A Field Theoretical Approach to Medical Natural Language Processing

Ricky K. Taira, Vijayaraghavan Bashyam, and Hooshang Kangarloo

***Abstract*—A parser for medical free text reports has been developed that is based on a chemistry/physics inspired "field theory" for word–word sentence-level dependencies. The transition from the linguistic world to the world of interacting particles with potential energies is guided by a psycholinguistics thought experiment related to the amount of "work" required to bring a reference word into an anchored configuration of words. Calibration experiments involving four and five grams were conducted. Data from these experiments were used as a knowledge source for estimating field conditions for words in sentences sampled from a corpus of medical reports. The result of the parser is a dependency tree that represents the global minimum energy state of the system of words for a given sentence. The system was trained and tested on a corpus of radiology reports. Preliminary performance, as quantified by link recall and precision statistics, is 84.9% and 89.9%, respectively.**

***Index Terms*—Knowledge representation, natural language processing (NLP), structured medical reporting.**

I. Introduction

THE GOAL of the medical natural language processing (NLP) is to transform the information content contained within a free text report (e.g., radiology) into a representation that is computer understandable. The representation is typically a first-order logic representation (without quantification) such as a conceptual graph or logic-based frames [1], [2]. Several medical NLP systems have been developed for a variety of high-end applications including automatic coding of patient reports [3], [4], extraction of findings documented in diagnostic reports [5]–[8], automatic flagging of alarm conditions [9], identification of patients with particular conditions [10], and analysis of co-occurrence relations among radiological findings [11].

Syntactic parsing is a subtask of NLP and generally agreed to be an important intermediate step toward the goal of a deep understanding of free text. Our "parser" combines the traditionally separate stages of syntactic parsing and semantic interpretation into a single unifying step, as discussed later. The input to the parser/semantic interpreter (henceforth, referred to as the parser) is word tokens from a single sentence tagged with basic part-of-speech and semantic class labels. Currently, there exist about 450 semantic tags and 12 part-of-speech categories within our lexicon [12], [13]. Table I shows an example input to the parser from this lexical analysis step. The output of the parser is a dependency graph that emphasizes semantics (Fig. 1) [14]–[16].

TABLE I
Input to Our Parser is the Results From a Lexical Analysis Step That Assigns Part-of-Speech and Semantic Tags to Each Word Token Within a Sentence

Word	Part-of-Speed	Semantic Class
A	art	pos.def_art
large	adj	pValue.size.assessment
heterogeneously	adv	pValue.uniformity
enhancing	gerund	pValue.signalStrength
mass	noun	physobj.finding.abnormal
is	_be	_is
seen	pastp	relation.exist.observed
in	prep	pos.in
the	art	pos.def_art
left	adj	pValue.spatial.direction
lower	adj	pValue.spatial.direction
pole	noun	physobj.selfReferenceLocation
of	prep	pos.of
the	det	pos.def_art
kidney	noun	physobj.anat
which	conn	pos.wh.which
is	_be	_is
consistent	adj	existence.possibility
with	prep	pos.with
renal	adj	physobj.anat
cell	noun	physobj.tissue
carcinoma	noun	condition.abnormal

Briefly, a *dependency graph* [17], [18] describes the structure of a sentence in terms of binary *head-modifier* (also called *dependency*) *relations*. A dependency relation is an asymmetric relation between a word called the *head* (also *governor* or *parent*), and a word called the *modifier* (also *dependent*, *daughter*, or *child*). A word in the sentence can play the role of the head in several dependency relations (i.e., it can have several modifiers) but each word can play the role of the modifier exactly once. One special word, named the *root*, does not play the role of the modifier in any relation. For consistency, we will use the convention that the root is linked to the end of the sentence marker, so that all the words in the sentence are used in a modifier role (i.e., form an attachment). The set of dependency relations that can be defined on a sentence form a tree, which is called the *dependency tree*.

The remainder of this paper is organized as follows. Section II briefly provides a background on the related work in the area of NLP, focusing on both medical and general approaches. Section III enumerates desiderata for an ideal medical NLP system. Section IV provides an introduction to our field theory paradigm, followed by a discussion of a set of experiments performed to explore the field patterns within a training corpus of reports (Section V). Section VI describes the implementation

Manuscript received October 19, 2005; revised May 27, 2006. This work was supported by the National Institutes of Health (NIH) under Program Project Grant (PPG) PO1-EB00216 and Grant RO1-EB000362.

The authors are with the Medical Imaging Informatics Group, University of California at Los Angeles, Los Angeles, CA 90024 USA (e-mail: rtaira@mii.ucla.edu; hkangarloo@mii.ucla.edu; vbashyam@ucla.edu).

Digital Object Identifier 10.1109/TITB.2006.884368

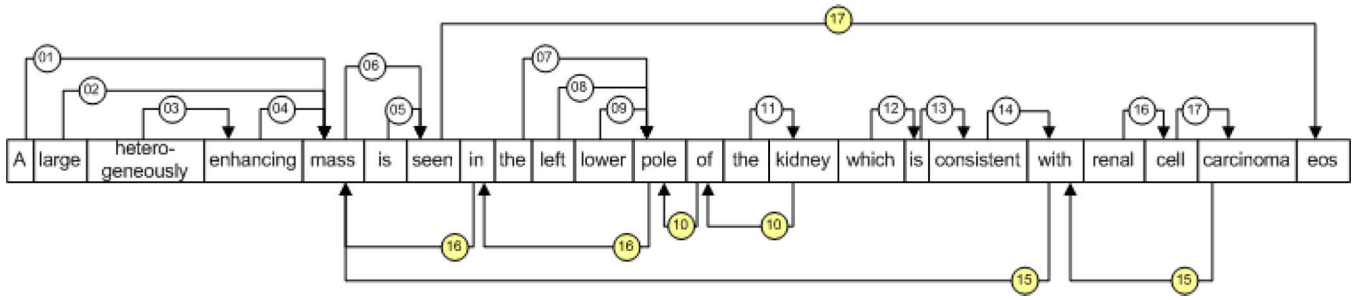

Fig. 1. Example of a word–word dependency diagram for the sentence tokenized in Table I. Individual link probabilities are not shown. eos: end of sentence.

of the parser execution using the empirically derived interaction tables. A preliminary evaluation of the parser performance, as summarized by recall and precision statistics within the domain of the radiology reports, is presented in Section VII. Finally, Section VIII concludes with a discussion of significant contributions of this work and future directions for this project.

II. Background and Related Work

Research in the cognitive sciences have demonstrated that language processing is extremely complex, involving a large number of fundamental intercommunicating processing centers (neurons) [19]. At one level, a cluster of neurons (memory structures) is activated by the individual words. These cluster centers induce perceptions of a number of possible concepts of direct and/or indirect correlation to the input word. Combinations of words quickly allow the mind to localize specific interpretations/contexts from which past referential experiences are loaded into the fore-memory.

Words, then, are central to the processing of language and serve as the fundamental processing units of our system. This design around word processors is consistent with the beliefs of other NLP investigators [20]–[22]. Small and Rieger treat each word of language as a complex procedural knowledge source that contains decision pathways reflecting the range of linguistic and other world knowledge about the word necessary to understand the word in a broad variety of contexts [21]. Their model's structure rests on the hypotheses that: 1) human knowledge about language is organized primarily as knowledge about words rather than as knowledge about rules and 2) language understanding is largely the coordination of information exchange among word experts as each examines its own involvement in its linguistic and conceptual environment. Rather than the rule application or grammatical pattern matching, the essence of the NLP is, therefore, one of mediating the passing of linguistic signals and concept fragments among the word experts as each goes about its own comprehension activities. They develop a theory around *control* (*processing*) regularities rather than around *data* regularities.

Language structure is observed to have a number of characteristics that can give clues as to how we can synthesize or analyze such configurations. For example, language consists of a set of reusable units that can combine hierarchically and recursively. Thus, context-free grammars using pushdown automata [23], [24] and probabilistic generative variants [25]–[27] have been the mainstream approach to language processing. Additionally, the formation rules of higher order constituents in language is strongly correlated to what the constituent means (i.e., function). Thus, semantically driven grammars that utilize hybrid bottom-up and top-down knowledge have emerged recently [28]. These semantically driven grammars may be especially appropriate for medical NLP systems, where the domain is limited and the types of communications are relatively confined. In fact, although the medical NLP systems have traditionally lacked the robust lower level statistical language models of the general NLP community, high-performance medical applications have been realized mostly as a result of a more comprehensive view of the final semantic representation structures for both concepts (e.g., UMLS, SNoMED-CT), and specific entity–relation–attribute models, e.g., radiology findings [1], [29], caBIG and CDE frames for cancer,[1] neuro-informatics (GENIE [30], BIRN [31]), and nursing informatics [2]. The Semantic Knowledge Representation project initiated by the National Library of Medicine is targeted for the biomedical free text.

Research in the area of complex systems science may be particularly relevant to language processing given the commonality of the hierarchical complexity of structures (words, phrases, thoughts, topics, etc.). Emergent-related approaches are being developed in various application areas that incorporate both bottom-up and top-down features. The trend toward such global optimization approaches is motivated by the possibility to create robust and adaptive solutions that may be important in medical NLP applications requiring both high recall and precision.

III. Desiderata

Our desiderata for robust medical NLP systems are as follows:

1) *Adaptability to New Domains*: Efforts to adapt state-of-the-art statistical parsers [32], [33] to medical corpora have mainly consisted of the incorporation of the medical lexicons (i.e., word-level knowledge) [34], [35]. These initial efforts to adapt general NLP engines to medical documents have resulted in good lexical coverage, yet, they have ignored the modification of grammar models. Grammatical styles and communicative goals in medicine are quite distinct from, for example, newswire articles from which these systems have been trained. Each subspecialty of medicine may have different language models

[1]National Cancer Institute's Cancer Biomedical Informatics Grid (caBIG). [Online]. Available: http://cabig.nci.nih.gov (last accessed Sep. 25, 2005).

(in the statistical sense). The relative frequency of words, the types of phrases used, and the grammatical constructs used to express specific communicative goals are different within targeted medical domains [36]. Methods to train existing NLP engines for new domains by individuals who are not the primary developers remain.

2) *Robustness to Unseen Patterns and Noisy Underlying Evidence*: We desire a grammatical model that can intelligently generalize decisions for which it has incomplete and/or uncertain knowledge.
3) *Guidance from Higher Order Knowledge*: A large effort in concept modeling is ongoing in medical informatics; these symbolic models can help guide an NLP parser as to linkages that are sanctioned and/or disallowed based on a phenomenological view. For example, a solid tumor would not have properties normally associated with a liquid (e.g., flow). These higher order constraints may assist in resolving ambiguities at the word (i.e., sense) or word–word (syntactic parse) levels. Globally consistent solutions are, thus, preferred over methods that utilize strict linear low- to high-level processing pipelines. It should be noted that a syntactic parse is an intermediate step within NLP.
4) *Intuitive Features*: A review of NLP feature descriptions indicate how complex and voluminous the feature space is for current state-of-the-art parsers [37]. The number of the state–space patterns, depending upon the level of detail (e.g., lexicalize context-free grammars, lexicalized semantic-role context-free grammars, etc.) can lead to hundreds of thousands of features incorporated into a single statistical model [38]. Features that are intuitive rather than clouded by technical jargon are desirable.
5) *Mechanistic Model*: Additionally, one can rarely visualize the connection between features; causal or influential diagrams that provide some understanding of how the available evidence propagates among variables would be helpful in organizing our knowledge. Some fundamental organizing concepts and principles of how words interact and link would help promote a more mechanistic theory for language.

IV. Field Theory Paradigm

A. Overview

Language processing is undoubtedly a complex signal-processing operation. We attempt, therefore, to formulate our theory in terms of this foundation. We start our signal processor model with the premise that words within a sentence should be viewed as active entities within an evolving environment, much like a cell in a biological system or an electron or atom in a quantum mechanical system. The nature of the word particles and the forces of nature within this virtual world are the important fundamental problems that we need to investigate. We use words as the central focus of abstraction for the system, knowing that other representations are also possible (e.g., a set of primitive semantic units).

We view parsing as a dynamics problem involving the "movement" and "bonding" of the atomic (i.e., single words) and molecular (phrase/compound) units. It is useful to conceptualize the movement of these word particles within a *linguistic world*. First, let us consider the probabilistic boundaries of a word. The boundaries for a reference word are defined by the range of signals it can receive from other words within the document. As a first perspective, we limit the signal propagation of words to a single sentence. That is, let us assume all words live in a world in which there is no communication of information beyond its sentence limits. This implies that the start and the end of a sentence impose such boundary conditions on the signals radiating from a word that such signals cannot escape the sentence. The end points of the sentence represent infinite impedance to all the signals emanating from within the sentence. If there is no communication of information beyond the sentence boundaries, we can say that the sentence *is* the "world" in which a word exists. The sentence comprises an isolated system. Our "system" for a word *is* the sentence.

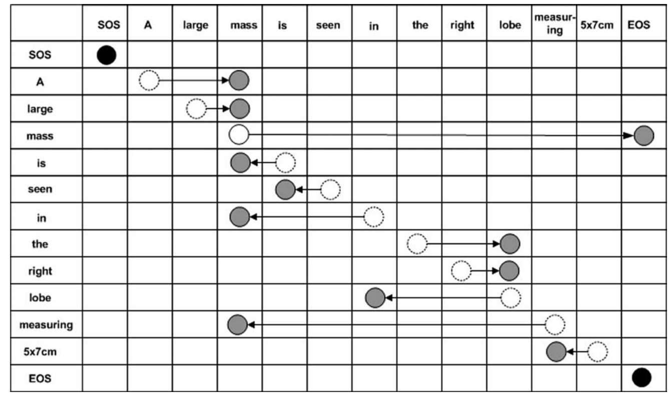

Fig. 2. Initial configuration of words is along the diagonal within our lattice representation. The final configuration corresponds to a minimum energy state, which represents the parse tree of the sentence.

Having defined the scope of our system, let us expand the details of the environment (space) in which words exist. In traditional linguistics, the "attachment" of the words is diagramed using a surface parse tree [39]. In our representation, we utilize a two-dimensional orthogonal matrix to assist in describing the state, word movement dynamics, and structure of our system of words (see Fig. 2). If the sentence contains N words, we define this matrix to be of dimensions $(N+2) \times (N+2)$. (As part of our formalization, we add two special words to each sentence: 1) sos, start-of-sentence word and 2) eos, end-of-sentence word.) This matrix defines the field coordinates within our representation and will impose some physical constraints on the movement of word particles.

The computing steps for our parser are as follows:

1) Initialize the system of words along the diagonal of the matrix. This represents a state characterized by the absence of forces and, hence, bonds between any of the constituent words of the system. Certainly, the identity matrix is not the "minimum energy," stable state of our system, unless there is absolutely no correlation (i.e., mutual information) between the words.
2) We now bring each word out of isolation and allow word–word interactions. In our representation, words may reside

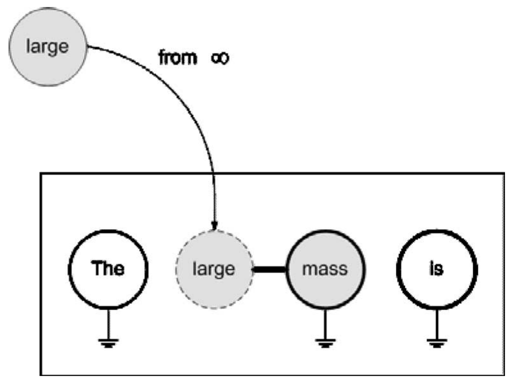

Fig. 3. The conceptualization of the potential energy of a reference word ("large") at the position of "mass" given a left and right context word is viewed to be related to the amount of "work" one needs to bring a reference particle (e.g., "large") into a system of anchored words.

only along their respective rows. Their column position indicates the head word it is interacting with (i.e., linking to).

3) A configuration for the system of words is proposed. Note that there are an exponentially large number of such configurations.
4) For a given configuration, we estimate each word's energy based on its local field conditions to be explained later; the energy of the system of words is the sum of the individual word energies.
5) We explore the configuration space either exhaustively for short sentences or via a genetic algorithm to search for a global minimum energy state for the sentence (sum of the energies of the individual words). This minimum energy configuration represents the final dependency tree for the sentence.

B. Word–Word Interactions

The key to the field theory approach is the estimation of the energy of a reference word w_i with respect to its sentence position x. Thus, we assume that some sort of hypothetical linguistic force field $F(x)$ exists. For example, we conceive of a system in which words are attracted or repelled from one another depending upon their *affinities* for one another. The thought experiment that motivates our work is as follows: in order to estimate the potential energy for a reference word w_i at position x (i.e., $word_i$ links to $word_x$), we conceptualize the notion of how much "work" U would be required to bring the reference word (e.g., "large") from infinity into an isolated system of n other anchored words (see Fig. 3). We define the energy of a particle at infinity to be zero

$$U(x) = -\int_{x_0}^{x} F(x)\,dx + U(x_0). \tag{1}$$

The potential energy of the particle relates to its attachment probability of the particle associated with the column position x; the lower the potential energy $U(x)$, the higher the probability for the particle to occupy the position x. Our problem can now be viewed as defining the energy of a traveling particle with respect to its position along the sentence (i.e., as it attempts to move to an off-diagonal position). The total sum system energy is what we need to minimize. The final topology of the distribution of words within the matrix, we claim, should reflect the configuration of the best dependency parse tree for a given sentence. Note that the definition of the potential energy is not confined to just words, but can involve reference entities that are larger constituents. This definition applies to all the involved particles (reference, target, left context, and right context). We discretize this energy into the following three categories (also see Fig. 4):

1) $U_i(x) > 0$: Energy is required to bring the reference word w_i at position x. This corresponds to a prototypical situation where there is a repulsion between two words (i.e., not semantically compatible).
2) $U_i(x) = 0$: No work is required to bring the reference word w_i at position x. This corresponds to a prototypical situation where the target word w_x is transparent (i.e., neutral) to the reference word w_i.
3) $U_i(x) < 0$: There is an attraction between the reference word w_i and the target attachment word w_x. This corresponds to a situation where there is an affinity between two words that indicates a more favorable configuration for the reference word.

A word particle in motion within the sentence lattice can, thus, be visualized as a ball rolling on a potential energy surface. Equilibrium phrasal structures or other complex constituencies correspond to the positions of the minima in the valleys on such a potential energy surface [see Fig. 4(a)]. Fig. 4 shows a conceptualized potential energy profile for the reference word "large" under three different field conditions described previously. Either left-to-right or right-to-left attachments can be modeled.

V. Experiments

Our first task for assessing the viability of the field theory approach was to investigate the reliability of the potential energy wells [Fig. 4(a)] as a strong indicator for word–word attachment. Since we do not fully understand the nature of our proposed word–word interactions, we first compile a table recording their behaviors performing what amounts to laboratory experiments. We consider two types of attachment scenarios, Type I and Type II, defined as follows:

- *Type 1 Links* are direct links, with a typical example being an *adjective–noun* link.
- *Type II Links* are mediated links, with typical examples being a noun1–preposition–noun2 link and a *noun1–transitive verb–noun2* link.

A. Construction of Type I Interaction Tables

Let us first consider, practically, how we can estimate an interaction probability for a given pair of words (A, B) given a sentence S and a proposed parse configuration PT. In this initial investigation, we seek to identify the features of Type I profiles that we hypothesize strongly indicate an A to B attachment. Specifically, a Type I forward-link profile is characterized by a free unstable reference word A experiencing a potential hill to

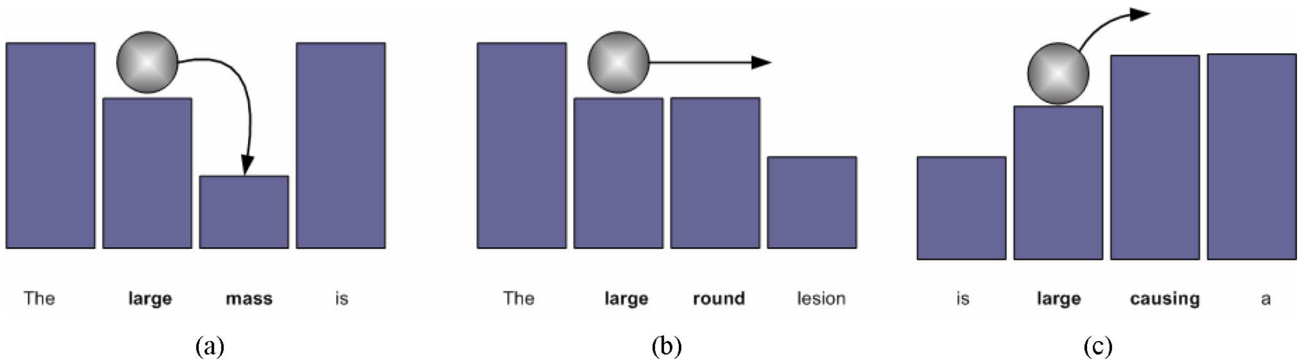

Fig. 4. Three basic energy landscapes. (a) Attraction. (b) Neutral. (c) Repulsion profiles for Type I links. Left and right context words provide improved determination of these profiles.

its left (L), a potential well B to its right, and a potential hill (R) to the right of B [see Fig. 4(a)]. We define the word stability to be directly related to the potential energy of the reference word at position x that is discretized into three energy states: negative attractive potential present at position x, neutral free state, or positive repulsive barrier potential. The most direct evidence associated with a Type I link is hypothesized to consist of four latent variables:

1) Is there a high potential barrier experienced by the reference word to the left?
2) Can the reference word A communicate with the target word B?
3) Is the link A $\rightarrow$ B semantically plausible? Is there a potential hill experienced by the reference word at a position to the right of the target word?
4) Is there a potential hill experienced by the reference word at a position to the right of the target word?

The observables in our model include the words of the sentence and any links present in the proposed configuration.

In our first set of experiments, we attempt to characterize the profiles using single words only (i.e., not constituent syntactic constructions). Thus, given four words $w_{\mathrm{L}}, w_{\mathrm{A}}, w_{\mathrm{B}}$, and w_{R}, our goal is to determine the interaction type, given these words are situated within the context of the thought experiment shown in Fig. 3

$$p(\alpha_i | w_{\mathrm{L}}, w_{\mathrm{A}}, w_{\mathrm{B}}, w_{\mathrm{R}}) \tag{2}$$

where

$$\alpha = \{\text{attraction, neutral, repulsion}\}.$$

That is, given three anchored words $w_{\mathrm{L}}, w_{\mathrm{B}}, w_{\mathrm{R}}$, how much "work" would be needed to bring the word w_{A} from infinity to link to word w_{B}. (*Note*: Right to left attachments can also be investigated by interchanging the roles of $w_{\mathrm{A}}, w_{\mathrm{B}}$.) The work states are quantized as either positive, zero, or negative corresponding to the interaction states of attraction, neutrality, and repulsion, respectively, as illustrated in Fig. 4. Recall, Type I interactions are typified by adjective–noun, adverb–verb, noun–noun, and collocation word–word linkages. A comprehensive investigation of Type I interactions would involve an exhaustive combinatorial statistical characterization. Thus, if there are M classes of words, the table consists of the upper order of $M \times M \times M \times M$ entries. (In radiology reports, the number of unique words is of the order of 10 000; the number of unique semantic classes of the order of 500; the number of part-of-speech tags of the order of 20 [40].) The actual experiments performed were conducted as follows.

TABLE II
(a) MINED TYPE I FOUR-GRAMS UNDER CONFIGURATION WITH NO EXISTING LINKS. IN THE CASE OF FORWARD LINKS, THE A-WORD IS THE REFERENCE WORD AND THE B-WORD THE ATTACHMENT SITE. (b) TYPE I FOUR-GRAMS UNDER CONFIGURATION WITH SOME EXISTING LINKS

Left - Context	Word A	Word-B	Right context	Dir	Energy State
SOS	*the*	*bright*	*density*	f	0
the	*bright*	*density*	*adjacent*	f	-1
bright	*density*	*adjacent*	*to*	f	+1
density	*adjacent*	*to*	*the*	f	-1
adjacent	*to*	*the*	*mass*	f	+1
to	*the*	*mass*	*is*	f	-1
the	*mass*	*is*	*increasing*	f	+1
mass	*is*	*increasing*	EOS	f	+1

(a)

Left - Context	Reference Word	Attachment Site	Right context	Energy State
SOS	*the*	*density*	*adjacent-to*	0
the	*density*	*adjacent-to*	*mass*	-1
density	*adjacent-to*	*mass*	*is*	+1
adjacent-to	*mass*	*is*	*increasing*	-1
mass	*is*	*Increasing*	*EOS*	+1

(b)

1) *Collection of Documents*: A total of 10 000 radiology reports collected from the UCLA Medical Center were retrieved from our radiology information system. All the documents of the patients were deidentified using the software described in [41]–[43]. A total of 8000 reports were set aside for training, 1000 for refinements, and 1000 for testing.
2) *Mining of Adjacent Four-Gram*: Near-exhaustive sampling of the connected four-grams within a sentence were compiled from the training corpus initially using a sliding window of four words. For each four-gram instance, we record the word string itself, the semantic and part-of-speech labels, as well as the originating sentence. Table II(a) shows the four-grams collected for the

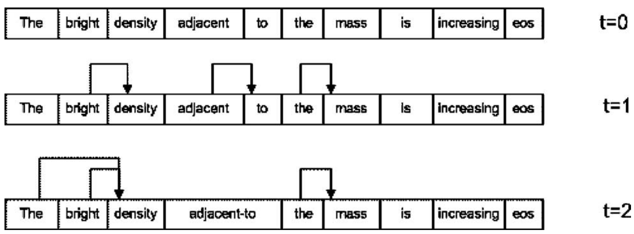

Fig. 5. During the process of mining new field conditions for reference words, we can perturb the configuration of the words within our matrix and identify new virtually adjacent n grams.

example sentence, "The bright density adjacent to the mass is increasing."

The four-gram represents the left context word, the reference word, the target attachment site, and the right reference word in Fig. 3.

3) *Manual Classification of Four-Grams*: A trained human who was guided by the thought experiment of Fig. 3 and knowledgeable of dependency tree representation for sentence parsing was asked to classify each four-gram in terms of one of the three interaction classes described previously (attraction, neutral, repulsion); truth was assisted by viewing the entire sentence for context. In some cases, the four-gram context was not sufficient to classify the interaction class; for example, commas and coordinating conjunctions used within the left or right context word positions. In these cases, an extra context word was added to the context.
4) *Perturbing the Sentence Configuration*: In mining for the field patterns, we can identify other virtually connected four-grams by perturbing the configuration for our system of words. For example, let us suppose that we are given a configuration shown in time step 1 in Fig. 5 that includes two instantiated links. We specify these links based on their favorable potential well profile as identified in the tagging step mentioned above. Recall that the instantiated links correspond to moving the reference word off diagonal and into the column corresponding to the head word (see Fig. 2). Under this new configuration, we must reestimate the field conditions for each reference word.
5) *Dynamic Recharacterization of Field Conditions*: Now, given a new configuration characterized by a proposed set of links, we must update the field conditions for each reference word. A few questions come to mind in our simple example.
   a) What should be the net effect of instantiating a link? What is the end product of such an operation? Is there some sort of convolution of information that occurs? Is the semantics of the linked pair dominated by the head word (e.g., *bright* $\rightarrow$ *density*) or do the linked words result in a semantics that is different than either of the constituent words (e.g., *adjacent* $\rightarrow$ *to* is a semantically coherent word sequence that conveys a single primitive relation [44])? To address these issues, semantic interpretation is integral to the parser algorithm. We compile a "reaction cookbook" as described later (see Section V-C). This allows us to address in part phenomena such as idiomatic phrases, multiple word senses, and nonsensical link propositions through world knowledge.
   b) What effect does a link have on the local field conditions? How do links affect communication paths between words within the sentence? Instantiated links can provide communication pathways between non-adjacent words. Alternatively, we can view some classes of links as having the property of being able to shield words that fall within its interval from those words that are outside its interval. We developed such communication rules based on the reference word's part-of-speech tag, the target-word's part-of-speech tag, and the link types that could be traversed between the reference and target word types.
6) *Mining of (Virtually) Adjacent Four-Grams*. Using the rules compiled for Type I communication pathways, we identify other four-grams as shown in Table II(b).
7) *Iterate:* Steps 3 through 6 are repeated until no new four-grams are found in a sentence.
8) *Loop*: The entire process is performed for each sentence within the training corpus.

The manual tagging step 3 was performed by two individuals; inter-tagger variability was not estimated in this exploratory work although it is an important evaluation, given our hope that the classification is relatively intuitive after initial training and can be performed by individuals with diverse backgrounds. The end result of this iterative mining process is a knowledge source that documents the Type I word–word interaction behavior as seen in a training corpus within our target medical domain.

Type I frequency statistics were compiled for each of the modes of interaction (2). Because there are several thousands of different words in medical text reports, it is likely that some words will not have appeared together in sentences in the training corpora. Unforeseen combinations of words will leave gaps in the system's ability to assess field conditions for a given reference word. Smoothed distributions were obtained by the following. 1) Relaxing word representations from strings to semantic class plus part-of-speech labels. Given the hierarchical organization of the semantic classes, a few levels of concept relaxation can be performed. For example, the Level 3 semantic class specification for the word "mass" is "physobj.abnormal.finding." The L2 specification, "physobj.abnormal" includes a larger number of entities (e.g., masses, artifacts, etc.). 2) Eliminating the left or right context word opposite to the link direction. Table III summarizes the number of unique patterns for different context definitions. A smoothed statistical model using interpolation over these various grain models is in development following the approaches described in [45] and [46].

B. Construction of Type II Interaction Tables

In Type II interactions involving the linking of three words, the major methodological differences compared to Type I interactions are as follows.

1) *Five-grams are Proposed*: This addresses the need to simultaneously satisfy semantic compatibility constraints involving relational words such as verbs and prepositions, which are two place predicates (e.g., tumor ← invading ← bone).
2) *Five-grams are Determined Based on the Notion of a Reduced Sentence:* A *reduced sentence* is defined as the subsequence of words in the sentence that is formed by reducing all the base noun phrases to their head-word alone. A simple set of rules can be found in [47] and [48] that identifies which of the children is the head-child of a phrase. An example sentence with its corresponding reduced sentence is presented next. Note that the use of a reduced sentence is only needed to facilitate the collection of training examples. The actual parser algorithm does not use the concept of a reduced sentence (see Section VI).
   - *Full sentence* (17 tokens):

$$\text{There are}|\text{two}|\text{low}|\text{density}|\text{lesions}|\text{in}|\text{the}|\text{left}|\text{kidney}|,\ |\text{the}|\ \text{largest}|\text{measuring}|\text{1.3 cm}|\text{which}|\text{is}|\text{unchanged.}$$

   - *Reduced sentence* (9 tokens):

$$\text{---}|\text{---}|\text{---}|\text{---}|\text{lesions}|\text{in}|\text{---}|\text{---}|\text{kidney}|,|\text{---}|\text{largest}|\text{measuring}\ |\text{1.3 cm}|\text{which} - \text{is}|\text{unchanged.}$$

We hypothesize that the identification of the constituents related to predicates with two arguments can be facilitated by the identification of a five-gram of the form $\{\mathrm{L\ A\ B\ C\ R}\}$. Thus, similar to Type I interactions, given five words $w_{\mathrm{L}}, w_{\mathrm{A}}, w_{\mathrm{B}}, w_{\mathrm{R}}$, our goal is to compute the probability of the interaction type, given that these words are situated within the context of now four context words:

$$p(\alpha_i | w_{\mathrm{L}}, w_{\mathrm{A}}, w_{\mathrm{B}}, w_{\mathrm{C}}, w_{\mathrm{R}}). \tag{3}$$

1) The automatic mining of Type II profiles is facilitated by proposing a reduced sentence representation that "deactivates" the reference words linked via Type I bonds.
2) Punctuation cannot fill the role of particle A or particle C.
3) The left word L cannot be of the syntactic class punctuation or conjunction. If this is the case, the L context is filled by concatenating the words from the punctuation to the left, halting when neither a punctuation nor a conjunction is encountered.
4) The right and left context words cannot be of the syntactic class punctuation or conjunction. If this is the case, then the R (L) context is filled by concatenating the punctuation/conjunction with the word(s) to the right (left) until a conjunction and/or a punctuation is no longer encountered.

The mining of Type II five-grams follows a similar procedure as described for Type I patterns. Similar smoothing operations involving relaxation of word-class representations and left/right context words in the direction opposite to the link were performed. Table III summarizes the number of unique five-gram entries for various grain representations.

TABLE III
NUMBER OF MINED UNIQUE N-GRAMS FOR VARIOUS WORD CLASS DEFINITIONS

Word Representation	Link Type	n-gram size	Number of Instances
Word string	I	4-gram	121,268
Word-string	I	3-gram	98,421
POS+Sem-L3	I	4-gram	92,770
POS+Sem-L3	I	3-gram	53,033
POS+Sem-L2	I	3-gram	46,597
POS+Sem-L1	I	3-gram	18,751
POS	I	4-gram	9,308
Word-string	II	5-gram	32,890
Word-string	II	4-gram	28,992
POS+Sem-L3	II	5-gram	24,700
POS+Sem-L3	II	4-gram	16,846
POS+Sem-L2	II	4-gram	16,002
POS+Sem-L1	II	4-gram	8,860
POS	II	5-gram	5,550

C. Construction of Semantic Interpretation Cookbook

Labeling of link semantics is equivalent to defining a logical relation between the parser-linked words. For example, in the phrase "bright density," we wish to describe the link as a description of the signal intensity of a radiological finding. A semantic interpretation cookbook was constructed that served as a knowledge source for guiding the labeling of parser links and in defining what concept modeling actions are to be taken as a link is instantiated (see later). The procedure for creating the cookbook was as follows:

1) *Identification of Unique Link Instances*: We scanned Type I and Type II interaction tables in search of all unique combinations of word linkages. We discovered 42397 Type I two-word link instances and 8368 Type II three-word link instances.
2) *Definition of Logical Relation Templates*: Mined linked instances were grouped according to the hand-crafted logical relation template definitions. Each template represented a primitive logical relation definition (e.g., *hasSize.assessment*). Template constraints were implemented using logic-base rules involving the membership of the specified logical relation argument slots (e.g., head, relation, value) to word-class memberships (e.g., a specific string, semantic class, or part-of-speech class). A total of 3214 templates were defined. Template definitions allowed fine grain control; for instance, the *hasSize* relation could be generalized as a property of any solid physical object, whereas the relation *hasCalcificationPattern* is applicable to a smaller class of objects, such as a lesion.
3) *Modeling a Response Function*: A modeling response function is attached to each template definition, providing individualized instructions as to how the presence of a given logical relation should be modeled. The response function syntax is declarative, and can contain a number of commands, including
   a) *Find*: Locate a data structure in the working memory such as the frame corresponding to a certain head word topic (e.g., mass) or the subframe data structure associated with a particular property (e.g., size).

b) *Assign*: This command is used to assign a value to a slot position identified by the corresponding *Find* command. More than one *Assign* command can follow a *Find* statement in situations where a logical relation implies multiple pieces of information, or where a number of default values can be specified.

c) *Reject*: In mining logical relation instances, erroneous links may be proposed by the parser (e.g., impossible interpretations such as "*the mass is normal*"). This command allows parser to reject a particular interpretation for a given sentence.

d) *Convolve*: This command changes the lexical features of the head, the relation, or the value words of the input logical relation. Words can be dynamically recharacterized, with the updated information sent back to the parser/semantic interpreter. For example, words with multiple senses can be resolved in this manner.

One result of mining the link instances within the training document domain is the automatic construction of a global syntactic–semantic dependency network that can serve as an initial definition of the ontology of the information space. We are currently exploring the possibility of creating such an ontology using this approach.

VI. Parsing Dynamics

The aim of the parser is to take a sentence that contains the basic part-of-speech and semantic class word labels as input and produce a parse tree indicating word dependency structure. We speculate that the probability of a parse tree, given a sentence, is related to our notion of the total sum of the potential energies of each word within the sentence for a given configuration. We, therefore, solve the problem

$$\text{Configuration}_{\text{best}} = \underset{\text{configuration}}{\operatorname{argmax}} \sum_{i=1}^{N} U_i(\text{configuration, sentence}). \tag{4}$$

A. Initial Ranking of Configurations

As noted previously, the configuration maps directly to a dependency tree. The key to the approach is the postulate that a parse tree can be viewed as a configuration of words within a sentence matrix, with each word i experiencing a potential energy U_i (configuration) at a word position defined by the configuration. The potential energy of the word in the proposed configuration, then, determines its positional stability.

In this paper, we will estimate the potential energy of each reference word within a sentence and a proposed parse configuration based solely on the empirical statistics compiled manually for our Type I and Type II interaction tables (see Section V). Development of an improved statistical model is left for future work. The parser algorithm is as follows.

1) Conservatively identifying possible attachment sites for each word. This step identifies all *possible* Type I or Type II link instances within a given sentence in a conservative fashion emphasizing very high recall (ideally strictly 100%) over precision. The global syntactic–semantic dependency network is used to guide this process. Fig. 6 shows an example of such a possibility map.

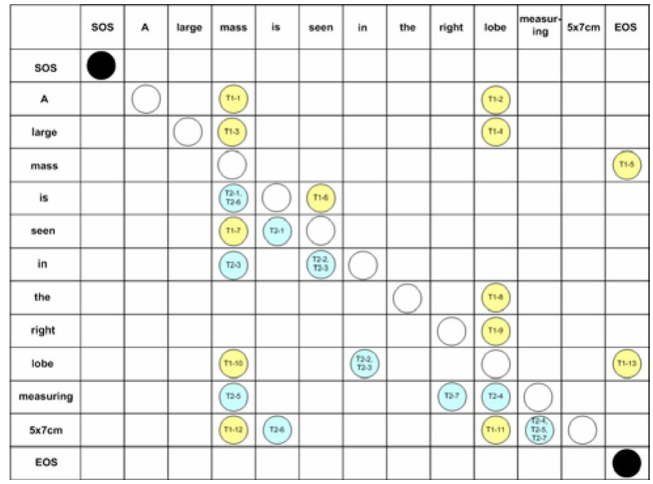

Fig. 6. Possibility space for Type I and Type II linkages.

2) Apply some conservative grammar rules to eliminate obvious false positive link proposals; for example, an article (e.g., the) cannot link past another article.
3) Initialize global minimum energy value to zero.
4) Based on the possibility map, propose a configuration for the words within the matrix. Note that, in general, there is an exponentially large number of such configurations. An estimate of the number of such configurations for dependency parsing can be found in [49].
5) Given now a configuration and the lexical analysis information (see Table I), compute the global energy (i.e., sum of the individual word energies) associated with the configuration. For Type 1 link propositions, the left context word, the reference word, the target head word, and the right context words are identified as described previously using the rules of virtual adjacency. (This amounts to a Markov blanket assumption for both the reference word and the target head word.) Given the four-gram(s), the reference word energy is estimated using the Type I interaction probability tables compiled previously. Energies of reference words associated with Type II links are similarly computed from five-gram context. (This amounts to a Markov blanket assumption for the reference word, the central relational word, and the target head word.)
6) If the global energy for this configuration is less than the current global minimum energy value, a note is made of this configuration and the global minimum value is updated. If the global energy is the same as the current global minimum energy, this configuration is stored within a list of top candidate parses.
7) Check for stopping conditions. If not met, go to step 4.
8) For sentences that include a relatively sparse possibility space, an exhaustive search of all possible configurations is performed. For sentences with prohibitively large number of configurations (e.g., long sentences with many

conjunctions and prepositions), a genetic algorithm is applied. For speed considerations, we limit the search to 500 generations, which, in our preliminary investigations, have lead to good convergence as compared to the exhaustive search. If the stopping conditions are satisfied, the algorithm lists the top scoring (i.e., minimum energy) parse configurations.

B. *Ambiguity Resolution*

In some sentences, the four- and five-grams are not sufficient to unambiguously identify the minimum energy well position for a given reference word. For example, in the sentence, "*The bright density adjacent to the mass is increasing*," the phrase "*is increasing*" could possibly attach to the word *mass* or to the subject *density*. Logically, this particular ambiguity arises because of the fact that the mass is used as a reference location. In order to assess the alternative link propositions, we essentially need to determine which of the following sequences is most likely:

$$\text{Density} \leftarrow \text{adjacent} - \text{to} \leftarrow \text{mass} \leftarrow \text{is} \leftarrow \text{increasing}$$
$$\text{Density} \qquad\qquad\qquad\qquad \leftarrow \text{is} \leftarrow \text{increasing}$$

Alternative pathways between two terminal words, thus, need to be ranked in these instances (i.e., one needs to find a common terminal head ancestor and assess which sequence of the link paths is most likely). Within our training corpus, we mine for these ambiguities (i.e., degenerate configurations) and again ask taggers to assign interaction modes for higher order n-grams. For example, in the ambiguity defined earlier, we present the following four-gram using higher order constituents:

density	(adjacent-to, mass)	(is, increasing)	EOS
Left Hill	Left Well Token	Reference Token	Right Hill

In this four-gram, we again want to gather the information whether the interaction type of the reference token (now the constituent (is, increasing)) with the target attachment site (now the constituent (adjacent-to, mass)) is of the mode: 1) no interaction (i.e., a transmission of signal); 2) repulsion; or 3) attraction. In this case, the correct marking for this four-gram would be transmission. This information can then be used to disambiguate between the two alternative parses seen earlier. Thus far, we have mined 9441 such patterns.

VII. Evaluation

A preliminary evaluation was conducted in order to estimate the recall and precision statistics for the task of dependency parsing. The evaluation procedure was guided by recommendations within [51] and [52], and is summarized as follows.

1) *Identification of Test Sentences*: First, all sentences originating from the "Findings" section of the reports from our test partition of 1000 reports were identified and pooled. We then randomly selected 300 sentences for evaluation. Any sentence that was less than ten word tokens was not selected.

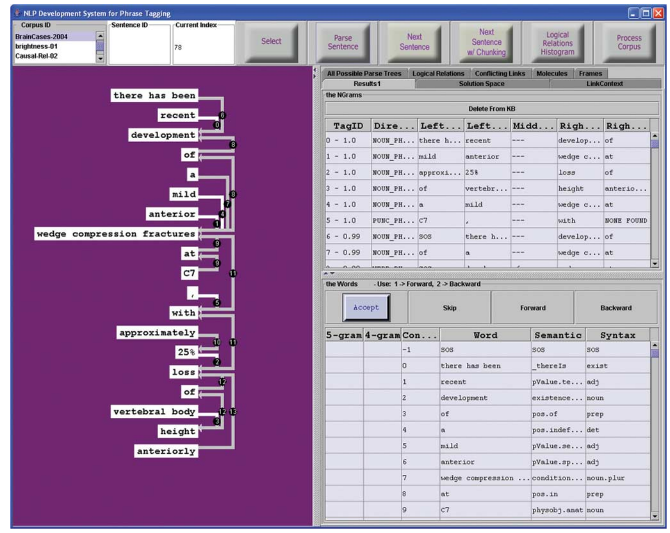

Fig. 7. Interface for visualizing and tagging parser links. The *left panel* shows the dependency tree, including Type I and Type II linkages. The *bottom right panel* lists the part of speech and semantic classes of individual words. The *upper right panel* shows the local field conditions for each reference word.

2) *Truth Determination*: A graphical tagging interface was created (Fig. 7) that allowed a human expert to assign dependency links between words. In this preliminary evaluation, a single tagger not involved with the development assigned all the link tags. Two other individuals reviewed the final dependency trees for each sentence. A committee involving the original tagger and two reviewers, then, either resolved their differences as to the correct parse or threw out the sentence entirely. Approximately 14% of the tagged examples were eliminated by the committee. New sentences were randomly selected to replace those thrown out. The final number of the tagged sentences was 306.
3) *Perform NLP*: The test sentences were then processed by the NLP system. For each sentence, the system recorded the Type I and Type II links associated with the final parse.
4) *Evaluate Results*: The NLP output was compared to the gold standard truth results. Performance was summarized in terms of: 1) recall defined by the ratio of the number of correct links and the number of all links as defined by the truth and 2) precision defined by the ratio of the number of the correct links and the number of defined links.

The average sentence length for the test sentences was 17.12 tokens; the average number of the unknown tokens per sentence was 0.086; the average number of conjunctions per sentence was 0.601; the average number of commas per sentence was 0.697; and the average number of prepositions per sentence was 1.858.

For Type I links, there were a total of 1825 links in the truth-determined parse trees. The recall performance was 95.01% while that of the precision it was 92.52%. For Type II links, there were a total of 1926 links in the truth-determined parse trees. The Type II recall performance was 75.49%; the precision was 87.06%. The combined Type I and Type II link performance statistics were as follows: recall, 84.99%; precision, 89.95%. The truth set had a total of 3751 links. A total of 141 out of the possible 306 sentences were parsed as completely correct (46.1%). The average processing time per sentence on a 2.0-GHz PC with 1 GB of memory was 15.1 s; there was no attempt to optimize the processing efficiency of the code.

False negative errors could be categorized according to the following knowledge deficits:

1) unknown words—currently, there is no mechanism for guessing the part-of-speech or semantic tag labels of unknown words;
2) idiomatic expressions;
3) terse reporting styles—for example, when punctuation, verbs, or conjunctions were left out (e.g., "a small catheter," a thoracotomy tube, left pleural space);
4) sentences or major clauses that begin with a prepositional phrase, as the system assumed that the heads of prepositional phrases were to the left of the preposition;
5) long sentences that included various connectives (e.g., but, so, nevertheless, however, etc.) were often left disjoint;
6) sentences that were dictated using poor grammar: "*The visualized osseous structures demonstrated full rib* ***destructive*** *as described above as well as osteopenia*."

False positive errors were seen in sentences that contained a number of prepositions where attachments were perceived as ambiguous. For example, *Small tubular areas of enhancement, isodense to vessels in lateral segment left lobe of liver, unchanged from prior studies and most likely a vascular anomaly/aberrhent vessel*. These types of ambiguities will require higher order semantic and discourse models in order to be resolved appropriately. Work on a global syntactic–semantic dependency network is a part of our future plans. Other types of errors were due to inadequacies in the shielding rules used to determine which words were allowed to communicate in Type I and Type II linkages.

VIII. Discussion

This paper presents initial exploratory work on the development of a natural language parser for medical reports that attempts to assign a stability metric of a reference word within a given sentence and proposed parse configuration. The parser is inspired by a physics-based model of word–word interaction and relies heavily on compiling the interaction statistics related to the notion of how much work is required to bring a reference word from infinity into a link position within a system of three (Type I) or four (Type II) other words. Our motivation for this approach is to develop a mechanistic paradigm toward medical NLP in which a community of developers can contribute to compiling these interaction tables. This, of course, relies heavily upon whether we can objectively agree on how to label such four-gram, five-gram, and higher order systems of words, and the degree to which we can minimize the tagging variability.

A. Relation to Other Work

It should be noted that there have been many hints to abstracting the NLP parsing problem to a physics-based paradigm: Befferman conceptualizes a lexical attraction and repulsion model in [53]; Yuret identifies word–word relations using a lexical attraction model [49]. The notion of the barrier words is used for phrase chunking by [54]. These analogies to physics systems, however, have been remotely and/or superficially explored.

Interestingly, the bioinformatics community working on the problem of protein folding have adopted many of the language-processing techniques developed in computational linguistics [55]; however, the mechanistic views of chemistry and physics that dictate protein-folding configurations have not been conceptualized and/or abstracted to the problem of NLP. This is an exploratory attempt to move in this direction and adapt some of the computational machinery already well developed in this area. The proposed field theory approach is our initial conceptualization of a causal approach to the problem of the sentence parsing that attempts to perform global optimization based on the features that span word level, word sequence level, link level, and logical relation level evidence.

B. Desiderata Revised

1) *Adaptability to New Domains*: Our long-term goal is to develop a global optimization model for medical language. This model must necessarily have many levels of evidence including syntactic–semantic, discourse, and pragmatic. Learning such a model will require a substantial amount of effort in terms of training examples and knowledge engineering for a comprehensive domain model to be developed. The limitation of this paper includes the need to hand tag and curate and large number of four- and five-grams that represent the minimal context for estimating the field conditions. Also recall that for link ambiguity, larger constituents were used in the definition of a word particle. Additionally, we define response functions for every class of unique semantic link types. Thus, the system can, likely, work within a limited domain such as radiology reports, but would require cooperative efforts to expand to larger information space domains.
2) *Robustness to Unseen Patterns*: Our system includes a relaxation method within the semantic description dimension for dealing with new four- and five-grams (see Table III). Statistical smoothing and clustering models are planned as part of the future work [46]. Current symbolic medical NLP approaches lack this ability to reason in this respect.
3) *Guidance from Higher Order Knowledge*: General NLP systems have traditionally followed a bottom-up approach mainly relying on part-of-speech preterminal tags. Medical NLP system typically includes improved word-level semantic tags [56]. Our current system uses a knowledge base of the primitive logical relations to both transform parser links to a conceptual representation and perform disambiguation.
4) *Intuitive Features*: The approach currently is based on a "communicative model" in which features are related to ability to communicate with other words, potential hills, and potential wells.

5) *Mechanistic Model*: We have tried to formulate the parser problem as an energy minimization problem with the energy potential of a reference word conceptualized by the traditional physics definition shown in Fig. 3. This is, in contrast, to purely data-driven approaches that do not propose any causal mechanisms for word–word attachments.

C. Future Work

Our future work includes the following.

1) Developing improved smoothed interpolative models for Type I and Type II interaction statistics.
2) Performing a more careful analysis of the types of links that allow communication between words.
3) Further developing and formalizing a global semantic model to guide parsing. For example, the development of a more comprehensive knowledge source to define sanctioned versus forbidden types of links.
4) Optimizing the efficiency of the search for the global minimum configuration; currently, some sentences take time of the order of minutes for processing.
5) Development of algorithms for handling coordinate conjunctions. Some initial experiments that propose initial grouping of coordinate conjunctive phrases and assessment of the field conditions associated with the phrase as a constituent has been promising.
6) Continuing our exploration of how we can differentiate the energy states of degenerative configurations. Exploration of elementary tree structures within the n-gram model is being explored.

Acknowledgment

The authors would like to thank P. S. Cho and S. Soderland for their stimulating discussion on issues related to this paper. They would also like to thank C. Demise for her tireless tagging efforts.

References

[1] C. Friedman, S. M. Huff, W. R. Hersh, E. Pattison Gordon, and J. J. Cimino, "The Canon Group's effort: Working toward a merged model," *J. Amer. Med. Inform. Assoc.*, vol. 2, pp. 4–18, 1995.

[2] S. Bakken, S. Hyun, C. Friedman, and S. B. Johnson, "ISO reference terminology models for nursing: Applicability for natural language processing of nursing narratives," *Int. J. Med. Inform.*, vol. 74, no. 7/8, pp. 615–622, 2005.

[3] N. Sager, M. Lyman, N. T. Nhan, and L. Tick, "Medical language processing: Applications to patient data representation and automatic encoding," *Methods Inform. Med.*, vol. 34, pp. 140–146, 1995.

[4] C. Friedman, L. Shagina, Y. Lussier, and G. Hripcsak, "Automated encoding of clinical documents based on natural language processing," *J. Amer. Med. Inform. Assoc.*, vol. 11, no. 5, pp. 329–402, 2004.

[5] C. Friedman, P. Alderson, J. Austin, J. Cimino, and S. Johnson, "A general natural-language text processor for clinical radiology," *J. Amer. Med. Inform. Assoc.*, vol. 1, pp. 161–174, 1994.

[6] U. Hahn, M. Romacker, and S. Schulz, "MEDSYNDIKATE—A natural language system for the extraction of medical information from findings reports," *Int. J. Med. Inform.*, vol. 67, no. 1–3, pp. 63–74, Dec. 2002.

[7] G. Hripcsak, C. Friedman, P. O. Alderson, W. DuMouchel, S. B. Johnson, and P. D. Clayton, "Unlocking clinical data from narrative reports: A study of natural language processing," *Ann. Intern. Med.*, vol. 122, pp. 681–688, 1995.

[8] M. Lyman, N. Sager, L. Tick, N. Nhan, F. Borst, and J.-R. Scherrer, "The application of natural-language processing to healthcare quality assessment," *Med. Decision Making*, vol. 11, no. 4, pp. S65–S68, 1991.

[9] L. M. Christensen, P. J. Haug, and M. Fiszman, "MPLUS: A probabilistic medical language understanding system," in *Proc. Workshop Nat. Lang. Process. Biomed. Domain*, Philadelphia, PA, Jul. 2002, pp. 29–36.

[10] C. Friedman, G. Hripcsak, L. Shagina, and H. Liu, "Representing information in patient reports using natural language processing and the extensible markup language," *J. Amer. Med. Inform. Assoc.*, vol. 6, pp. 76–87, 1999.

[11] G. Hripcsak, J. H. M. Austin, P. O. Alderson, and C. Friedman, "Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports," *Radiology*, vol. 224, pp. 157–63, 2002.

[12] R. K. Taira and S. A. Soderland, "A statistical natural language processor for medical reports," in *Proc. AMIA Fall Symp.*, Washington, DC, Nov.1999, pp. 970–974.

[13] R. K. Taira, S. Soderland, and R. Jakobovits, "Automatic structuring of radiology free text reports," *Radiographics*, vol. 21, pp. 237–245, 2001.

[14] L. Tesnière, *Elements de Syntaxe Structurale*, 2nd ed. Paris, France: Klincksieck, 1965.

[15] R. C. Schank, "Conceptual dependency: A theory of natural language understanding," *Cogn. Psychol.*, vol. 3, pp. 552–631, 1972.

[16] R. C. Schank and L. G. Tesler, "A conceptual parser for natural language," in *Proc. IJCAI-69*, Bedford, MA, pp. 569–578.

[17] I. Melčuk, *Dependency Syntax: Theory and Practice*. Albany, NY: State Univ. New York Press, 1988.

[18] D. D. Sleator and D. Temperley, "Parsing English with a link grammar," Carnegie Mellon Univ., School Comput. Sci., Pittsburgh, PA, Tech. Rep. CMU-CS-91-196, 1991.

[19] A. H. Bond, "A psycholinguistically and neurolinguistically plausible system-level model of natural-language syntax processing," *Neurocomputing*, vol. 65, pp. 833–842, 2005.

[20] S. Small, "Word expert parsing: A theory of distributed word-based natural language understanding" Ph.D. dissertation, Comput. Sci. Dept., Univ. Maryland, Tech. Rep. TR-954, 1980.

[21] S. Small and C. Rieger, "Parsing and comprehending with word experts (A theory and its realization)," in *Strategies for Natural Language Processing*, W. G. Lehnert and M. H. Ringle, Eds. Hillsdale, NJ: Erlbaum, 1982, pp. 89–147.

[22] U. Hahn, S. Schacht, and N. Broker, "Concurrent, object-oriented natural language parsing: The parsetalk model," *Int. J. Human–Comput. Stud.*, vol. 41, pp. 179–222, 1994.

[23] A. V. Aho and J. D. Ullman, *The Theory of Parsing, Translation, and Compiling. Vol 1: Parsing*. Englewood Cliffs, NJ: Prentice Hall, 1972.

[24] J. E. Hopcroft and J. D. Ullman, *Formal Languages and Their Relation to Automata*. Menlo Park, CA: Addison-Wesley, 1979.

[25] E. Charniak, *Statistical Language Learning*. Boston, MA: MIT Press, 1994.

[26] J. Eisner, "Bilexical grammars and a cubic-time probabilistic parser," in *New Development in Natural Language Parsing*, H. C. Bunt and A. Nijholt, Eds. Boston, MA: Kluwer, 2000.

[27] S. Geman and M. Johnson, "Probability and statistics in computational linguistics, a brief review," in *Mathematical Foundations of Speech and Language Processing* (IMA Volumes in Mathematics and its Applications), vol. 138, M. Johnson, S. Khudanpur, M. Ostendorf, and R. Rosenfeld, Eds. New York: Springer-Verlag, 2003, vol. 138, pp. 1–26.

[28] R. Ge and R. J. Mooney, "A statistical semantic parser that integrates syntax and semantics," in *Proc. 9th Comput. Nat. Lang. Learn.*, Univ. Michigan, Ann Arbor, MI, Jun. 29–30, 2005, pp. 9–16.

[29] W. D. Bidgood. Jr., "Documenting the information content of images," in *Proc. AMIA Annu. Fall Symp.*, Washington, DC, Nov. 1997, pp. 424–428.

[30] D. Gardner, Y. Xiao, M. Abato, K. Knuth, and E. Gardner, *BrainML and GENIE: Neuroinformatics Schemas for Neuroscience Data Sharing*. vol. 28, Washington, DC: Society for Neuroscience Abstracts, 2002.

[31] A. Gupta, B. Ladascher, J. S. Grethe, and M. E. Martone, "Towards a formalization of disease-specific ontologies for neuroinformatics," *Neural Netw.*, vol. 19, no. 9, pp. 1277–1292, 2003.

[32] D. Grinberg, J. Lafferty, and D. Sleator, "A robust parsing algorithm for link grammars," presented at the 4th Int. Workshop Parsing Technol., Prague, Czech Republic, Sep. 1995.

[33] D. Klein and C. D. Manning, "Accurate unlexicalized parsing," in *Proc. 41st Meeting Assoc. Comput. Linguistics*, Sapporo, Japan, 2003, pp. 423–430.

[34] Y. Huang, H. J. Lowe, D. Klein, and R. J. Cucina, "Improved identification of noun phrases in clinical radiology reports using a high-performance

statistical natural language parser augmented with the UMLS specialist lexicon," *J. Amer. Med. Inform. Assoc.*, vol. 12, no. 3, pp. 275–285, 2005.
[35] P. Szolovits, "Adding a medical lexicon to an English parser," in *Proc. AMIA Fall Symp.*, Washington, DC, 2003, pp. 639–643.
[36] F. M. Hall, "Language of the radiology report: Primer for residents and wayward radiologist," *Am. J. Roentgenol.*, vol. 175, pp. 1239–1242, 2000.
[37] F. J. Och, D. Gildea, S. Khudanpur, A. Sarkar, K. Yamada, A. Fraser, S. Kumar, L. Shen, D. Smnith, K. Eng, V. Jain, Z. Jin, and D. Radev, "A smorgasbord of features for statistical machine translation," in *Proc. Human Lang. Technol. Conf. N. Am. Chap. Assoc. Comput. Linguistics (HLT/NAACL)*, Boston, MA, May 2–7, 2004, pp. 161–168.
[38] M. Collins and T. Koo, "Discriminative reranking for natural language parsing," *Comput. Linguistics*, vol. 31, no. 1, pp. 25–70, 2005.
[39] R. Huddleston, *Introduction to the Grammar of English*. Cambridge, U.K.: Cambridge Univ. Press, 1984.
[40] V. Bashyam and R. K. Taira, "A study of lexical behavior of sentences in chest radiology reports," in *Proc. AMIA Annu. Symp.*, Washington, DC, Oct.2005, p. 891.
[41] A. A. Bui, G. S. Weinger, S. J. Barretta, J. D. N. Dionisio, and H. Kangarloo, "DataServer: An XML gateway for medical research applications," *Ann. N. Y. Acad. Sci.*, vol. 980, pp. 236–246, 2002.
[42] A. A. Bui, J. D. N. Dionisio, C. A. Morioka, U. Sinha, R. K. Taira, and H. Kangarloo, "DataServer: An infrastructure to support evidence-based radiology," *Acad. Radiol.*, vol. 9, pp. 670–678, Jun. 2002.
[43] R. K. Taira, A. A. Bui, and H. Kangarloo, "Identification of patient name references within medical documents using semantic selectional restrictions," in *Proc. AMIA Fall Symp.*, Washington, DC, Nov. 2002, pp. 757–761.
[44] D. Marcu, *The Theory and Practice of Discourse Parsing and Summarization*. Boston, MA: MIT Press, 2000, p. 241.
[45] P. F. Brown, V. J. Della Pietra, P. V. deSouza, J. C. Lai, and R. L. Mercer, "Class-based n-gram models of natural language," *Comput. Linguistics*, vol. 18, no. 4, pp. 467–479, 1992.
[46] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 2000.
[47] D. Magerman, "Statistical decision-tree models for parsing," in *Proc. 33rd Annu. Meeting Assoc. Comput. Linguistics*, Cambridge, MA, 1995, pp. 276–283.
[48] F. Jelinek, J. Lafferty, D. Magerman, R. Mercer, A. Ratnaparkhi, and S. Roukos, "Decision tree parsing using a hidden derivation model," in *Proc. Human Lang. Technol. Workshop*, Plainsboro, NJ, 1994, pp. 272–277.
[49] D. Yuret, "Discovery of linguistic relations using lexical attraction" Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, May 1998.
[50] J. H. Holland, *Adaptation in Natural and Artificial Systems. An Introductory Analysis With Applications to Biology, Control and Artificial Intelligence*. Ann Arbor, MI: Univ. Michigan Press, 1975.
[51] *Proceedings of the Sixth Message Understanding Conference (MUC-6)*. San Francisco, CA: Morgan Kaufmann, 1995.
[52] C. Friedman and G. Hripcsak, "Evaluating natural language processors in the clinical domain," *Methods Inform. Med.*, vol. 37, pp. 334–344, 1998.
[53] D. Beeferman, A. Berger, and J. Lafferty, "A model of lexical attraction and repulsion," in *Proc. ACL-EACL'97 Joint Conf.*, Madrid, Spain, 1997, pp. 373–380.
[54] G. W. Moore, R. E. Miller, and G. M. Hutchins, "Indexing by MeSH titles of natural language pathology phrases identified on first encounter using the barrier word method," in *Computerized Natural Medical Language Processing for Knowledge Representation*, J. R. Scherrer, R. A. Cote, and S. H. Mandil, Eds. Amsterdam, The Netherlands: Elsevier, 1989, pp. 29–39.
[55] D. B. Searls, "The language of genes," *Nature*, vol. 420, pp. 211–217, 2002.
[56] P. Spyns, "Natural language processing in medicine: An overview," *Methods Inform. Med.*, vol. 35, no. 4/5, pp. 285–301, 1996.

Ricky K. Taira, photograph and biography not available at the time of publication.

Vijayaraghavan Bashyam, photograph and biography not available at the time of publication.

Hooshang Kangarloo, photograph and biography not available at the time of publication.