

# ONYX: A System for the Semantic Analysis of Clinical Text

Lee M. Christensen, Henk Harkema, Peter J. Haug,  
Jeannie Y. Irwin, Wendy W. Chapman

Department of Biomedical Informatics

University of Pittsburgh

Pittsburgh, PA 15214, USA

lmc61 heh23 rey3 wec6 @pitt.edu

University of Utah

Salt Lake City, Utah, 84143, USA

Peter.Haug@intermountainmail.org

## Abstract

This paper introduces ONYX, a sentence-level text analyzer that implements a number of innovative ideas in syntactic and semantic analysis. ONYX is being developed as part of a project that seeks to translate spoken dental examinations directly into chartable findings. ONYX integrates syntax and semantics to a high degree. It interprets sentences using a combination of probabilistic classifiers, graphical unification, and semantically annotated grammar rules. In this preliminary evaluation, ONYX shows inter-annotator agreement scores with humans of 86% for assigning semantic types to relevant words, 80% for inferring relevant concepts from words, and 76% for identifying relations between concepts.

## 1 Introduction

This paper describes ONYX, a sentence-level medical language analyzer currently under development at the University of Pittsburgh. Since ONYX contains a number of innovative ideas at an early stage of development, the objective of this paper is to paint a broad picture of ONYX and to present preliminary evaluation results rather than analyzing any single aspect in detail.

ONYX is being developed as part of a project aimed at extracting information from spoken dental examinations. Currently, dental findings must be charted after an exam is completed or may be charted by an assistant who acts as a transcriptionist during the exam. Our goal is to design a system capable of automatically extracting chartable find-

ings directly from spoken exams, potentially also supporting automated decision support and quality control. We are also developing tools to enable the system to be ported to other clinical domains and settings.

Extracting information from unedited speech transcriptions presents a number of challenges. Sentences may be fragmented or telegraphic, and much of the speech may be irrelevant for our purposes. The following example illustrates some of these difficulties:

*"Okay. Okay. Open. Okay. No. I is missing. Two occlusal distal amalgam. Actually, make that occlusal. Also, one palatal amalgam. Can you close just slightly? And perfect. Okay, now open again."*

The relevant findings in this example are that tooth number one is missing and tooth number two has amalgam fillings on the occlusal and palatal surfaces. Our ultimate challenge is to create a system that can recognize relevant sentences and perform competently in the face of the inherent ambiguity and noise commonly found in conversational speech. ONYX does not yet address all of these challenges, although we have clear directions we are pursuing as described in the Future Work section of this paper. Our goal in this paper is to describe the current state of ONYX and the innovations we feel will enable it to be adapted to complex NLP tasks in the future.

## 2 Overview of ONYX

ONYX is the middle component of a pipelined architecture as illustrated in figure 1. The entry point to this architecture is a speech-to-text analyzer, which takes input from a microphone worn

by the dentist and produces a transcription that ONYX analyzes for semantic content. ONYX's output is then passed to a discourse analyzer that applies dental knowledge to assemble ONYX's sentence-level semantic representations into chartable exam findings.

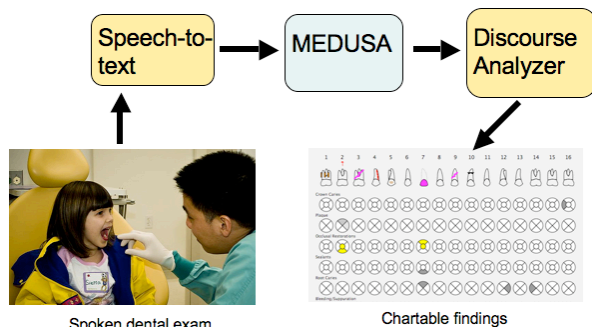


Figure 1. Speech-to-chart pipeline.

ONYX looks for dental conditions such as caries, fractures and translucencies; restorations such as fillings and crowns; tooth locations; and modifiers such as tooth part, tooth surface, and condition extent. It produces templates of words and concepts. Table 1 shows a summary of four templates (Dental Condition, Tooth Location, Surface and State) representing the meaning of "eight mesio might have a slight translucency."

Dental Condition	
Condition Concept	*translucency
Condition Term	"translucency"
Severity Concept	*superficial
Severity Term	"slight"
Tooth Location	
Location Concept	*numberEight
Tooth Number	"eight"
Surface	
Surface Concept	*mesial
Front/Back Term	"mesio"
State	
State Concept	*possible
State Term	"might"

Table 1: ONYX templates for "eight mesio might have a slight translucency." Terms with an \* are inferred concepts.

ONYX's interpretations are represented as binary predicates that take the templates as arguments (for convenience, only the summary concepts from the templates are shown):

*ConditionAt(\*translucency, \*numberEight) & LocationHasSurface(\*numberEight, \*mesial) & StateOf(\*translucency, \*possible)*

ONYX builds on ideas from MPLUS (Christensen et al, 2002), which was used primarily to interpret radiology reports. MPLUS uses Bayesian networks (BNs) to produce filled templates. Through a training process, words from the corpus of training documents are manually associated with states of terminal nodes in a BN, and concepts are associated with states of nonterminal nodes. When MPLUS interprets a sentence, it instantiates the BNs with words from the sentence and infers the most probable concepts consistent with those words. It then generates templates filled with those words and concepts.

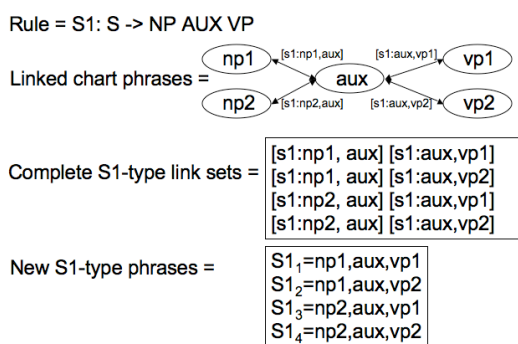
BNs have proven useful in semantic analysis (e.g. Ranum 1989, Koehler 1998, Christensen 2002); their performance degrades gracefully in the face of various types of lexical and syntactic noise. The main disadvantage with using BNs is their inherent computational complexity. ONYX employs a semantics-intensive form of parsing, interpreting each phrase as it is constructed rather than waiting until the syntactic analysis is completed to do the interpretation. For this reason we have developed an experimental probabilistic classifier for ONYX called a Concept Model (CM). CMs support a tree-structured representation of related words and concepts (figure 5), structurally similar to the BNs used by MPLUS, but using a more efficient model of computation. In essence CMs are trees of Naïve Bayes classifiers, although they contain enhancements, not described in this study, which in general make them more accurate than strict Naïve Bayes. Each node together with its children constitutes a single classifier. When a CM is applied to words in a sentence, word-level CM states are assigned a probability based on training data. Probabilities are propagated upwards through the CM, calculating probabilities for all concepts that depend directly or indirectly on the words of the sentence.

### 3 ONYX Syntactic Analyzer

For this project we desired a parser that was fast, flexible and robust. We designed a variation on a bottom-up chart parser (Kay, 1980) and hand-crafted an initial set of 52 context-free grammar

rules. Chart parsers based on Kay’s algorithm maintain an agenda of “edges,” which correspond to partially or completely instantiated grammar rules. In the original algorithm, for each new phrase added to the chart an edge is created for each rule that can begin with that phrase. In addition, each existing edge that abuts and can be extended with that phrase is duplicated with a pointer to the new phrase. When an edge has no more unmatched components, it is regarded as a new phrase that can begin or extend other edges. Since edges are used to anticipate all possible continuations of phrases vis-à-vis the grammar, the number of edges grows quickly relative to the number of words in the sentence. Charniak et al. (1998) noted that exhaustively parsing maximum-40-word sentences from the Penn II treebank requires an average of 1.2 million edges per sentence.

ONYX’s parse algorithm replaces edges with binary links. We briefly describe this new algorithm. A set of binary link templates is defined for each grammar rule. For instance, the rule  $S \rightarrow NP\ AUX\ VP$  (labeled S1) would produce the templates  $[s1:np,aux]$  and  $[s1:aux,vp]$ . When a phrase is added to the chart, binary links for all applicable rules are added from that phrase to juxtaposed phrases to the left and right on the chart. When a right or left-terminating link is added (all links for rules with two or three components are right or left terminating), a quick search is done in the other direction for links belonging to the same rule. Each complete set of links defines a new phrase of the target type, as shown in figure 2.



**Figure 2.** Binary links for the rule S1:S->NP AUX VP used to generate new phrases of type S1 from juxtaposed NP, AUX, and VP phrases on the chart.

Although we have not analyzed the time and space complexity of this algorithm, it has proven to be

more efficient than the edge-based parser used by MPLUS. Time and space complexity for chart parsers is calculated based on the number of edges produced, which has been shown to be  $O(n^3)$ , with  $n$  words in a sentence. Since binary links, unlike edges, are only used to record grammatical relations between juxtaposed phrases on the chart (rather than anticipating possible continuations), are not duplicated, and can participate in the creation of multiple new phrases, the number of binary links grows more slowly than the number of edges. On the other hand, the need to search for completed link sets increases processing time. We plan to formally analyze the time and space requirements of this algorithm in a future study.

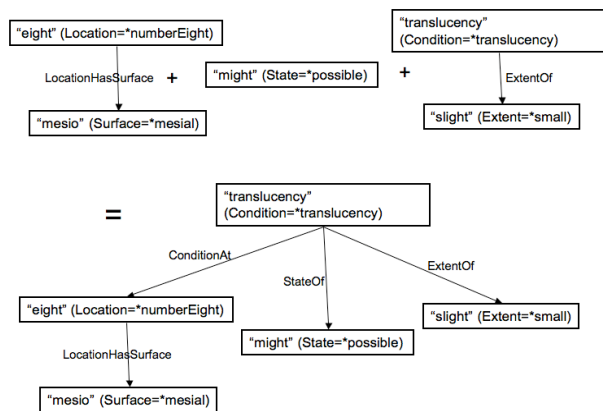
#### 4 □ ONYX Semantic Analyzer

In ONYX syntax and semantics are highly integrated. Rather than waiting for a completed parse tree to begin the interpretation process, ONYX semantically interprets each phrase as it is created and before it is placed on the chart. Each phrase is assigned a “goodness” score based in part on the goodness of its semantic interpretation, and this score is used in determining the order in which phrases are expanded, resulting in a semantically guided best-first search.

To represent semantic relations between templates, ONYX uses a custom-built first-order predicate language with a syntax based roughly on the Knowledge Interchange Format (Genesareth & Fikes, 1992). ONYX interpretations are conjuncts of binary predicates formulated in this language, with templates as arguments. This language is for internal use only; ONYX will use standard language protocols for communicating with external systems. We decided to implement our own language rather than using an existing implementation in order to have access to the underlying data structures, which we use in three ways not traditionally applied to symbolic languages: 1- We have extended our language to include Java objects as constants and Java methods as functions and relations. In particular, CM templates are treated as constants in the language, and CMs are semantically typed functions that map words to templates. 2- As described next, ONYX’s default mode of semantic interpretation is based on a form of graph unification. Binary predicates are treated as unifi-

able links in a graph as shown in figure 3. 3-ONYX uses the predicate structure of an interpretation to pass information between CMs. For instance, if an interpretation contains the relation *ConditionAt(Condition, Location)*, ONYX inserts the summary concept from the Location CM into the Condition CM. This allows the Condition CM to factor tooth location into its determination of the most probable Condition concept.

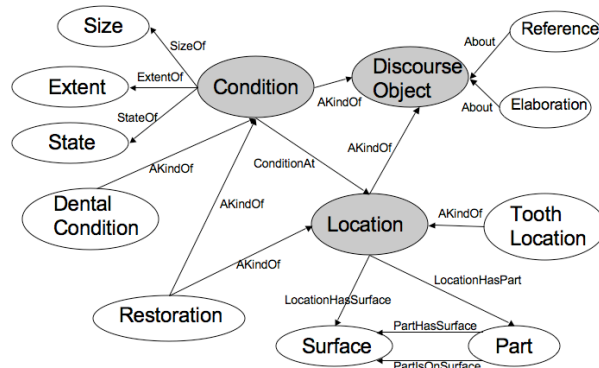
Figure 3 illustrates ONYX’s unification-based interpretation process. ONYX relies on a semantic network that defines types and relations in the dental domain (figure 4). As dental concepts are brought together in a phrase, links connecting those concepts are extracted from the semantic network and formulated into binary predicates in an interpretation. As phrases are joined together in larger phrases, their relations and templates are merged, resulting in an interpretation tree denoting a dental object (e.g. dental condition, tooth location, tooth surface) with possibly multiple levels of modifiers. For instance, the interpretation for "eight mesio might have a slight translucency" can be generated from the partial interpretations of the phrases "eight mesio", "might" and "slight translucency" as shown in figure 3.



**Figure 3.** Interpreting “eight mesio might have a slight translucency” using graph unification.

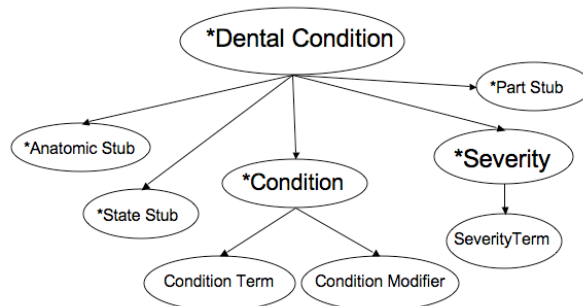
There are two primary justifications for using unification in this way. First, conjoined phrases, particularly noun phrases, often contain unifiable partial descriptions of a single object. Second, if concepts appear together in a phrase, there is a good chance that relations connecting those concepts in the semantic network are captured, explicitly or implicitly, in the meaning of the phrase.

The dental semantic network is shown in figure 4. Terminal (white) nodes define concrete semantic types associated with dental CMs. For instance, the DentalCondition type is associated with the concept model shown in figure 5.



**Figure 4.** Semantic network for dental exams.

Nonterminal (gray) nodes represent abstract types with no associated CMs. A concrete type may have more than one abstract parent type. For instance, a Restoration, such as a crown, is both a Condition and a Location. As such, it can exist at a tooth location, e.g., "the crown on tooth 5," and it can be the location of condition, e.g., "the crack on the crown on tooth 5." Since a concrete type can have multiple parent types, ONYX often produces multiple alternative interpretations over words of a sentence. For instance, ONYX may produce two interpretations for "mesial amalgam"—one referring to the mesial surface of an amalgam filling, and one referring to an amalgam filling on the mesial surface of some unspecified tooth. ONYX uses probabilities derived from training cases to prefer the latter interpretation, which is the more likely of the two.



**Figure 5.** Dental Condition Concept Model.

Each concept model has a tree structure as illustrated in figure 5, which shows the structure of the Dental Condition CM. Nonterminal nodes repre-

sent concepts, and terminal nodes represent words, with the exception of stub nodes. The value of a stub node is the summary concept (i.e., root node) from the CM of the same name.

One problem with ONYX's graph-based model of interpretation is that the semantic network does not capture all relations that might be expressed in a dental exam. The network was deliberately kept simple by including mostly relations that are categorically true (e.g., all teeth have surfaces) or that are frequently talked about (e.g., restorations are frequently mentioned as being locations of other conditions). This restriction helps keep the unification process tractable and minimizes ambiguity, but interpretations may miss important points. For instance, the ONYX interpretation of "15 occlusal amalgam" is *ConditionAt(\*filling, \*toothFifteen) & LocationHasSurface(\*toothFifteen, \*occlusal)* which can be paraphrased as "a filling at tooth 15 and tooth 15 has an occlusal surface". This interpretation misses the important fact that the filling is *on* the occlusal surface of tooth 15, which we would normally infer from the fact that "occlusal" adjectivally modifies "amalgam." Another limitation is that although the semantic network as it stands can describe single objects with their modifiers, it cannot be used to build up complex descriptions involving multiple objects of the same type.

To address these limitations we have added a second, more specialized mode of interpretation that is contingent on lexical and syntactic information from the parse and that can introduce into an interpretation predicates that do not exist in the semantic network. This mode of interpretation uses semantic types and patterns attached to grammar rules. As an example, the rule  $NP \rightarrow AP NP$  can be semantically annotated thus:

$NP\langle Restoration \rangle \rightarrow AP\langle Surface \rangle NP\langle Restoration \rangle$   
 $\Rightarrow OnSurface(Restoration, Surface)$

This rule captures the idea that if a Surface-type adjectival phrase modifies a Restoration-type noun phrase, the restoration exists on that surface. Applied to "occlusal amalgam" this rule would produce an interpretation *OnSurface(\*filling, \*occlusal)*, which is the relation missing from the previous example. Semantically annotated grammar rules

can also connect objects of the same semantic type. For instance, we might define a rule

$NP\langle Condition \rangle \rightarrow NP\langle Condition1 \rangle \text{ "caused by" } NP\langle Condition2 \rangle$   
 $\Rightarrow CausedBy(condition1, condition2)$

This rule can match phrases such as "leakage caused by a crack along the lingual surface", and link the two conditions (leakage and crack) with a *CausedBy* relation. This mechanism enables ONYX to construct complex descriptions with multiple objects.

We have added a mechanism to the ONYX training tool that allows semantically annotated grammar rules to be generated semi-automatically during training. A human annotator with sufficient linguistic background can view the parse trees generated by ONYX for corpus sentences, repair those parse trees and/or add new semantic relations if necessary, then apply a function that creates copies of the rules embodied in those trees with semantic types and predicates attached.

## 5 □ Integrating Syntax and Semantics

Although most NLP systems apply semantic analysis to completed parse trees, in humans the two processes are more integrated. Syntactic expectations are greatly influenced by word meanings, as illustrated by "garden path" sentences such as "The man whistling tunes pianos." In ONYX, syntax and semantics are highly interleaved. This is accomplished in several ways:

1- ONYX's parse algorithm permits words to be processed in any order, rather than strictly left-to-right, since binary grammar links can be added to the phrase chart in any order. This allows ONYX to be instructed to focus on semantically interesting words first, which can be used, among other things, to gather useful information from ungrammatical speech or run-on sentences where attempting to look for complete sentences in strict left-to-right fashion would be unsuccessful.

2- ONYX implements a variation on a probabilistic context free grammar (PCFG) (Charniak, 1997) that associates grammar rules with semantic types. Based on training, a conditional probability is calculated for each  $\langle \text{rule}, \text{type} \rangle$  pair given specific

<rule, type> assignments to the rule's components. The probability of a phrase is then calculated as the product of the probabilities of the phrase rule and its semantic type, given the rule and type of each of its child phrases. ONYX is then able to prefer phrases that best accommodate the semantic types of their constituents. Specifically,

$$\text{prob}(\text{phrase}) = \prod (\text{prob}(\text{rule}(\text{phrase}) + \text{semtype}(\text{phrase}) | \text{rule}(\text{childPhrase}) + \text{semtype}(\text{childPhrase})))$$

3- One hard problem in parsing is determining the correct structure of conjunctive noun phrases. ONYX applies semantic guidance to solve this problem. For instance, in a chest radiology report the words "right and left lower lobe opacity" can be grouped in several different ways, and different groupings can produce different interpretations. The correct grouping should be something like: [[[right and left] [lower lobe]] opacity], rather than [[right and [left lower]] [lobe opacity]]. ONYX currently employs a simplistic representation of the meaning of a conjunctive phrase as a list of interpretations. The correct interpretations for "right and left lower lobe opacity" would be two predicate expressions covering the words (right, lower, lobe, opacity) and (left, lower, lobe, opacity). ONYX generates a measure of the similarity of these expressions based on the cosine similarity of the lists of non-null nodes in their CM templates. This measure is factored into the phrase's goodness score under the heuristic that semantically balanced conjunctive phrases are more likely to be correct than imbalanced ones.

4- As mentioned earlier, ONYX can utilize grammar rules annotated with semantic types and patterns. Semantically annotated rules constrain phrases to match particular semantic types, and can contribute predicates to the interpretation of those phrases. This gives ONYX's grammar the character of a semantic grammar.

5- Phrases are weighted and preferred by ONYX according to their goodness score, which is based on three measures: the probability of the phrase as determined by the PCFG formula, the conjunct cosine similarity score, if applicable, and the goodness score of the phrase's semantic interpretation. The PCFG and conjunct similarity formulas

are based on semantic criteria, as mentioned earlier. Interpretation goodness scores are calculated as a simple product of the probabilities of the semantic relation predicates they contain. Relation probabilities are in turn derived from training data, and are conditioned on the concepts they contain. The probability of a relation is calculated as the number of times a pair of concepts appears together in the target relation divided by the number of times they appear together in any set of relations. The goodness score of a phrase is thus highly semantically determined.

$$\begin{aligned} \text{goodness}(\text{phrase}) &= F(\text{prob}(\text{phrase}, \text{PCFG}), \\ &\quad \text{conjunctSimilarity}(\text{phrase}), \\ &\quad \text{goodness}(\text{interp}(\text{phrase}))) \\ \text{goodness}(\text{interp}(\text{phrase})) &= \\ &\quad \prod \text{prob}(\text{relations}(\text{interp}(\text{phrase}))) \\ \text{prob}(\text{relation}) &= \\ &\quad \text{count}(\text{relation} + \text{concepts}(\text{relation})) / \\ &\quad \text{count}(\text{anyConnection}(\text{concepts}(\text{relation}))) \end{aligned}$$

## 6 Evaluation

We performed a preliminary evaluation of ONYX for the extraction of relevant dental concepts and relations on a set of twelve documents in our current training corpus.

**Reference Standard.** Each document was independently annotated by three human annotators (authors LC, JI and HH), who used the ONYX training tool to fill in templates representing dental conditions, tooth locations and other relevant concepts, as well as to select the semantic relations linking those templates. The annotators then reviewed disagreements and by consensus created a reference standard set of templates and relations. Where the annotators did not have sufficient dental knowledge to reach an agreement they consulted dental clinicians.

**Outcome Metrics.** To evaluate ONYX on the relatively small corpus of documents, we applied a leave-one-out approach: for each sentence in the reference standard, ONYX was trained using the templates from the remaining reference standard sentences. ONYX was then applied to the target sentence, and the resulting templates and relations were compared to the reference standard. We measured inter-annotator agreement (IAA) between ONYX and the reference standard using the formula described in Roberts et al (2007):

$$IAA = (2 * \text{correct}) / (\text{spurious} + \text{missing} + \text{correct})$$

We calculated IAA separately for CM words, concepts, and semantic relations. A correct match is a word, concept or relation generated by both the reference standard and ONYX; a spurious item is one ONYX generated that did not exist in the reference standard; and a missing item is one that existed in the reference standard but was not generated by ONYX. In addition to IAA we identified the concepts and relations most commonly in error and calculated percentages for those errors.

We compared ONYX’s performance on the target documents with that of a simple baseline parser we created for this purpose. The baseline parser processes the words of a sentence from left to right, creating phrases for sets of juxtaposed words that can be interpreted together using the semantic network. No grammar rules are employed, there is no analysis of conjunctive phrases, and goodness scores are not calculated. Our goal was to get a feel for how much these factors contribute to generating correct interpretations. There is no precedence for this particular approach as far as we are aware, so we regard this comparison as informative but not definitive.

## 7 Results

IAA results for ONYX and the baseline parser are shown in table 2. ONYX performs best at inserting words into appropriate nodes in the CMs, with IAA of 86%, and less well for inferring the best concept (80%) and identifying relations among concepts (76%). ONYX consistently out-performs the baseline parser.

Table 2: IAA for assignment of words, concepts, and relations.

		IAA
Words (n = 904)	ONYX	86%
	Baseline	57%
Concepts (n = 1186)	ONYX	80%
	Baseline	53%
Relations (n = 297)	ONYX	76%
	Baseline	41%

Although this study does not examine all the reasons for the differences in performance between ONYX and the baseline parser, some reasons can

be illustrated with an example. Conjunctive phrases are common in dental discourse, and a failure to handle conjuncts can result in both concept and relation errors. For instance, given the sentence "4, 5, 6, 7 fine" ONYX generates separate interpretations covering the word groupings (4, fine), (5, fine), (6, fine), and (7, fine), which would yield four ConditionAt relations, four Location concepts (\*numberFour, \*numberFive, \*numberSix, \*numberSeven) and one Condition concept (\*normalTooth) appearing in each relation. The baseline parser in contrast does not discover this distribution of terms and so omits all but the ConditionAt relation over (7, fine). Trying to merge juxtaposed tooth numbers, the baseline parser also infers that at least some of these denote tooth ranges instead of individual teeth (e.g. interpreting "4, 5" as "4 to 5" instead of "4 and 5"), which causes it to misclassify Location concepts. The ability to generate correct parse trees and to use the structure of those parse trees in the interpretation process is important in generating correct interpretations.

Tables 3 and 4 show breakdowns by percentage of the concepts and relations most commonly in error in ONYX’s interpretations (errors accounting for more than 15%).

Table 3: Per-concept error percentages

Dental Condition Summary Concept	18%
Tooth Location Summary Concept	17%
Dental Condition Intermediate Concept	16%
Surface Summary Concept	15%
Total	66%

Table 4: Per-relation error percentages.

Surface of Part	47%
Location of Condition	23%
Total	70%

## 8 Related Work

ONYX is a new application inspired by SPRUS (Ranum, 1989), Symtext (Koehler, 1998), and MPLUS (Christensen, 2002), which all used Bayesian Networks to infer relevant findings from text. Other medical language processing systems implement different approaches to encode clinical concepts and their modifiers, along with relations between concepts, including MedLEE (Friedman,

1994), a largely statistical system by Taira and colleagues (Taira, 2007), and MedSyndikate (Hahn, 2002).

Many of ONYX's components leverage research in the general and clinical NLP domains, including the use of chart parsing (Kay, 1980) and probabilistic context free grammars (Charniak, 1997). ONYX's use of semantically annotated grammar rules was inspired in part by MedLEE (Friedman et al, 1994), which uses a semantic grammar.

Although incorporating ideas and approaches from others, we feel that ONYX is unique in several ways, including its high level of syntactic/semantic integration and the ways in which it blends symbolic and probabilistic representations of domain knowledge. We plan to make ONYX available through open source when the system is more complete.

## 9 Limitations

There are several limitations to this study. Although ONYX introduces several innovations, these are not described in detail in this study and are not individually evaluated for their effect on ONYX's performance. Instead, this study presents a broad overview of ONYX and evaluates ONYX's overall performance against a reference standard on a small test sample. Another limitation of our study is the baseline system—because similar systems generate different output than ONYX and do not model the same domain, finding a competitive baseline application is difficult. In spite of its imperfection, we believe the baseline we implemented to be reasonable.

## 10 Future Work

One limitation of a system like ONYX is the overhead of manually creating complex training cases. To address this shortcoming, the ONYX training tool invokes ONYX to automatically create templates and relations for corpus sentences, and human trainers correct any mistakes. A semi-automated approach greatly speeds up the training process and facilitates agreement among human trainers. We plan to further automate this process using an approach derived from Thelen & Riloff (2002), which uses a classifier with features based

on extraction patterns derived from Autoslog (Riloff, 1996). We plan to adapt this approach to automatically classify CM word assignments, and also to automatically classify semantic relations between CM templates. We will add this functionality to the training tool to enable it to find and annotate relevant sentences automatically where possible. We will also apply this functionality to enable ONYX to recognize relevant sentences in new documents based on their similarity to training sentences, and we will use semantic patterns stored with training sentences to aid in interpreting noisy segments of text that ONYX cannot parse. We plan to compare the performance of grammar-based and feature-based semantic analysis in future studies. With more fully automated training, we also hope to make ONYX more easily portable to new domains and clinical settings in the future.

## Conclusions

This paper describes ONYX, which is being developed as part of a system for extracting chartable findings from spoken dental examinations. ONYX contains a number of innovative ideas including a novel adaptation of Kay's (1980) parse algorithm; a symbolic language extended to include probabilistic and procedural elements; an integration of syntax and semantics that includes a semantically weighted probabilistic context free grammar and interpretation based both on a semantic network and a semantic grammar. Considering ONYX's early stage of development it performed reasonably well in this limited evaluation but must be extended to address challenges in extracting findings from spoken dental exams.

## Acknowledgments

This work was funded by NIDCR 1 R21DE018158-01A1 "Feasibility of a Natural Language Processing-based Dental Charting Application.

## References

- E. Charniak. 1997. *Statistical parsing with a context-free grammar and word statistics*. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*, pp. 598-603.
- E. Charniak, S. Goldwater and M. Johnson. 1998. *Edge-Based Best-First Chart Parsing*. In *Proceedings of*



- the Sixth Workshop on Very Large Corpora*, pp. 127-133.
- Lee M. Christensen, Peter J. Haug, and Marcelo Fiszman. 2002. MPLUS: A Probabilistic Medical Language Understanding System. *Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain, Philadelphia*, pp. 29 – 36.
- Carol Friedman, Phil Alderson, John Austin, James Cimino, & Stephen Johnson. 1994. A general natural language text processor for clinical radiology. *Journal of American Medical Informatics Association* 1(2), pp. 161–174.
- M. R. Genesereth and R. E. Fikes. Knowledge Interchange Format, Version 3.0 Reference Manual. Technical Report Logic-92-1, Stanford, CA, USA, 1992.
- Hahn U, Romacker M, Schulz S. 2002. Medsyndikate-a natural language system for the extraction of medical information from findings reports. *Int J Med Inf.* 67(1-3), pp. 63-74.
- M. Kay. 1980. Algorithm schemata and data structures in syntactic parsing. In *Readings in Natural Language Processing*, pp. 35 – 70. Morgan Kaufmann Publishers Inc.
- Koehler, S. B. 1998. SymText: A natural language understanding system for encoding free text medical data. Ph.D. Dissertation, University of Utah.
- Ranum D.L. 1989. Knowledge-based understanding of radiology text. *Comput Methods ProBiomed.* Oct-Nov;30(2-3) pp. 209-215.
- Ellen Riloff, 1996. Automatically Generating Extraction Patterns from Untagged Text. *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pp. 1044 – 1049. The AAAI Press/MIT Press.
- Angus Roberts, Robert Gaizauskas, Mark Hepple, Neil Davis, George Demetriou, Yikun Guo, Jay Kola, Ian Roberts, Andrea Setzer, Archana Tapuria, Bill Wheelidin. 2007. The CLEF Corpus: Semantic Annotation of Clinical Text. *AMIA 2007*, pp. 625 – 629.
- Taira R, Bashyam V, Kangarloo H. 2007. *A field theory approach to medical natural language processing.* *IEEE Transactions in Inform Techn in Biomedicine* 11(2).
- Michael Thelen and Ellen Riloff. 2002. *A Bootstrapping Method for Learning Semantic Lexicons using Extraction Pattern Contexts.* *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, pp. 214 – 221.