	Anatomic Location Mod2 :	upper (1.0)
1018	Anatomic Location Mod3 :	null (1.0)
1019	Anatomic Location Mod4 :	null (0.969697)
1020	Anatomic Location Mod5 :	null (1.0)
1021	*Change With Time :	*diminished (1.0)
1022	Change Term :	decreased~v (1.0)
1023	Change Degree :	null (0.423528)
1024	Change Quality :	null (0.811764)
1025	*Subanatomic Information :	*null (0.909091)
1026	Subanatomic Link :	null (0.913621)
1027	Subanatomic Loc Modifier :	null (0.957066)
1028	Subanatomic Location :	null (0.907561)
1029	Subanatomic Mod1 :	null (0.980288)
1030	Subanatomic Mod2 :	null (0.99949)
1031	Subanatomic Mod3 :	null (0.969697)
1032	Subanatomic Mod4 :	null (0.981818)
1033	Subanatomic Mod5 :	null (1.0)

These concepts are the part of the output that will ultimately be stored in the clinical data base as a representation of the content of the report. Figure 1 illustrates the output of SymText for the sentence "The mild hazy infiltrate in the right upper lobe has decreased since yesterday."

In this example the concept slots are indicated by an asterisk (*) at the beginning of the slot name and of the slot value. Other slots are for words taken from the sentence.

Our efforts to test the effectiveness of the techniques used in SymText have been focused on a formative approach to evaluation and development. In this approach we select groups of 10 chest x-ray reports, process them with SymText, and evaluate the results. The evaluation is carried out by a team member (PH) who compares the concepts produced with those he assigns to sentences from the reports. Following each evaluation we alter SymText to correct any flaws discovered or to test a new approach and we repeat the review process.

Two types of analysis are done. In the first, the emphasis is on the main conceptual underpinnings in each medical fact. In our example above "*localized infiltrate" from the Observation slot and "*present" from the State slot combine to summarize the principal medical fact abstracted by the parser. If the system fails to recognize this pair of concepts it has missed the overall theme of the sentence or the constituent sub-sentence which the output represents.

Version*	Description
Version 2.05	The initial version thought adequate for testing.
Version 3.0	New transformations were added to better handle conjunctions. A new ATN grammar was introduced. The syntactic parsing strategy was revised so that redundant definitions were not produced. A new flow was tested that used transformations to get alternate parses of a sentence instead of backtracking in the ATN. In this approach, the ATN grammar is required to produce only a minimal parse. The transformations then continue to reformat the parse tree based on that parse.
Version 3.3	Additional grammar and transformations. Enhanced handling of ambiguity (when there are multiple competing word level slot fillers). First attempt to use the absence of words in a sentence in the semantics. To do this we inspect sentences for possible fillers for critical slots. If these fillers are not present we instantiate the slot as "null". Belief network structure used in parsing. Previously only the probabilities generated by the network were considered during the parse (structural information came from the grammar and transformations).
Version 3.5	Previous versions were shown to be too sensitive to single unusual terms. Changed processing so that no single word could drive the analysis to completion.
Version 3.6	Modest improvements to Bayesian Network. New ATN grammar is tested complimented by more transformations. Testing done without "null" instantiation process.

*Not all versions were tested with 10 reports. Some proved disappointing on cursory testing with a group of test sentences and were skipped. These versions are the subset that were tested fully.

In the second type of analysis undertaken, the accuracy of the secondary concepts (Topic, Severity, Topic/location Link, Anatomic Information, Change with Time, and Subanatomic Information) are evaluated in a similar way.

To aggregate information concerning the accuracy of each output listing, we categorize each concept abstracted from the report in one of four ways (in the case of the main concept we treat it and the state concept as a single composite medical fact). First, a concept filling a slot may accurately reflect the information in a sentence. Second, the slots are filled but the results may incorrectly interpret the real concepts in the sentence. Third, information present in the sentence may warrant encoding (i.e. contain relevant concepts) but the system may fail to recognize them and no slot fillers (in fact, no output at all) exists. This is a complete failure of SymText.

A fourth category was also recorded to reflect the fact that SymText is expected to capture information only for families of concepts where it has been trained. This expectation reflects a common dilemma in expert system development: these systems often cannot tell when a problem is outside of their range of knowledge. In the realm of NLUSs, the principal manifestation of this difficulty occurs when a system attempts to interpret utterances whose concepts are not represented in its underlying knowledge base. As an example, the system's current knowledge base does not include information about surgical history. Therefore the system would be expected to disregard the sentence "The patient is previously status post median sternotomy for CABG." If it generated an output from this sentence it would be in error and the concepts in the bad output listing would be marked as incorrect. Sentences containing foreign concepts occur with varying frequency in our test sets. In order to measure the opportunity for these errors and the success of SymText in avoiding them, we chose to record information for those sentences where SymText correctly concluded that no interpretation was possible. In these cases SymText fails appropriately to create an event instantiation

The results reported here represent five iterations on SymText's syntactic and semantic models exploring different strategies for improving its behavior. Table 1 describes the changes in approach that characterized each version.

RESULTS

The results of five passes through the evaluation procedure are shown below.

Figure 2: Key concept recognition behavior of five versions of SymText.

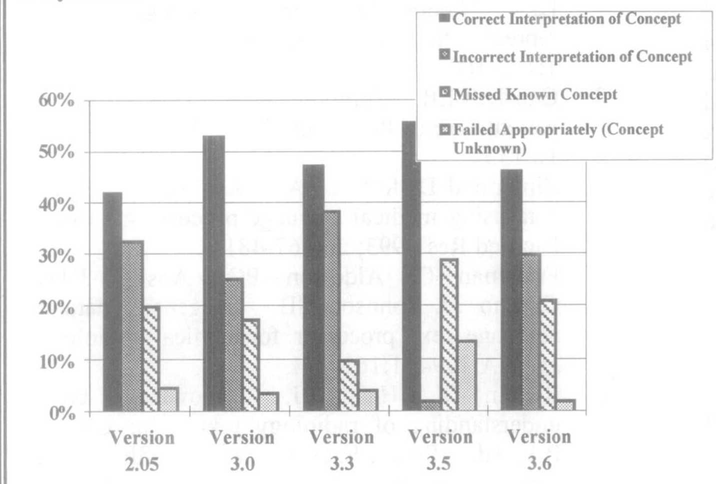


Figure 3: Combined comparison of five versions of SymText (key concepts).

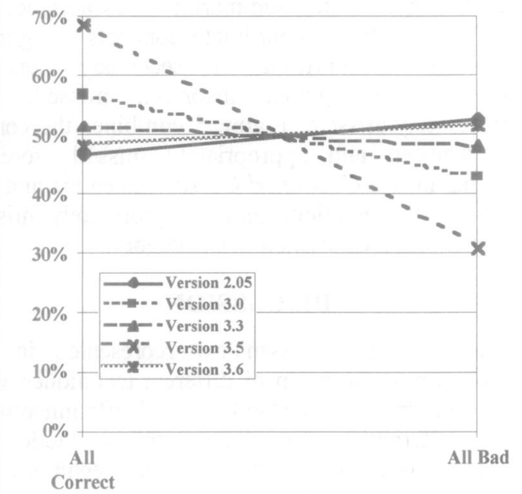


Figure 2 and 3 summarize accuracy data for the main concept/state slots in two ways. First, the concepts found in the outputs data from each version are categorized in the four ways described above. A simple percentage for each outcome category is graphed for each of the five versions of the system. This allows us to compare the effects of different parsing strategies on each of the four types of behavior.

The second display (figure 3) combines the two acceptable behaviors, "correct" and "appropriately missed", into *Combined Correct* and the two bad behaviors, "incorrect" and "missed known concepts", into *Combined Bad* to summarize the effects of the changes in the system with different versions. It is apparent from this graph that the approach taken in version 3.5 represents the most successful of these five approaches. However, aspects of the approaches in other versions also hold promise. We hope to show in the future that a combination of the best features of these algorithms can further improve the accuracy.

The information displayed above focuses on the main concept represented in the output from SymText and the status (present, absent, or possible) associated with that concept. In the semantic model used in this work, this is the concept that coordinates the estimation of probabilities for the nodes represented and that summarizes the data elements which will ultimately be stored.

Figure 4: Secondary concept recognition behavior of five versions of SymText.

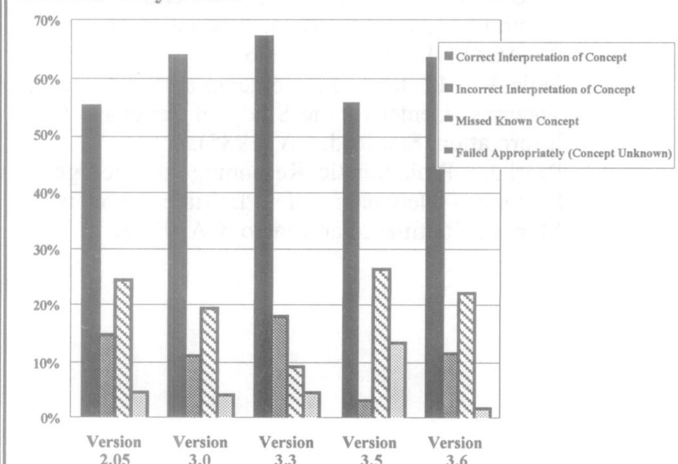
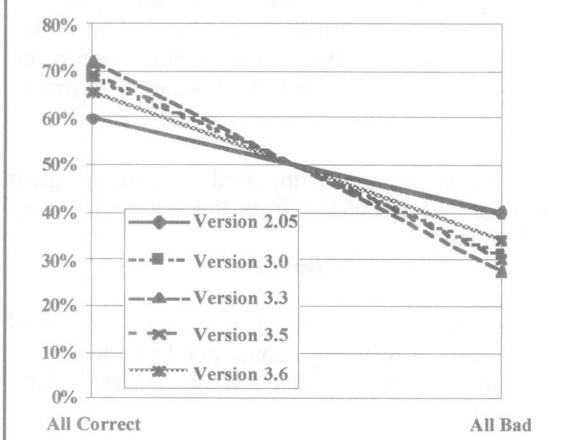


Figure 5: Combined comparison of five versions of SymText (secondary concepts).



However, there are up to 6 additional secondary concepts produced for each sentence. Above we use similar displays to summarize the accuracy of SymText for these subordinate concepts. Figure 4 shows the proportions of secondary concepts categorized as correct, incorrect, missed, and appropriately missed. Figure 5 combines the correct interpretations and appropriately missed (foreign) concepts into a *Combined Correct* category and the incorrect interpretations and inappropriately missed concepts into a *Combined Bad* category.

DISCUSSION

In the versions of SymText represented in the displays above, a group of different techniques were tested to help us understand the tradeoffs inherent in several alternatives for linking the syntactic and semantic aspects of the system. Each technique had advantages as well as disadvantages. Other approaches appear attractive and are being incorporated into new models for testing.

Ultimately a collection of the most promising techniques will be woven into a version of SymText which will be integrated into the HELP system. Unlike its predecessor, this NLUS will not reside in the Tandem mainframe computer that houses the HELP system. Instead it will reside on a separate computer constituting a natural language server. As the capabilities of this system grow, we anticipate sending it a variety of reports for encoding. The results will be returned to the originating system for storage and will serve as the substrate for a collection of applications that require data abstracted from medical reports to provide decision support and other services.

Once a version of SymText exists combining the best of the techniques evaluated for this paper as well as the best of the additional techniques that we are exploring, we will test the accuracy of this system in a large scale summative evaluation. Until that system is defined, we have found significant advantage in the iterative testing scheme described here.

*This publication was supported in part by grant number 5 R01 LM05323 from the National Library of Medicine.

References

1. Sager N, Friedman C, Lyman MS. Medical Language Processing: Computer Management of Narrative Data. Addison-Wesley, Menlo Park, CA (1987).
2. Sager N, Margaret L, Bucknall C, Nhan N, Tick LJ. Natural language processing and the representation of clinical data. JAMIA (1994); 1:142-160.
3. Gabrielli ER. Computer assisted assessment of patient care in the hospital. J Med Syst(1988); 12:135.
4. Zingmond D, Lenert LA. Monitoring free-text data using medical language processing. Comp Biomed Res 1993; 26:467-481.
5. Friedman C, Alderson PO, Austin JHM, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. JAMIA 1994; 1:161-174.
6. Ranum DL, Haug PJ. Knowledge based understanding of radiology text. In: Greenes RA, ed. Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care. (1988): 141-145.
7. Haug PJ, Ranum DL, and Frederick PR. Computerized Extraction of Coded Findings from Free-Text Radiology Reports. Radiology (1990); 174:543-548.
8. Evans RS, Gardner RM, Bush AH, Burke JP, Jacobsen JA, Larsen RA, Meier FA, Warner HR. Development of a computerized infectious disease monitor (CIDM). Comput Biomed Res (1985) 18:103-113.
9. Haug PJ, Koehler S, Lau, LM, Wang P, Rocha R, Huff S. A Natural Language Understanding System Combining Syntactic and Semantic Techniques. Eighteenth Annual Symposium on Computer Applications in Medical Care (1994), pp. 247-251.
10. Huff SM, Haug PJ, Stevens LE, Dupont RC, Pryor TA. HELP The New Generation: A New Client-Server Architecture. Eighteenth Annual Symposium on Computer Applications in Medical Care (1994), pp. 271-275.
11. Woods WA. Transition Network Grammars for Natural Language Analysis. Communications of the ACM (1970); 13:591-606.
12. Sells P. Lectures on Contemporary Syntactic Theories. Center for the Study of Lanugage and Information, Stanford, CA (1985).
13. Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufman, San Mateo, CA, 1988.