Research Paper ■

Experiments in Concept Modeling for Radiographic Image Reports

DOUGLAS S. BELL, MD, EDWARD PATTISON-GORDON, MS, ROBERT A. GREENES, MD, PhD

Abstract **Objective:** Development of methods for building concept models to support structured data entry and image retrieval in chest radiography.

Design: An organizing model for chest-radiographic reporting was built by analyzing manually a set of natural-language chest-radiograph reports. During model building, clinician-informaticians judged alternative conceptual structures according to four criteria: content of clinically relevant detail, provision for semantic constraints, provision for canonical forms, and simplicity. The organizing model was applied in representing three sample reports in their entirety. To explore the potential for automatic model discovery, the representation of one sample report was compared with the noun phrases derived from the same report by the CLARIT natural-language processing system.

Results: The organizing model for chest-radiographic reporting consists of 62 concept types and 17 relations, arranged in an inheritance network. The broadest types in the model include FINDING, ANATOMIC LOCUS, PROCEDURE, ATTRIBUTE, and STATUS. Diagnoses are modeled as a subtype of FINDING. Representing three sample reports in their entirety added 79 narrower concept types. Some CLARIT noun phrases suggested valid associations among subtypes of FINDING, STATUS, and ANATOMIC LOCUS.

Conclusions: A manual modeling process utilizing explicitly stated criteria for making modeling decisions produced an organizing model that showed consistency in early testing. A combination of top-down and bottom-up modeling was required. Natural-language processing may inform model building, but algorithms that would replace manual modeling were not discovered. Further progress in modeling will require methods for objective model evaluation and tools for formalizing the model-building process.

■ J Am Med Informatics Assoc. 1994;1:249-262.

An impediment to achieving a comprehensive computer-based patient record is the difficulty of representing clinicians' observations.¹ For observational data to be used across computer applications for decision support, quality assurance, abstraction for research, and other tasks, the *meaning* of recorded observations must be interpretable by computer. Two possible strategies for acquiring computer-interpretable observational data are "structured data entry," in which

clinicians enter observations by selecting concepts directly from a controlled vocabulary, and "natural-language processing," in which clinicians enter observations using natural language and their words are transformed into a controlled-vocabulary representation by an algorithm. Although these two strategies present very different user interfaces, both depend on an underlying model of unambiguous medical concepts that can be shared among applications using patient-record data. For example, to handle observations of lung nodules on chest x-rays, both natural-language processing and structured data-entry systems can make use of a model of chest radiography that gives the subtypes and attributes of lung nodules.

Sharable concept models are difficult to develop.² All models are abstractions of the real world that omit details in order to focus on the aspects of reality important to a given task.³ Models developed for different tasks therefore omit different kinds of de-

Affiliation of the authors: Harvard Medical School, Brigham and Women's Hospital, Boston, MA.

Supported in part by National Library of Medicine grants LM04572 and LM07092, and contract N01-LM-3539.

Correspondence and reprints: Douglas S. Bell, MD, Decision Systems Group, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115.

Received for publication: 11/16/93; accepted for publication: 1/14/94.

tails. Building a model that is adequate for more than one task will require a clear articulation of the demands that each task places upon the model. Determining the modeling needs of an individual task is an iterative process, involving both "top-down" and "bottom-up" activity.⁴ Each iteration requires insight into the conceptual structures that are important to the given task (top-down) in conjunction with validation and testing against the needs of applications that perform the task (bottom-up). The result of each iteration is a working model, documented, wherever possible, by the reasons that individual conceptual structures meet the demands of the task.

This paper presents the methods and results of an initial manual modeling iteration aimed at meeting the needs of reporting and retrieving radiologists' interpretations of chest radiographs. We chose to focus on the limited domain of chest radiography in order to move quickly toward functioning applications that could show practical use of concept modeling.

Background

Structured Data Entry

The idea that clinicians' observations could be entered into the medical record by direct selection from a controlled, structured concept set is not new. Early structured data-entry systems were implemented in a hypertension clinic,⁵ a radiology department,^{6,7} and a hospital general-medical ward.⁸ These systems did not have lasting success. Recent efforts have been directed at "predictive data entry" based on deeper models of the domain of discourse,⁹ at taking full advantage of a graphical user interface,¹⁰⁻¹⁵ and at leveraging the contribution of nonmedical personnel.¹⁰ A recent endoscopy reporting system has shown that structured data entry typically captures more complete data than does natural-language dictation.¹⁶ Structured data-entry systems have generally been more successful in narrower knowledge domains. One factor limiting the generalizability of structured data-entry systems is the lack of a widely accepted standard model of clinical findings. Although it might be argued that structured data entry should be based upon currently available controlled medical vocabularies,¹⁷ no currently available controlled vocabulary meets the criteria for a general system of medical semantics.¹⁸

Concept Models of Medical Information

Worldwide interest is now focusing on the development of models that represent valid relationships

among concepts more deeply than do traditional controlled vocabularies.¹⁹⁻²⁷ Instead of representing a term such as "calcified right lung nodule" with a single code, a deeper model might represent an atomic concept, "nodule," along with relationships that allow it to be linked to the modifying concept "calcified" and the location concept "lung," which in turn can be modified by the laterality "right."

Most recent modeling efforts attempt to restrict the kinds of knowledge represented in order to avoid computational intractability. It has been shown, however, that an aggressively restricted representation was not sufficient to represent knowledge for reasoning about acid-base and electrolyte disorders.²⁸ Rector defines an intermediately restricted *conceptual knowledge* model as one that represents only the valid combinations of concepts that form medically sensible expressions. This kind of model is held in contrast to a *general inferential knowledge* model, which would also represent associations among specific findings, diseases, and treatments.²⁹ The GALEN knowledge base is an example of a conceptual knowledge model.²² It contains the knowledge that a bone is a valid location for a fracture, fractures may be caused by pathologic conditions, a femur is a kind of bone, and osteoporosis is a kind of pathologic condition, so that it makes sense to talk about a fracture of the femur caused by osteoporosis. The DXplain³⁰ knowledge base, on the other hand, is an example of a general inferential knowledge model. It contains a list of specific diseases that could cause a long-bone fracture, including osteoporosis, Paget's disease, and Albright's Hereditary Osteodystrophy. The term *concept model* has recently been used by Rector in place of the term *conceptual-knowledge model*.²²

The Findings-Diagnosis Continuum

Recent work in our group has focused on a broad conceptual framework for representing image descriptions.³¹ In our view, no sharp distinction exists between findings and diagnoses. Instead, findings exist in a continuum from the most direct observations to highly interpretive diagnoses. All findings are essentially interpretations of patterns of raw perceptions. Some findings are composite, however, in that they can be asserted as interpretations of other findings. In chest radiography, for example, "congestive heart failure" is a composite finding because it could be asserted as an interpretation of the more low-level findings "pleural effusion," "increased interstitial markings," and "enlarged heart." "Pleural effusion," in turn, could be asserted as an interpretation of the elemental finding "blunted costophrenic angle." The latter finding is elemental in that it would

only be asserted directly, not as an interpretation of another finding.

In a prior modeling experiment, the findings–diagnosis continuum model was able to represent the knowledge of a real reporting application. Eight broad concept types and eight relations from the findings–diagnosis continuum model³¹ were used as the base to build a model for pelvic-ultrasound reporting.²⁶ This core model included the concept types FINDING, COMPOSITE FINDING, ELEMENTAL FINDING, and ETIOLOGIC DIAGNOSIS. These concept types are called “broad” because they subsume most of the other concept types in the model and they lay out the basic relations by which concepts may be linked. The domain content for this prior experiment was taken from the standardized terms used by UltraSTAR, a structured reporting application that is now in routine use in the Ultrasound Section of the Radiology Department, Brigham and Women’s Hospital.¹⁰ The resulting pelvic-ultrasound model contained 103 concept types and ten relations. A structured data-entry application that would directly use this model remains under development, the original UltraSTAR application having been built without the use of a formal conceptual taxonomy.

Automated Model Discovery

Computer-based analysis of natural language holds the promise of deriving concept models empirically from the sublanguage that clinicians actually use in a particular medical domain.^{32–34} Tools for sublanguage analysis are not widely available, however.

The CLARIT system makes selective use of natural-language processing techniques, and has been applied primarily to automatic indexing of documents.^{35,36} CLARIT can process a large corpus of text from a particular knowledge domain into a “first-order thesaurus”—a list of terms in the domain ranked by a measure of their importance in the domain. CLARIT thesaurus generation involves two major steps. The first step extracts noun phrases from the source text using heuristic natural-language processing and a large lexicon. Noun phrases are “morphologically normalized,” meaning that lexical variants are mapped to a single form. To illustrate, from the sentence “There are a few scattered small nodular opacities that are not clearly vascular in origin, possibly representing granulomata” CLARIT extracts the noun phrases “scattered small nodular opacity,” “origin,” and “granulomata.”

CLARIT’s second step scores all normalized noun phrases based on several statistics, including their

distribution within the domain corpus and their uniqueness compared with general English-language terms. The highest-scoring terms are selected out, forming a first-order thesaurus. CLARIT, as it is currently available, does not attempt to build a “second-order thesaurus,” which would represent the valid relationships among terms.

Representing Concept Models

Cimino has described requirements for medical-concept representations.¹⁸ Sowa’s conceptual-graph formalism³⁷ is a representation system that can support these requirements. We use a simplification of Sowa’s formalism. Its basic units are concept types, concept instances, and relations.

Concept types are arranged in a subsumption hierarchy of supertypes and subtypes (sometimes called an “is-a” hierarchy). The fact that LEFT LUNG is a subtype of LUNG may be asserted in conceptual-graph notation using the “<” symbol:

LEFT LUNG < LUNG.

Instances of concept types are represented in conceptual-graph notation by the type name enclosed in brackets. Concept instances may be *individual* or *generic*. [LUNG: #2021] is an *individual* instance of the concept type LUNG in some database. The number “#2021,” called an *individual marker*, gives the concept instance a unique identity among other instances of LUNG in the database. *Generic* concept instances refer to some unspecified instance of a concept type. These are noted by omitting the individual marker. Thus [LUNG] is a *generic* instance of the concept type LUNG; it may be read as meaning “some lung.”

Concept instances may be linked by relations. Each relation is directional and can only link concept instances from a restricted set of types. Relations are represented in conceptual-graph notation by the relation name enclosed in parentheses. *Canonical graphs* connect generic concepts with relations, to show which concept types may be validly linked by a relation. Thus the canonical graph

[LUNG] —> (HAS PART) —> [UPPER LOBE].

indicates that an upper lobe is a concept that may be sensibly expressed as a part of a lung. Relations in canonical graphs are inherited along the type hierarchy. Thus [LEFT LUNG] inherits the (HAS PART) relation to [UPPER LOBE]. We say that canonical graphs set out “semantic constraints” because concepts that are not linked in a canonical graph may not be sen-

sibly combined in a single expression. Notice that these conceptual graphs use only standard American Standard Code for Information Interchange (ASCII) characters, so they may be transmitted easily by electronic mail.

The conceptual-graph formalism can represent both a concept model for patient data and the actual patient data.^{20,27} It has also been shown that a semantic-network concept model¹⁹ can be represented separately from a patient database that stores references into the model.^{38,39}

Methods

Chest-radiograph Report Corpus

Twenty megabytes of natural-language reports had been collected previously from four hospitals; 55,411 unique, normalized noun phrases had been parsed from this corpus by the CLARIT^{35,36} selective natural-language processing system. These noun phrases had been further scored, ranked, and filtered by the CLARIT system, creating a first-order thesaurus of 21,108 candidate terms.*

We selected a sample of 18 reports from this corpus for manual examination. Reports were selected based on their being intermediate in length (containing 15 to 25 noun phrases parsed by CLARIT) and on their containing interesting findings and difficult-to-model concept types such as enumeration and temporal comparisons. From these 18 reports, we selected the first three reports for complete sentence-by-sentence representation using our model.

Manual Modeling of Chest-radiograph Reports

A semantic network for chest-radiograph reporting was built in group meetings involving two to four medical informaticians. The group always included at least one clinician. A structure of broad, organizing concept types was created by abstracting patterns from statements in the sample chest-x-ray reports and by applying clinical knowledge. When appropriate, concept types and relations were taken from our own previous modeling work. Twelve concept types and ten relations were taken from a previously

developed model of the "findings–diagnosis continuum."³¹ An additional 15 concept types and three relations, primarily modeling general principles of anatomy, were taken from a previously developed model for pelvic-ultrasound reporting.²⁶ (The four concept types taken from reference 31 that were not included in reference 26, as described in the Background section, modeled features that were specific to describing plain radiographs, and had thus been excluded from the pelvic-ultrasound model.) To test the organizing model, three reports were manually represented sentence by sentence.

The actions of model building included creating new concept types and relations, and renaming or removing concept types when an alternative representation was found to be preferred for content that was already represented. Alternative representations were judged by four qualitative criteria: 1) the degree to which a report generated from the representation would communicate the clinically relevant information in the source text, 2) the degree to which the model would provide important semantic constraints (potentially facilitating structured data entry⁴¹), 3) the degree to which a single form would exist in the model for expressing a single meaning (a "canonical form," potentially facilitating image retrieval), and 4) the simplicity of the model as a whole. CLARIT noun-phase parses of the sample sentences and the CLARIT thesaurus were both available during these meetings. Concept types were named using words from the corresponding CLARIT noun phrases when possible.

We kept track of our model using a combination of our own semantic-network software (Thenetsys⁴²), ASCII files of conceptual graphs, and graphic drawings.

Exploration of Automated Model Discovery

To begin exploring how natural-language processing might automate model building, the noun phrases parsed by CLARIT for each sentence in one sample report were compared with the representations of the same sentences in the manually built model. When a CLARIT noun phrase corresponded with more than one concept instance in the model representation, the words within the noun phrase were examined for patterns of concept types that tend to co-occur within noun phrases.

Results

Manually Built Semantic Model

The organizing model resulting from our model-building process consisted of 62 concept types and

*This work was generously contributed by Professor David A. Evans, Director of the Laboratory for Computational Linguistics, Carnegie Mellon University, Pittsburgh, Pennsylvania. The work is part of a project to compare modeling approaches among a number of collaborating groups.⁴⁰ The four hospitals that contributed reports were Brigham and Women's Hospital, Columbia-Presbyterian Hospital, LDS Hospital, and Oregon Health Sciences University Hospital.

17 relations. Most concept types are subtypes of the broad concept types FINDING, ANATOMIC LOCUS, PROCEDURE, ATTRIBUTE, STATUS, and QUALIFIER. Our modeling of each of these broad types is explained in turn. Using this organizing model to represent three sample reports in their entirety resulted in 79 narrower concept types being added to the model. The majority of these narrower concept types were subtypes of FINDING or ANATOMIC LOCUS, but many new ATTRIBUTES, STATUSES, and QUALIFIERS also needed to be created. Representing the first sample report resulted in many adjustments to the organizing model. The remaining two reports were representable without any changes to the organizing model. The full representation of one sample report is shown in the appendices.

Representation of Findings

The broadest concept types in our model are organized around the type FINDING. A FINDING is any concept that can be independently asserted in the interpretation of a DIAGNOSTIC PROCEDURE. In other words, a FINDING may be sensibly asserted as having been observed without any associated supporting concepts. Thus NODULE is a FINDING, whereas CALCIFIED is not. The most important relations of FINDING

are shown both graphically and in conceptual-graph ASCII notation³⁷ in Figure 1. Any instance of FINDING may be associated with a set of DIAGNOSTIC PROCEDURE subtypes by which the FINDING can be revealed, a set of ANATOMIC LOCUS subtypes at which the FINDING can be located, and a set of possible QUALIFIERS that may further describe it.

The broadest subtypes of FINDING are shown in Figure 2. A COMPOSITE FINDING is a FINDING that may be asserted as an interpretation of a set of lower-level FINDINGS. An ELEMENTAL FINDING such as OPACITY, on the other hand, would never be asserted as a higher-level interpretation. The HAS INTERPRETATION relation provides the link from a lower-level FINDING (either elemental or composite) to a COMPOSITE FINDING. Each assertion of a COMPOSITE FINDING may have an associated CERTAINTY. An example COMPOSITE FINDING is CONGESTIVE HEART FAILURE, which may be asserted as an interpretation of the FINDINGS INCREASED INTERSTITIAL MARKINGS, ENLARGED CARDIAC SILHOUETTE, and PLEURAL EFFUSION (among others) as shown below in conceptual-graph ASCII notation.

```
INCREASED INTERSTITIAL MARKINGS < ELEMENTAL FINDING.
ENLARGED CARDIAC SILHOUETTE    < COMPOSITE FINDING.
```

Figure 1 The broadest concept types and relations in our model. In the graphic representation, generic concepts are shown as labelled rectangles and relations are shown as labelled directional arcs. The conceptual-graph ASCII notation below the figure is explained in the Background section of this paper. For a detailed explanation of this notation, which Sowa calls the conceptual-graph linear form, see reference 37.

PLEURAL EFFUSION < COMPOSITE FINDING.
 CONGESTIVE HEART FAILURE < COMPOSITE FINDING.
 [INCREASED INTERSTITIAL MARKINGS] -
 (HAS INTERPRETATION) -> [CONGESTIVE HEART FAILURE].
 [ENLARGED CARDIAC SILHOUETTE] -
 (HAS INTERPRETATION) -> [CONGESTIVE HEART FAILURE].
 [PLEURAL EFFUSION] -
 (HAS INTERPRETATION) -> [CONGESTIVE HEART FAILURE].

The model also provides for linkage of findings via formal diagnostic *INFERENCE PROCEDURES*. An *INFERENCE PROCEDURE* would provide a Bayesian or Boolean association from lower-level FINDINGS to a COMPOSITE FINDING. *INFERENCE PROCEDURE* is shown in italics because it is not a concept type but rather is a placeholder for interaction between the model and potentially external knowledge modules.

The type ETIOLOGIC DIAGNOSIS allows certain COMPOSITE FINDINGS to be linked to an ETIOLOGIC AGENT. ETIOLOGIC DIAGNOSIS inherits all of the relations of FINDING and COMPOSITE FINDING, so an ETIOLOGIC DIAGNOSIS can have further interpretations or can contribute to further inferences. ETIOLOGIC AGENTS are entities capable of causing disease or resulting in particular findings. ETIOLOGIC AGENTS mentioned in the sample reports include RADIATION THERAPY and LOBECTOMY.

Representation of Qualifiers, Attributes, and Statuses

Findings may be modified by ATTRIBUTE-STATUS pairs via the QUALIFIER concept, as shown in Figure 1, with examples given below. An OPACITY, for example, may be described as PLATELIKE via a SHAPE OBSERVATION (a QUALIFIER), which pairs SHAPE (an ATTRIBUTE) with its valid STATUSES, including PLATELIKE. The same

Figure 2 Subtypes and relations of FINDING. As in Figure 1, generic concepts are shown graphically as rectangles and relations are labelled arcs. In addition, subtype relationships are shown by bold arcs labelled "(subtype)." *Inference Procedure* is not a concept type but rather a placeholder for interaction between the semantic model and potentially external knowledge modules. Note that [ETIOLOGIC DIAGNOSIS] inherits all of the relations of [FINDING] and [COMPOSITE FINDING].

pairing might alternatively have been modeled by a HAS SHAPE relation from OPACITY directly to a set of valid STATUSES, including PLATELIKE. We chose to use the QUALIFIER concept, however, to allow better expression of semantic constraints. QUALIFIERS allow ATTRIBUTES and STATUSES to be paired for each FINDING at exactly the desired level in their respective type hierarchies. QUALIFIER concept types also facilitate preassembly of commonly used ATTRIBUTE-STATUS combinations. An example of a preassembled QUALIFIER is MULTIPLE, which marks a concept as plural without further specifying a quantification. MULTIPLE preassembles QUANTITY (an ATTRIBUTE) with GREATER THAN ONE (a STATUS).

OPACITY < ELEMENTAL FINDING.
 SHAPE OBSERVATION < QUALIFIER.
 SHAPE < ATTRIBUTE.
 IRREGULAR < STATUS.
 PLATELIKE < STATUS.
 ROUND < STATUS.
 MULTIPLE < QUALIFIER.
 QUANTITY < ATTRIBUTE.
 GREATER THAN ONE < STATUS.

[OPACITY] -
 (HAS QUALIFIER) -> [SHAPE OBSERVATION].

[SHAPE OBSERVATION] -
 (HAS ATTRIBUTE) -> [SHAPE]
 (HAS STATUS) -> [PLATELIKE]
 (HAS STATUS) -> [ROUND]
 (HAS STATUS) -> [IRREGULAR].

[MULTIPLE] -
 (HAS ATTRIBUTE) -> [QUANTITY]
 (HAS STATUS) -> [GREATER THAN ONE].

Representation of Anatomic Loci

We have grouped all anatomic concepts that might be referred to in a report under the concept type ANATOMIC LOCUS. Subtypes and relations of ANATOMIC LOCUS are shown below. BILATERAL ANATOMIC LOCUS allows paired anatomic loci to be referenced individually. Any ANATOMIC LOCUS can be divided into somewhat indistinct SPATIAL SUBREGIONS based on one of the body's axes. Shown are regionalizations along left-to-right and superior-to-inferior axes. We reserved the HAS PART relation for sharply distinct ANATOMIC COMPONENTS, such as the lobes of the lung.

BILATERAL ANATOMIC LOCUS < ANATOMIC LOCUS.
 CHEST < ANATOMIC LOCUS.
 SPATIAL SUBREGION < ANATOMIC LOCUS.
 ANATOMIC COMPONENT < ANATOMIC LOCUS.
 LUNG < BILATERAL ANATOMIC LOCUS.

LEFT_RIGHT AXIS REGION < SPATIAL SUBREGION.
 SUPERIOR_INFERIOR AXIS REGION < SPATIAL SUBREGION.
 LUNG FIELD < SUPERIOR_INFERIOR AXIS REGION.
 LUNG ZONE < SUPERIOR_INFERIOR AXIS REGION.
 SUPERIOR < SUPERIOR_INFERIOR AXIS POSITION.
 MIDDLE < SUPERIOR_INFERIOR AXIS POSITION.
 INFERIOR < SUPERIOR_INFERIOR AXIS POSITION.
 LOBE < ANATOMIC COMPONENT.
 UPPER LOBE < LOBE.
 MIDDLE LOBE < LOBE.
 LOWER LOBE < LOBE.
 LEFT < LATERALITY.
 RIGHT < LATERALITY.
 BILATERAL < LATERALITY.

[ANATOMIC LOCUS] -
 (HAS SUBREGION) -> [SPATIAL SUBREGION]
 (HAS PART) -> [ANATOMIC COMPONENT].

[BILATERAL ANATOMIC LOCUS] -
 (OF LATERALITY) -> [LATERALITY].

[LEFT_RIGHT AXIS REGION] -
 (OF LATERALITY) -> [LATERALITY].

[SUPERIOR_INFERIOR AXIS REGION] -
 (HAS POSITION) -> [SUPERIOR_INFERIOR AXIS POSITION].

[LUNG] -
 (HAS SUBREGION) -> [LUNG ZONE].
 (HAS PART) -> [LOWER LOBE].

Representation of Procedures

A PROCEDURE is a TECHNIQUE applied to a particular ANATOMIC LOCUS. The relations and subtypes of PROCEDURE are shown below. A DIAGNOSTIC PROCEDURE applies a TECHNIQUE to answer questions regarding a patient's condition, while an INTERVENTION is aimed at improving a patient's condition. Example PROCEDURES from the sample report include CHEST X_RAY and LOBECTOMY.

DIAGNOSTIC PROCEDURE < PROCEDURE.
 INTERVENTION < PROCEDURE.
 PLAIN RADIOGRAPH < DIAGNOSTIC PROCEDURE.
 CHEST X_RAY < PLAIN RADIOGRAPH.
 LOBECTOMY < INTERVENTION.
 EVALUATION TECHNIQUE < TECHNIQUE.
 PLAIN RADIOGRAPHY < EVALUATION TECHNIQUE.
 AP < X_RAY DIRECTION.
 PA < X_RAY DIRECTION.
 LATERAL < X_RAY DIRECTION.
 SUPINE < PATIENT POSITION.
 UPRIGHT < PATIENT POSITION.

[PROCEDURE] -
 (HAS TECHNIQUE) -> [TECHNIQUE]
 (APPLIED TO) -> [ANATOMIC LOCUS].

[DIAGNOSTIC PROCEDURE] -
 (HAS TECHNIQUE) -> [EVALUATION TECHNIQUE]
 (COMPARED WITH) -> [DIAGNOSTIC PROCEDURE].

[PLAIN RADIOGRAPH] -
 (HAS TECHNIQUE) -> [PLAIN RADIOGRAPHY].

[PLAIN RADIOGRAPHY] -
 (HAS X-RAY DIRECTION) -> [X-RAY DIRECTION]
 (HAS PATIENT POSITION) -> [PATIENT POSITION].

Exploration of Automated Model Discovery

Appendix B shows a sentence-by-sentence comparison of the results of CLARIT noun-phrase parsing with our model's representation. No one-to-one correspondence exists between CLARIT noun phrases and the concept types used from our model. Manual modeling produced 76 concept instances for this sample report, while CLARIT produced 24 noun phrases. Although 22 of these 24 noun phrases had at least partial correspondence with one or more concept instances in the manually built model of the same sentence, no algorithm is apparent that would transform these noun phrases into representations in our model. We do note the following correspondences, however. Four noun phrases matched a concept exactly. Three of these exact matches were FINDINGS (SURGICAL CLIP, ATELECTASIS, and POST OPERATIVE CHANGES) and one was an INTERVENTION (CORONARY ARTERY BYPASS GRAFT). Three other noun phrases contained words matching one or more STATUSES followed by words matching a FINDING ("new surgical clip," "intact sternotomy wire," and "new plate like opacity"). Two noun phrases contained words matching anatomic locations followed by words matching a FINDING ("left pleural effusion" and "left low lobe atelectasis").

Twenty-six concept instances in the representation of these sentences do not correspond to words in either the CLARIT noun phrases or the original text. An example is RADIATION REACTION (an ETIOLOGIC DIAGNOSIS) in the representation of Sentence 4 in Appendix B. This concept has to be represented to connect an observed REGIONAL INCREASED OPACITY (an ELEMENTAL FINDING) to its stated possible etiology, RADIATION THERAPY. Thus, the diagnosis is only implied in the original text by the mention of an etiologic agent. Three ANATOMIC LOCUS concept instances that we represented were only implied in the text. All of the ATTRIBUTES and QUALIFIERS in our model were left implied in the text. We found it necessary to explicitly represent these implied concepts to achieve a complete and connected representation for the sentence. Complete, connected representations using the

most atomic concepts possible are necessary to meet the criterion of having a single form for a single meaning (a canonical form).

Discussion

Concept modeling is central to sharing the meaning of information among computer applications and even within an individual application. The shortcomings of available controlled medical vocabularies as systems of meaning have been described.^{18,43} Increasing effort is being devoted to developing semantic-network models of medical subdomains and controlled vocabularies.¹⁹⁻²⁷ This report differs from most prior reports of modeling efforts in that the process by which the model was built is described in detail and the broadest organizing concept types in the model are described comprehensively.

The Process of Manual Modeling for Radiologic Reporting

Building semantic models rigorously is difficult. Our modeling process was aided by combining top-down design by domain experts with bottom-up abstraction from examples of domain content. Nonetheless, we considered many alternative structures and made many revisions to the model. Our experience would indicate that two groups modeling the same knowledge domain are likely to build quite different models unless they share both methods for model building and criteria for judging their model. We used four explicitly stated criteria to decide among alternative representations. These criteria have so far been applied only subjectively, but they increased our confidence in the resulting model.

We encountered many of the same difficulties that Masarie et al. encountered when building a model that mapped cardiopulmonary findings among the controlled vocabularies of three existing expert systems.⁴⁴ Their model, mapping a total of 522 terms from the expert systems, took two years to develop despite the assistance of "automatic instantiator" software. Challenges they identified included a labor-intensive modeling process and difficulty achieving consensus among the authors regarding the definitions and most intuitive names for concepts in their system. Although we encountered the same difficulties, we found that the modeling process became easier as the model grew and our confidence in the process increased. After representing one sample report, extending the model based on two additional sample reports did not require any changes to the broader, organizing concept types and relations.

Robustness of the Current Model

Our modeling started with a previously developed core model of a "findings–diagnosis continuum." Although the remainder of the organizing representation was altered somewhat during modeling, this core model changed very little. Extensive validation of this model is not yet possible because applications that will directly use it to perform structured reporting and image indexing remain under development. Completion of these applications was not possible, however, without an initial model.

Our experience suggests that modeling radiologists' interpretations from low-level to higher-level findings is important to provide both semantic constraints and canonical forms. Radiologists often omit from their reports descriptions of the low-level findings from which they inferred higher-level findings. They may, for example, simply report "congestive heart failure" rather than list the lower-level findings that they observed, such as "blunted costophrenic angles," "upper zone vascular redistribution," and "increased interstitial markings." Conversely, they may list the lower-level findings they see without mentioning the higher-level findings that should be inferred from them. Clinicians will probably always leave some important concepts unstated as long as they can assume that the readers of their reports have medical knowledge and can perform seemingly obvious inferences. Abstraction and retrieval of radiographic results may be enhanced if systems are able to suggest or fill in these often-unstated concepts. As an example, an epidemiologist may need to retrieve all patients in a particular patient database whose chest radiographs showed a pleural effusion. Some reports, however, may mention both "blunted costophrenic angles" and "congestive heart failure" but may not mention "pleural effusion." A system that is able to suggest or fill in the unstated intermediate interpretation, either at the time that reports are generated or at the time that queries are executed, would increase the completeness of retrieval. The knowledge required for this, however, would go beyond a pure "concept model" as defined by Rector,²² to include valid associations among *specific* lower- and higher-level findings.

Exploration of Automated Model Discovery

Our modeling process relied on manual examination of natural-language reports for guidance. Automated natural-language processing holds the promise of at least assisting the expert modeler. To examine how representations in our model might have been automatically generated from source noun phrases, we

looked for patterns of correspondence between concept instances in our representation of sample sentences and noun phrases parsed by the CLARIT system from the same sentences.

Many noun phrases corresponded loosely to one or more concept instances in our representation, but few noun phrases matched a concept instance completely. Some noun phrases specifically paired instances of FINDING and STATUS, or FINDING and ANATOMIC LOCUS. Further work with natural-language processing tools may therefore enhance "bottom-up" modeling by nominating specific associations between findings and modifiers. Parsing algorithms that yield more complex noun phrases (connecting noun phrases across prepositions, for example) might show richer concept associations than did the simple noun phrases used in this study. More complex noun phrases, however, could also be expected to increase the number of false-positive associations that would need to be examined. At this stage it seems unlikely that sublanguage analysis based on noun-phrase parsing could replace the "top-down" contribution of domain experts.

Next Steps in Modeling

Development of an organizing model using explicitly stated criteria for making modeling decisions has made the extension of our model increasingly fast and consistent, with improving agreement about the meaning and purpose of individual concepts. One avenue of future work would be to continue extending the model in chest radiography, using our current process and new sample reports, until no new concepts are being added to the model.

Our judgment, however, is that the most important next step in model building is to convert our criteria into more objective methods of model evaluation. Communication of clinically relevant information could be tested by comparing a natural-language report with text generated from a model-based representation of the same report. Testing the quality of semantic constraints in our model will require a prototype structured data-entry application that uses a semantic model. Testing whether a single expression exists for a single meaning could be performed manually by asking two different modelers to represent the same report independently. A model-based query and retrieval system would also be necessary to test the model's support for practical image indexing. Simplicity of the model (the least important criterion) might be measured by the ratio of organizing concept types to the more specific concept types that are actually referred to in reports.

A model authoring tool would enhance the reproducibility of model building and the maintainability of model content. Minimum requirements of such a tool would include 1) browsing of concepts and their relations; 2) adding and editing concepts; 3) input and output of a standard interchange format for medical-concept representation, which currently could be conceptual graphs; and 4) maintaining of documentation of modeling decisions. It should also assist an author in objectively evaluating a proposed change to the model. Authoring tools will be particularly important as collaboration in modeling becomes wider and attempts are made to extend models across task domains. Authoring tools would also facilitate using the same modeling process in a different knowledge domain, such as clinic notes. Any model that is sharable across more than one task domain will be a significant achievement. We are taking the first steps toward that goal.

Conclusion

We report a manual modeling process that combines top-down insight into the structure of concepts in a subdomain with bottom-up analysis of natural-language reports. Modeling decisions that might have been made arbitrarily were instead guided by explicitly stated criteria. Applied to chest radiography, this process produced a small organizing model that showed consistency in early testing. Further progress in report modeling will be enhanced by the development of software applications that directly use the model.

The authors thank Dr. David Evans for collecting the chest radiography corpus and providing CLARIT analysis; Dr. Robert McClure and Dr. David Wormuth for helpful discussions during model development; Dr. Stephen Johnson for helpful discussions of sublanguage analysis; and Ron Rouse and Dr. Charles Kahn, Jr., for their thoughtful reviews of this manuscript.

The authors' full model in conceptual graph ASCII notation is available by anonymous ftp at [dsg.harvard.edu](ftp://dsg.harvard.edu), and will remain so for at least one year from the date of publication.

References ■

1. Institute of Medicine Committee on Improving the Patient Record. *The Computer-Based Patient Record: An Essential Technology for Health Care*. Washington, DC: National Academy Press, 1991.
2. Musen MA. Dimensions of knowledge sharing and reuse. *Comput Biomed Res*. 1992;25(5):437-67.
3. Rumbaugh J, Blaha M, Premerlani W, Eddy F, Lorenson W. *Object-Oriented Modeling and Design*. Englewood Cliffs, NJ: Prentice Hall, 1991.
4. Barr CE, Komorowski HJ, Pattison-Gordon E, Greenes RA. Conceptual modeling for the unified medical language system. In: Greenes RA, ed. *Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care*. Silver Springs, MD: IEEE Computer Society Press, 1988:148-51.
5. Greenes RA, Barnett GO, Klein SW, Robbins A, Prior R. Recording, retrieval and review of medical data by physician-computer interaction. *N Engl J Med*. 1970;282(Feb. 5):307-15.
6. Leeming BW, Simon M. Computerized radiologic reporting: The CLIP system. In: *Proceedings of the Sixth Conference on Computer Applications in Radiology & Computer Aided Analysis of Radiological Images*. Los Angeles: IEEE Computer Society Press, 1979:414-9.
7. Leeming BW, Simon M, Jackson JD, Horowitz GL, Bleich HL. Advances in radiologic reporting with computerized language information processing (CLIP). *Radiology*. 1979;133:349-53.
8. Fischer PJ, Stratmann WC, Lundsgarrde HP. User reaction to PROMIS: issues related to acceptability of medical innovations. In: *Proceedings of the Fourth Annual Symposium on Computer Applications in Medical Care*. Los Angeles: IEEE Computer Society Press, 1980:1722-30.
9. Rector AL, Nowlan WA, Kay S. Foundations for an electronic medical record. *Methods Inf Med*. 1991;30:179-86.
10. Bell DS, Greenes RA, Doubilet PD. Form-based clinical input from a structured vocabulary: initial application in ultrasound reporting. In: Frisse M, ed. *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1992:789-91.
11. Bernauer J, Gumrich K, Kutz S, Lindner P, Pretschner DP. An interactive report generator for bone scan studies. In: Clayton P, ed. *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care (SCAMC)*. New York: McGraw-Hill, 1991:858-60.
12. Campbell KE, Wieckert K, Fagan LM, Musen MA. A computer-based tool for generation of progress notes. In: Safran C, ed. *Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1993:284-8.
13. Kuhn K, Zemmler T, Reichert M, Heinlein C, Roesner D. Structured data collection and knowledge-based user guidance for abdominal ultrasound reporting. In: Safran C, ed. *Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1993:311-5.
14. Naeymi-Rad F, Almeida FD, Trace D. IMR-entry (intelligent medical record-entry). In: Frisse M, ed. *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1992:783-4.
15. Lussier YA, Maksud M, Desruisseaux B, Yale PP, St.-Arsenault R. PureMD: a computerized patient record software for direct data entry by physicians using a keyboard-free pen-based portable computer. In: Frisse M, ed. *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1992:261-4.
16. Kuhn K, Gaus W, Wechsler JG, et al. Structured reporting of medical findings: evaluation of a system in gastroenterology. *Methods Inf Med*. 1992;31:268-74.
17. Board of Directors of the American Medical Informatics Association. Standards for medical identifiers, codes, and messages needed to create an efficient computer-stored medical record. *J Am Med Informatics Assoc*. 1994;1:1-7.
18. Cimino JJ, Hripcsak G, Johnson SB, Clayton PD. Designing an introspective, multipurpose, controlled medical vocabulary. In: Kingsland L, ed. *Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care*. Washington, DC: IEEE Computer Society Press, 1989:513-8.
19. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-

- based approaches to the maintenance of a large controlled medical terminology. *J Am Med Informatics Assoc.* 1994;1(1):35-50.
20. Campbell KE, Musen MA. Representation of clinical data using SNOMED III and conceptual graphs. In: Frisse M, ed. *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1992.
 21. Rossi-Morri A, Bernauer J, Pakarinen V, et al. CEN/TC251/PT003 models for representation of terminologies and coding systems in medicine. In: *Proceedings of the Seminar Opportunities for European and U.S. Cooperation in Standardization in Health Care Informatics*. Geneva, 1992.
 22. Rector AL, Nowlan WA, Glowinski A. Goals for concept representation in the GALEN project. In: Safran C, ed. *Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1993:414-8.
 23. Volot F, Zweigenbaum P, Bachimont B, et al. Structuration and acquisition of medical knowledge using UMLS in the conceptual graph formalism. In: Safran C, ed. *Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1993:710-4.
 24. Baud R, Lovis C, Alpay L, et al. Modelling for natural language understanding. In: Safran C, ed. *Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1993:289-93.
 25. Schroeder M. Knowledge based analysis of radiology reports using conceptual graphs. In: *Proceedings of the Seventh Annual Workshop on Conceptual Graphs*, Las Cruces, New Mexico, July 8-10, 1992.
 26. Bell DS, Greenes RA. Building a semantic network for radiologic records. In: Kahn M, ed. *Proceedings of the 1993 Spring Congress of the American Medical Informatics Association*, Bethesda, MD: American Medical Informatics Association, 1993:58.
 27. Bernauer J. Conceptual graphs as an operational model for descriptive findings. In: Clayton P, ed. *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1991:214-8.
 28. Haimowitz JJ, Patil RS, Szolovits P. Representing medical knowledge in a terminological language is difficult. In: Greenes R, ed. *Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care*. Silver Spring, MD: IEEE Computer Society Press, 1988:101-5.
 29. Rector AL, Nowlan WA, Kay S. Conceptual knowledge: the core of medical information systems. In: Lun K, Degoulet P, Piemme T, Rienhoff O, eds. *Proceedings of the MEDINFO '92*. Amsterdam: North-Holland, 1992:1420-6.
 30. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain: an evolving diagnostic decision-support system. *JAMA*. 1987;258:67-74.
 31. Greenes RA, McClure RC, Pattison-Gordon E, Sato L. The findings-diagnosis continuum: implications for image descriptions and clinical databases. In: Frisse M, ed. *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1992:383-7.
 32. Johnson SB, Gottfried M. Sublanguage analysis as a basis for a controlled medical vocabulary. In: Kingsland L, ed. *Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care*. Washington, DC: IEEE Computer Society Press, 1989:519-23.
 33. Velardi P, Pazienza M, Magrini S. Acquisition of semantic patterns from a natural corpus of texts. *SIGART Newsletter*. 1989;108:115-23.
 34. Grishman R, Kittredge R, ed. *Analyzing Language in Restricted Domains: Sublanguage Description and Processing*. Hillsdale, NJ: L. Erlbaum Associates, 1986.
 35. Evans DA, Hersh WR, Monarch IA, Lefferts RG, Handerson SK. Automatic indexing of abstracts via natural-language processing using a simple thesaurus. *Med Decis Making*. 1991;11(suppl):S108-15.
 36. Evans DA, Ginther-Webster K, Hart M, Lefferts RG, Monarch IA. Automatic indexing using selective NLP and first-order thesauri. In: *Proceedings of the RIAO '91*. Barcelona: Automa University of Barcelona, 1991:624-44.
 37. Sowa JF. *Conceptual Structures: Information Processing in Mind and Machine*. The Systems Programming Series. Reading, MA: Addison-Wesley, 1984.
 38. Johnson SB, Friedman C, Cimino JJ, Clark T, Hripsak G. Conceptual data model for a central patient database. In: Clayton P, ed. *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care*. Washington, DC: IEEE Computer Society Press, 1991:381-5.
 39. Friedman C, Hripsak G, Johnson SB, Cimino JJ, Clayton PD. A generalized relational schema for an integrated clinical patient database. In: Miller RA, ed. *Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care*. Los Alamitos, CA: IEEE Computer Society Press, 1990:335-9.
 40. Evans DA, Chute GG, Cimino JJ, et al. CANON: toward a medical-concept representation language. In: Kahn M, ed. *Proceedings of the 1993 Spring Congress of the American Medical Informatics Association*. Bethesda, MD: American Medical Informatics Association, 1993:26.
 41. Nowlan WA, Rector AL. Medical knowledge representation and predictive data entry. In: Stefanelli M, Hasman A, Fieschi M, Talmon J, eds. *Proceedings of the AMIE-91*. Berlin: Springer-Verlag, 1991:1043-9.
 42. Pattison-Gordon E. Thenetsys: A semantic network system. Decision Systems Group Technical Report DSG-93-02, 1993.
 43. Feinstein AR. Unsolved scientific problems in the nosology of clinical medicine. *Arch Intern Med*. 1988;148:2269-74.
 44. Masarie FE, Miller RA, Bouhaddou O, Giuse NB, Warner HR. An interlingua for electronic interchange of medical information: using frames to map between clinical vocabularies. *Comput Biomed Res*. 1991;24(4):379-400.

(The appendixes follow)

APPENDIX A—Full Text of a Sample Report

PA view is compared to the previous examination dated 10-22-91. Surgical clips are again seen along the right mediastinum and right hilar region. There are new surgical clips in the distribution of the circumflex artery as well as 4 intact sternotomy wires. There is persistent increased right paramediastinal opacity, possibly related to previous radiation therapy. New plate like opacities are seen in the left and mid lower lung zones, compatible with atelectasis. There has been some interval improvement in the left pleural effusion.

1. Slight interval decrease in left pleural effusion.
2. Left lower lobe atelectasis.
3. Post-operative changes consistent with coronary artery bypass graft as well as previous lobectomy on the right.

APPENDIX B—Representation of a Sample Report

Each sentence of this report is shown followed by the noun phrases parsed from the sentence by CLARIT and a conceptual graph representation of the sentence. Conceptual graphs for an individual report consist of concept instances linked by relations. In a functioning reporting system, concept instances would be represented in a patient database as references to the corresponding concept types in the model. Many concept instances could be simply generic references in the model (generic concept instances), but instances that link a particular set of other instances would require individual records in the patient database. These specific concept instances are designated here with arbitrary *individual markers* (such as “#2201”) simply to show that they would have identity within a patient database...

Concept types that were not introduced in the Results section of this paper are introduced here with a subtype assertion that relates the new concept to its closest supertype.

Sentence 1: PA view is compared to the previous examination dated 10-22-91.

CLARIT: “pa view,” “previous examination”

Model:

```
[CHEST_X_RAY:#22] -
  (HAS_TECHNIQUE) -> [PLAIN_RADIOGRAPHY]
  (HAS_X_RAY_DIRECTION) -> [PA],
  (COMPARED_WITH) -> [CHEST_X_RAY:#21].
```

Sentence 2: Surgical clips are again seen along the right mediastinum and right hilar region.

CLARIT: “surgical clip,” “right mediastinum,” “right hilar region”

Model:

```
SURGICAL_CLIP < COMPOSITE_FINDING.
TEMPORAL_STATUS < ATTRIBUTE.
PERSISTENT < STATUS.
SIDE < LEFT_RIGHT_AXIS_REGION.
HILAR_REGION < SPATIAL_SUBREGION.
MEDIASTINUM < ANATOMIC_LOCUS.
```

```
[SURGICAL_CLIP:#2201] -
  (HAS_QUALIFIER) -> [MULTIPLE]
  (HAS_QUALIFIER) -> [QUALIFIER:#22011] -
    (HAS_ATTRIBUTE) -> [TEMPORAL_STATUS]
    (HAS_STATUS) -> [PERSISTENT],
  (HAS_LOCATION) -> [LUNG:#22012] -
    (OF_LATERALITY) -> [RIGHT]
    (HAS_SUBREGION) -> [HILAR_REGION],
  (HAS_LOCATION) -> [MEDIASTINUM:#22013] -
    (HAS_SUBREGION) -> [SIDE:#22014] -
    (OF_LATERALITY) -> [RIGHT].
```

Sentence 3: There are new surgical clips in the distribution of the circumflex artery as well as 4 intact sternotomy wires.

CLARIT: “new surgical clip,” “distribution,” “circumflex artery,” “intact sternotomy wire”

Model:

NEW < STATUS.
 STERNOTOMY WIRE < ELEMENTAL FINDING.
 CONDITION QUALIFIER < QUALIFIER.
 CONDITION < ATTRIBUTE.
 INTACT < STATUS.
 INTEGER QUANTITY < QUALIFIER.
 INTETER < STATUS.
 CIRCUMFLEX CORONARY ARTERY DISTRIBUTION < ANATOMIC LOCUS.

[SURGICAL CLIP:#2202] -
 (HAS QUALIFIER) -> [MULTIPLE]
 (HAS QUALIFIER) -> [QUALIFIER:#22021] -
 (HAS ATTRIBUTE) -> [TEMPORAL STATUS]
 (HAS STATUS) -> [NEW],
 (HAS LOCATION) -> [CIRCUMFLEX CORONARY ARTERY DISTRIBUTION].

[STERNOTOMY WIRE:#2203] -
 (HAS QUALIFIER) -> [MULTIPLE]
 (HAS QUALIFIER) -> [CONDITION QUALIFIER:#22031] -
 (HAS ATTRIBUTE) -> [CONDITION]
 (HAS STATUS) -> [INTACT],
 (HAS QUALIFIER) -> [INTEGER QUANTITY:#22032] -
 (HAS ATTRIBUTE) -> [QUANTITY]
 (HAS STATUS) -> [4],
 (HAS QUALIFIER) -> [QUALIFIER:#22033] -
 (HAS ATTRIBUTE) -> [TEMPORAL STATUS]
 (HAS STATUS) -> [NEW].

Sentence 4: There is persistent increased right paramediastinal opacity, possibly related to previous radiation therapy.

CLARIT: "persistent increased right paramediastinal opacity," "previous radiation therapy"

Model:

REGIONAL INCREASED OPACITY < ELEMENTAL FINDING.
 PARAMEDIASTINAL REGION < SPATIAL SUBREGION.
 RADIATION REACTION < ETIOLOGIC DIAGNOSIS.
 RADIATION THERAPY < ETIOLOGIC AGENT.
 POSSIBLE < CERTAINTY.

[REGIONAL INCREASED OPACITY:#2204] -
 (HAS QUALIFIER) -> [QUALIFIER:#22041]
 (HAS ATTRIBUTE) -> [TEMPORAL STATUS]
 (HAS STATUS) -> [PERSISTENT],
 (HAS LOCATION) -> [LUNG:#22042] -
 (OF LATERALITY) -> [RIGHT]
 (HAS SUBREGION) -> [PARAMEDIASTINAL REGION],
 (HAS INTERPRETATION) -> [RADIATION REACTION:#22043] -
 (HAS ETIOLOGY) -> [RADIATION THERAPY]
 (HAS CERTAINTY) -> [POSSIBLE].

Sentence 5: New plate like opacities are seen in the left and mid lower lung zones, compatible with atelectasis.

CLARIT: "new plate like opacity," "left mid low lung zone," "compatible," "atelectasis"

Model:

ATELECTASIS < COMPOSITE FINDING.
 LOWER < SUPERIOR-INFERIOR AXIS POSITION.

[OPACITY:#2205] -
 (HAS QUALIFIER) -> [MULTIPLE]
 (HAS QUALIFIER) -> [QUALIFIER:#22051] -
 (HAS ATTRIBUTE) -> [TEMPORAL STATUS]

```

                (HAS STATUS) -> [NEW],
(HAS QUALIFIER) -> [SHAPE OBSERVATION:#22052] -
                (HAS ATTRIBUTE) -> [SHAPE]
                (HAS STATUS) -> [PLATELIKE],
(HAS LOCATION) -> [LUNG:#22053] -
                (OF LATERALITY) -> [LEFT]
                (HAS SUBREGION) -> [LUNG FIELD:#22054] -
                        (HAS POSITION) -> [MIDDLE],
                (HAS SUBREGION) -> [LUNG FIELD:#22055] -
                        (HAS POSITION) -> [LOWER],,
(HAS INTERPRETATION) -> [ATELECTASIS].

```

Sentences 6 & 7: There has been some interval improvement in the left pleural effusion.

1. Slight interval decrease in left pleural effusion.

CLARIT: "interval improvement," "left pleural effusion,"
 "slight interval decrease," "left pleural effusion"

Model:

```

PLEURAL SPACE < BILATERAL ANATOMIC LOCUS.
INTERVAL CHANGE < ATTRIBUTE.
IMPROVEMENT < STATUS.
DECREASE < STATUS.

```

```

[PLEURAL EFFUSION:#2206] -
  (HAS QUALIFIER) -> [QUALIFIER:#22061] -
        (HAS ATTRIBUTE) -> [INTERVAL CHANGE]
        (HAS STATUS) -> [IMPROVEMENT],
  (HAS QUALIFIER) -> [QUALIFIER:#22062] -
        (HAS ATTRIBUTE) -> [INTERVAL CHANGE]
        (HAS STATUS) -> [DECREASE],
  (HAS LOCATION) -> [PLEURAL SPACE:#22063] -
        (OF LATERALITY) -> [LEFT].

```

Sentence 8: Left lower lobe atelectasis.

CLARIT: "left low lobe atelectasis"

Model:

```

[ATELECTASIS:#2207] -
  (HAS LOCATION) -> [LUNG:#22071] -
        (OF LATERALITY) -> [LEFT]
        (HAS PART) -> [LEFT LOBE].

```

Sentence 9: Post-operative changes consistent with coronary artery bypass graft as well as previous lobectomy on the right.

CLARIT: "post operative change," "coronary artery bypass graft," "previous lobectomy," "right"

Model:

```

POST OPERATIVE CHANGES < ETIOLOGIC DIAGNOSIS.
LOBECTOMY < ETIOLOGIC AGENT.
LOBECTOMY < INTERVENTION.
CORONARY ARTERY BYPASS GRAFT < ETIOLOGIC AGENT.
CORONARY ARTERY BYPASS GRAFT < INTERVENTION.

```

```

[POST OPERATIVE CHANGES:#2208] -
  (HAS ETIOLOGY) -> [CORONARY ARTERY BYPASS GRAFT].

```

```

[POST OPERATIVE CHANGES:#2209] -
  (HAS ETIOLOGY) -> [LOBECTOMY:#22091] -
        (APPLIED TO) -> [LUNG:#22092] -
                (OF LATERALITY) -> [RIGHT].

```