Lexically-Driven Natural-Language Access To The INTERNIST-I Knowledge Base

David A. Evans and Sandra Katz

December 1986

Report No. CMU-LCL-86-4

Laboratory for Computational Linguistics

139 Baker Hall
Department of Philosophy
Carnegie Mellon University
Pittsburgh, PA 15213

Contents

List of Figures

APPENDICES

Appendix 1: Phrasal Entries in the Lexicon

Appendix 2: List of Bound Morphemes in the Lexicon

Appendix 3: Examples of Lexical Mapping

Appendix 4: List of INTERNIST-I Terms According to Frequency

Appendix 5: A Proposal to Enhance the Electronic Textbook

INTRODUCTION

The following is a final report on our first efforts to produce a practical natural-language interface to the INTERNIST-I knowledge base as embodied in the QUICK MEDICAL REFERENCE (QMR) system.[1] The work reported on in this document was supported in large measure by a grant from Paul Mongerson and the CAMDAT Foundation, Inc., in response to a proposal to enhance a PC-based version of the INTERNIST-I system.[2] The original proposal is appended to this document as Appendix 5 for reference and for more detailed information about the objectives of our efforts. Additional support was derived from a grant from the National Library of Medicine (MedSORT-II) and by a grant from the Josiah Macy. Jr., Foundation.

Short of complete natural-language understanding, any application designed to develop a natural-language front end to a computerized system must decide what constraints in performance will be tolerable and what balance to strike between efficiency in processing and completeness in natural-language coverage. In our case, especially because we were restricted to the development of a system that would operate on a PC/AT under Turbo Pascal, we chose to develop a method of sophisticated string matching rather than attempt to develop a complete natural-language processing system. The emphasis in our work then has been on the creation of a detailed lexicon/thesaurus and on an analysis of the structure of concepts in the knowledge base we wish to access. The results are a practical but limited system that nevertheless affords users considerable flexibility in interacting with the QMR system. This technical report describes the components of our system and the considerations that have informed their design.

Much of the initial part of the report, discussing the design of the lexicon and the principles of semantic mapping we exploit, reproduces material that appears in [Evans and Katz, to appear]. In addition, this report discusses the actual scoring algorithm we have used in a prototype system to access findings in the INTERNIST-I knowledge base;[3] and also includes several appendices containing detailed information on features of the lexicon and prototype system performance.

---

[1] QMR is under development by Drs. R.A. Miller, F.E. Masarie, and J.D. Myers and discussed in Miller *et al.* 1986a,b.

[2] In its original version, the PC-based system was referred to as the "Electronic Textbook." It has subsequently been named the QUICK MEDICAL REFERENCE, or QMR system.

[3] Findings are clinically significant observations, usually involving several atomic biomedical concepts, that contribute to the identification of diagnoses. In the INTERNIST-I knowledge base, findings are subclassified further as *patient-history, patient-symptoms, physical-exam-signs,* and *laboratory-test-results.*

LEXICALLY-DRIVEN NATURAL-LANGUAGE ACCESS TO THE INTERNIST-I KNOWLEDGE BASE

David A. Evans and Sandra Katz

1. Constrained Natural-Language Processing.

With the proliferation of large-scale electronic databases and knowledge-based expert systems of all kinds, there is increasing demand for *practical* natural-language processing (NLP). But what constitutes *practical* natural-language processing? The straightforward answer is ***any NLP that is both efficient and sufficient to the task at hand.*** **Of course, this begs two** further questions: What determines efficiency? What sufficiency? Different systems will impose different *specific* requirements on the answers to these questions, but, in general, efficiency will be measured by the speed and accuracy of processing, sufficiency by the ability to accomplish user-determined goals. For example, in searching through very large-scale bibliographic databases (with millions of records), efficiency may be judged high if the system returns all the relevant citations—and few irrelevant ones—in thirty seconds. Thus, *real-time* 'dialogue' may not be an important criterion. The same system might be judged sufficient if it succeeds in mapping most user-formulated expressions of goals—a desire for articles on a certain topic, for example—into representations and actions that facilitate efficiency.

To effect such performance, *practical* NLP almost certainly implies *incomplete* or, at least, highly *constrained* NLP. This is made feasible, in part, by the explicit control of the contexts of man-machine interaction imposed by many systems. For example, a system can announce that it is prepared to accept any user-entered statement as the identification of a topic, to be used in searching for bibliographic citations; and it could control changes of state by always requiring that the user choose from among an explicit set of alternative interaction contexts ***(e.g., add-qualification-to-topic, browse, start-again, etc.)*** **via a menu of choices. In the case** of the system we describe in this paper, practical NLP has been limited to *lexical-semantic processing.* It has been possible to ignore discourse-level pragmatics, including phenomena associated with turn-taking, speech acts, implicature, presupposition, inference, and even reference; and, to a large extent, to ignore syntax *as* well. The important questions, of course, concern the adequacy and generality of our approach.

We suggest that, for well-structured knowledge bases in restricted domains, the relatively simple semantic network of a natural-language lexicon can provide the basis for effective NLP. In the balance of this paper, we present a description of the domain we have investigated, a discussion of the structure of the lexicon that resides at the heart of our system, examples of system performance, and reflections on the potentials and limitations of our approach.

2. INTERNIST-I as a Constrained Semantic Domain.

INTERNIST-I and its progeny (including CADUCEUS and the QMR system) are a family of more-or-less expert systems, designed for consultation in developing diagnoses in internal medicine. (Cf. [Masarie 1985], [Miller *et al.* 1982, 1986a,b], for information on the QMR System; [Pople 1982] for discussion of CADUCEUS.) By many measures, INTERNIST-I is one of the most ambitious and successful medical expert systems yet developed. The system's knowledge base currently contains information about 570 diseases and more than 4000 associated findings, with each diagnosis 'profiled' in terms of (typically) dozens of findings.

A chief obstacle to the widespread use of systems like INTERNIST-I is the difficulty new users have in understanding the fine distinctions among clinical observations that the system encodes in its many findings. Almost every physician who might use the system as an aid in formulating diagnoses knows *what* the clinically important observations are that must be communicated to the system; almost no one is prepared to learn enough about the idiosyncratic distinctions in the system's understanding of findings to learn *how* to do so efficiently.

Our efforts have been directed toward building a natural-language interface to the INTERNIST-I knowledge base as embodied in the QMR system (cf. Miller *et al.* [1986a, b]). The special focus of our work has been on developing a means of interpreting user-entered, natural-language characterizations of clinical observations as one or more of the clinically significant findings that the system is prepared to accept. While we can require that whatever the user enters be intended to identify a finding, we cannot prevent the user from using 'non-standard' or vague terminology or from formulating incomplete expressions. Thus, our special problems have focused on mapping non-canonical expressions to candidate canonical ones, and on identifying sources of vagueness and incompleteness, to direct the user's attention to the appropriate contrasts that must be drawn.

Since developing a list of findings is prelimiary to performing the many analytical tasks in the system, there is a great premium on ease of entry and speed of processing. And while there are numerous other possible modes of user-system interaction, entering findings presents the greatest problem for new users because of the rigidity of form and stringency of precision in the system's representations of findings. As an example of the problem, consider the following findings (as presented in Pople [1986:272ff]):

- **SKIN NODE <S> ARM <S> EXTENSOR AREA <S>**
- **SKIN SUBCUTANEOUS NODULE <S> NEAR OLECRANON**

It would be relatively easy for a user who wanted to describe a skin lesion near the elbow to choose either one of these and feel satisfied that the correct identification had been made. However, in the system, the second is recognized as having special significance in the identification of rheumatoid arthritis, the first is not. The situation is, of course, more complicated when the user states the condition in natural language. The apparently adequate

description, *skin lesion near the elbow*, would have to map to both findings and offer the user a choice.

To take another example (following Evans [1986:285]), the general description of a patient as having a *gallop rhythm* would be insufficient to determine an *internist-i* finding, since the following findings form a contrast set:

- HEART SOUND <S> S3 LEFT VENTRICULAR GALLOP
- HEART SOUND <S> S3 RIGHT VENTRICULAR GALLOP
- HEART SOUND <S> S4 LEFT ATRIAL GALLOP
- HEART SOUND <S> S4 RIGHT ATRIAL GALLOP
- HEART SOUND <S> SUMMATION GALLOP LEFT VENTRICULAR
- HEART SOUND <S> SUMMATION GALLOP RIGHT VENTRICULAR

If we take our central problem to be that of effecting a semantic mapping between two domains—the user's domain of knowledge of internal medicine and the system's domain of findings—what do we take as our fixed field of reference? For our purposes, the answer has been a natural-language lexicon that reflects the special semantic/pragmatic structure of the INTERNIST-I knowledge base. As a *natural-language* lexicon, it conveniently serves as a filter on user-generated expressions, while making explicit just those relations that are recognized in the knowledge base.

To illustrate the structure imposed by a lexicon that is derived from a well-defined domain such as the INTERNIST-I knowledge base, we should consider the practical problem of building a lexicon *just for the terms that appear in the knowledge base itself.* In *form*, findings such as the ones above seem to be phrases of English words interlaced with special symbols (*e.g.*, "<S>"). In fact, while there are some 5000 terms employed in representing findings in INTERNIST-I, all findings are treated as atomic, so individual terms (such as "*skin*" and "*arm*" in the examples above) have no interpretation. Thus, if one wishes to create a lexicon of terms, an immediate problem is to decide what relations to assign various terms so that contexts from which they were removed (*viz.*, the holistic findings) can be restored when necessary. One approach—indeed, our own—is to decompose the findings into a network of frames that identify individual terms while assigning explicit relations among them. In effect, we identify the *concepts* that comprise the findings when we do this; and any lexicon we develop will take the frame-based network as its semantic domain.

As an example of the results of such a decomposition, consider the following set of INTERNIST-I findings involving *dyspnea*, i.e. shortness of breath:

- DYSPNEA
- DYSPNEA ABRUPT ONSET

- DYSPNEA ACUTE RECURRENT ATTACK HX
- DYSPNEA AT REST
- DYSPNEA AT REST RELIEVED BY RECUMBENCY
- DYSPNEA EXERTIONAL
- DYSPNEA IMPROVEMENT AFTER HEMOPTYSIS HX
- DYSPNEA PAROXYSMAL NOCTURNAL

A corresponding frame-based representation might be given as follows:

```
Frame: dyspnea

Clincial-Time-Course-Duration: acute,
                               paroxysmal
Clincial-Time-Course-Variability: abrupt-onset,
                                  recurrent-attack-hx,
                                  nocturnal,
                                  improvement-after-hemoptysis-hx
Patient-Circumstance: at-rest,
                      relieved-by-recumbency,
                      nocturnal
```

In such an exercise, it is apparent that decomposing the findings does not yield individual *terms*, rather, unit *concepts*, which may be expressed as phrases of terms. Hence, the key to mapping natural-language expressions into relevant candidate findings depends on the identification of appropriate *clusters* of concepts.

When we examine the entire knowledge base in this fashion, we discover numerous contrast-sets, corresponding to the partial semantic fields in which concepts are arrayed. The semantic fields, in turn, reflect the super-structure of the frame-based representation schemata—as seen in the names of the slots on the frames for findings, for example. In the frame given above, *Patient-Circumstance* defines a semantic field, partially populated by the concepts associated with *at-rest, relieved-by-recumbency*, and *nocturnal*

One difficulty in exploiting the structure of semantic fields involves controlling concept links, which, in turn, requires paying careful attention to the way concepts are organized across semantic fields. As an illustration, consider two examples of potentially bad consequences of adding information to the lexicon that is not immediately represented in the findings, as when we add synonyms and synonym-like terms to the canonical vocabulary.

The first potential problem stems from the fact that terms that are near in meaning may sometimes differ in degree of generality, leading to a difference in potential for linking

to other terms. If we supplement the term *finger* with the terms *digit, index-finger,* and *pinky* without specifying further structure, we invite each of these to be linked, in turn, to terms that are closely related to any of the set. In particular, because *digit* can be linked to *toe* (as a synonym), *pinky* can be related to *toe,* which is not desirable. We could simply declare synonymy to be a reflexive, jut non-transitive relation; but that does not address the problem here. In fact, none 01 ihe relations among these examples reflects synonymy; rather, the terms differ in degree of generality on a classification scale, reflecting *is-a/kind-of* relations. Such relations establish new contrast-sets, which may not always reflect useful distinctions in the knowledge base. Thus, *finger* and *toe* are in the contrast set defined by ***kind-of-digit]*** **and** ***index-finger*** **and** ***pinky*** **are in the contrast-set defined by** ***kind-of-finger.*** Neither of these contrasts has significance for medical diagnosis.

The second problem stems from simple polysemy. If we propagate the synonyms for the added term *digit,* we will generate, perhaps, *number,* a term that has a legitimate role to play in building certain diagnostically significant concepts, such as *acalculia.* But we would not want to relate *finger* and *number* through the term *digit.* Worse, we would not want to **relate** ***hand*** **to** ***acalculia*** **via a** ***finger-digit-number*** **link.**

The point these two cases underscore is that any supplemental information included in a lexical-semantic field, however legitimate, can undermine the coherence of the original semantic domain. We can avoid this infelicitous consequence by requiring that only semantic relations defined by the original domain be allowed; and by requiring that relational links be explored only among the canonical, or original, terms. In fact, the principles we exploit in designing the lexicon for our system insure that the semantic relations of the original domain are preserved.

3. The Practical Lexicon as a Semantic Network.

In the lexical processing of natural-language expressions in our system, there are, conceptually, only two important distinctions: we want to capture all the words that map onto concepts that are relevant in the semantic domain of findings and to ignore all others. We do this, effectively, by building the lexicon "backwards" from the terminology and atomic concepts in the findings, first by developing morphological variants of the canonical expressions; then by adding synonyms; then terms that bear an important (typically pragmatic or contrast-set) relation to the canonical expression. Any expression that can be analyzed as morphologically-identical to, synonymous with, or significantly-related to a canonical expression should have some potential for mapping to (or invoking) the canonical expression.

Consider the case of single canonical terms that correspond to atomic concepts, for example, the term *jaundice.* We would want any occurrence of *jaundiced,* a morphological variant, to have, potentially, the same semantic effect as the word *jaundice* itself. We would also want any occurrence of *icterus,* a synonym of *jaundice,* to have, potentially, the same semantic effect. Finally, we would want occurrences of *yellow, bilirubin, hyperbilirubinemia, hepatitis,* and *liver,* all to have some potential—not as great as that of synonyms—for having

similar semantic effects as *jaundice.* To summarize, we might develop the following lexical entry for the canonical term *jaundice:*

```
Item:    jaundice
Morph.:  j aundiced
Syn:     icterus
Rel-to:  yellow, bilirubin, hyperbilirubinemia,
         hepatitis, liver
```

Here, the *Morph:-fie\d* lists morphological variants; the 5yn;-field, synonyms; and the *Rel-*to;-field, expressions that may bear a special (but unspecified) relation to the lexical item. We should note that the terms that appear in the relational fields can, themselves, have morphological variation, synonyms, and other "related" expressions; and that we would want each of these, also, to have some (diminished) potential for mapping to *jaundice.* Thus, for every item introduced under *jaundice,* we can construct another lexical entry, such as the following for *icterus:*

```
Item:    icterus
Morph:   icteric, icteritious, icteroid
Syn:     jaundice
Rel-to:  icterogenic, icteroanemia, icterohematuric,
         icterohemoglobinuria, icterohepatitis
```

At this point, too, we could continue to develop further lexical items, but that would lead us down paths that are ever more remote from the canonical source, *jaundice.* Thus, at this level—once removed from the original concept—we stop. Recall that we are interested in exploiting the lexical-semantic network as a device to facilitate a mapping *from* natural-language *to* canonical forms. This highlights a design feature of our lexicon: networks have *directionality* and the end-points of all mappings are canonical expressions.

The end-point of a mapping need not be a single canonical term, of course; it can be a phrase, such as would be appropriate for the concept *resistance-to-flexion.* This concept appears in the finding NECK RESISTANCE TO FLEXION, the clinically significant observation more colloquially expressed as *stiff neck,* or in medical jargon, *nuchal rigidity.* We could give a lexical entry as follows:

```
Item:    resistance-to-flexion
Syn:     stiff, rigidity
```

For phrases, morphological variation is vacuous, of course, and any important actual natural-language variation, such as when the expression *resistant to flexion* is used, will be captured by the individual lexical items that comprise the phrase. In general, all terms in canonical phrasal expressions will be represented in the lexicon as canonical single terms, as well.

Thus, the items *resistance~to-flexion, resistance,* and *flexion* all would appear as canonical lexical entries. (Appendix 1 contains a listing of phrasal entries.)

How is directionality effected? We distinguish canonical terms from *derived* or non-canonical terms by labeling the former "Val: +" and the latter "Val: -." We compile the full lexicon from the list of canonical entries, each of which is marked "Val: +," by generating new entries (if required) and new links as specified by the relations in the canonical entries. The effects of this are manifold. First, all *new* entries are labeled "Val: -," and each contains at least one canonical item among its *Morph:, Syn:,* or *Rel-to:* fields. Second, multiple references to items result in additional relational links being built, whether between canonical and canonical terms or between canonical and non-canonical terms. For illustration, consider the effect of expanding the canonical entries, *jaundice* and *yellow* (which occurs in such findings as SKIN YELLOW BROWN DISCOLORATION). Initially, we would have the following:

```
Item:    jaundice
Val:     +
Morph:   jaundiced
Syn:     icterus
Rel-to:  yellow, bilirubin, hyperbilirubinemia,
         hepatitis, liver

Item:    yellow
Val:     +
Morph:   yellowish
Rel-to:  race-oriental, color, pigmentation, ochre, jaundice,
         colored, discolored, aurantiasis-cutis, caxotenemia
```

Under *jaundice,* the terms *liver, hepatitis, yellow,* and *bilirubin* are canonical; *icterus* and *hyperbilirubinemia* are not. Under *yellow,* the terms *pigmentation, race-oriental, colored,* and *jaundice* are canonical; *yellowish, color, ochre, discolored, aurantiasis-cutis,* and *carotenemia* are not. In generating the expanded lexicon, the following would be among the new items added:

```
Item:    yellowish
Val:     -
Morph:   yellow

Item:    color
Val:     -
Morph:   colored
Rel-to:  yellow
```

The entry for *yellowish* would have only one canonical anchor—*yellow.* The entry for *colored*

would have two—one for *yellow*, and one for *color*, presumably, because the canonical entry *color* would have "Morph: colored" as one of its specifications.

After morphological variants, the synonyms of a canonical form are most important for mapping non-canonical forms to the appropriate target. Besides the usual sorts of synonyms, we have included the following domain-specific types of "concept-equivalents" under the label *Syn.*:

```
Abbreviations:    E.K.G. <==> electro-cardiogram <==> EKG

Jargon:           functional <==> psychological
                  acute <==> severe (NOT => sharp)

Phrasal:          stiff-neck <==> nuchal-rigidity

Bound-Morphemes: dys= <==> difficulty, problem
```

The last category of synonyms, *bound morphemes*, introduces another feature of the lexicon that facilitates practical natural-language processing. [4] In a domain such as biomedicine with many Greek and Latin neologisms it is natural to find numerous concept-equivalents packaged in quite varying morphological wrappers. Users could very easily intend to indicate the clinically significant observation of *dysphagia* by any of a large number of phrases, including *trouble swallowing*, *difficulty swallowing*, *swallowing difficulty*, *problems with swallowing*, *etc.* To handle such variation, we have found it important to develop lexical entries for the more common bound morphemes among the canonical terms. Some examples are the following:

```
Item:   dys=
Val:    +
Compl:  =arthria, =esthesia, =pareunia, =phagia, =phonia,
        =pnea, =uria
Syn:    difficulty, defect, impaired, abnormal, perverted,
        lessened, problem, pain, trouble, decreased, bad

Item:   =phagia
Val:    +
Compl:  dys=, a=
Syn:    swallowing, drinking, eating
```

Along with bound morphemes we introduce a new field in the lexicon, *Compl:*, for "completer(s)." Among the completers, we list explicitly those bound morphemes that can com-

[4] We indicate the point of attachment of bound morphemes with the symbol "=". This serves for notational convenience in proofing the lexicon, but has no other processing significance. Indeed, the "=" is completely ignored.

bine with the present one to yield a canonical word. Effectively, this permits us to distinguish canonical *words* from canonical *terms.* A word is always a string of characters bounded by spaces; a term—since it is anchored to a canonical *concept*—can be a sub-string of characters in a word, a single word, or a phrase of words.

Note also that the entry for a bound morpheme can use any of the usual lexical fields *(Morph:, Syn:, Rel-to:)* as needed or appropriate. Thus morphological variation among bound morphemes, such as *a=/an=* or *hemo=/hem=,* could be recorded; and relations less direct than synonymy could be noted. In fact, we do not need either sort of information in our present application, as we utilize bound morpheme forms only as standardized links in the semantic network, not as active linguistic forms. Recall that we exploit bound morphemes to facilitate a mapping between a user's phrasal characterization of an observation when the canonical form is a single term. We do not use bound morphemes in the morphological analysis of user terminology. Hence, for example, even though we might record the bound-morpheme forms for *clot-of-blood (thrombo=)* and *destruction (=lysis),* we would not be able to recognize the non-canonical term *thrombolysis* as involving either of the bound morphemes. (Appendix 2 contains a listing of bound morphemes in the lexicon.)

While the information we have recorded in the lexicon is predominantly *semantic,* not all the peculiarities of the INTERNIST-I findings can be captured in strictly semantic relations. In particular, we have found it useful to include a field, *Add:,* in the lexical entry of some items that is designed to effect *pragmatic* distinctions, required in the proper formulation of selected INTERNIST-I findings. For illustration, consider the following problem in clinical observation.

At the level of symptoms, a valid observation might be that the patient reports weakness in his or her arm. But "arm weakness" is not a finding; it is merely the starting point from which further observations can be made to identify the clinical significance of the observation, with a potential to be realized as one or more of many quite different findings. For example, one would have to determine whether the weakness were due to a local phenomenon. If local, whether it were principally neurological, neuro-muscular, muscular, or related to the joints. If non-local, whether it were a symmetric or non-symmetric phenomenon; unilateral or bilateral; proximal (in the shoulders) or distal (in the forearm or hands). Furthermore, an apparently local phenomenon could be unrelated to an actual medical problem, as when the disorder is functional (i.e., psychological). Each potential refinement of the original observation calls quite different sets of findings to mind. Some require distinctions that are implicated in the normal use of English expressions in the medical domain, but not directly related to lexical semantics.

In the case of the example above, the use of the word *arm*—singular—might signal distinctions that are captured in INTERNIST-I by the addition of qualifiers such as *distal, unilateral.* or *asymmetric.* In contrast, the use of *arms*—plural—might suggest potential findings in which *bilateral* and *symmetric* are required. The following actual INTERNIST-I findings illustrate the contrast:

- **ARM <S> MUSCLE ATROPHY BILATERAL**
- **ARM <S> MUSCLE ATROPHY UNILATERAL**

It is unlikely that a user would add the term "unilateral" to an expression in which "arm" appeared, because to do so would be redundant. However, we want to make the potential distinction available to the system during the processing of user-formulated expressions, so we include under the lexical entries for "arm" and "arms" the terms "unilateral" and "bilateral," respectively, as follows:

```
Item: arm
Add:  unilateral

Item: arms
Add:  bilateral
```

The "add" field tells the system that when the word "arm" appears in the input string, "unilateral" should be paired with it, so that *arm. . .unilateral* findings will have a greater possibility of being retrieved than *arm . . .bilateral* findings.

In summary, then, we have used the lexicon to record canonicality of form, and pragmatic qualifications (special to the INTERNIST-I domain) for each lexical item. And we have recorded (1) morphological variation, (2) relevant synonymy, and (3) other useful relations among lexical items. As detailed in the following sections, such lexical information is sufficient to allow a "lexically-driven parser" to map user expressions to a small number of candidate, INTERNIST-I findings (often only one), effecting a practical natural-language front end to the INTERNIST-I knowledge base.

4. Examples of Constrained NLP.

By what route does the system map a user-generated expression of a finding to one of the findings that INTERNIST-I is prepared to accept? The first and perhaps most important stage is to build a semantic network of links between input terms and canonical terms. Each input "token" that the system recognizes is connected to one or several canonical terms via the main semantic relations in the lexicon: *morphological variant, synonym,* and *related-to.* To take a simple example, consider the brief string *patient has chills.* Reading each token from left to right, the system recognizes *patient* and *has* as general English terms ("Val 0" items), so they are not analyzed further. *Chills,* however, appears in the lexicon as a non-canonical ("Val -") item containing two types of links to canonical terms—"synonym" and "related-to." It is synonymous with *rigor* and related to *fever.* The latter can be further linked to several canonical terms via its "related-to" field: *degrees, heat, hot,* and *temperature.* We can represent this semantic network schematically as shown in Figure 1. The analysis of *chills* yields two semantic fields—one (single-item) field for *rigor* and a second one for

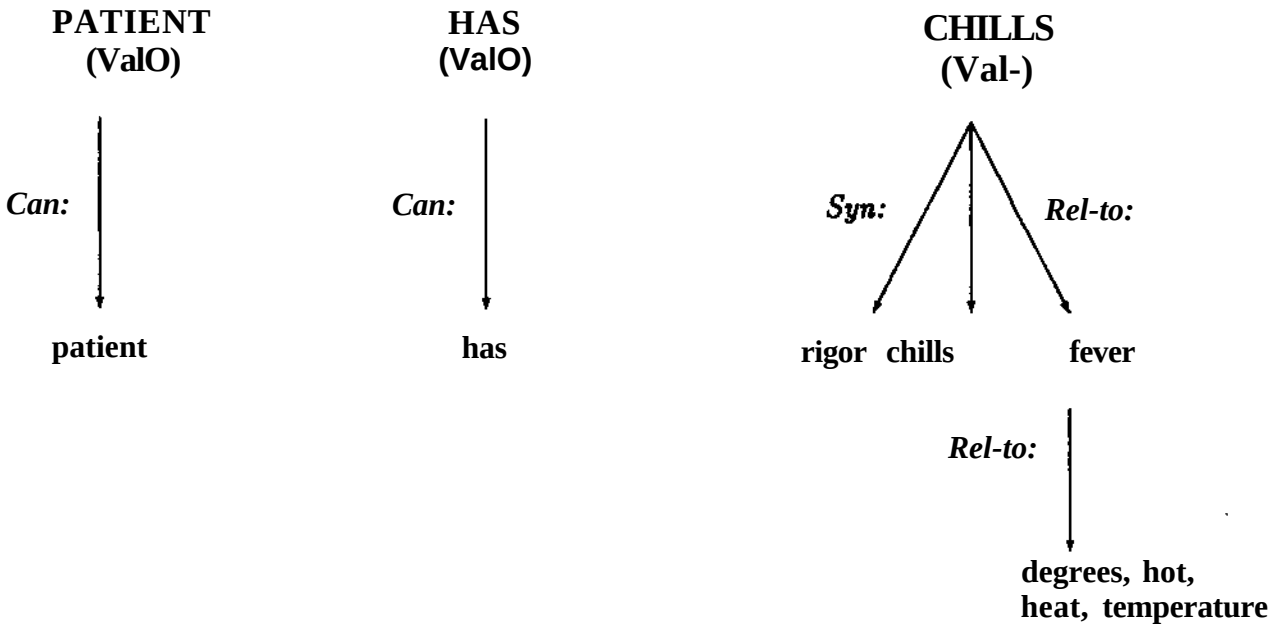

Figure 1: LEXICAL ANALYSIS OF *patient has chills*

fever **and all of the canonical terms associated with it. Even though the latter contains more members than the former, it stems from a weaker bond between items than the field for *rigor.* "Synonym" links are stronger than "related-to" links, so *rigor* will carry more weight in choosing among candidate INTERNIST-I concepts (findings) than will *fever* and, even less so, its associated terms. Consequently, the one-word finding *rigor* is returned as an interpretation of the user's observation.**

To gain a better understanding of how semantic fields associated with input terms converge as pointers to concepts in the INTERNIST-I knowledge base, consider another, more complex example, based on the string *stomach pains worse after meals,* as shown in Figure 2. Notice that the search for semantic fields associated with each input token takes place at two levels. First, the canonical terms associated with each token are retrieved, and then all of the canonical terms further associated with these items are retrieved. For example, once *meal* (the canonical term for *meals)* is found, all of the canonical terms listed in its "related-to" field are retrieved, yielding a more complete semantic field for *meals* than would be achieved by stopping the search at *meal.*

A variety of methods can be used to identify the 'best-fit' findings, once the lexical processing has been completed. The simplest involves a brute-force search of all finding-strings for co-occurrences of terms that appear in different branches of the lexical-semantic network, along with some method of scoring 'distance' from the user's original expression. One scoring scheme we have tested places high value on any user-entered word or phrase that matches a canonical term in the lexicon either directly or via morphologial variation (assigning a

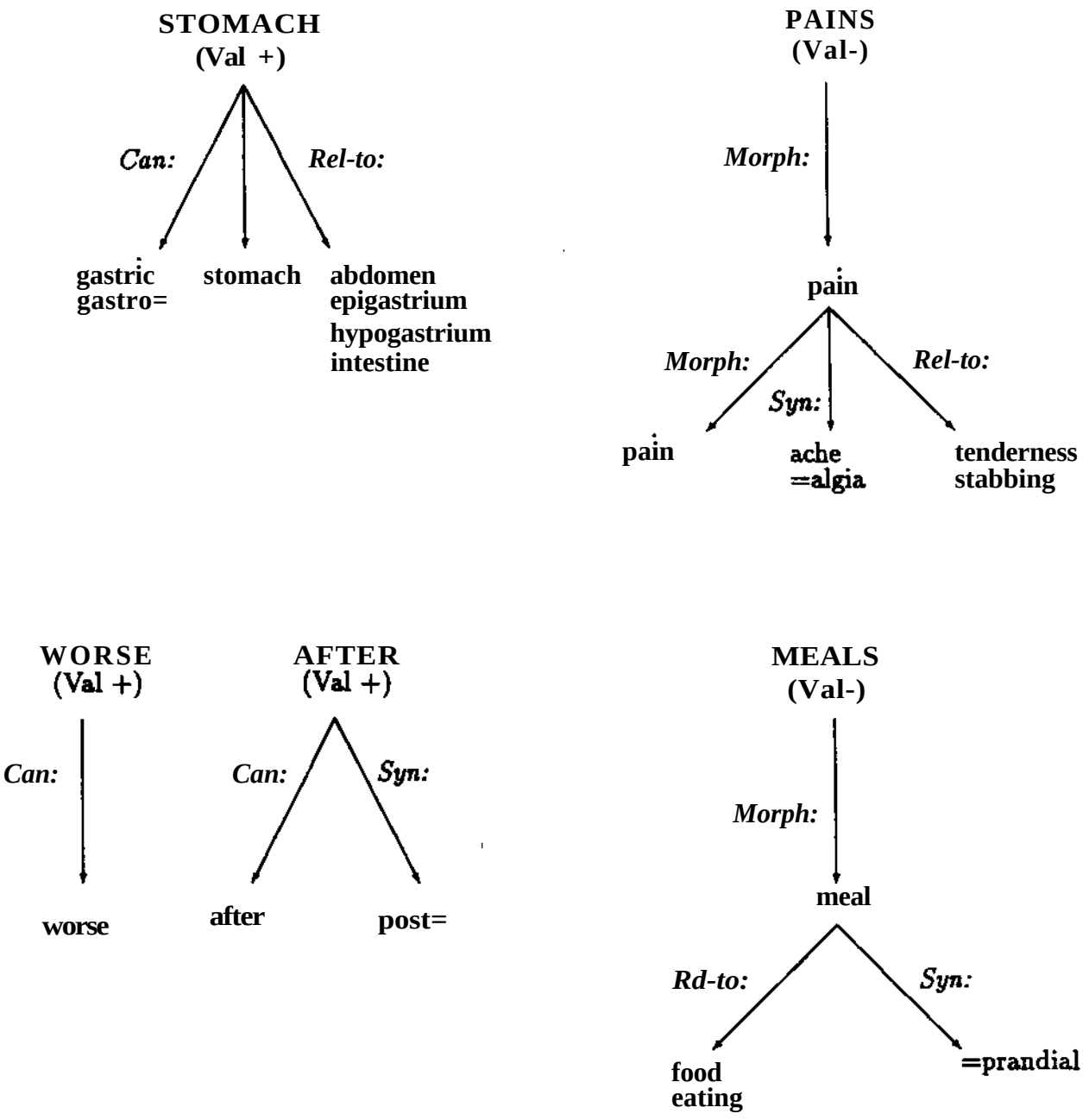

Figure 2: LEXICAL ANALYSIS OF *stomach pains worse after meals*

score of "9" to such expressions); places lower value on direct synonyms ("6"); and places least value on searches that originate with non-canonical words or phrases, reflecting other combinations of links (including possible scores of "4," "3," "2" and "1"). The value of a candidate finding is just the sum of all the scores associated with terms it includes. Using this approach, the system returns ABDOMEN PAIN GTR THAN 1 HR POSTPRANDIAL as 'best guess' for the above phrase. Note also that, in doing so, it has identified *postprandial* as a canonical term by combining the bound morphemes *post=* and *=prandial.* (Additional examples of lexical mapping can be found in Appendix 3.)

5. The Search and Scoring Algorithm.

The system processes user input in three main stages: *lexical analysis, search and scoring.* In essence, it identifies canonical terms that map onto terms in the input string, and uses them to retrieve findings that potentially match the input string. These candidate findings are then scored and those that fall within a cut-off specified by the user are returned. The following discussion elaborates upon each stage of the algorithm.

During *lexical analysis,* words in the input string are "tokenized"—*i.e.,* matched against forms that appear as entries in the lexicon. In the process, adjacent words that can be analyzed as phrases in the lexicon are combined into phrasal tokens. The lexical entry for each token is then searched for canonical terms that are directly associated with it via *Morph:, Syn:,* and *Rel-to:* links. Furthermore, the canonical terms associated with these terms are retrieved. One might look at the token as the "parent" term whose "children" are the canonical terms directly associated with it. These terms are, in turn, "parents" of the next level of associated terms.

Each "child" term gets assigned a score based on two factors: (1) the relationship it bears to its "parent" and (2) the score assigned to its "parent." A *Morph:* links holds a value of "3," a *Syn:* link a value of "2," and a *Rel-to:* link a value of "1." Canonical terms have an implicit *Morph:* link with themselves, thus holding a value of "3." The score is simply the product of the values associated with the "parent" and "child" terms. Thus, a morphological variant of a canonical term has a score of "9" (3 x 3), a synonym of a synonym a score of "4" (2 x 2), a morphological variant of a related term a score of "3" (1 x 3), a related term of a related term a score of "1" (1 x 1), *etc.*. As an illustration, consider our previous example, *stomach pains worse after meals,* as shown in Figure 3. We can now augment our diagram with the scores associated with each link. One of the main points to note is that each token is first mapped to a canonical entry in the lexicon before other links are propagated from that term. Thus, in the diagram, the tokens *stomach, worse,* and *after,* being canonical, are directly mapped to *stomach, worse,* and *after,* respectively. Then all of the canonical terms associated with these terms through the various links are retrieved. Since there is an implicit *Morph:* link between a canonical term and itself, a "copy" of the term gets brought down to a terminal node and assigned a score of "9." The tokens *Pains* and *meals,* being morphological variants of the canonical terms *pain* and *meal,* respectively, are first mapped

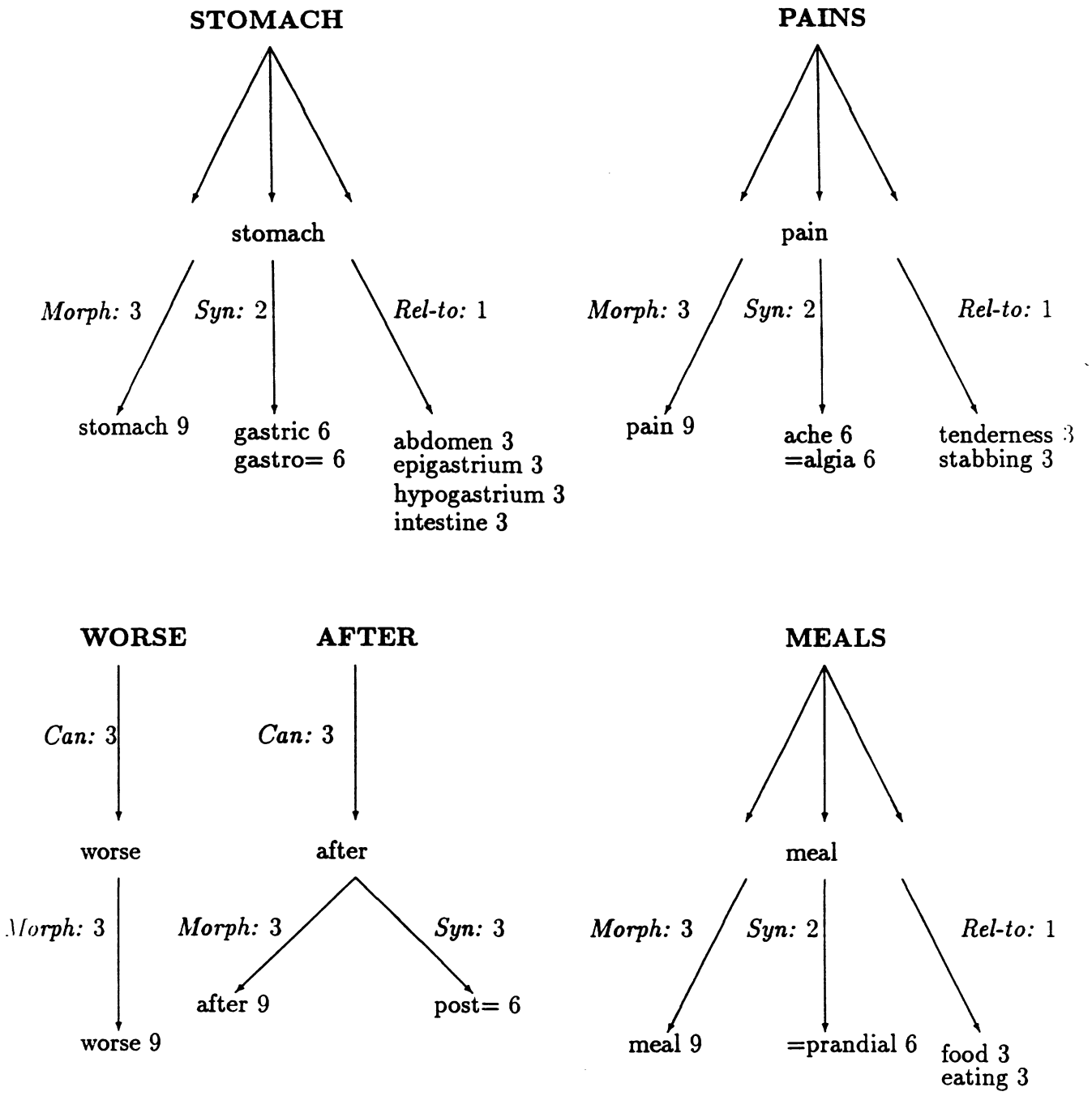

Figure 3: SCORED CANONICAL TERMS IN TRACE FOR *stomach pains worse after meals*

to these terms and additional links are propagated. The score assigned to the "child"' terms is the product of the links leading to them.

It often happens that a token is a morphological variant of a morphological variant, as would have been the case had our sample phrase been *stomach painfully worse after meals.* The same principles of lexical analysis described above would apply. The first goal would be to map to a canonical term, the second to "sprout" additional terms from this term. The main difference is an additional level of indirection. *Painfully* would first be mapped to *painful,* which would in turn be mapped to the canonical term *pain* via a *Morph:* link. From this point on, the analysis would be identical to that of *pains.*

Some lexical entries contain the pragmatic field *Add:,* which references other canonical terms that are implied by the corresponding lexical item, as in:

```
Item: arm
Add:  unilateral

Item: arms
Add:  bilateral
```

During lexical analysis, canonical terms that appear in the *Add:* field are appended to the list of tokens, and all of the canonical terms associated with them are retrieved and scored. For example, the input string *pain in left arm* would be associated with five token entries: *pain, in, left, arm,* and *unilateral.* The lexical entries for *left* and *arm* both contain an *Add:* field that references *unilateral.*

Candidate findings are selected by appealing to the set of canonical terms associated with each token. These sets can best be schematized in columnar form, as shown below for the sample phrase *family history of diabetes.* The canonical terms are listed in descending order based on their score.

FAMILY	HISTORY	DIABETES
•family 9	*hx 6	•diabetes 9
•familial 9	chronic 1	•diabetic 9
•relative 6	disease 1	carbuncle 3
	intermittent 1	insulin 3
	time 1	mellitus 3
		tolbutamide 3

Starred ("*") items are those which are actually used to retrieve candidate findings. We might refer to them as "anchors." Anchors must hold a score higher than 4 and can not

be derived through a *Rel-to:* link.[5] The anchor with the highest score is used to select the first set of candidate findings, which axe all of the findings that contain this term. In our example, *family* and *diabetes* both have a score of "9," so one, say *diabetes* is arbitrarily selected. Here is the list of diabetes findings:

DIABETES INSIPIDUS FAMILY HX

DIABETES MELLITUS FAMILY HX

DIABETES MELLITUS HX

Next, the system searches the other token columns *{family* and *diabetes)* for the highest-scored term that maps onto a term in one or more findings. That would be *family,* in our example, since this word appears in two of the three findings. Then, the next highest-scored term, *fix,* is selected and mapped against the findings. This process continues until the highest potential score of any candidate finding is less than the cut-off distance below the highest matched finding. For each cycle, the highest potential score is the score of the anchor plus the scores of the remaining highest-scored items from each column. At this point, a new anchor might be selected from the column that held the original one, if there is one, and the whole process would be repeated. In our example, all findings that contain the word *diabetic* would be retrieved. There is only one:

GLUCOSE TOLERANCE TEST DIABETIC

The *actual* score of each candidate finding is the sum of the scores associated with each of its terms. Not all terms, of course, will have a score—only those that matched words in the token columns. Let's see how this works in our example:

```
   9        3       9     6
DIABETES  INSIPIDUS FAMILY HX    27

   9        3       9    6
DIABETES  MELLITUS FAMILY HX     27

   9        3     6
DIABETES  MELLITUS HX            18
                           9

GLUCOSE TOLERANCE TEST DIABETIC  9
```

[5]We have discovered that anchors derived through *Rel-to:* links or through two *Syn:* links cause the system to retrieve candidate findings that fail to match the input string closely.

Assuming that the user requests only the top ten percent of scored findings, only the first two would be returned.[6]

6. Limitations and Future Directions of Lexically-Driven NLP.

A fair description of the approach we take to processing natural-language Miput is "sophisticated string matching"—*string matching* because each token in the input string is mapped to a series of canonical terms via *Morph:, Syn:* and *Related-to:* links; *sophisticated* because these links make it possible for terms to match even if they are not exactly the same. The preceding, representative example shows that this approach succeeds in mapping natural language to a highly restricted domain like the INTERNIST-I corpus of findings. It enables the system to tolerate non-standard terminology, even when the user's formulation of a finding is more or less specific than compatable concepts in the knowledge base. However, there axe certain types of input that the current system can not handle adequately, particularly **those involving *quantification* and *negation.***

Measurements are extremely common in the medical domain—especially those associated with laboratory tests: white blood cell counts, glucose levels, ph and titer values, *etc.* It is not surprising, then, that the INTERNIST-I knowledge base contains many concepts involving measurement, for example:

- **COLD ANTIBODY TITER GTR THAN 1:1000**
- **JOINT FLUID WBC 3000 TO 20000**
- **T3 SERUM DECREASED**

Some of these measurements are expressed as a range of values *(e.g., 3000 to 20000)* while others cite a plateau from which the patient deviates (e.</., *gtr than 1:1000).* Clearly, string-matching alone can not handle such variability in quantities. Users may be precise in expressing a quantified finding; they might say, for example, *white blood cell count in joint fluid was 2500.* Or, at the opposite extreme, they may be much more vague and use evaluative expressions like "normal" and "abnormal," as in *white blood cell count in joint fluid was abnormal.* To handle both types of expressions, the system needs to have available to it domain-specific knowledge about what constitutes normal values, as well as procedures to map a specific value to a range of values and to interpret a given value as "normal" or "abnormal," "positive" or "negative," "present" or "absent," *etc.*

[6] **An alternative scoring scheme (which we have not implemented) might exploit other features of the mapping between canonical terms and findings. For example, terms that appear only once or twice among findings are more "salient" than those that appear more frequently. It might be possible to bias scoring to more salient concepts *(e.g., amcnorrhca, bisulfite,* and *decubitus,* all of which occur in only one finding) over less salient ones (such as *pain, biopsy,* and xmy, which occur in many findings). For reference, we include as Appendix 4 a list of terms according to frequency of occurrence among findings.**

We have already modified some of the frame-based representations of findings from the INTERNIST-I knowledge base. The frames contain the domain-specific information that is needed to handle some kinds of quantification and negation. For example, consider the frame for *joint fluid wbc count*:

```
JOINT FLUID WBC COUNT
    status-spec: NUMERICAL VALUE
    norm:        LESS THAN 10
    qualifiers:  BASOPHIL EOSINOPHIL NEUTROPHIL LYMPHOCYTE
                 MONOCYTE PLASMA-CELL
    method:      ARTHROCENTESIS
    phys-range:  BTW 0 1000000
```

The *status specification (status-spec)* field indicates that the result of a wbc test involving joint fluid will be expressed as a number. The *normal (norm)* field provides a means of determining whether that number is normal. *Qualifiers* lists the types of white blood cells that might be counted in more precise *joint-fluid-wbc-count* findings, while *method* specifies the laboratory procedure used to acquire this information. Finally, the *physiological range* field could enable the system to determine if the value entered by a user is legitimate. In this case, any value above 100,000 (or, less likely, below zero) should trigger an error message to the user, informing him or her of the inappropriateness of his input.

Once these frames are completed, as well as the more specific frames that correspond to individual findings (*e.g.,* the *joint-fluid-wbc-3000-to-20000* frame), the system should also be able to handle some of the problems involving *negation*, particularly when terms signaling explicit negation, such as *absent* or *negative*, are used. A procedure could be triggered to compare these quantifiers to the value in a frame's *norm* field. Since the concepts in the INTERNIST-I knowledge base are typically expressions of abnormalities, user input that contains *absent* or *negative* for findings with a *norm* value of *absent* or *negative* should be considered expressions of normal results, so the corresponding findings should be negated. For example, the knowledge base only contains an expression for the abnormal result of a botulinum antitoxin neutralization test: BOTULINUM ANTITOXIN NEUTRALIZATION TEST POSITIVE. The normal value, as expressed in the frame for *botulinum antitoxin neutralization test*, is *negative*. Had the user entered *negative* in this context, a procedure could be invoked to return BOTULIN ANTITOXIN NEUTRALIZATION TEST POSITIVE with a "negative" flag attached to it. This information is required for the diagnostic system that our program is designed to access, since normal findings are often as clinically significant to INTERNIST-I as abnormal findings.

A straightforward way of handling negation would be to instruct the user to prepend expressions of abnormalities with "no" if he or she wished to express a normal condition. This would certainly work for simple cases that do not require special procedures to interpret the domain-specific knowledge encoded in frames. For example, *no family history of diabetes* would directly retrieve *diabetes insipidus family hx* and *diabetes mellitus family hx*, but each

would be passed to the diagnostic system with a "negative" flag. However, this approach has its limitations. For example, a user might fail to see two findings embedded in the following phrase, rather than one: *no shortness of breath, but difficulty swallowing.* Without the syntactic processing strategies needed to determine scope of negation, the system would return several negatively-flagged findings, such as the following:

- APNEA DURING SLEEP
- APNEA EPISODIC
- DYSPHAGIA INTERMITTENT
- DYSPHAGIA PROGRESSIVE

Only the first two findings should be negated, since the scope of negation ends at *but* This suggests that we need to add a syntactic-processing component to the system, as well as procedural access to frames.

Implicit negation promises to be much more difficult than either explicit negation or quantification—probably because it often involves components of both. Consider the following mapping:

```
Enter a phrase —> he drank little alcohol

lymph node <s> pain with alcohol
face flushing after alcohol ingestion
abdomen pain exacerbation with alcohol
alcohol ingestion heavy recent hx
alcohol illicit ingestion hx
```

Clearly, the system needs to be able to treat *little* as the opposite of *heavy,* and therefore return the finding ALCOHOL INGESTION HEAVY RECENT HX with a "negative" or "absent" flag. One way this might be effected is by adding a field to the lexicon called "opposite-of" that would be filled for modifiers that determine whether a finding is present or absent. When the search tree contains a term that is the opposite of one in a candidate finding, a procedure could be evoked to negate that finding.

Concerning sufficiency, or accuracy of retrieval, the addition of *semantic role* information to items in the lexicon along with the use of frames to represent findings would enable the system to better assess the degree of certainty attached to a user's observation. This would obviate a subtle problem in the current system, where, by accident of lexical-semantic association an observation may get linked to a finding that implicates a higher degree of clinical certainty than was intended. For example, observations based on a patient's statement of symptoms carry less clincal validity than observations based on a physician's examination of a patient. If a user reported the observation *leg numbness and tingling,* the system could

return a finding such as LEG <S> SENSATION VIBRATORY DECREASED, by following links **between *numbness* and *decreased sensation,* and between *tingling* and *vibratory*. The statement** as given, however, clearly indicates a *patient-symptom report,* while the finding is the proper expression of the result of a *physical examination.* Noting in the lexicon that *tingling* is a *symptom* and in the frame for LEG <S> SENSATION VIBRATORY DECREASED that it is a *physical-exam* finding would permit the system to make finer discriminations, via the use **of rules such as *never identify a symptom-term with a physical-exam finding.***

With the modifications suggested above to handle aspects of quantification (in laboratory findings, for example) and explicit negation, the limits of strictly-lexical processing will be reached. All further refinements will depend on the development of increasingly comprehensive knowledge bases for the grounding of lexical search and increasingly powerful procedures for interpreting co-occurring concepts. The next stage of development, then, will emphasize frame-based knowledge networks, with, perhaps, some syntactic pre-processing to suggest special modification relations.

REFERENCES

Evans [1986] Evans, David A., "Problems and Potentials in Existing Systems—The INTERNIST Case," in Levy, Allan H., and Williams, Ben T., (Editors), *Proceedings of the AAMSI Congress 86,* Anaheim, California, May 8-10, 1986, 282-287.

Evans and Katz [to appear] Evans, David A., and Katz, Sandra, "A Practical Lexicon **for Contsrained NLM," in *Proceedings of the Third Eastern States Conference on Linguistics** (ESCOL), 1986,* Ohio State University, in press.

Masarie *et al.* **[1985]** Masarie, F.E., Jr., R.A. Miller, J.D. Myers, "INTERNIST-I Properties: Representing Common Sense and Good Medical Practice in a Computerized Medical Knowledge Base," *Computers and Biomedical Research,* vol. 18 (1985), pp. 458-479.

Miller *et* *al.* [1982] Miller, R.A., H.E. Pople, and J.D. Myers, [a]INTERNIST-I, An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine," *New England Journal of Medicine,* vol. 307 (1982), pp. 468-476.

Miller *et al.* **[1986a]** Miller, R.A., Masarie, F.E., and Myers, J.D., "Quick Medical Reference for Diagnostic Assistance," *MD Computing,* 1986a: 3: 34-48.

Miller *et al.* **[1986b]** Miller, R.A., McNeil, M.A., CaUinor, S., Masarie, F.E., Myers, J.D., "Status Report: The INTERNIST-l/Quick Medical Reference Project," in press, *West J. Med.,* Dec. 1986b.

Pople [1982] Pople, H.E., "Heuristic Methods for Imposing Structure on Ill-Structured Problems: The Structuring of Medical Diagnostics," in Peter Szolovits (ed.) *Artificial Intelligence in Medicine,* Westview Press: Boulder, Colorado, 1982, pp. 119-190.

Pople [1986] Pople, H.E., "Freeing the Language of Discourse for Medical Consultation Systems," in Levy, Allan H., and Williams, Ben T., (Editors), *Proceedings of the AAMSI Congress 86,* Anaheim, California, May 8-10, 1986, 272-273.

APPENDIX 1: PHRASAL ENTRIES IN THE LEXICON

I7-keto
17-ketogenic
<rat-tail>
a-wave
acanthosis-nigricans
accessory-muscle
acid-fast-bacteria
adrenal-cortex
aeranoias-hydrophilia
agglutination-test
air-crescent
alpha-1
alpha-1-antitrypsin
alpha-galactosidase
alpha-rhythm
alveolar-ridge
american–indian
aminole^Allinic-acid
ainmonium-chloride
connordim-chloride-test
amyl-nitrite
angioid-streak
animal-husbandry
antigen-antibody
antithrtsnbin-iii
antitoxin-neutxalization-test
aorta-thoracic
aorta-thoracic-artery
aorta-valve
ap-diameter
arterial-blood-pressure
artery-superficicL1-peLLpable
asbestos-body
ascitic-fluid
ascorbic-acid
ashkenazi-jew
atlanto-occipital
attenuated-artery
auer-rod
back-pain
barium-enema
bee-sting
bence-jones
beta-globulin
beta–gluccsidase
bile-<iuct
bile-duct-ctetruction
•biliary-tract
blood-pressure
blood-product
blood–streaked
blood-transfusion
blue-line
border-of-heart
bcwel-sound
bnidzinski-sign
bullous-inyringitis
c-peptide
calcaneal-spur
calcium-oxalate
cat-scratch
cat-scratch-disease
caustic-injury
celiac-sprue
central-lucency
centrifugal-flew
cepiialad-flcw
cerebral-artery
cervical-rib
charcot-leyden
cheyne-stokes
chocolate-cyst
chorionic-villi
chrornatin-dot
chvostek-sign
closing-motion
clostridium-sordellii
coal-dust
coffee-ground
cogwheel-rigidity
colchicine-therapy
cold-hemolysin-test
collateral-artery
conmon-csu^tid
cuuiiinjnicating-cavity
cxa^xtterized-tcracpgrajiiy
conduction-velocity
congestive-heart-failure
conjugate-deviation
construction-worker
contact-dermatitis
coambs-test
corkscrew-vessels
corrigan-pulse
costovertebral-angle
cranial-nerve
creatine-lcinase
crescentic-air-shadcw
cross-matching
crying-attacks
ajrschmann-spiral
cutaneous-nodule
cutaneous-plaque
cyclic-anp
cystic-fibrosis
cystic-pool
cytotoxic-drug
decubitus-ulcer
dehydrationr-syndrcme
delusions-of-grandeur
diabetes-^fya^J.I/Hies
diploe-suture
doppler-ultrasonography
draining-sinus
dt^uytrens–contracture
e-coli
ehler-danlos
ehler-danlos-syndrcme
end-diastolic
endocarditis-infective

campylobacter-jejuni
caput-medusae
carbon-dioxide
carbon-tetrachloride
cardiac-silhouette
carotid-siphon
erlenmeyer-flask
erythema-chronicum-migrans
erythema-marginatum
erythema-multiforme
erythema-nodosum
ethnic-background
evoked-potential
expiratory-phase
familial-disorder
familial-mediterranean-fever
family-history
fetor-hepaticus
filling-defect
flame-shaped
flight-of-ideas
foam-cell
focal-atrophy
foreign-body
francisella-tularensis
friction-rub
gamma-globulin
gastric-wall
gauchers-disease
genu-valgum
germinal-center
ghon-complex
giant-cell
glucose-tolerance-test
gold-salt
goodpasture-syndrome
gram-stain
grand-mal
gross-blood
gross-inspection
growth-hormone
growth-retardation
heart-disease
heel-pad
hemagglutination-test
hemoglobin-f
hemoglobin-ss
hepatic-artery
hepatic-vein
hepatitis-b
herpes-zoster
hiatal-hernia
high-density-wall
hoffman-sign
homans-sign
horner-syndrome
howell-jolly
htlv-iii
hydroxyindoleacetic-acid
immunodiffusion-test
environmental-temperature
epicardial-fat
epidermal-fragility
epinephrine-provocative-test
epithelial-cell
ergonovine-maleate
kaposi-sarcoma
kayser-fleischer-ring
kerley-b-line
kernig-sign
knife-like
kuru-plaque
kveim-test
lacrimal-gland
lactose-tolerance-test
lamina-dura
le-cell
le-test
leonine-facies
light-chain
lightening-pain
loading-test
lupus-erythematosus
macular-star
mandibular-overbite
manic-depressive-disease
marfan-syndrome
mass-palpable
mesenteric-vein
mid-diastolic
mitral-valve
monckeberg-sclerosis
monoamine-oxidase-inhibitor
mononuclear-cell
motor-activity
moulage-sign
mucosal-fold
mycobacterium-tuberculosis
myocardial-infarction
neisseria-gonorrheae
neisseria-meningitidis
nephrotic-syndrome
neurologic-sign
obturator-sign
onion-skin
optic-disc
osmotic-fragility
otitis-media
oxyphilic-cell
p-wave
paracoagulation-test
parietal-cell
pas-positive
pasteurella-multocida
patent-ductus-arteriosus
pelger-huet
perchlorate-discharge-test
pericardial-fluid
perioral-paresthesia
peripheral-branch

immunoelecxrophoresis-tesz
inclusion-bodies
inclusion-body
infusion-test
insect-bite
interosseus-space
intranuclear-inclusion-bodies
intrauterine-device
inverted-3-sign
ipsi–unilateral
jack-knife-position
jakob-creutzfeldt
photic-stimilation
physical-exercise
pill-rolling
placenta-abruptio
placental–fragment
placental-fragments
plain-film
plantar-surface-of-foot
plasma-cell
pleural-fluid
pneunatcpsis-cystoides-intestiralis
poison-ivy
polycythemia-vera
polyphasic>-potential
postictal-sleeping
postural-abnormality
poverty-of-content
pr-interval
precipitin-test
pressure-deformity
pressure^radient
pronator-sign
pilmnnary-arterlogrcgfay
pulmonary-artery
puimonary-embolus
pil·monary-infarction
pulse-cxantour-abnorinali^
pulsus-alternans
pulsus-bisferiens
pulsus-paradoxus
punctate-keratitis
q-wave
qt-interval
guincke-pulse
r-rickettsii
radi2d-artery
radiates—to
radiation-therapy
radioiodine-scan
radioiodine-treatment
radius-ulna
range-of-flexion
range–of–inotiai
range-of-rotation
rat-tail
raynauds-phencraenon
rebound-tenderness
rectal-shelf
peripheral-lung-zone
perirectal-abscess
peritoneal-aspirate
peritonsillar-abscess
periumbilical-region
periungual-erythema
periungual-necrosis
perivascular-pigmentation
periventricular-area
peyronies-diiease
phenotype-zz
philadelphia-chromcsome
rose-bengal-stctining
sacroiliac-joint
saddle-area
salt-craving
scarlet-fever
schilling-test
scintillating–scotoma
secondary-bronchus
self-induced
sella-turcica
semi-synthetic-penicillin
senile-plaque
sensory-level
sentinel-locp
septal-defect
severi^-of-il·lness
sickle-cell
sickle-cell-disorder
skeletal-defect
skin-test
sarali.-intestine
smooth-muscle
snalce-bite
social-withdrawal
southern-ejrcpean
^Decific-gravity
spider-angicroa
spike-like
spina-bifida
splenic-artery
splenic-flexure
staphylococcus-aureus
staphylococcus-epidermidis
sternal-border
stimulation-test
strai^it-leg-raising-test
stra^fcerry-tangue
stxeptococcus-anaerobic
streptococcus-bovis
str^ptocxxxxxis-epidemic
streptococcus-pyogenes
streptococcus-viridans
str^tolysin-o
striated-muscle
subareolar^-area-of-breast
sudan-stain
sulfur-colloid-iv
sumnation-gallop

red-brown
reed-sternberg-cell
reflux-test
refracfcory-to-therapy
relaxation-time
resistance-to-flexion
resistant-to-oonventional-therapy
resuscitation-cardiopilmonary
reticulin-fiber
retrobulbar-area
retroverted-uterus
rheumatic-fever
rheumatoid-factor
rib-cage
ring-shaped
ring-sideroblast
rokitansky-aschoff
rose-bengal
tolbutamide-test
tourniquet-test
toxemia-of-pregnancy
toxoid-i.Tnmiinization
tracheal-deviation
transfusion-reaction
transverse-colon
tricuspid-valve
triple-furrtpwed
trousseau–sign
tubular-cell
turner^syndroDe
two-thirds
ulcei^crater
ulcerative-colitis
upper-arm
upper-^ctremity
i^jper-respiratory-tract-infection
uretero-pelvic
uric-acid
use-of-tanpan
v-wave
vaginal-bleeding
suniighr-exposure
supraclavicular-area
suprapubic-region
suprasellar-area
suprasterral-pulsation
swan-neck-deformity
systolic-arterial-blood-pressure
systolic-click
systolic-pulse
systolic-retraction
t-wave
target-cell
technetium-99m
terminal-ileum
thoracic-spine
thoa^it-content
thcoght-content-abnormality
tinel-sign
vanillylmandelic-acid
variability-with-time
varicose-vein
vena-cava
venous-filling-time
venous-hum
venous-pattern
venous-pressure
vesicular^-rash
villous-border
visual-acuity
vitamin-b
vitamin-bl2
vitamin-d
vitamin-k
vitreous-hanorrhage
von-willebrand's
wilsons-difsease
wing-beating
wire-loop
wolff-parkinson-v*iite
yersinia-pestis

APPENDIX 2: LIST OF BOUND MORPHEMES IN THE LEXICON

a=
ad=
adrenal=
aer=
amb=
amph=
amyo=
an=
ana=
angio=
ante=
anti=
apo=
append=
aut=
auto=
bi=
bis=
blepharo=
brachy=
brady=
burs=
cac=
caco=
cat=
cata=
cath=
cellul=
cent=
cephal=
chrom=
circum=
co=
com=
con=
conjunctiv=
contra=
crypt=
cyst=
cysto=
dacry=
dactyl=
de=
deca=
deci=
demi=
dent=
derma=
dermat=
di=
dia=
dipl=
dipla=
diplo=
dis=
dys=
ec=
echo=
ecto=
ef=

episcler=
episscler=
erythro=
es=
eu=
ex=
exo=
extra=
fore=
galact=
galacto=
gaster=
gastr=
gastro=
glomerullonephr=
glomerulonephr=
glosso=
hem=
hemat=
hemato=
hemi=
hepa=
hepar=
hepato=
hetero=
holo=
homeo=
homo=
hydra=
hydro=
hyp=
hyper=
hyph=
hypo=
idio=
ileo=
in=
infra=
inter=
intra=
intro=
iridocycl=
iso=
juxta=
karyo=
kata=
kath=
kera=
kerat=
kinisi=
lact=
laparo=
laryng=
laryngo=
latero=
lepto=
leuko=
lith=
lymph=
lymphang=

mega=
megal=
melan=
melano=
mes=
meso=
meta=
metra=
metro=
micro=
mio=
mono=
multi=
my=
myel=
myelo=
myo=
myxa=
myxo=
neo=
nephr=
nephra=
nephro=
neu=
neuro=
niter=
nitro=
non=
not=
nucleo=
o=
ob=
oculo=
odont=
oligo=
omo=
omphal=
oophoron=
opisth=
orchid=
ortho=
os=
oste=
osteo=
osteomyel=
oxy=
pachy=
pan=
par=
para=
path=
per=
peri=
pharyng=
plur=
pneu=
poly=
post=
pre=
pro=

en=
endo=
endocard=
entero=
ento=
epi=
pyr=
re=
retro=
rhino=
saccar*=
sacro=
salping=
salpingo=
sarco=
sclero=
sebo=
semi=
sialo=
macro=
maculo=
mal=
mast=
me =
rosdi=
SCIUII=
splerr=
spleno-
sub=
supers
supra=
syn=
tachy=
tele=
tetra=
thic*=
thyro=
thyroid
proto=
pseud=
pseudo=
psych=
py=
pyo=
trans=
tri=
tropho=
\mi=
iorethr=
urin=
urino=
uro=
vagin=
vaso=
venter=
ventro=
xanth=

APPENDIX 3: EXAMPLES OF LEXICAL MAPPING

```
          Synonym: "ens".
          Synonym: "within".
          Related to: "location".
          Related to: "position".
  Token = "stools"
      Synonym
        Canonical: "feces".
          Related to: "excretion".
          Related to: "colon".
          Related to: "constipation".
          Related to: "defecation".
          Related to: "diarrhea".
          Related to: "enema".
          Related to: "incontinence".
          Related to: "intestine".
          Related to: "rectum".


results:

(21)  604: feces gross blood
(18)  602: feces black tarry
```

```
EXAMPLE 7

Enter a phrase_____> 21 year old black man

Tree structure ox tokens read in:
  Token = "year-old"
      Related to
        Canonical: "age".
          Related to: "adult".
          Related to: "immature".
          Related to: "mature".
          Related to: "menopause".
          Related to: "old".
          Related to: "time".
          Related to: "year".
  Token = "black"
    Canonical
        Canonical: "black".
          Synonym: "colored".
          Synonym: "melan=".
          Synonym: "melanos".
          Synonym: "negro".
          Synonym: "tarry".
          Related to: "race".
          Related to: "color".
          Related to: "dark".
  Token = "man"
      Synonym
        Canonical: "male".
          Related to: "sex".
          Related to: "testis".
      Related to
        Canonical: "sex".
          Morph. Var.: "sexual".
          Synonym: "=pareunia".
          Related to: "female".
          Related to: "libido".
          Related to: "male".


results:

(13)  1414: sex male
(13)  1296: race negro
```

Comments:

This trace demonstrates the fact that the system can not handle ranges. The IHTERNIST-I age findings indicate ranges, such as "age 16 to 25," "age 25 to 55," and "age gtr than 55." Although these findings were all included in the list of candidate matches (but are not shown here because they did not score within the cut-off range), the system was unable to select "age 16 to 25" as the best match. This problem will be solved when the lexicon interacts with a frame-based representation of findings.

EXAMPLE 8

```
Enter a phrase_____> recently bitten by a cat

Tree structure of tokens read in:
  Token = "recently[11]
      Morph. Var.
        Canonical: "recent".
          Synonym: "current".
          Synonym: "neo=".
          Related to: "current".
          Related to: "time".
  Token = "bitten"
      Morph. Var.
        Canonical: "bite".
          Related to: "husbandry".
          Related to: "animal".
          Related to: "baa-sting[11].
          Related to: "insect-bite".
          Related to: "mouth".
          Related to: "snake".
          Related to: "snake-bite".
          Related to: "sting".
          Related to: "tick".
  Token = "cat"
    Canonical
        Canonical: "cat".


results:

(27)  293: cat scratch or bite recent hx
```

EXAMPLE 9

```
Enter a phrase____> skipped periods
Tree structure ox tokens read in:
  Token = "skipped-periods"
      Synonym
        Canonical: "oligomenorrhea".
          Related to: "amenorrhea".
```

results:

(6) 1140: oligomenorrhea

Comments:

This trace brings out the need to include phrases in the lexicon, in addition to single-word entries. The t^o-word input string is considered one token that maps onto the canonical term "oligomenorrhea.>a

```
EXAMPLE 10

Enter a phrase_____>  sore liver when palpated

Tree structure of tokens read in:
  Token * "sore[11]
      Synonym
        Canonical: "tender".
          Morph. Var.: "tenderness".
      Related to
        Canonical: "pharyngitis".
          Related to: "throat".
          Related to: "inflammation".
          Related to: "pharynx".
  Token = "liver"
    Canonical
        Canonical: "liver".
          Synonym: "hepa=".
          Synonym: "hepar=".
          Synonym: "hepatic".
          Synonym: "hepato=".
          Related to: "abdomen".
          Related to: "abdominal".
          Related to: "biliary-tract[11].
          Related to: "blood".
          Related to: "eztrahepatic".
          Related to: "fetor-hepaticus".
          Related to: "gallbladder".
          Related to: "hepatocellular".
          Related to: "hepatomegaly".
          Related to: "intrahepatic".
          Related to: "rose-bengal-staining".
          Related to: "hepatojugular".
  Token = "when"
    Canonical
        Canonical: "when".
  Token = "palpated"
      Morph. Var.
        Canonical: "palpable".
          Morph. Var.: "palpation".
          Synonym: "expansile".
          Related to: "touch".
      Morph. Var.
        Canonical: "palpation".
          Morph. Var.: "palpable".
          Related to: "exam".
          Related to: "fremitus".
          Related to: "touch".
```

results:

(24) 991: liver tender on palpation

EXAMPLE 11

```
Enter a phrase____> renal insufficiency

Tree structure of tokens read in:
  Token = "renal"
    Canonical
        Canonical: "renal".
          Morph. Var.: "rena".
          Synonym: "kidney".
          Related to: "indocyanine".


results:

(9)  2905: kidney <s> arteriography rena artery <ies> poststenotic dilatation
(9)  2894: kidney <s> arteriography aberrant renal artery obstructing uretero pelvic junction
(9)  2897: kidney <s> arteriography complete obstruction  of main renal artery or major branch
<es>
(9)  2902: kidney <s> arteriography main renal artery diffuse enlargement
(9)  2906: kidney <s> arteriography stenosis main renal artery bilateral
(9)  2907: kidney <s> arteriography stenosis main renal artery unilateral
(9)  2908: kidney <s> arteriography stenosis major branch <es> renal artery <ies>
(9)  2922: kidney <s> ivp notching of renal pelvis and/or proximal ureter
(9)  2933: kidney <s> ivp renal displacement
(9)  2946: kidney <s> plain film renal displacement
(9)  2973: kidney <s> venography renal vein occlusion
```

```
EXAMPLE 12

Enter a phrase____> losing weight

Tree structure of tokens read in:
  Token = "losing"
      Morph. Var.
        Canonical: "loss".
          Morph. Var.: "loose".
          Synonym: "lack".
          Synonym: "absence".
          Related to: "a=".
          Related to: "amount".
  Token = "weight"
    Canonical
        Canonical: "weight".
          Related to: "emaciation".
          Related to: "epicardial-fat".
          Related to: "gm".
          Related to: "gram".
          Related to: -gravity".
          Related to: "heavy".
          Related to: "obesity".


results:

18)  1717: weight loss gtr than 10 percent
```

```
EXAMPLE 13

Enter a phrase____> mass in right kidney by cat scan

Tree structure of tokens read in:
  Token = "mass[11]
    Canonical
        Canonical: "mass".
          Synonym: "lesion[11]".
          Synonym: "knob".
          Related to: "granuloma".
  Token = "in"
    Canonical
        Canonical: "in".
          Synonym: "en=".
          Synonym: "within".
          Related to: "location".
          Related to: "position".
  Token = "right"
    Canonical
        Canonical: "right".
          Related to: "^lateral".
          Related to: "axis".
          Related to: "unilateral".
  Token = "kidney"
    Canonical
       Canonical: "kidney".
          Synonym: "nephr=".
          Synonym: "nephra=".
          Synonym: "nephro=".
          Synonym: "renal".
          Synonym: "rena".
          Related to: "abdomen".
          Related to: "abdominal".
          Related to: "adrenal".
          Related to: "anuria".
          Related to: "glomerulonephr=".
          Related to: "ivp".
          Related to: "nephrocalcinosis".
          Related to: "nephrogram".
          Related to: "nephrolithiasis".
          Related to: "nephrotomography".
          Related to: "ureter".
          Related to: "urine".
          Related to: "suprarenal".
  Token = "cat-scan"
      Morph. Var.
        Canonical: "computerized-tomography".
```

Synonym: "computerized".

results:

(39) 1735: abdomen computerized tomography right lower quadrant extrapelvic low density mass

```
EXAMPLE 14

Enter a phrase____> metastatic renal cell carcinoma

Tree structure of tokens read in:
  Token = "renal"
    Canonical                          "
        Canonical: "renal".
          Morph. Var.: "rena".
          Synonym: "kidney".
          Related to: "indocyanine".
  Token = "cell"
    Canonical
        Canonical: "cell".
          Synonym: "=cyte".
          Related to: "rbc".
          Related to: "wbc".
          Related to: "myelocyte".
          Related to: "lymphocyte".
          Related to: "aaisocyte".
          Related to: "basophil".
          Related to: "cellulitis".
          Related to: "culture".
          Related to: "cytology".
          Related to: "hypocellular".
  Token = "carcinoma"
    Canonical
        Canonical: "carcinoma".


results:

(18)  3938: urine sediment renal tubular cell <s>
(18)  3939: urine sediment renal tubular cell cast <s>
```

```
EXAMPLE 15

Enter a phrase_____> chest x-ray revealed bilateral effusions

Tree structure of tokens read in:
  Token = "chest*1
    Canonical
        Canonical: "chest".
          Related to: "breast".
          Related to: "breathing".
          Related to: "fremitus".
          Related to: "lung".
          Related to: "pectoriloquy".
          Related to: "rales".
          Related to: "rib-cage".
          Related to: "substernal".
  Token = "x-ray"
      Synonym
        Canonical: "lymphangiography".
      Synonym
        Canonical: "plain-film".
      Synonym
        Canonical: "xray".
          Synonym: "=gram".
          Synonym: "=graphy".
          Related to: "myelogram".
         Related to: "pyelography".
          Related to: "radiolucency".
          Related to: "tomography".
          Related to: "venography".
      Related to
        Canonical: "angiocardiography".
          Related to: "heart".
      Related to
        Canonical: "bronchogram".
          Related to: "lung".
      Related to
        Canonical: "cardiac-silhouette".
          Related to: "heart".
      Related to
        Canonical: "cholecystography".
          Related to: "gallbladder".
      Related to
        Canonical: "contrast".
      Related to
        Canonical: "hysterosalpingography".
          Related to: "uterus".
      Related to
```

```
      Canonical: "nephrogram".
         Related to: "kidney".
  Token = "bilateral"
    Canonical
        Canonical: "bilateral".
```

```
results:

(24)  350: chest xray diaphragm elevated bilateral
(24)  352: chest zray diaphragm low bilateral
(24)  2223: chest xray hilar adenopathy bilateral
(24)  2245: chest zray perihilar soft alveolar density <ies> bilateral
```

EXAMPLE 16

Eater a phrase____> can not move right lower extremity

```
Tree structure of tokens read in:
  Token = "can"
      Morph. Var.
        Canonical: "canned".
  Token = "not"
    Canonical
        Canonical: "not".
          Synonym: "de=".
          Synonym: "in=".
          Synonym: "non".
          Synonym: "non=".
  Token = "move"
      Synonym
        Canonical: "=kinesia".
          Synonym: "motion".
          Synonym: "movement".
          Related to: "muscle".
      Synonym
        Canonical: "=kinetic".
          Synonym: "movement".
          Synonym: "motion"•
          Related to: "muscle".
     Related to
        Canonical: "spraxia".
          Related to: "movement".
          Related to: "control".
  Token = "right"
    Canonical
        Canonical: "right".
          Related to: M=lateral".
          Related to: "axis".
          Related to: "unilateral".
  Token = "lower-extremity"
      Synonym
        Canonical: "leg".
          Related to: "achilles".
          Related to: "call".
          Related to: "claudication".
          Related to: "erythema-nodosum".
          Related to: "extremity".
          Related to: "lemur".
          Related to: "foot".
          Related to: "hip".
          Related to: "kernig-sign".
```

```
          Related to: "knee".
          Related to: "peripheral".
          Related to: "thigh[11]".
          Related to: "walking".


results:

(24)  908: leg <s> cyanosis unilateral not relieved by elevation
(21)  189: arm <s> weakness unilateral with ipsi unilateral leg weakness
(21)  904: leg <s> ankle clonus sustained unilateral
(21)  906: leg <s> cyanosis dependent unilateral
(21)  907: leg <s> cyanosis non dependent unilateral
(21)  909: leg <s> cyanosis unilateral relieved by elevation
(21)  914: leg <s> edema unilateral massive
(21)  915: leg <s> edema unilateral slight or moderate
(21)  917: leg <s> flaccid unilateral
(21)  920: leg <s> homans sign present unilateral
(21)  926: leg <s> muscle <s> atrophy unilateral
(21)  971: leg <s> weakness unilateral monoplegic
(21)  967: leg <s> weakness bilateral
```

EXAMPLE 17

```
Enter a phrase ---> left facial twitching

Tree structure of tokens read in:
  Token = "left"
    Canonical
        Canonical: "left".
          Related to: "=lateral".
          Related to: "axis".
          Related to: "unilateral".
  Token = "facial"
    Canonical
        Canonical: "facial".
          Morph. Var.: "face".
  Token = "twitching"
    Canonical
        Canonical: "twitching".
          Synonym: "spasm".
          Related to: "chorea".
          Related to: "fasciculation".


results:

(18)  1052: muscle <s> facial twitching
```

EXAMPLE 18

```
Enter a phrase____> decreased appetite

Tree structure ox tokens read in:
  Token = "decreased"
    Canonical
        Canonical: "decreased".
          Morph. Var.: "decrease".
          Synonym: "less".
          Synonym: "=p«nia".
          Synonym: "a=".
          Synonym: "bradys".
          Synonym: "hyp=".
          Related to: "impaired".
          Related to: "abnormal".
          Related to: "absent".
          Related to: "amount".
   Token = "appetite"
     Canonical
         Canonical: "appetite".
           Related to: "diet".
           Related to: "anorexia".
           Related to: "food".
           Related to: "pica".
           Related to: "taste".


 results:

 (15)  155: appetite increased hx
```

```
EXAMPLE 19

Enter a phrase ---> bladder incontinence

Tree structure of tokens read in:
  Token = "bladder"
    Canonical
        Canonical: "bladder".
          Synonym: "cysto=".
          Synonym: "cyst=".
          Synonym: "vesical".
          Synonym: "vesicle".
          Related to: "incontinence".
          Related to: "urine".
          Related to: "ivp".
          Related to: "prostate".
          Related to: "ureter".
          Related to: "urination".
  Token = "incontinence"
    Canonical
        Canonical: "incontinence".
          Related to: "bladder".
          Related to: "feces".
          Related to: "urine".


results:

(15)  605: feces incontinence
(15)  1696: urine incontinence
```

EXAMPLE 20

```
Enter a phrase_____> moderated cigarette abuse

Tree structure of tokens read in:
  Token = ''cigarette[11]
    Canonical
        Canonical: "cigarette[11]".
          Related to: "abuse".
          Related to: "smoking".
  Token = "abuse"
    Canonical
        Canonical: "abuse".
          Related to: "cigarette".
          Related to: "alcoholism".
          Related to: "alcohol".


results:
(12)  359: cigarette smoking hx
```

EXAMPLE 21

Eater a phrase_____> history of peptic ulcer disease

```
Tree structure of tokens read in:
  Token = "history"[11]
    Canonical
        Canonical: "history".
          Synonym: "hx".
          Related to: "chronic".
  Token = "peptic"
    Canonical
        Canonical: "peptic".
  Token = "ulcer"
    Canonical
        Canonical: "ulcer".
          Morph. Var.: "ulceration".
          Related to: "inflammation".
          Related to: "lesion".
          Related to: "skin".
  Token = "disease"
    Canonical
        Canonical: "disease".
          Synonym: "=pathy".
          Synonym: "disorder".
          Synonym: "mal".
          Synonym: "path=".
          Related to: "acute".
          Related to: "chronic".
          Related to: "emaciation".
          Related to: "epidemic".
          Related to: "exacerbation".
          Related to: "illness".
          Related to: "immunization".
          Related to: "infarction".
          Related to: "infection".
```

results:

```
(30)  1683: ulcer peptic family hx
(30)  1684: ulcer peptic hx
```

EXAMPLE 22

```
Enter a phrase____> nasal polyposis

Tree structure of tokens read in:
  Token = "nasal"
      Morph. Var.
        Canonical: "nose".
          Synonym: "rhino=".
          Related to: "=osmia".
          Related to: "epistaxis".
          Related to: "nasopharyngeal".
          Related to: "olfactory".
          Related to: "pharynx".
          Related to: "polyp".
          Related to: "rhinorrhea".
          Related to: "sinus".
          Related to: "features".


results:

(9)  1111: nose mucosa nodule <s>
(9)  1112: nose sucosal polyp <s>
(9)  1116: nose tip ulcer <s>
(9)  1117: nose turbinate <s> necrotic
```

```
EXAMPLE 23

Enter a phrase____> gastrointestinal bleed

Tree structure of tokens read in:
  Token = "gastrointestinal[11]
    Canonical
        Canonical: "gastrointestinal".
          Related to: "esophagus".
          Related to: "stomach".
  Token = "bleed"
      Morph. Var.
        Canonical: "bleeding".
          Synonym: "hemorrhage".
          Related to: "blood".
          Related to: "injury".
          Related to: "menorrhagia".
          Related to: "metrorrhagia".


results:

(15)  812: hemorrhage gastrointestinal acute recent hx
```

```
EXAMPLE 24

Enter a phrase____> pupils equal round

Tree structure of tokens read in:
  Token = "pupils"
      Morph. Var.
        Canonical: "pupil".
          Related to: "eye".
          Related to: "iris".
  Token = "equal"
      Morph. Var.
        Canonical: "equally".
      Synonym
        Canonical: "iso=".
  Token = "round"
    Canonical
        Canonical: "round".
          Morph. Var.: "rounded[11]".
          Synonym: "rounded".
          Related to: "bulb".
          Related to: "spheroid".


results:

(15)  558: eye <s> pupil <s> argyll robertson
(15)  559: eye <s> pupil <s> constricted unreactive
(15)  560: eye <s> pupil <s> dilated unreactive
(15)  561: eye <s> pupil <s> unequal
```

```
EXAMPLE 25

Enter a phrase____> pupils are reactive to light and accommodation

Tree structure of tokens read in:
  Token = "pupils[11]
      Morph. Var.
        Canonical: [M]pupil[M].
          Related to: "eye[11]".
          Related to: "iris".
  Token = "reactive"
    Canonical
        Canonical: "reactive".
          Morph. Var.: "reaction".
  Token = "light"
    Canonical
        Canonical: "light".
          Related to: "color".
          Related to: "fluorescence".
          Related to: "lucency".
          Related to: "lucid".
  Token = "and"
      Synonym
        Canonical: "with".
          Synonym: "during".
          Synonym: [H]syn=".
         Synonym: "containing".


results:

(16)  558: eye <s> pupil <s> argyll robertson
(16)  559: eye <s> pupil <s> constricted unreactive
(16)  560: eye <s> pupil <s> dilated unreactive
(16)  561: eye <s> pupil <s> unequal
```

```
EXAMPLE 26

Enter a phrase –> left  chest  nodule

Tree structure of tokens read in:
  Token = "left"
    Canonical
        Canonical: "left".
          Related to: "^lateral".
          Related to: "axis".
          Related to: "unilateral".
  Token = "chest"
    Canonical
        Canonical: "chest".
          Related to: "breast".
          Related to: "breathing".
          Related to: "fremitus".
          Related to: "lung".
          Related to: "pectoriloquy".
          Related to: "rales".
          Related to: "rib-cage".
          Related to: "substernal".
  Token = "nodule"
    Canonical
        Canonical: "nodule".
          Horph. Var.: "node".
          Morph. Var.: "nodular".
          Synonym: "knob".
          Related to: "cutaneous-nodule".
          Related to: "miliary".
          Related to: "zanthomata".


results:

(18)  2266:  chest zray solitary pulmonary nodule
(18)  2240:  chest zray lymph node <s> calcified
(18)  2220:  chest zray diffuse nodular density <ies> non calcified
(18)  2251:  chest zray pleural mass or nodular thickening
```

```
EXAMPLE 27

Enter a phrase ---> jugular venous distension

Tree structure of tokens read in:
  Token = "jugular"
    Canonical
        Canonical: "jugular".
          Related to: "throat".
          Related to: "neck".
  Token = "venous"
    Canonical
        Canonical: "venous".
          Morph. Var.: "vein".


results:

(18)  895: jugular venous hum
(18)  3484: pulse jugular venous a wave increased
(18)  3485: pulse jugular venous v wave increased
```

```
EXAMPLE 28

Enter a phrase ---> increased withdrawal

Tree structure of tokens read in:
  Token = "increased"
    Canonical
        Canonical: "increased".
          Morph. Var.: "increase".
          Morph. Var.: "increasing".
          Synonym: "augmentation".
          Synonym: "abundant".
          Synonym: "enlarged".
          Synonym: "hyper=".
          Synonym: "augmented".
          Related to: "abnormal".
          Related to: "amount".
  Token = "withdrawal"
    Canonical
        Canonical: "withdrawal".
          Related to: "dependency".
          Related to: "dependent".


results:

(9)  233: behavior social withdrawal
```

EXAMPLE 29

Eater a phrase____> hepatic bruit

Tree structure of tokens read in:
  Token = "hepatic"
    Canonical
        Canonical: "hepatic".
          Synonym: "liver".
          Related to: "indocyanine".
  Token = "bruit"
    Canonical
        Canonical: "bruit".
          Related to: "murmur".
          Related to: "sound".
          Related to: "artery".
          Related to: "vein".

results:

(12) 2202: celiac arteriography hepatic artery enlarged
(12) 2203: celiac arteriography hepatic artery single aneurysm
(12) 2204: celiac arteriography hepatic artery visualized gtr than 20 second <s>
(12) 1256: pressure hepatic vein wedge increased
(12) 1257: pressure hepatic vein wedge normal
(12) 1743: abdomen ultrasonography hepatic vein or hepatic portion of inferior vena cava echogen:
material
(12) 3131: liver computerized tomography <enhanced> hepatic vein <s> non visualized
(12) 3152: liver venography hepatic vein numerous interlacing collateral <s>
(12) 3153: liver venography hepatic vein obstruction
(12) 3955: vena cava inferior venography obstruction at hepatic vein level

EXAMPLE 30

```
Eater a phrase____> mouth ulcers

Tree structure of tokens read in:
  Token * "mouth"
    Canonical
        Canonical: "mouth".
          Synonym: "oriiice".
          Synonym: "os=".
          Related to: "canal".
          Related to: "bite".
          Related to: "=stomy".
          Related to: "esophagus".
          Related to: "gingiva".
          Related to: "jaw".
          Related to: "mandible".
          Related to: "mastication".
          Related to: "oral".
          Related to: "palatal".
          Related to: "periodontal".
          Related to: "perioral".
          Related to: "perioral-paresthesia",
          Related to: "pharynx".
          Related to: "regurgitation".
          Related to: "respiratory".
          Related to: "speech".
          Related to: "taste".
          Related to: "teeth".
          Related to: "tongue".
          Related to: "trismus".
          Related to: "features".
  Token = "ulcers"
      Morph. Yar.
        Canonical: "ulcer".
          Morph. Var.: "ulceration".
          Related to: "inflammation[11]".
          Related to: "lesion".
          Related to: "skin".


results:

(18)  1044: mouth superficial ulcer <s>
(18)  1046: mouth ulcer <s> with necrotic membrane <s>
```

EXAMPLE 31

```
Enter a phrase_____> currently taking penicillin

Tree structure of tokens read in:
  Token = "currently"
      Horph. Var.
        Canonical: "current".
          Synonym: "present".
          Synonym: "recent".
          Related to: "recent".
          Related to: "potential".
          Related to: "flow".
          Related to: "time".
          Related to: "voltage".
      Synonym
        Canonical: "recent".
          Synonym: "current".
          Synonym: "neo=".
          Related to: "current".
          Related to: "time".
  Token = "penicillin"
    Canonical
        Canonical: "penicillin".
          Related to: "drug".


results:

(15)  1176: penicillin or semi synthetic penicillin administration recent hx
```

```
EXAMPLE 32

Enter a phrase_____> weak

Tree structure of tokens read in:
  Token = "weak"
      Morph. Yar.
        Canonical: "weakness".
          Related to: "adynamic".
          Related to: "rigidity".
          Related to: "strength".
      Related to
        Canonical: "diluted".
          Synonym: "attenuated".


results:

(9)  184: arm <s> weakness bilateral
(9)  18S: arm <s> weakness distal only
(9)  186: arm <s> weakness proximal and distal
(9)  187: arm <s> weakness proximal only
(9)  188: arm <s> weakness unilateral monoplegic
(9)  189: arm <s> weakness unilateral with ipsi unilateral leg weakness
(9)  S73: eye <s> weakness of downward and outward gaze
(9)  S74: eye <s> weakness of lateral gaze
(9)  S7S: eye <s> weakness of upward and inward gaze
(9)  680: hand <s> muscle <s> weakness thenar
(9)  967: leg <s> weakness bilateral
(9)  ,968: leg <s> weakness distal only
(9)  '969: leg <s> weakness proximal and distal
(9)  970: leg <s> weakness proximal only
(9)  971: leg <s> weakness unilateral monoplegic
(9)  1042: mouth palatal weakness
(9)  1053: muscle <s> facial weakness bilateral
(9)  1054: muscle <s> facial weakness unilateral including forehead
(9)  1055: muscle <s> facial weakness unilateral lower two thirds only
(9)  1068: muscle <s> weakness exacerbation with rapid repetitive movement <s>
(9)  1069: muscle <s> weakness generalized
(9)  1070: muscle <s> weakness masseter <s>
(9)  1071: muscle <s> weakness pelvic girdle
(9)  1072: muscle <s> weakness shoulder girdle
(9)  1073: muscle <s> weakness trapezius and/or sternomastoid
```

EXAMPLE 33

```
Eater a phrase____> dark lesion in posterior pharynx

Tree structure of tokens read in:
  Token = "dark"
    Canonical
        Canonical: "dark".
          Related to: "black[1]'.
  Token = "lesion"
    Canonical
        Canonical: "lesion".
          Morph. Var.: "lesions".
          Synonym: "injury".
          Synonym: "growth".
          Synonym: "mass".
          Synonym: "mass-palpable".
          Synonym: "wound".
          Synonym: "ulceration".
          Related to: "burn".
          Related to: "dermatitis".
          Related to: "injury".
          Related to: "miliary".
          Related to: "rash".
          Related to: "ulcer".
  Token = "in"
   Canonical
        Canonical: "in".
          Synonym: "en*".
          Synonym: "within".
          Related to: "location".
          Related to: "position".
  Token = "posterior"
    Canonical
        Canonical: "posterior".
          Synonym: "back".
          Synonym: "buttock".
          Synonym: "dorsal".
          Related to: "anal".
  Token « "pharynx"
    Canonical
        Canonical: "pharynx".
          Synonym: "pharyng*".
          Related to: "mouth".
          Related to: "nasopharyngeal".
          Related to: "nose".
          Related to: "pharyngitis".
          Related to: "tonsil".
```

results:

(21) 1187: pharynx hyperemia diffuse
(18) 3283: mouth lesion culture histoplasma

EXAMPLE 34

```
Eater a phrase_____> congestive heart failure recently

Tree structure of tokens read in:
  Token = "congestive[11]
    Canonical
        Canonical: "congestive[11].
          Morph. Var.: "congestion[11].
  Token = "heart"
    Canonical
        Canonical: "heart".
          Morph. Var.: "precordial".
          Synonym: "cardiac".
          Synonym: "=cardia".
          Synonym: "=cardial".
          Synonym: "=cardium".
          Related to: "atrial".
          Related to: "angiocardiography".
          Related to: "aorta".
          Related to: "aorta-thoracic".
          Related to: "aorta-valve".
          Related to: "arterial-blood-pressure".
          Related to: "artery".
          Related to: "atrium".
          Related to: "blood".
          Related to: "blood-pressure".
          Related to: "cardiac".
          Related to: "cardiac-silhouette".
          Related to: "cardiopulmonary".
          Related to: "coronary".
          Related to: "diastole".
          Related to: "diastolic".
          Related to: "endocarditis-infective".
          Related to: "flutter".
          Related to: "gallop".
          Related to: "myocardial".
          Related to: "pacemaker".
          Related to: "palpitation".
          Related to: "pericardial".
          Related to: "pericardium".
          Related to: "precordial".
          Related to: "pressure".
          Related to: "pulse".
          Related to: "pulsus".
          Related to: "regurgitation".
          Related to: "resuscitation-cardiopulmonary".
          Related to: "systolic".
          Related to: "vena-cava".
```

```
          Related to: "ventricle".
  Token = "failure"
    Canonical
        Canonical: "failure".
  Token = "recently"
      Morph. Var.
        Canonical: "recent".
          Synonym: "current".
          Synonym: "neo=".
          Related to: "current".
          Related to: "time".
```

results:

(27) 697: heart failure congestive hx

Appendix 4: List of internist-i Terms According to Frequency

001	a1	001	appendectomy	001	bsp
001	a3	001	appearance	001	butcher
001	a8	001	aphasia	001	burst
001	abstract	001	apathy	001	bursa
001	absence	001	ap	001	burning
001	abruptio	001	arsenic	001	bullous
001	above	001	argyll	001	bullae
001	abnormality	001	architecture	001	bulky
001	aberrant	001	arachnodactyly	001	bulging
001	acuity	001	arabic	001	bypass
001	activated	001	astrocytosis	001	c3
001	action	001	asthma	001	c4
001	acromioclavicul	001	asterixis	001	caustic
001	acral	001	astereognosis	001	caudate
001	acne	001	association	001	catecholamine
001	acicular	001	aspirin	001	cataract
001	aching	001	ashkenazi	001	cataplexy
001	acetaminophen	001	aschoff	001	casoni
001	accessory	001	ascending	001	cartilage
001	acanthosis	001	attempts	001	carpopedal
001	acalculia	001	atonic	001	cardiopulmonary
001	adventitial	001	ataxic	001	carbuncle
001	adult	001	autumn	001	caput
001	adrenalectomy	001	automatism	001	captopril
001	adherent	001	autoamputation	001	capillary
001	aeromonas	001	auricular	001	canned
001	agraphia	001	auer	001	caliber
001	agglutinin	001	awakening	001	calculi
001	akinetic	001	axillary	001	calcaneal
001	allergy	001	b17	001	cage
001	alexia	001	b27	001	cadmium
001	amyotrophic	001	b8	001	cervicodorsal
001	amphoric	001	battery	001	ceruloplasmin
001	amount	001	barefoot	001	cercaria
001	amorphous	001	barbiturate	001	cephalosporin
001	ammonium	001	balding	001	cephalization
001	ammonia	001	balanitis	001	centrolobular
001	aminopyrine	001	benzene	001	centrifugal
001	aminolevulinic	001	bentonite	001	cellulitis
001	aminoglycoside	001	being	001	cellularity
001	aminoaciduria	001	beef	001	chvostek
001	american	001	bee	001	chunky
001	amenorrhea	001	bisulfite	001	chronicum
001	ameboid	001	bisferiens	001	chromosome
001	anxious	001	birefringent	001	chorioretinitis
001	anuria	001	birds	001	chorionic
001	antrum	001	bipolar	001	chorea
001	antithrombin	001	biots	001	chondrocalcinos
001	anticoagulant	001	bifida	001	cholescintigrap
001	anosmia	001	bifid	001	cholecystectomy
001	anorexia	001	biconcave	001	chocolate
001	ano	001	blunted	001	chloroquine
001	annulus	001	blindness	001	chloramphenicol
001	anisocyte	001	blepharoptosis	001	chlamydia
001	angiotensin	001	blenorrhagicum	001	childhood
001	angiokeratoma	001	bland	001	cheyne
001	angioid	001	bowing	001	cheilosis
001	angina	001	bovis	001	citrobacter
001	anesthesia	001	bossing	001	cisternography
001	androgen	001	bosselated	001	circumscribed

001 anal
001 anacrotic
001 ana
001 apraxia
001 apprehensive
001 appetite
001 cleaner
001 cm
001 cotton
001 corrigan
001 corrected
001 cornea
001 corkscrew
001 oopiropouptiyrin
001 converting
001 convergence
nm conventional
001 controlling
001 controlled
001 oontralateral
001 *j^nBi?tj^pj~^ic?n
001 cjunsLxictive
001 consuricceci
001 constipation
001 consanguineous
001 conjunctivitis
001 conjugated
001 coni^l.
001 conglcmerate
001 cuuyescive
001 congestion
001 congenital
001 confluent
001 confined
001 confabulation
001 concentric
001 uuk*-«iilXciulOn
001 t*mtnig-imrt
001 compliance
001 ccnmunity
001 cccmLmìcating
001 columnar
001 collapse
001 colic
001 colchicine
001 cogwheel
001 coffee
001 coal
001 cryptorchidism
001 cryoprecipitate
001 crycglctaulin
001 cryofibrinogen
001 crushing
001 cross
001 crescentic
001 crepitus
001 crepitant
001 cremasteric
001 creatine
001 craving
001 brudzinsJci
001 bronchorrhea
001 brcmide
001 brittle
001 bradykinesia
001 brachial
001 detachment
001 desquamative
001 e^Agrpifwymflfrjon
001 dermal
001 derealization
001 deprivation
001 depressive
001 deposition
001 depleted
001 depigmented
001 depersonalizati
001 dependency
001 dental
001 demonstrative
001 delirium
001 dehydration
001 degradation
001 deficit
001 decubitus
001 decorticate
001 decision
001 decidua
001 decerebrate
001 decalcified
001 diuretic
001 diuresis
001 distortion
001 distorted
001 distended
001 distance
001 disseminated
001 disposal
001 disorientation
001 dislocation
001 diploe
001 diphyllcbothriu
001 dip
001 dimorphic
001 diluted
001 digested
001 difficulty
001 difficile
001 diastole
001 diabetic
001 downward
001 dorsal
001 doom
001 donath
001 dock
001 doca
001 drainage
001 dr4
001 dr2
001 dust
001 cigarati^a
001 cluster
001 clump
001 clanipr;ne
001 clitoris
001 clindanycin
001 dysarthria
001 earlofce
001 earache
001 echolalia
001 eccentric
001 edta
001 edrophonium
001 efferent
001 effective
001 egqphony
001 ehler
001 elliptocyte
001 elicited
001 electrical
001 emprosthotonos
001 embolism
001 emaciation
001 enzyme
001 environment
001 encpthalmos
001 enhancement
001 engorged
001 endothelial
001 endocarditis
001 encephalitis
001 enanthem
001 epithelium
001 ,ithel io d
001 epithelialized
001 episcleritis
001 ^ i h seal
001 epinephrine
001 epiderrooFhytosi
001 epidemic
001 epicardial
001 equally
001 erythrojphagocyt
001 erythrotnpialgia
001 erythrocytic
001 erythematosus
001 erysipelas
001 erlenmeyer
001 ergot
001 ergonovine
001 esophagitis
001 esophageal
001 eschar
001 ethanol
001 evening
001 extrarectal
001 extrapelvic
001 extinction
001 extensor

001	cramp	001	duroziez	001	extending
001	curve	001	dura	001	explosive
001	curschmann	001	dupuytrens	001	expiratory
001	curling	001	duplication	001	expansion
001	cuboidal	001	duplicated	001	exostoses
001	cytotcadc	001	dw3	001	exocrine
001	cystoides	001	dysuria	001	exhibitionistic
001	cystine	001	dysphonia	001	exfoliative
001	cycloserine	001	dyspareunia	001	exact
001	dark	001	dysfunction	001	f
001	danlos	001	dysesthesia	001	fqea-l[illegible]
001	device	001	dysdiadochokine	001	farm
001	familial	001	glycosuria	001	hypopauathyroid
001	falciparum	001	gluteal -	001	hypokinesia
001	failure	001	glutamine	001	hypoglycemia
001	factitious	001	glucosidase	001	hyped[illegible]ic
001	facilitation	001	glucocerebrosid	001	hypocellular
001	fevl	001	glcmerulonephri	001	hyphae
001	fetus	001	glabellar	001	hyperventilatio
001	fetor	001	gm	001	hypertrophic
001	fetoprotein	001	[illegible]	001	hypertonic
001	festinating	001	gonorrhea	001	hypersegmented
001	fenoprofen	001	gold	001	hyperpnea
001	female	001	goiter	001	hyperpigmented
001	feculent	001	group	001	hyperparathyroi
001	fecalith	001	ground	001	hyperostosis
001	fecalis	001	grimacing	001	hyperlucency
001	features	001	gravidarum	001	hyperkinetic
001	fearful	001	grasp	001	hyperkinesia
001	fear	001	granule	001	hyperemesis
001	filtrate	001	grandeur	001	hyperconcentrat
001	filipino	001	graded	001	hyperactive
001	fibrous	001	gustatory	001	hydrcphilia
001	fibrosa	001	guilt	001	hydrogen
001	fibroma	001	guaiac	001	hydrocarbon
001	fitorillary	001	gyrate	001	hydranmios
001	fluorescence	001	gyneccmastia	001	hydralazine
001	fluctuating	001	hard	001	hydatidiform
001	flocculaticn	001	haptoglobin	001	hyaluronidase
001	flexibility	001	halothane	001	hyalinization
001	fleischer	001	half	001	1131
001	flask	001	heterophile	001	ichthyosis
001	flame	001	herpetiformis	001	ideal
001	foul	001	hepatomegaly	001	ige
001	forward	001	hepatojugular	001	igd
001	formation	001	[illegible]	001	ii
001	forehead	001	heparin	001	illogical
001	forced	001	hemophilia	001	illicit
001	foramen	001	hemodialysis	001	iliac
001	fold	001	hemochrcmatosis	001	ileal
001	foci	001	heroatoxylin	001	impotence
001	fruity	001	[illegible]	001	irnpetigo
001	fruits	001	helper	001	impending
001	frothy	001	height	001	iscedance
001	frost	001	health	001	inpaction
001	frigidity	001	healed	001	imounodiffusion
001	friable	001	histrionic	001	umunocyte
001	fresh	001	histamine	001	imninization
001	fragmentation	001	hilum	001	imnersion
001	fragment	001	hiccups	001	immediately

001 fsh
001 fused
001 furuncle
001 fungus
001 gauchers
001 gastrointestina
001 gastrocolic
001 garbage
001 gangrenosum
001 gallstone
001 galactosidase
001 galactorrhea
001 gag
001 crenninal
001 genu
001 genitalia
001 general
001 ghon
001 insecticide
001 insect
001 inncminate
001 inner
001 ink
001 inhalation
001 infiltrated
001 infertility
001 infective
001 indocyanine
001 indistinct
001 indian
001 india
001 indentable
001 increasing
001 incongruous
001 inooagulable
001 including
001 lnarpropriate
001 inadequate
001 inability
001 iodine
001 icecac
001 iris
001 iridocyclitis
001 isoproterenol
001 isoniazid
001 ischemia
001 ivy
001 ixories
001 janeway
001 jacksonian
001 jacknife
001 jew
001 jet
001 jolly
001 juxtaglanerular
001 junctional
001 junction
UUI judgement
001 kayser
001 kaposi

001 howell
001 horseshoe
001 homer
001 horizontal
001 hookworm
001 hcmsspxiiality
001 home
001 hoarseness
001 husbandry
001 hump
C01 humerus
001 hum
001 huet
001 hysterosa1pinan
001 hypotonia
001 hypothermia
001 hypotension
001 hypoplasia
001 laminar
001 lamina
001 l^ir^llated
001 lagcphthalmos
001 lag
001 lactate
001 lactase
001 lacrimal
001 lacking
001 lack
001 laboratory
001 1abol1fti
001 labeled
001 lb
001 lethargy
001 lesicns
001 leptoneningeal
001 leprosy
001 lepromin
001 leonine
001 lenticular
001 lengthened
001 leaning
001 lhermittes
001 lh
001 livedo
001 lithium
001 1ipoidica
001 lipase
001 lip
001 lining
001 lincanycin
001 ligation
001 loose
001 lobulation
001 lotaulatfrt
001 lobular
001 lobar
001 lupus
\J\JJm [illegible]
001 lucid
001 lysozyme

001 image
001 inward
001 invagination
001 mtrarecxai
001 intralobular
001 intraepithelial
001 intestinalis
001 intjfyrnhai arap^i
001 interpedicular
001 interlacing
001 intention
001 insufflation
001 instrumentation
001 iHFpisf^t"p<i
001 insomnia
001 insipidus
001 insight
001 insertion
001 macrocytic
001 mchc
001 mexican
001 methysergide
001 methyldopa
001 methotrexate
001 methacholine
001 metaplasia
001 metanephrine
001 metamyelocyte
001 rostranvj^rj!viosis
001 metallic
001 metacarpophalan
001 nv?t^c^I'pa1
001 mesventeric
UUI mesothelicma
001 mesangial
001 mercury
001 msntai
001 menopause
001 meningitis
001 membranous
001 TnfVTaoo"!on
001 medusae
001 medium
001 medical
001 media
001 mitochoridri^1
001 mississippi
001 minimal
001 miner
001 military
001 miliary
001 miliaria
001 migrating
001 migrans
001 microunit
001 microspherocyte
001 micnDcytic
001 microaneurysm
001 mm
001 moulage

001 ketonuria
001 ketone
001 ketogenic
001 kernig
001 kerley
001 keratosis
001 keratopathy
001 keratodenna
001 kinase
001 knuckle
001 known
001 knob
001 knife
001 koilonychia
001 kussmaul
001 kuru
001 kveim
001 laxative
001 laughing
001 laterally
001 last
001 lanugo
001 landsteiner
001 lamp
001 myxedema
001 myringitis
001 myositis
001 mycglobinuria
001 myoglcbin
001 myodonus
001 myoclonic
001 myelofibrosis
001 myelocytic
001 ny obacterlal
001 inycelia
001 inyalgia
001 nausea
001 nasopharyngeal
001 narcolepsy
001 nalfon
001 newborn
001 iyn.I'puff'j. ff 001
001 neurone
001 neuronal
001 neurofibroma
001 iy^.TTT>?wn'P1
001 neuralgia
001 nephrotcncgraph
001 nephrotic
001 nephrolithiasis
001 nephrogram
001 nephrocalcinosi
001 neovascularizat
001 neoplastic
001 neologisms
001 negro
001 negligible
001 needle
001 necrotizing
001 necrolysis
001 lynphocytcsis
001 lymphatic
001 lynphangitis
001 mature
001 matted
001 matrix
001 matching
001 mastoid
001 mastication
001 mastectcny
001 masseter
001 maskli3ce
001 marriage
001 marginatum
001 marfans
001 manic
001 mandibular
001 mammils
001 malnutrition
001 mallory
001 Tnaformation
001 maleate
001 making
001 maculopapular
001 olfactory
001 omphalitis
001 onycholysis
001 onion
001 opisthotonos
001 opalescent
001 orthotonos
001 orthostatic
001 orthopnea
001 orifice
001 oriental
001 organized
001 organism
001 organ
001 osteosclerotic
001 osteosclerosis
001 Oeiteoporosis
001 osteomyelitis
001 osteoarthropath
001 osteitis
001 osmotic
001 otitis
001 other
001 outward
001 outline
001 overbite
001 ovary
001 oxyphilic
001 oxidase
001 oxalate
001 ax
001 pathologic
001 patchy
001 patch
001 partial
001 part
001 mottling
001 mottled
001 motor
001 morphine
001 more
001 moon
001 monoxide
001 monotonal
001 monocyte
001 monoarticular
001 monoamine
001 mandceberg
001 mole
001 mitism
001 mt,ishi:^jm
001 muramidase
001 mumps
001 maltiforme
001 mucus
001 mucous
001 mucor
001 mucopurulent
001 mucinosa
001 macin
001 periventricular
001 peritubular
001 peritonsilar
001 perirectal
001 perioral
001 periodontal
001 perineal
001 perilobular
001 perihilar
001 perig!omerular
001 periductal
001 "pericarditis
001 periapical
001 perianal
001 percussable
001 perchlorate
001 penicillamine
001 penetrating
001 pendular
001 pelger .
001 pedal
001 pectoris
001 pertoriloquy
001 photosensitivit
001 photic
001 phosphorus
001 phobia
001 Philadelphia
001 phecxdircroocytcm
001 phenytoin
001 phenylbutazone
001 phentolamine
001 phenotype
001 phenothiazine
001 phenindione
001 pharynx

001	necrobiosis	001	parotitis	001	pistol
001	nitrofurantoin	001	parkinsan	001	pink
001	nigricans	001	parenteral	001	pinguecula e
001	nightmares	001	paratracheal	001	pineal
001	night	001	parasite	001	pigmentosa
001	nicotine	001	paraortic	001	pigmented
001	nmol	001	paranoia	001	pigment
001	notching	001	parallel	001	piecemsal
001	notch	001	paradoxus	001	pica
001	northern	001	paracoagulation	001	plugging
001	nodosum	001	paracentesis	001	plication
001	nodal	001	papule	001	plethoric
001	nocturia	001	papular	001	plecnorphic
001	nullipara	001	papillitis	001	platybasia
001	nuclei	001	papilledema	001	plateau
001	nucleated	001	papilla	001	plate
001	nuclear	001	pansystolic	001	placental
001	nystagmus	001	pancreatitis	001	placenta
001	obturator	001	pancreatic	001	pneumomediastin
001	cfcstructing	001	palpitation	001	pneumaturia
001	obstipation	001	palmanental	001	pneumatosis
001	obliteration	001	paiTrar	001	pneumatocoele
001	occurring	001	pallidum	001	poverty
001	occlusive	001	painless	001	posttetanic
001	odor	001	pacemaker	001	postprandial
001	offspring	001	peyronies	001	postpartum
001	Ohio	001	artussig	001	portacaval
001	oily	001	perseveration	001	porphyria
001	oliguria	001	peroxidase	001	porphcbilinogen
001	K.I j<jcfnoT>7TThg^	001	T^yTTliciC'ig	001	pork
001	polyphasic	001	rapidly	001	scarlet
001	polypeptide	001	raising	001	scapular
001	polycythemia	001	TTVli^vLiir^ncy	001	scanning
001	polydonal	001	radial	001	scalenus
001	poliomyelitis	001	rabbits	001	sewer
001	poisoning	001	reversible	001	serous
001	poison	001	reversed	001	serosanguincus
001	poikilocyte	001	xetruversion	001	serology
001	protuberant	001	retrcperitoneal	001	septate
001	protrusion	001	retrobulbar	001	sentinel
001	protrude	001	retiring	001	sensorineural
001	protcporptayrin	001	reticuloendothe	001	senile
001	prothranbin	001	T^ti^il.in	001	semitrcpical
001	protease	001	•pprHriilaTig	001	semi
001	protamine	001	reticular	001	self
001	prosthetic	001	|ioca igC?-j ^t.ion	001	seductive
001	prostatitis	001	resolution	001	seborrhea
001	pronator	001	resistance	001	seasonal
001	prominence	001	resection	001	season
001	proliferans	001	L'GSl	001	shot
001	prolactin	001	religiosity	001	shiny
001	projection	001	relative	001	shin
001	proinsulin	001	related	001	shifting
001	productive	001	regenerative	001	shellfish
001	production	001	regenerated	f\f\n 001	snelr
001	prooonvertin	001	refractory	001	sheathing
001	processes	001	reference	001	shallow
001	prarsiiTynflj^	001	EPOini^ncY	001	shadow
001	probenecid	001	recruit	001	size
001	proaccelerin	001	recording	f\01	sittii^

001 primagravida
001 priapism
001 pretibial
001 pressured
001 preserved
001 preoccupation
001 preference
001 precipitation
001 pr
001 psychosis
001 psoas
001 pseudopolyposis
001 pseudofracture
001 psammoma
001 punctate
001 punched
001 pulsating
001 pulsatile
001 puffy
001 pubic
001 pyuria
001 pyrimethamine
001 pyridoxine
001 pyrazinamide
001 pyramidal
001 pyoderma
001 pyloric
001 pyelogram
001 q
001 qrs
001 quinine
001 quinidine
001 quincke
001 queckenstedt
001 raynauds
001 rare
001 spider
001 spheroid
001 sphenoid
001 spastic
001 ss
001 stupor
001 stuart
001 stridor
001 striae
001 streptolysin
001 streptokinase
001 streaked
001 streak
001 strawberry
001 straw
001 strabismus
001 stooped
001 stokes
001 sting
001 stimulus
001 stimuli
001 stiffness
001 sternomastoid
001 stellate
001 reasoning
001 reactive
001 reaction
001 rhinorrhea
001 rheumatoid
001 rheumatic
001 ritualized
001 rise
301 rigor
001 rightward
001 ridging
001 ria
001 routine
001 rouleaux-
001 roth
001 rosette
001 ranberg'
001 rakitansky
001 rodents
001 robertson
001 rpr
001 rubra
001 rubor
001 saturation
001 sarcoma
001 sanitation
001 saline
001 scrotum
001 scraping
001 sderanalacia
001 sderodactyly
001 sderitis
001 scintillating
001 schizocyte
001 schirroer
001 scars
001 terminated
001 tensilon
001 tenosynovitis
001 tpnpsmi.is
001 [illegible]
001 tearing
001 teardrop
001 thyroidectany
001 through
001 tlii.uiuboplastin
001 thranbophlebiti
001 thranbin
001 throbbing
001 thready
001 thorotrast
001 thoracentesis
001 thiouracil
001 thickness
001 thiazide
001 thallium
001 tinnitus
001 tinel
001 tick
001 toxoid
001 site
001 siphon
001 sinusoidal
001 sinusitis
001 simultaneous
001 silver
001 silly
001 sighing
"01 sialorrhea
·01 sialography
:oi sialectasis
001 skeletal
001 slit
001 smooth
001 smoking
001 snake
001 southwestern
001 sour
001 sordellii
001 sonolucent
001 somnolence
001 somnambulism
001 somatoserisory
001 somatic
001 social
001 spur
001 spring
001 spontaneous
001 spongiform
001 splinter
001 splenectany
001 splash
001 spirochete
001 spiral
001 spindle
001 spina
001 urethritis
001 used
001 vasoactive
001 varicose
001 wwX VCLLXV^rxxCl
001 vani1lylirranrteli
001 valley
001 valgum .
001 vaginitis
001 vagina
001 vacuole
001 vacuolated
001 vaocination
001 vdrl
001 veterinarian
001 vessels
001 vessel
001 vesicle
001 vertigo
001 vera
001 venule
001 vegetables
001 vitreous
001 vitiligo

001 stature
001 star
001 staining
001 stained
001 stabbing
001 suture
001 surreptitious
001 sural
001 supraventricula
001 supraclavicular
001 suppressor
001 supine
001 sunlight
001 sunned
001 sulfonamide
001 suicide
001 suffusion
001 succussion
001 submandibular
001 subhyaloid
001 subepithelial
001 subarachnoid
001 subacute
001 sweat
001 swan
001 swallow
001 system
001 synthetic
001 [illegible]
001 symptom
001 tartrate
001 tarry
001 target
001 tangles
001 [illegible]
001 tambour
001 tachypnea
001 table
001 tetracycline
001 tetrachloride
001 [illegible]
001 testosterone
002 accumulation
002 adynamic
002 adh
002 aggressive
002 alveoli
002 altitude
002 alopecia
002 alcoholism
002 albumin
002 amnesia
002 antitrypsin
002 antitoxin
002 angiana
002 angioinvasive
002 angiocentric
002 analgesic
002 apophyseal
002 apnea

001 toxin
001 tourniquet
001 tophus
001 tophi
001 tolbutamide
001 trunk
001 truncal
001 trousseau
001 tropical
001 trisnus
001 triple
001 trimester
001 triglyceride
001 trigeniinal
001 triamterene
001 trapezius
001 transverse
001 transit
001 transferrin
001 transaminase
001 turner
001 turbinate
001 turbid
001 tumefaction
001 tufting
001 tubercle
001 two
001 tyramine
001 tympanites
001 ulnar
001 unusual
001 untreated
001 unpasteurized
001 unit
001 unequal
001 uncooked
001 uncontrollable.
001 unconsciousness
001 upward
001 uroporphyrin
001 uriniferous
001 uricase
002 cooing
002 contracture
002 contact
002 constriction
002 [illegible]
002 conjugate
002 confirmation
002 [illegible]
002 complex
002 compensated
002 coma
002 colored
002 colloid
002 collapsed
002 colicky
002 coccidioidin
002 coat
002 coarctation

001 vital
001 visualization
001 viridans
001 violent
001 vincristine
001 villi
001 vii
001 view
001 von
001 voltage
001 vulvar
001 watery
001 washout
001 walled
001 walking
001 waddling
001 wed
001 whorl
001 whispered
001 when
001 wheezing
001 withdrawal
001 wire
001 wilsons
001 wild
001 widened
001 worthlessness
001 wool
001 wolff
001 writhing
001 x
001 xanthcnata
001 [illegible]
001 xerostomia
001 xylose
001 yawning
001 year
001 zz
002 abundant
002 abortion
002 ability
002 achilles
002 exogenous
002 ferritip
002 fenur
002 fish
002 fiber
002 flew
002 flight
002 foreign
002 follicle
002 folate
002 friability
002 frequent
002 fragility
002 fusion
002 furrcwed
002 gastrin
002 gastrectdoy
002 ganglia

002	arteriosus	002	crypt	002	gametocyte
002	arrhythmia	002	crying	002	gel
002	arch	002	curvilinear	002	glucagon
002	assay	002	cystoscopy	002	glomerular
002	aseptic	002	d	002	gout
002	ascorbic	002	dexamethasone	002	groin
002	augmented	002	destructive	002	green
002	auditory	002	dermatitis	002	grand
002	av	002	depressed	002	guarding
002	axis	002	defecation	002	haustra
002	basophil	002	death	002	hair
002	basilar	002	diverticula	002	herpes
002	bacterial	002	disorder	002	hepatocyte
002	babinski	002	discrete	002	henle
002	beta	002	discoloration	002	hemoptysis
002	bengal	002	diplopia	002	hemolysin
002	bence	002	dioxide	002	hemoglobinuria
002	bitemporal	002	difference	002	hematuria
002	biliary	002	dot	002	hemagglutinatio
002	bicuspid	002	dose	002	heel
002	bicarbonate	002	doppler	002	heavy
002	blue	002	dog	002	heat
002	boggy	002	droplet	002	healing
002	broad	002	drop	002	histoplasmin
002	bradycardia	002	dullness	002	hirsutism
002	buttock	002	dull	002	hiatal
002	bursitis	002	ductus	002	hormone
002	burn	002	early	002	honeycomb
002	bulge	002	ectopic	002	homans
002	buffy	002	ectasia	002	hoffman
002	cavitation	002	echogenicity	002	hypoglycemic
002	carotenoids	002	edge	002	hypertonia
002	carcinoma	002	edematous	002	hypersensitivit
002	capillaries	002	elongation	002	hyperresonant
002	candidiasis	002	emptying	002	hyperkeratosis
002	canal	002	enteroenteric	002	hyperemia
002	calcitonin	002	enterococcus	002	hydroxyproline
002	cerebellar	002	endometriosis	002	hydroxyindoleac
002	cephalad	002	epistaxis	002	iii
002	center	002	episodic	002	ileum
002	chromatin	002	erythroblast	002	ileocecal
002	christmas	002	erythematous	002	immobilization
002	charcot	002	european	002	inverted
002	clot	002	euphoric	002	intussusception
002	closing	002	every	002	intrusive
002	clearance	002	extrauterine	002	intravascular
002	clawing	002	extrahepatic	002	intrauterine
002	coombs	002	expansile	002	intracerebral
002	interosseous	002	mu	002	purpura
002	internuclear	002	mycoplasma	002	purple
002	inhibitor	002	neutrophilic	002	pyogenic
002	inguinal	002	neutralization	002	qt
002	infusion	002	nitroglycerin	002	r
002	infundibular	002	nicking	002	range
002	influenza	002	notched	002	rales
002	infiltrate	002	normoblast	002	raised
002	indium	002	o	002	radioimmunoassa
002	indentation	002	obstructed	002	race
002	incontinence	002	obsession	002	retinitis
002	ix	002	olecranon	002	reticulocyte

002 jerk
002 jones
002 keratitis
002 knee
002 kyphosis
002 kyphoscoliosis
002 lasting
002 laryngoscopy
002 lacrimation
002 labile
002 ldh
002 leyden
002 leukemia
002 le
002 line
002 lightning
002 lift
002 lid
002 libido
002 longer
002 location
002 lobe
002 lymphoblast
002 marginal
002 mandible
002 male
002 malaria
002 malar
002 mal
002 major
002 magnesium
002 macrophage
002 mcv
002 mch
002 metyrapone
002 metrorrhagia
002 metatarsophalan
002 metatarsal
002 menses
002 menorrhagia
002 mellitus
002 mediterranean
002 medial
002 meat
002 milky
002 mild
002 migratory
002 migraine
002 middle
002 microvascular
002 monoplegic
002 moist
002 muscularis
002 mucoid
002 sulfur
002 suitable
002 suicidal
002 subperiosteal
002 ofiftfyrtqai
002 syphilis
002 optic
002 opening
002 opacity
002 orbital
002 osteophyte
002 osteoid
002 out
002 p
002 p2
002 patient
002 patent
002 parotid
002 paravertebral
002 parasternal
002 paraaortic
002 palpation
002 painful
002 perivenous
002 perivascular
002 perithyroid
002 peripartum
002 periostitis
002 periorbital
002 perifollicular
002 rxyrjf^yrij\irq
002 perfusion
002 peptide
002 peptic
002 penicillin
002 pg
002 physical
002 phenomenon
002 phase
002 pharyngitis
002 plague
002 pneumothorax
002 pneumonia
002 postmenopausal
002 portion
002 poor
002 pool
002 polydipsia
002 polyarticular
002 polarized
002 provocative
002 pranyelocyte
002 prcncnocyte
002 profuse
002 primary
002 presystolic
002 precordial
002 precipitin
002 psoriasis
002 pustule
003 arteriovenous
003 agVvacrfcQg
003 attenuated
003 azotemia
003 b
003 basophilic
002 retardation
002 resting
002 responsive
002 resistant
002 resin
002 residual
002 relaxation
002 rebound
002 rhonchi
002 rigid
002 ridge
002 rickettsii
002 rotation
002 rose
002 root
002 ruddy
002 S3
002 S4
002 saddle
002 scoliosis
002 schizophrenia
002 schizont
002 sexual
002 sex
002 severity
002 serosa
002 serial
002 sequence
002 sella
002 secretion
002 seam
002 sharp
002 sinusoid
002 silhouette
002 sideroblast
002 skip
002 snap
002 southern
002 solitary
002 sprue
002 sphincter
002 species
002 sparing
002 squatting
002 string
002 stress
002 store
002 stone
002 stippling
002 sternum
002 surface
002 suprasternal
002 suprapubic
002 summation
003 frcm
003 frequency
003 fuscbacterium
003 gamma
003 grwth
003 h^>atocellular

This Appendix collects examples of lexical tracing and lexical mapping between user expressions and the INTERS1ST-I knowledge base of findings. These traces were generated on a system designed to test both the content of the lexicon and the efficacy of lexically-driven NLP. The details of the actual scoring of results and retrieval of findings are omitted as they are test-application specific and do not affect the question of lexical design or the relevance of terminology to IITERIIST-I findings. Implementation of the test system has been done both in C on a Vax ll\780 and in Turbo Pascal on an IBM PC\AT.

Words in the input string are "tokenized"—i.e., matched against forms that appear as entries in the lexicon. In the process, adjacent words that can be analyzed as phrases in the lexicon are combined into phrasal tokens. The canonical terms that are directly linked to each token are listed beneath it, and the type of links ("Canonical," "Morph:," "Syn:$_t$" "Rel-to:") are identified. Furthermore, the canonical terms associated with these terms are listed, and are signalled graphically by an additional level of indentation. In each trace, only the top 10% of scored findings that the system retrieved as candidate matches are displayed.

EXAMPLE I

```
Enter a phrase ---> grandmother had diabetes

Tree structure of tokens read in:
  Token = "grandmother"
      Related to
        Canonical: "family".
          Morph. Var.: "familial".
          Synonym: "relative".
      Related to
        Canonical: "relative".
          Synonym: "family".
          Related to: "family-history".
  Token = "diabetes"
    Canonical
        Canonical: "diabetes".
          Morph. Var.: "diabetic".
          Related to: "insipidus".
          Related to: "carbuncle".
          Related to: "insulin".
          Related to: "mellitus".
          Related to: "tolbutamide".


results:

(12)  420: diabetes insipidus family hx
(12)  421: diabetes mellitus family hx
```

Comments:

Note how the system prompts the user to distinguish between related findings. The user entered the general term "diabetes" and was asked to select between two types of diabetes.

EXAMPLE 2

```
Eater a phrase____> pitting edema in thighs

Tree structure of tokens read in:
  Token = "pitting"
    Canonical
        Canonical: "pitting".
          Synonym: "depression".
          Related to: "spleen".
          Related to: "teeth".
  Token = "edema"
    Canonical
        Canonical: "edema".
          Morph. Var.: "edematous".
          Synonym: "tumefaction".
          Related to: "tumefaction".
          Related to: "papilledema".
  Token = "in"
    Canonical
        Canonical: "in".
          Synonym: "en=".
          Synonym: "within".
          Related to: "location".
          Related to: "position".
  Token = "thighs"
      Morph. Var.
        Canonical: -thigh".
          Related to: "femoral".
          Related to: "groin".
          Related to: "leg".


results:

(27)  913: leg <s> edema tense non pitting
(24)  503: extremity <ies> moist warm edematous non pitting distal only

Comments:
--------

This trace helps to illustrate a couple of problems with the current
system.  First, it is unable to distinguish between multiple meanings
of a word in different contexts.  Although only the best guesses have
been returned, the system listed all findings that contain the word
"depression" as candidate matches, even those in which the sense of
"depression" is "unhappy" or "lethargic."  Second, the system can not
handle negation.  It returned "non pitting" findings when the user
said "pitting."
```

EXAMPLE 3

```
Enter a phrase_____>  reduced libido

Tree structure of tokens read in:
  Token = "reduced[11]
      Synonym
        Canonical: "decreased".
          Norph. Var.: "decrease".
          Synonym: "less".
          Synonym: "-penia".
          Synonym: "a=".
          Synonym: "brady=".
          Synonym: "kyp=".
          Related to: "impaired".
          Related to: "abnormal".
          Related to: "absent".
          Related to: "amount".
  Token = "libido"
    Canonical
        Canonical: "libido".
          Related to: "sex".


results:

(15)  974: libido decreased
```

EXAMPLE 4

```
Enter a phrase_____> palpable skull mass

Tree structure of tokens read in:
  Token = "palpable"
    Canonical
        Canonical: "palpable".
          Morph. Var.: "palpation".
          Synonym: "expansile".
          Related to: "touch".
  Token = "skull"
    Canonical
        Canonical: "skull".
          Synonym: "cranial".
          Related to: "head".
          Related to: "diploe".
          Related to: "cerebral".
  Token = "mass"
    Canonical
        Canonical: "mass".
          Synonym: "lesion".
          Synonym: "knob".
          Related to: "granuloma".


results:

(27)  1547: skull palpable mass <es>
```

EXAMPLE 5

```
Enter a phrase____> difficulty breathing

Tree structure of tokens read in:
  Token = "difficulty"
    Canonical
        Canonical: "difficulty".
          Synonym: "dys=".
  Token = "breathing"
    Canonical
        Canonical: "breathing".
          Morph. Var.: "breath".
          Synonym: "=pnea".
          Synonym: "respiratory".
          Related to: "tachypnea".
          Related to: "stridor".
          Related to: "apnea".
          Related to: -chest".
          Related to: "dyspnea".
          Related to: "expiratory".
          Related to: "lung".
          Related to: "movement".
          Related to: "orthopnea".
          Related to: "respiratory".
          Related to: "wheezing".


results:

(12)  452: dyspnea abrupt onset
(12)  453: dyspnea acute recurrent attack <s> hx
(12)  454: dyspnea at rest
(12)  455: dyspnea at rest relieved by recumbency
(12)  456: dyspnea exertional
(12)  457: dyspnea improvement after hemoptysis hx
(12)  458: dyspnea paroxysmal nocturnal

Comments:
--------

This trace demonstrates the system's ability to handle input phrases
that coincide with single-word medical terms having Greek or Latin
morphemes. In this case, the bound morpheme "dys-," a synonym of
"difficulty" can be combined with the bound morpheme "=pnea," meaning
"breathing" to call up "dyspnea" findings.  Mote, once again, the
discriminations that the returned findings prompt the user to make.
```

```
EXAMPLE 6

Enter a phrase____> blood in stools

Tree structure of tokens read in:
  Token = "blood"
    Canonical
        Canonical: "blood".
          Morph. Var.: "bloody".
          Synonym: "artery".
          Synonym: "=aemia".
          Synonym: "=emia".
          Synonym: "hem=".
          Related to: "blood-streaked".
          Related to: "transfusion".
          Related to: "bleeding".
          Related to: "hematemesis".
          Related to: "agglutinin".
          Related to: "albumin".
          Related to: "anemia".
          Related to: "angio=".
          Related to: "anticoagulant".
          Related to: "antihemophilic".
          Related to: "artery".
          Related to: "azotemia".
          Related to: "blood-pressure[11].
          Related to: "blood-transfusion".
          Related to: "culture".
          Related to: "heart".
          Related to: "hemato=".
          Related to: "hemoglobin".
          Related to: "hemophilus".
          Related to: "hemoptysis".
          Related to: "hemorrhage".
          Related to: "hyperemia".
          Related to: "leukemia".
          Related to: "liver".
          Related to: "plasma".
          Related to: "platelets".
          Related to: "polyp".
          Related to: "pool".
          Related to: "pulsus".
          Related to: "serosanguinous".
          Related to: "valve".
          Related to: "vein".
  Token = "in"
    Canonical
        Canonical: "in".
```

002	syndesmcphyte	003	basal	003	hemosiderin
002	syncope	003	band	003	h«nrphi"l\is
002	taste	003	binding	003	hemianopsia
r\no	tall	003	bilirubin	003	hematocQi^^ls
uuz	tail	003	bluish	003	henaternesis
002	tetani	003	bloody	003	histolytica
002	terminal	003	blastanycin	003	hilar
002	tense	003	blast	003	hot
002	temperature	003	bladder	003	iicBoonymcus
002	thyroglobulin	003	bundle	003	hypogastirxutn
002	thought	003	bulb	003	hypejrtrcphy
002	thin	003	c	003	hyperesthesia
002	thenar	003	cat	003	ideas
002	tightness	003	c^nqnylofvict^r	003	irltracranial
002	toxemia	003	c*Jek1c^L\IlIS	003	intracavitary
002	tortuosity	003	cnolecystograpn	003	intimate
002	treponema	003	wicLxn	otn UVJ	intimal
002	transplantation	003	circulation	003	intestinal
002	turcica	003	clubbed	003	^T*\"-^^"s^^pn`i ar
002	tuft	003	closure	003	inocculation
002	twitching	003	dick	003	injury
002	umbilicus	003	corneal	003	inflammatory
002	umbilical	003	copper	003	infection
002	unreactive	003	content	003	increase
002	urticaria	003	confusion	003	jejuni
002	urobilinogen	003	crescent	003	jugular
002	urination	003	current	003	keto
002	urxc	003		003	latex
002	urgsncv	003	cillr^rvarXpy	003	
002	using	003	cyclic	003	liter
002	lose	003	V4JCEX51JUKSIIL.	003	lipcproteinemia
002	us	003	dementia	003	long
002	vascularity	003	degeneration	003	loading
f\f\o uuz	voxxcLuxan	UUJ	decrescendo	AM UU3	maximal
002	vesical	003	deafness	003	macular
002	vertical	003	diffusion	003	megakaryocyte
002	velocity	003	<iiab»tes	003	median
002	visible	003	double	003	minute
002	viscosity	003	echogenic	003	minor
002	virus	003	eehirEQCoccal	003	midline
002	vibratory	003	ecchymosis	003	TiOY+ilast
002	voluble	003	elisa	003	nul'tocida
002	waxy	003	enterob^cter	003	multinucleated
002	week	003	entameba	003	mv^o^^?^ip^Ij^\J.
002	web	003	epidermidis	003	myelocyte
002	wiring	003	epidermal	003	obesity
002	widening	003	erythiodenua	003	occipital
003	a2	003	estrogen	003	ophthalmoplegia
003	adenopathy	003	evoked	003	Das^>c*!rrella
003	agnosia	003	exudate	003	pas
003	age	003	extravasation	003	parietal
003	alternans	003	^xtr2ffliedul1ary	003	paresthesia
003	aldosterone	003	Avnnhthal11CS	003	palm
003	anputation	003	excretion	003	palatal
003	amp	003	fatty	003	pad
003	antihencphilic	003	fibrinogen	003	petechiae
003	ankle	003	fibrillation	003	periungual
003	anemia	003	forceful	003	peritoneum
003	apex	003	follicular	003	peristalsis
003	arthritis	003	foamy	003	periodic

003	periarticular	003	zoster	004	glad
003	plantar	004	abuse	004	granular
003	potassium	004	absorption	004	hernia
003	poststenotic	004	activity	004	hepatitis
003	postictal	004	acidophilic	004	hip
003	positional	004	acidity	004	hypesthesia
003	pruritus	004	adrenal	004	igm
003	prostate	004	alkaline	004	iga
003	prominent	004	amebic	004	improvement
003	preparation	004	ameba	004	intrinsic
003	pseudomembrane	004	antacid	004	intervertebral
003	pubis	004	ankylosis	004	interval
003	ratio	004	angle	004	intercostal
003	rate	004	aspiration	004	inspiratory
003	retraction	004	atrioventricula	004	irregularity
003	repetitive	004	aura	004	leptospira
003	remission	004	augmentation	004	lead
003	reflux	004	avascular	004	liquid
003	redness	004	bacteroides	004	like
003	rhythm	004	black	004	lordosis
003	rigidity	004	botulinum	004	longitudinal
003	rumbling	004	cardiac	004	lucency
003	rugae	004	carbon	004	lymphangiograph
003	s1	004	chloride	004	macule
003	salt	004	clonus	004	mentation
003	sad	004	claudication	004	megaloblast
003	scratch	004	costovertebral	004	microscopic
003	scotoma	004	cortex	004	motile
003	schilling	004	contraceptive	004	morning
003	septum	004	conduction	004	myelomonocyte
003	secretin	004	collateral	004	narrow
003	sgpt	004	coli	004	nocturnal
003	sgot	004	creatinine	004	osteolytic
003	shift	004	cutaneous	004	osmolality
003	sided	004	culdocentesis	004	oxygen
003	sickle	004	descending	004	pattern
003	sleeping	004	delayed	004	parathormone
003	sole	004	degree	004	paralysis
003	spot	004	decrease	004	paradoxic
003	space	004	diverticulum	004	persistent
003	straight	004	distention	004	pentagastrin
003	stepwise	004	disc	004	penis
003	standing	004	drug	004	phosphate
003	suprasellar	004	duration	004	phenylephrine
003	sudden	004	duodenal	004	phalanges
003	sudan	004	e	004	pitting
003	subluxation	004	ear	004	popliteal
003	subclavian	004	effusion	004	polyp
003	swallowing	004	elevated	004	proteus
003	synovium	004	endobronchial	004	proteinuria
003	symphysis	004	epithelial	004	proliferation
003	symmetrical	004	extraluminal	004	prolapse
003	tension	004	exertional	004	precipitated
003	tendon	004	feet	004	pseudomonas
003	temporal	004	first	004	pupil
003	thumbprinting	004	fine	004	pulsus
003	touch	004	filled	004	pulsation
003	travel	004	flushing	004	radiolucent
003	transient	004	fluctuant	004	radioiodine
003	trachea	004	fourth	004	radiation

003	t"M^TT5'-'l in
003	unrelieved
003	uniform
003	water
003	within
003	worse
004	soft
004	SpiellKJCfcaJdly
004	spike
004	spasticity
004	striated
004	steroid
004	SUi.i.v;Ail"Ju6Q,
004	t
004	tachycardia
004	tortno11^
004	transfusion
004	tsh
004	urethra
004	urea
004	urate
004	uterus
004	vasculitis
004	variable
004	valvular
004	valsalva
004	vitamin
004	villous
004	warm
004	weight
004	wrist
005	a
005	abscess
005	adnexal
005	adhesion
005	alpha
005	alcohol
005	amyl
005	animal
005	approx
005	before
005	bite
005	brucftlla
005	jUj-\7l1L3.B^-ift?fl.\
005	breast
005	f^lciuin
005	character
005	changing
005	ocugh
005	cortisol
005	ccnplete
005	cranial
005	[illegible]
005	differentiated
005	dry
005	enteric
005	field
005	flutter
005	food
005	friction
004	folds
004	frontal
004	gaze
004	gait
004	girdle
004	giant
005	interlobular
005	inspection
005	infarction
005	indirect
005	jaw
005	late
005	larynx
005	lactose *
005	laceration
005	lumen
005	massive
005	marking
005	medullary
005	mean
005	milk
005	midsystolic
005	myelcblast
005	near
005	ng
005	nitrogen
005	nitrite
005	papillary
005	pallor
005	periumbilical
005	platelets
005	potential
005	position
005	product
005	response
005	renin
005	nab
005	s2
005	sacroiliac
005	s9ftnsi*frHiivi
005	scalp
005	septal
005	separation
005	sensory
005	shoulder
005	slow
005	.sofiium
005	^>littii^
005	splenic
005	spinal
005	sustained
005	suprarenal
	u
005	t4
005	tactile
005	testis
005	teeth
005	tonsil
005	toe
005	tricuspid
004	retention
004	saccular
004	stiymental
004	secondary
004	slight
004	sleep
005	zone
006	air
006	anylase
006	asymmetrical
006	attack
006	bl2
006	block
006	bleeding
006	cava
006	cervix
006	clear
006	ooroiiciry
006	cocci
006	deviation
006	dense
006	diarrhea
006	draining
006	dysphagia
006	emg
006	enboli
006	exercise
006	excessive
006	fallopian
006	film
006	flat
006	flank
006	foot
006	fremitus
006	.gallop
006	gingiva
• 006	h°^/^^rin^
006	hour
006	hyperplasia
006	impaired
006	involuntary
006	inclusion
006	klebsiella
006	^inesur •
006	marcrin
006	manometry
006	meningitidis
006	month
006	nail
006	rfr-rt-;*-i TV=>H
006	ova
006	patellar
f\r\c	peripheral
uuo	phosphatase
006	pleura
006	pladn
006	portcil
006	polyuria
006	progressive
006	pulmonic

005 gangrene
005 hallucinations
005 [illegible]
005 hematocrit
005 hypervascularit
005 hypertension
005 igg
005 illness
005 [illegible]
005 intramural
005 intolerance
005 into
006 thickened
006 therapy
006 unresponsive
006 urinary
006 varices
006 vaginal
006 vesicular
006 vena
006 visual
006 viii
006 wedge
007 abrupt
007 acth
007 agglutination
007 arteritis
007 basement
007 behavior
007 bowel
007 brcwn
007 breathing
007 cavity
007 capacity
007 calvarium
007 dostridium
007 cord
007 conjunctiva
007 coarse
007 crater
007 cyst
007 cyanosis
007 destruction
007 dermatomal
007 deposit
007 deep
007 dyspnea
007 electrcphoresis
007 erosion
007 external
007 fasting
007 facies
007 facial
007 [illegible]
007 flaccid
007 for
007 free
007 globulin
007 gravity
007 homogeneous
005 trichinella
005 tremor
005 treatn^rrt
005 transmitt&i
005 trabeculae
005 typhi
005 V
005 volume
005 vocal
005 well
005 wound
005 yellcw
007 periportal
007 plaque -
007 pneumocystis
007 pneumococcus
007 posture
007 premature
007 rest
007 [illegible]
007 ring
007 rib
007 salmonella
007 single
007 spleen
007 spasm
007 st
007 subdiaphragmat±
007 syndrooB
007 thigh
007 tip
007 tube
007 [illegible]
007 vcndting
007 white
007 worker
008 anplitude
008 calyceal
008 cerebral
008 change
008 cortical
008 colonoscopy
008 depression
008 diplococci
008 edema
008 eosinophil
008 femoral
008 fixation
008 gastric
008 gonorrheae
008 head
008 holosystolic
008 hyaline
008 immature
008 involvement
008 intrahepatic
008 leukocyte
008 light
008 lumbar
008 ch
006 resorprcion
006 residence
006 reciprocal
006 round
006 shunt
006 short
006 stricture
006 subcutaneous
006 [illegible]
006 sweating
006 thronbi
006 throat
008 tolerance
008 visualized
009 anyloid
009 arteriolar
009 background
009 brash
009 calf
009 cholesterol
009 corticosteroid
009 delusions
009 displaced
009 direct
009 diet
009 diaçciragm
009 diameter
009 during
009 erythema
009 ethnic
009 fever
009 fixed
009 flexion
009 foam
009 •franciF^I1r\
009 fracture
009 histiocyte
009 hla
009 impulse
009 internal
009 neurologic
009 nerve
009 protein
009 reed
009 red
009 rectum!
009 [illegible]
009 sternberg
009 swelling
009 telangiectasia
009 thrill
009 thick
009 time
009 ulceration
009 venography
010 anaerobic
010 alveolar
010 common
010 cast
010 displacement

007	insulin	008	oral	010	discharge
007	inferior	008	pancreatography	010	eosincphilic
007	indurated	008	palpable	010	enhanced
007	induced	008	pestis	010	gradient
007	iv	008	pelvis	010	isolation
007	jaundice	008	ph	010	iron
007	loop	008	posterior	010	innnurcblast
007	lyiiifcjA 'lUllkA	008	prolonged	010	loss
007	material	008	purulent	010	moderate
007	meq	008	pyogenes	010	main
007	necrotic	008	rapid	010	onset
007	not	008	remote	010	pigmentation
007	numerous	008	shaped	010	peritoneoscopy
007	*on*	008	significant	010	rounded
007	output	008	suppression	010	respiratory
007	over	008	tender	010	recurrent
007	paroxysmal	008	tissue	010	radiating
007	parenchymal	008	total	010	specific
010	sigmoidosccpy	013	superior	019	branch
010	segment	013	type	019	cystic
010	tract	013	trophozoite	019	endoscopy
010	ureter	013	times	019	fluorescent
010	wide	013	without	019	ml
010	yersinia	014	absent	019	TraaHj^gtlrral
Oil	aureus	014	chronic	019	nodule
on	atrial	014	dl	020	ciiolangiography
Oil	aneurysm	014	fat	020	carotid
Oil	bodies	014	factor	020	diastolic
Oil	continuous	014	interstitial	020	mitral
Oil	[illegible]	014	mucosal	020	«[illegible]
on	cold	014	nocardia	020	transtiwt\<r&\1
Oil	duodenum	014	neisseria	020	ventricular
on	distribution	014	obstruction	021	atrim
Oil	intermittent	014	staphylococcus	021	apicad
on	lateral	014	sclerosis	021	.aortograptiy
on	mnnonucl^^r	015	antigen	021	cryptococcus
on	nose	015	anterior	021	exacerbation
on	pyelography	015	angiography	021	focal
on	peritoneal	015	body	021	glcmeruli
on	rod	015	candida	021	hand
on	surgery	015	gross	021	pulse
on	speech	015	low	021	percent
on	solid	015	myelogram	021	tuberculosis
on	seizure	015	mucosa	021	wall .
on	tularensis	015	mg	022	atrophy
on	tubular	015	nodular	022	aspergillus
on	titer	015	neoplasm	022	dilatation
012	aortic	015	percussion	022	fast
012	bile	015	rectal	022	h^>atic
012	back	015	sinus	022	leaflet
012	cytology	015	tongue	022	large
012	OutILLGLS^ulOn	015	uptaJce	022	mouth
012	cervical	015	venous	022	plasha
012	epigastrium	016	abnozmal	022	thickening
012	enlargement	016	central	022	wave
012	ejection	016	calcified	023	at
012	gas	016	disease	023	angiocardiograp
012	gallbladder	016	finger	023	coccidioides
012	hemorrhage	016	face	023	exposure
012	[illegible]	016	glucose	023	only

012	movement	016	legionella	023	radioisotope
012	motion	016	malignant	023	substernal
012	rash	016	neutrophil	023	sediment
012	stenosis	016	necrosis	027	after
012	severe	016	neck	027	bacteria
012	trauma	016	renal	027	feces
012	toxoplasma	016	streptococcus	027	open
012	third	017	ascitic	027	upper
012	vertebral	017	acute	027	ulcer
013	affect	017	contrast	028	extremity
013	actinomyces	017	fibrosis	028	gram
013	breath	017	marked	028	irregular
013	containing	017	negative	029	enema
013	deformity	017	proximal	029	in
013	elevation	017	relieved	029	mycobacterium
013	eeg	017	second	029	skull
013	fistula	018	crystal	029	ventricle
013	high	018	dilated	030	pericardial
013	inflammation	018	ingestion	030	pelvic
013	mesenteric	018	normal	031	brain
013	membrane	018	occlusion	031	duct
013	mcg	018	pregnancy	031	distal
013	occupation	018	superficial	031	generalized
013	pancreas	019	atypical	031	predominant
031	quadrant	034	present	104	urine
032	arterial	034	thyroid	105	liver
032	esophagus	035	reflex	110	pain
032	retina	035	tenderness	114	positive
032	rbc	036	granuloma	119	gtr
032	vein	037	multiple	119	chest
032	valve	037	sound	121	recent
033	acid	038	smear	133	fluid
033	border	039	area	138	or
033	calcification	039	intestine	142	blood
034	arm	039	localized	144	kidney
034	csf	090	barium	157	than
034	catheterization	091	with	184	decreased
034	caseating	095	of	188	skin
034	ivp	096	bone	219	increased
034	infiltration	096	stain	265	culture
034	lower	103	left	378	hx

APPENDIX 5: A PROPOSAL TO ENHANCE THE ELECTRONIC TEXTBOOK

A Proposal to Enhance the Electronic Textbook:

Making Access to Medical Knowledge Robust and Natural

24 November 19S5

Submitted to

Paul Mongerson and the CAMDAT Foundation, Inc.

790 Faxmington Avenue

Fannington, Connecticut 06032

From Carnegie-Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

Principal Investigator:

David A. Evans, Ph.D.
Department of Philosophy
Carnegie-Mellon University

Proposal to Enhance the Electronic Textbook:
Making Access to Medical Knowledge Robust and Natural

David A. Evans, Ph.D.
Carnegie-Mellon University

Abstract

We propose to enhance the Electronic Textbook by adding a user-interface that would enable physicians to access diagnostic information in a variety of modes via natural-language expressions. In particular, we plan to develop (1) a natural-language parser that would translate English-language expressions of medical findings and diagnoses into their system-canonical representations and (2) an "expert-reference" module that would engage the user in a brief dialogue to clarify a user's intentions whenever a chosen natural-language expression was ambiguous, vague, or incomplete. Interactions with the system will proceed via a combination of dialogue-like exchanges of information and pop-up menus of choices. In all cases, the user will be expected either to respond with an English phrase or to move a cursor to an indicated choice; all statements of choice will also be in natural English. Our design will be modular to insure that future refinements can be easily added to the system.

1. Background

The Electronic Textbook (ET), developed by Jack D. Myers, Randolph A. Miller, and colleagues, represents a practical extension to the INTERNIST-I expert medical diagnostic system, as it affords users direct access to the INTERNIST-I knowledge base on a personal computer. The ET provides physicians with the comprehensive and detailed information required to enable them to test their hypotheses and confirm their understanding of the features of a case as they formulate differential diagnoses. In this sense, then, the ET can be regarded as an electronic encyclopedia focused on the specific problem of diagnosis.

The ET provides information in several ways, including diagnostic profiles, lists of diagnoses that contain specific findings, and relationships among diagnoses that can give rise to specific findings. Currently, approximately 75% of the diagnoses projected for the INTERNIST-I knowledge base axe available in the ET (amounting to over 550 individual diagnoses); and over 4000 findings are recognized as elements in a differential diagnosis. The present system can return for any of its diagnoses a comprehensive list of the findings that have reliably been reported to be associated with that diagnosis along with a measure (weighting) of their importance in the diagnosis. For any *n positive* findings and any *m negative* findings *{n,m* < 61), the present system can also return a list of diagnoses or patterns of related diagnoses consistent with those findings. Thus, the system is already capable of providing a physician with information to supplement his medical knowledge or to remind him of aspects of a diagnosis or pattern of findings that he might have overlooked.

Users of the ET currently gain access to its knowledge bases only via a fixed hierarchy of choices, usually presented to the user in the form of a menu or as a one-stage instruction to the user to enter a key term that corresponds to the cannonical form of the manifestation or diagnosis name that the system is prepared to accept. This would be a difficult word-entry task were it not for the ability of the system to accept abbreviated input (searching for correspondences in initial strings of characters of words, rather than for whole words) and word-order-free concatenations of terms. (See the appendix for examples of ET menus and diagnostic profiles, including manifestation names.) However, the current system, while generally well-designed to guide the user towards the kind of response (information) that the system is prepared to offer, cannot accept user-generated, English-language descriptions of findings or diagnoses. Providing the ET with such an ability would represent a major step in the development of a useful medical resource, since an essential desideratum for the widespread acceptance of the ET is ease and naturalness of interaction.

2. The Natural-Language Problem

There are several characteristics of natural-language usage that will have to be accommodated if we are to make the ET a robust and efficient resource for physicians. We consider each category of nat:iral-language problem in turn, below.

(1) Mapping NL expressions into fixed expressions

The names of diagnoses and findings in the INTERNIST system and in the ET database are typically given not in single words but in phrases of several words. For the most part, the phrases are composed from a vocabulary that is completely familiar to the physician, but they are constructed according to the needs of clarity in classification, and not in the syntax of natural English. Consider the following examples (taken from the list of possible findings):

ABDOMEN COMPUTERIZED TOMOGRAPHY AORTA CURVILINEAR CALCIFICATION

ABDOMEN COMPUTERIZED TOMOGRAPHY AORTA SACCULAR ANEURYSM

ABDOMEN COMPUTERIZED TOMOGRAPHY AORTA TORTUOSITY AND/OR DILATATION

ABDOMEN COMPUTERIZED TOMOGRAPHY GAS IN BILIARY TRACT

ABDOMEN COMPUTERIZED TOMOGRAPHY GAS IN SUPERIOR MESENTERIC AND/OR PORTAL VEIN <S>

ABDOMEN COMPUTERIZED TOMOGRAPHY PARA-AORTIC AND/OR RETROPERTTONEAL LYMPH NODE <S> ENLARGED

ABDOMEN COMPUTERIZED TOMOGRAPHY RIGHT LOWER QUADRANT EXTRAPELVIC LOW DENSITY MASS <ES>

ABDOMEN COMPUTERIZED TOMOGRAPHY SMALL INTESTINE INTRAMURAL GAS

ABDOMEN COMPUTERIZED TOMOGRAPHY SUBDIAPHRAGMATIC LOW DENSITY MASS <ES>

ABDOMEN PAIN EPIGASTRIUM RECURRENT ATTACK <S> HX

ABDOMEN PAIN EPIGASTRIUM RELIEVED BY ANTACID

ABDOMEN PAIN EPIGASTRIUM RELIEVED BY FOOD

ABDOMEN PAIN EPIGASTRIUM SEASONAL HX

ABDOMEN PAIN EPIGASTRIUM UNRELIEVED BY ANTACID

ABDOMEN PAIN EXACERBATION WITH ALCOHOL

ABDOMEN PAIN EXACERBATION WITH BREATHING

ABDOMEN PAIN EXACERBATION WITH COUGH

ABDOMEN PAIN EXACERBATION WITH EXERCISE

ABDOMEN PAIN EXACERBATION WITH FOOD

ABDOMEN PAIN EXACERBATION WITH MOTION
ABDOMEN PAIN GENERALIZED

It is easy to discern the intended concept behind most of these expressions, but it is clear that no one would volunteer any of them as the description of a clinical finding in natural discourse. A list of the "non-natural" features of these expressions would include the following:

- **The expressions mix 'real' language with abbreviations**

  "HX" for "history"

- **The expressions don't tolerate morphology (i.e., inflectional and other grammatical variations in the form of words)**

  "MASS <ES>" **or** "ATTACK <S>" **for "mass," "masses," "attack," and "attacks"**

- **The expressions don't have natural word-order**

  "ABDOMEN PAIN EPIGASTRIUM SEASONAL HX" **for "a history of seasonal epigastric abdominal pain"**

- **The expressions are 'rigid' - not allowing word/phrase variants**

  "COMPUTERIZED TOMOGRAPHY" **is used;** "CT" **or** "C.T." **can't be used**

- **The expressions collapse phrases that would naturally have prepositional phrases (PPs) and verb phrases (VPs) intermixed**

  "ABDOMEN COMPUTERIZED TOMOGRAPHY SMALL INTESTINE INTRAMURAL GAS" **represents what might be given in a much longer phrase or sequences of phrases: "CT scan of the abdomen showed intramural gas in the small intestine"**

- **""The expressions show arbitrary (unpredictable) variations**

  •^VASCULAR DISEASE OF SMALL INTESTINE" **is the label for a class of disorders, while** "SMALL INTESTINAL NEOPLASM" **is the label of a sister class. Why** "OF SMALL INTESTINE" **in one case and** "SMALL INTESTINAL" **in the other? Both should be acceptable in either case.**

- **The expressions are insensitive to synonymous phrasing**

  "ABDOMEN PAIN EXACERBATION WITH FOOD" **should be the equivalent of "abdominal pain worse after eating"**

This list could continue, but it suffices to present a representative sample of problems that are amenable to linguistic analysis and are redressable through parsing techniques that we have developed. This whole problem, generically, is the many-to-one mapping problem.

(2) Mapping specific information into appropriate generic category

It is most natural to report findings in full detail. If a physician receives a laboratory report on white blood cell count, he records the precise count in his records, and not a 'grading' of the count or a qualitative evaluation (such as "high" or "low"). The ET, however, has many graded findings, often expressed as intervals or ranges. Thus, for example, all white blood cell counts made on joint fluid must be assigned to one of three findings: "JOINT FLUID WBC 3000 TO 20000" or "JOINT FLUID WBC GTR THAN 20000" or "JOINT FLUID WBC LESS THAN 3000". Currently, a user who wants to report a finding of joint fluid wbc at 12000 must select the appropriate canonical finding from this set of choices. In a more robust system, that assignment would be made automatically.

(3) Mapping negative/positive/qualitative expressions into appropriate patterns

The current system does not allow users to enter significant "normal" findings, where what counts as normal corresponds to the absence of one or more positive findings. This can be effected, occasionally, by asserting an appropriate canonical finding. But it would be more effective if the system had information about accepted 'normals' that it could appeal to in context. This point supplements the one made in (2), above. Frequently, findings incorporate judgments that 'wash-out' quantitative details; and that is especially important where discrimination between normal and abnormal conditions must be made. Consider just the following instances of findings that employ such measures:

EKG Q WAVE <S> ABNORMAL

FINGER TO NOSE TEST ABNORMAL

HEART XRAY CARDHC SILHOUETTE ABNORMAL LOCALIZED BULGE

HEEL TO KNEE TEST ABNORMAL

LIVER BIOPSY SINUSOID <S> INFILTRATED WITH ABNORMAL UNIFORM MONONUCLEAR CELL <S>

SKULL XRAY MASTOID <S> ABNORMAL

SPINE COMPUTERIZED TOMOGRAPHY ABNORMAL CALCIFICATION WITHIN SPINAL CANAL

THYROID PERCHLORATE DISCHARGE TEST ABNORMAL

WBC ABNORMAL MONONUCLEAR CELL <S> WITH VILLOUS BORDER <S>

XYLOSE ABSORPTION TEST ABNORMAL

ALKALINE PHOSPHATASE BLOOD GTR THAN 2 TIMES NORMAL

ALKALINE PHOSPHATASE BLOOD INCREASED NOT OVER 2 TIMES NORMAL

AMYLASE BLOOD GTR THAN 2 TIMES NORMAL

ANTITHROMBIN HI LESS THAN 50 PERCENT OF NORMAL

CORTICOSTEROID <S> 17 OH URINE GTR THAN 4 TIMES NORMAL

CORTICOSTEROID <S> 17 OH URINE INCREASED NOT OVER 4 TIMES NORMAL

LEUKOCYTE ALPHA GALACTOSIDASE A LESS THAN 30 PERCENT OF NORMAL

LEUKOCYTE BETA GLUCOSIDASE ACTIVITY LESS THAN 30 PERCENT OF NORMAL

LEUKOCYTE GLUCOCEREBROSIDASE ACTIVITY LESS THAN 30 PERCENT OF NORMAL

LUNG <S> PERFUSION SCAN LOCALIZED DEFECT... WITH NORMAL LUNG <S> BY CHEST XRAY

PRESSURE HEPATIC VEIN WEDGE NORMAL

RENIN BLOOD 2 TO 10 TIMES NORMAL

RENIN BLOOD GTR THAN 10 TIMES NORMAL

SCHILLING TEST WITH INTRINSIC FACTOR B12 ABSORPTION NORMAL

STEROID <S> 17 KETO URINE GTR THAN 4 TIMES NORMAL

STEROID <S> 17 KETO URINE INCREASED NOT OVER 4 TIMES NORMAL

Qualitative evaluations do not stop with "normal" and "abnormal". There are many 'hidden' scales among the findings and many terms that suggest evaluations of those scales:

SIGHING FREQUENT

YAWNING FREQUENT

DEXAMETHASONE TEST LOW DOSE SUPPRESSION, etc.

It would be desirable for the system to be able to use information such as "normal blood pressure," reported of a 20-year-old woman, to determine an appropriate range and to convert that range into a canonical finding, if possible.

(4) Interaction in the face of uncertainty and inadequacy

The current set of findings naturally reflects the judgments and discriminations of the creators of INTERNIST-I. In particular, the findings reflect the sensibilities of Dr. Jack Myers, the principal expert diagnostician associated with the INTERNIST project. This is both an asset and a liability in making the ET widely useful. It is an asset in the sense that the findings are precise, consistent, and "fine-grained" enough to capture most of the manifestations that one might expect to encounter in the examination of any case. But it is a liability when

the categories of discrimination used by the physician-user do not mesh with those of the system. For example, a practicing physician might describe a patient as having a "gallop," and consider the observation to be adequate. The ET, however, recognizes several types of gallop rhythm, depending on whether the 'gallop' is right, left, ventricular, or atrial, etc. The exact set of discriminations that occur in the ET is the following:

HEART SOUND <S> S3 LEFT VENTRICULAR GALLOP
HEART SOUND <S> S3 RIGHT VENTRICULAR GALLOP
HEART SOUND <S> S4 LEFT ATRIAL GALLOP
HEART SOUND <S> S4 RIGHT ATRIAL GALLOP
HEART SOUND <S> SUMMATION GALLOP LEFT VENTRICULAR
HEART SOUND <S> SUMMATION GALLOP RIGHT VENTRICULAR

In the face of the incomplete report "gallop rhythm," the system should initiate a procedure to refine the input (probably best handled by a limited menu of choices), directing the user's attention to the appropriate set of contrasts.

Similarly, when the physician reports a finding that is unrepresented among the current set, the system should inform the physician (and ignore the finding) or attempt to offer the physician a choice of similar findings from among a representative set.

3. Proposal

We propose to augment the current version of the ET with a natural-language processing capacity by July, 1986. Our efforts in this first-phase project will be directed at increasing the ease of user-system interaction and the ability of the system to accept natural language input. We do not plan to abandon the basic mechanics of the current version of the ET; and we certainly propose to retain the ability to accept fragmentary, partial words as input. But we believe both that users should have the freedom to enter expressions in what they consider to be acceptable natural language phrasing (even if the words themselves are only partially complete) and that there should be greater feedback from the system when the user's choices are problematic or ambiguous. This last point—providing the user with information about the internal semantics of the system—is essential to rapid learning and user confidence.

In brief, we propose to redesign the front-end interaction shell of the ET to allow for more "contexts" of interaction than the current system, and in particular, to facilitate the entry of

'whole' cases in a single sequence of English phrases. To achieve this goal, we plan to write a special-purpose, natural-language, bottom-up parser to 'pre-process' the input expressions into a form that the ET will accept, prior to passing the refined input on to a modified version of an existing natural-language, top-down parser, to map the input expressions into target canonical phrases. We also plan to begin development of a modest "expert-semantics" module that will identify problematic expressions and elicit clarification from the user. We describe each of the principal components of our proposal in greater detail, below.

(1) Development of a (minimal) front-end shell

The optimal input from the user will be a list of one or more English phrases, where each phrase corresponds to the expression of a single finding or diagnosis. Initially, the user will be asked what his goals are (i.e., whether he wants to review a diagnosis profile, list findings in differential diagnosis, etc.), then he will be asked to offer phrasal descriptions. We want the system to be robust enough, eventually, to eschew misspellings and maybe provide automatic 'completion' of input. At this stage of development, we will geax the system to accept word-initial segments (with maximum length truncated at four or five characters) in linguistically-meaningful combinations. There will be no automatic completion, but after the user has entered a list of expressions, the system will process the list serially and either accept an expression or question it. We do not want to engage in actual dialogue with the user at either of these junctures, but do wish to provide the user with information about the system's interpretation of the user's intentions. When the system has mapped a user-offered expression successfully, it will show the user the proposed mapping, for example, by returning the message:

"Interpreting <<User expression>> as <<canonical expression>>"

Here, the user will have the opportunity to accept the interpretation (by hitting a carriage-return) or to veto it, thereby initiating a clarification of interpretation. Where it is necessary to refine input, either as the result of a user veto or as the result of a failure to find a successful mapping, the system should attempt to direct the user to acceptable alternative choices, perhaps by keying on relevant, recognized concepts in the defective input. Again, this will not involve actual dialogue with the user, but will proceed via menus as in the current version of ET.

The principal role of the shell is to direct the user toward his stated goal while informing him of his many options at each choice point. We will need a path-checking facility to keep

track of where the user is and where he is going.

(2) Two-pass parsing of natural language phrases

The actual parsing of the user-generated English expressions will take place in two stages. The first stage will involve a bottom-up identification of lexical items and the second stage will involve a top-down identification of the linguistic relations among the lexical items. Put slightly differently, the first-stage processing will involve a string parser that will map words or word-fragments into their appropriate lexical categories; and the second-stage processing will involve a frame parser that will map concatenations of category-labeled words into their appropriate canonical expressions.

In order to give the ET such a language-processing capability, it will be necessary (a) to augment the current word-list of acceptable terms (taken from the canonical expressions) with an expanded lexicon of synonyms and 'equivalent' expressions (e.g., "after eating" = "with food"); (b) to build a bottom-up parser to identify lexical items (perhaps by looking at lists of equivalent expressions); and (c) to modify an existing top-down, case-frame parser (DYPAR-4) to identify equivalence of whole expressions.

In this first-stage effort to enhance the ET, our work on the lexicon of synonyms will remain relatively limited. Eventually the lexicon will be made sensitive to individual user-preferences, but initially, it will be a straightforward expansion of synonyms based on the terms that occur in the canonical expressions. Furthermore, the lexical items will be labeled for categories that reflect their *conceptual* roles in the canonical expressions, not necessarily their syntactic categories. Thus, we can expect to use labels such as "location-modifier" or "quantity" instead of the more traditional, semantically-neutral linguistic categories "adjective" or "noun".

The important observation about the parsing task more generally is that the semantic domain of the ET is highly controlled and rather limited. Every canonical manifestation name, for example, has a central manifestation-type concept - such as 'laboratory-finding', 'technique', 'physical-observation', etc. - preceeded or succeeded by a small range of possible concept-modifiers. This sort of regularity in the semantics allows us to write a fairly simple 'semantic-grammar' for the canonical expressions. Given the top-down, case-frame parser we plan to use, it will be relatively easy to build 'frames' for most of the canonical expressions.

(3) Development of procedures for resolving imprecision

Our plans for developing an "expert-semantics[71] module must be modest in this stage of development.

- **For incomplete expressions - the system must respond with a menu of choices to complete the input**
- **For problematic expressions (e.g., qualitative measures, etc.) - the system must initiate procedures to resolve the 'quality' judgment quantitatively, or must offer 'canned' explanations or requests for rephrasing. In time, this could involve a question-answer dialogue with the user.**

4. Personnel

The principal investigator, David A. Evans, Assistant Professor of Linguistics, Carnegie-Mellon University, will direct the research and development of the parser. He will consult and collaborate with Randolph A. Miller, Associate Professor of Medicine, the University of Pittsburgh, one of the developers of the Electronic Textbook, on all aspects of the project, but especially in identifying the semantics underlying the findings terminology and in matters related to the design of the user/system interface. Both Evans and Miller will donate approximately 10 percent of their academic-year time to the project, though Evans will work full-time on the project in June 1986, and will be paid for that time. Jaime G. Carbonell, Associate Professor of Computer Science at Carnegie-Mellon University, and Fred E. Masarie, Research Assistant Professor of Medicine at the University of Pittsburgh, will act as unpaid consultants to the project, each donating (10%) of his time. The project will employ a research assistant (full-time), who will be responsible for producing the lexicon and grammar for the canonical expressions, and a research programmer (part-time), who will be responsible for modifying DYPAR-4 and developing a user-interaction shell. 'Brief biographical sketches of the named participants are given below.

Evans

Evans graduated from Stanford University with a Ph.D. in Linguistics in 1982. He is currently Assistant Professor of Linguistics in the Department of Philosophy, Carnegie-Mellon University, and Co-Director of the Laboratory for Computational Linguistics. He has an extensive background in discourse analysis and computational linguistics, in particular, natural language processing. Before coming to CMU in 1983, he was a Research Associate at

the Center for Advanced Study in the Behavioral Sciences (Stanford), working with a group of linguists on the computational implementation of a "situation semantics" as a model of discourse (19S2), and a Post-Doctoral Fellow in Cognitive Science at the University of California, Berkeley (19S2-19S3).

From September, 1983, until September, 1984, Evans was the Principal Investigator on a project on Doctor-Patient Communication; from October, 1984 until November, 19S5, a Co-Principal Investigator on a project for the National Library of Medicine to automate the indexing and retrieval of texts in large-scale databases; and since February, 19S5, a Co-Principal Investigator on a three-year project designed to provide the CADUCEUS system - a successor to INTERNIST-I - with a natural-language explanation facility. He is especially interested in the development of computationally tractable models of discourse and in the production of differential summaries of natural language texts.

Miller

Miller is currently an Associate Professor of Medicine in the Department of Medicine, Division of General Medicine, at the University of Pittsburgh. He received his M.D. from the University of Pittsburgh School of Medicine and completed his residency training in internal medicine at the University Health Science Center of Pittsburgh during 1976-79. He has been an active member on the INTERNIST-I project since 1973. Next to Dr. Jack Myers, Miller has been the other major contributor of medical expertise to the project over the past decade. He is a board certified internist and maintains a clinical ^practice with hospital teaching responsibilities. He was recipient of an NLM New Investigator Award 1981-S3 during which he developed a patient case simulator program based on the INTERNIST-I knowledge base. He recently received a Research Career-Development Awaxd for a five year period during which he will work on the Electronic Textbook.

Carbonell

Carbonell is an Associate Professor of Computer Science, and a member of the Artificial Intelligence group of the Computer Science Department, at Carnegie-Mellon University. He has worked in computational linguistics at BBN, Yale University, and Carnegie-Mellon University over the past dozen years. His work has included developing syntactic and semantic parsing theories (case-frame analysis), constructing practical natural language interfaces (XCALIBUR, PLUME), building general-purpose robust parsing algorithms (DYPAR, MULTIPAR), and implementing natural-language generators.

Masarie

Masarie is currently a Research Associate in the Department of Medicine, Division of General Medicine, at the University of Pittsburgh. He received his M.D. from the University of Oregon Health Sciences Center in 1982. He completed a surgical internship and one year of general surgery residency at the university of Pittsburgh during 1982-84. Since July 1984, he has been an active member of the INTERNIST-I project, involved with the ongoing development of the computerized medical knowledge base. He primary interest revolves around the representation of medical knowledge in a computerized form.

5. Budget

Omitted.

Appendix: Examples of ET Information

A. Introductory Caution

```
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
                CAUTION    CAUTIOH   CAUTIOH    CAUTIOI
   This knowledge base is still under development and is INCOMPLETE.
   This is particularly relevant when generating the differential dx of a
   manifestation. ALWAYS keep the incompleteness of the knowledge base in
   mind! This program is EDUCATIONAL in nature and has not been validated
   for clinical use with patients. The user is responsible for verifying
   any information before applying it in the clinical setting.

   The creators of the program deny any and all responsibility for misuse
   of the program in caring for patients, and cannot be held responsible
   for any harm to patients resulting from the use of this program.
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

B. Top-level menu of choices

```
  1 General help in using this program
> 2 Display a disease profile's findings/Iinks
  3 Display the differential diagnosis of a finding
  4 Show the combined differential dx of 2 findings
  5 Find associations of finding to diagnoses
  6 STOP the program
```

C. The disease hierarchy (top-level)

```
  MORE SELECTABLE IODES BEYOND THIS SCREEN

> 1 Go BACK TO TOP level of hierarchy
  2 Go BACK ONE level of hierarchy
  3   LIVER AND BILIARY SYSTEM DISEASE 1.
  4   DISEASE OF ENDOCRINE SYSTEM 2.
  5   KIDNEY AND URINARY TRACT DISEASE 3.
  6   PULMONARY DISEASE 4.
  7   DISEASE OF GASTROINTESTINAL SYSTEM 5.
  8   CARDIOVASCULAR DISEASE 6.
  9   HEMATOLOGIC DISEASE 7.
```

10 PERITONEAL DISEASE 8.
11 JOINT DISEASE 9.
12 NEUROLOGIC DISEASE 10.
13 METABOLIC DISEASE 11.
14 SYSTEMIC DISEASE 12.
15 RETINOPATHY 13.
16 SXII DISEASE 14.
17 BONE DISEASE 15.
18 GYNECOLOGICAL DISEASE 16.

D. Following the hierarchy dora to a specific diagnosis

MORE SELECTABLE NODES BEYOND THIS SCREEN

1 Go BACX TO TOP level of hierarchy
2 Go BACK OKE level of hierarchy
3 LIVER AND BILIARY SYSTEM DISEASE 1.
4 DISEASE OF ENDOCRINE SYSTEM 2.
5 KIDNEY AND URINARY TRACT DISEASE 3.
6 PULMONARY DISEASE 4.
> 7 DISEASE OF GASTROINTESTINAL SYSTEM 5.
8 CARDIOVASCULAR DISEASE 6.
9 HEMATOLOGIC DISEASE 7.
10 PERITONEAL DISEASE 8.
11 JOINT DISEASE 9.
12 NEUROLOGIC DISEASE 10.
13 METABOLIC DISEASE 11.
14 SYSTEMIC DISEASE 12.
15 RETINOPATHY 13.
16 SKIN DISEASE 14.
17 BONE DISEASE 15.
18 GYNECOLOGICAL DISEASE 16.

Current hierarchy depth is 2

1 Go BACK TO TOP level of hierarchy
2 Go BACK ONE level of hierarchy
3 ESOPHAGEAL DISEASE 5.1.
> 4 GASTRODUODENAL DISEASE 5.2.
5 DISEASE OF SMALL INTESTINE 5.3.
6 DISEASE OF LARGE INTESTINE 5.4.
7 PANCREATIC DISEASE 5.5.

Currant hierarchy depth is 3

1 Go BACK TO TOP level of hierarchy
2 Go BACK QIE level of hierarchy
3 PEPTIC ULCER DISEASE 5.2.1.
4 * FUICTIOIAL DYSPEPSIA 5.2.2.
> 5 GASTRIC IEGPLASM 5.2.3.
6 • GASTRITIS GIAIT HYPERTROPHIC <MEHETRIERS> 5.2.4.
7 * MALLQRY VEISS SYIDROME 5.2.5.
8 * GASTRQDUQDEIAL PERFORATIOI 5.2.6.
9 * PYLORIC OBSTRUCTIOI 5.2.7.
10 * GASTROIITESTIIAL SARCOIDOSIS 5.2.8.

Current hierarchy depth is 4

> 1 Go BACK TO TOP level of hierarchy
2 Go BACK OIE level of hierarchy
3 * GASTRIC CARCIIOMA 5.2.3.1.
4 * GASTRIC LYNPHONA 5.2.3.2.

E. A **disease** profile **(selected** via **the** disease hierarchy)

MORE SELECTABLE IODES BEYOID THIS SCREES

1 Go BACK TO TOP level of hierarchy
2 Go BACK OIE level of hierarchy
3 LIVER AID BILIARY SYSTEM DISEASE 1.
4 DISEASE OF EIDOCRIIE SYSTEM 2.
5 KIDIEY AID URIIARY TRACT DISEASE 3.
6 PULMOIARY DISEASE 4.
7 DISEASE OF GASTROIITESTIIAL SYSTEM 5.
8 CARDIOVASCULAR DISEASE 6.
9 HEMATOLOGIC DISEASE 7.
10 PERITOIEAL DISEASE 8.
11 JOIIT DISEASE 9.
12 IEUROLOGIC DISEASE 10.
> 13 METABOLIC DISEASE 11.
14 SYSTEMIC DISEASE 12.
15 RETIIOPATHY 13.
16 SKII DISEASE 14.
17 BOIE DISEASE 15.
18 GYIECOLOGICAL DISEASE 16.

Current hierarchy depth is 2

1 Go BACK TO TOP level *o*± hierarchy
2 Go BACK OHE l*xel of hierarchy
3 • DIABETES MELLITUS 11.1.
4 * DIABETIC KETOACIDOSIS 11.2.
5 • PORPHYRIA CUTAHEA TARDA 11.3.
> 6 * PORPHYRIA ACUTE IHTERMITTEHT 11.4.

Disease Profile for: Part 1 of 10
PORPHYRIA ACUTE IITERMITTEIT

4 5 PORPHOBILIHOGEI URINE INCREASED
3 3 PORPHYRIA FAMILY HI
2 5 AMIIOLEVULIIIC ACID URIHS GTR THAI 4 MG PER DAY
2 3 UROPORPHYRII URIIE GTR THAI 40 MCG PER DAY
1 4 ABDOMEN PAII ACUTE
1 4 ABDOMEN PAII COLICKY
1 4 ABDONEI PAII SEVERE
1 4 IIDOCYAIIIE GREEN RETENTION INCREASED
1 4 T4 BINDING GLOBULIN INCREASED
1 4 URIIE DARK HI
1 4 UROBILIIOGEI URIIE INCREASED
1 3 ABDONEI PAII GENERALIZED
1 3 ABDOMEN TENDERNESS GENERALIZED
1 3 ABDOMEN XRAY COLON DISTENTION WITH GAS
1 3 ABDOMEI XRAY SMALL IITESTIIE GAS FILLED LOOP <S>
1 3 BOVEL SOUND <S> INCREASED
1 3 CHLORIDE BLOOD DECREASED
1 3 CHOLESTEROL SERUM INCREASED
1 3 COLON BARIUM ENEMA DILATATION

Disease Profile for: Part 2 of 10

PORPHYRIA ACUTE IITERMITTEIT

1 3 DYSARTHRIA OR DYSPHOIIA
1 3 EEG DIFFUSE 101 SPECIFIC CHANGE <S>
1 3 EKG ST SEGMENT DEPRESSIOI WITHOUT RECIPROCAL ELEVATION
1 3 GLUCOSE TOLERANCE DECREASED
1 3 HYPERTENSION ABRUPT ONSET
1 3 MUSCLE <S> TENDERNESS DIFFUSE
1 3 NEUROLOGIC SIGI <S> TRANSIENT

1 3 POTASSIUM BLOOD DECREASED
1 3 PRESSURE ARTERIAL DIASTOLIC 95 TO 125
1 3 SODIUM BLOOD DECREASED
1 2 ABDOMEN DISTENTION
1 2 ABDOMEN RESPIRATORY MOVEMENT DECREASED
1 2 ABDOMEN SURGERY RECENT HX
1 2 ABDOMEN SURGERY REMOTE HX
1 2 ABDOMEN URINARY BLADDER PALPABLE OR PERCUSSABLE
1 2 ABDOMEN XRAY SMALL INTESTINE FLUID LEVEL <S>
1 2 ACALCULIA
1 2 AMNESIA POSTICTAL
1 2 ARM <S> WEAKNESS PROXIMAL AND DISTAL

Disease Profile for: Part 3 of 10

PORPHYRIA ACUTE INTERMITTENT

1 2 ARM <S> WEAKNESS PROXIMAL ONLY
1 2 BACK PAIN MIDLINE SEVERE
1 2 BARBITURATE ADMINISTRATION RECENT HX
1 2 BOWEL SOUND <S> DECREASED
1 2 BOWEL SOUND <S> INCREASED SYNCHRONOUS WITH COLIC
1 2 CONFUSION POSTICTAL
1 2 DEHYDRATION
1 2 DIARRHEA ACUTE
1 2 DIARRHEA INTERMITTENT
1 2 DYSPHAGIA LIQUID <S>
1 2 DYSPHAGIA SOLID <S>
1 2 ESTROGEN ADMINISTRATION RECENT HX
1 2 EXTREMITY <IES> PIGMENTATION INCREASED DISTAL
1 2 EYE <S> BLEPHAROPTOSIS
1 2 EYE <S> CONVERGENCE IMPAIRED
1 2 EYE <S> NYSTAGMUS
1 2 EYE <S> PUPIL <S> UNEQUAL
1 2 EYE <S> VISUAL ACUITY DECREASED
1 2 EYE <S> WEAKNESS OF DOWNWARD AND OUTWARD GAZE

Disease Profile for: Part 4 of 10

PORPHYRIA ACUTE INTERMITTENT

1 2 EYE <S> WEAKNESS OF LATERAL GAZE
1 2 EYE <S> WEAKNESS OF UPWARD AND INWARD GAZE
1 2 FASTING RECENT HX

1 2 FECES INCONTINENCE
1 2 HEADACHE SEVERE
1 2 HEMATOCRIT BLOOD LESS THAN 3S
1 2 HEMOGLOBII BLOOD LESS THAH 12
1 2 LEG <S> WEAKNESS PROXIMAL AND DISTAL
1 2 LEG <S> WEAKNESS PROXIMAL ONLY
1 2 LUNG <S> FORCED VITAL CAPACITY DECREASED
1 2 MAGNESIUM BLOOD DECREASED
1 2 MENSES PRECIPITATION OR EXACERBATION OF DISEASE HX
1 2 MOUTH MUCOSA DRY <XEROSTOMIA>
1 2 MOUTH MUCOSA DRY HX
1 2 MOUTH PALATAL WEAKNESS
1 2 MUSCLE <S> FACIAL WEAKNESS BILATERAL
1 2 MUSCLE <S> FACIAL WEAKNESS UNILATERAL INCLUDING FOREHEAD
1 2 MUSCLE <S> WEAKNESS GENERALIZED MARKED
1 2 MUSCLE <S> WEAKNESS MASSETER <S>

Disease Profile for: Part 5 of 10

PORPHYRIA ACUTE INTERMITTENT

1 2 MUSCLE <S> WEAKNESS PELVIC GIRDLE
1 2 MUSCLE <S> WEAKNESS SHOULDER GIRDLE
1 2 OXYGEN TENSION BLOOD ARTERIAL LESS THAN 80
1 2 PREGNANCY EXACERBATION OF DISEASE HX
1 2 PREGNANCY POSTPARTUM IMPROVEMENT OF DISEASE HX
1 2 PRESSURE ARTERIAL ORTHQSTATIC HYPOTENSION
1 2 PRESSURE ARTERIAL SYSTOLIC LESS THAN 90
1 2 PROTEINURIA
1 2 PSYCHOSIS
1 2 REASONING ABSTRACT ABILITY IMPAIRED
1 2 REFLEX <ES> DEEP TENDON INCREASED GENERALIZED
1 2 REFLEX GAG ABSENT
1 2 RETINA ARTERIOLAR SPASM
1 2 SEIZURE <S> GRAND HAL
1 2 SENSORY LEVEL ABDOMINAL
1 2 SCOT 40 TO 119
1 2 SGPT 40 TO 199
1 2 STOMACH BARIUM MEAL DILATED ATONIC
1 2 SULFONAMIDE ADMINISTRATION RECENT HX

Disease Profile for: Part 6 of 10

PORPHYRIA ACUTE INTERMITTENT

1 2 TRAISAMIIASE <S> BLOOD IICREASED
1 2 URIIE IICQITIIEICE
1 2 VBC LESS THAI 4000
1 1 ABDOMEI PAII CHROMIC
1 1 ABDONEI PAII EPIGASTRIUM
1 1 ABDOMEa PAII EPIGASTRIUM RECURREIT ATTACK <S> HX
1 1 ABDOMEI PAII EPIGASTRIUM UIRELIEVED BY AITACID
1 1 ABDOMEI PAII HYPOGASTRIUM
1 1 ABDOMEI PAII LEFT LOVER QUADRAIT
1 1 ABDOMEI PAII LEFT UPPER QUADRAIT
1 1 ABDOMEI PAII 101 COLICKY
1 1 ABDOMEI PAII PERIUMBILICAL
1 1 ABDOMEI PAII RADIATIIG TO BACK
1 1 ABDOMEI PAII RADIATIIG TO IIGUIIAL AREA <S> OR GEIITALIA
1 1 ABDOMEI PAII RIGHT LOVER QUADRAIT
1 1 ABDOMEI PAII RIGHT UPPER QUADRAIT
1 1 ABDOMEI PAII RIGHT UPPER QUADRAIT RECURREIT ATTACK <S> HX
1 1 ABDOMEI PAII SUPRAPUBIC
1 1 ABDOMEI TEIDERIESS EPIGASTRIUM <IOI HEPATIC>

Disease Profile for: Part 7 of 10

PORPHYRIA ACUTE IITERMITTEIT

1 1 ABDOMEI TEIDERIESS HYPOGASTRIUM
1 1 ABDOMEI TEIDERIESS LEFT LOVER QUADRANT
1 1 ABDOMEI TEIDERIESS LEFT UPPER QUADRAIT
1 1 ABDOMEI TEIDERIESS PERIUMBILICAL
1 1 ABDOMEI TEIDERIESS RIGHT LOVER QUADRAIT
1 1 ABDOMEI TEIDERIESS RIGHT UPPER QUADRAIT <IOI HEPATIO
1 1 ALCOHOL IIGESTIOI HEAVY RECEIT HX
1 1 ALCOHOLISM CHROMIC HX
1 1 ALKALIIE PHOSPHATASE BLOOD IICREASED IOT OVER 2 TIMES IORMAL
1 1 AMYLASE BLOOD GTR THAI 2 TIMES IORMAL
1 1 ARM <S> VEAKIESS UIILATERAL MOIOPLEGIC
1 1 BREATHIIG ACCESSORY MUSCLE <S> USED
1 1 COMA
1 1 DYSPIEA AT REST
1 1 EYE <S> BLIIDIESS
1 1 EYE <S> DIPLOPIA
1 1 EYE <S> OPTIC DISC ATROPHY
1 1 FEVER IITERMITTEIT
1 1 GLUCOSE BLOOD FASTIIG 130 TO 300

Disease Profile lor: Part 8 of 10

PORPHYRIA ACUTE IHTERMITTEHT

1 1 GLYCOSURIA
1 1 LEG <S> VEAKHESS UNILATERAL MOIOPLEGIC
1 1 METHYLDOPA ADMIIISTRATIOI HI
1 1 MUSCLE <S> ATROPHY GEIERALIZED
1 1 MUSCLE <S> CRAMP <S>
1 1 MUSCLE <S> VEAKIESS TRAPEZIUS AND/OR STERHOMASTOID
1 1 PHEIYTOII ADMIIISTRATIOI HX
1 1 PRESSURE ARTERIAL DIASTOLIC LESS THAI 60
1 1 REFLEX BABIISKI SIGI PRESEIT BILATERAL
1 1 REFLEX BABIISKI SIGI PRESEIT UIILATERAL
1 1 REGURGITATIOI OF LIQUID <S> THROUGH IOSE
1 1 SEIZURE <S> FAMILY HX
1 1 SEIZURE <S> GRAID MAL HX
1 1 SEISORY LEVEL CERVICAL
1 1 SEISORY LEVEL THORACIC
1 1 TOIGUE PROTRUSIOI WITH MARKED DEVIATIOI
0 4 ABDOMEI PAII PRESEIT
0 4 ETHHIC BACKGROUID IORTHERI EUROPEAI
0 4 OISET ABRUPT

Disease Profile for: Part 9 of 10

PORPHYRIA ACUTE IITERMITTEIT

0 4 SEX FEMALE
0 3 ABDOMEI TEIDERIESS PRESEIT
0 3 AFFECT APPREHEISIVE
0 3 AGE 26 TO 55
0 3 COISTIPATIOI
0 3 EKG SIIUS TACHYCARDIA
0 3 MYALGIA
0 3 TACHYCARDIA
0 3 VOMITIIG RECEIPT
0 2 AFFECT DEPRESSED
0 2 AGE 16 TO 25
0 2 AGE GTR THAI 55
0 2 AIOREXIA
0 2 COIFUSIOI OR DISORIEITATIOI
0 2 COITRACEPTIVE ORAL ADMIIISTRATIOI RECEIT HX

0 2 DELIRIUM
0 2 FEVER
0 2 IEUROLQGIC EXAM CRANIAL IERVE <S> ABNORMAL
0 2 IEUROLOGIC EXAM OF EXTREMITY <IES> SENSATION ABNORMAL

Disease Profile for: Part 10 of 10

PORPHYRIA ACUTE INTERMITTENT

0 2 IEUROLOGIC EXAM OF TRUNK ABIORMAL
0 2 PRESSURE ARTERIAL SYSTOLIC 90 TO 110
0 2 SEX MALE
0 2 SKII SVEATIIG INCREASED GENERALIZED
0 2 VBC 14000 TO 30000
0 2 VBC 4000 TO 13900 PERCENT IEUTROPHIL <S> INCREASED
0 2 WEIGHT LOSS GTR THAI 10 PERCENT
0 1 EMACIATION
0 1 EXTREMITY <IES> EXAM MOTOR OR MUSCLE ABIORMALITY
0 1 IEUROLOGIC EXAM OF EXTREMITY <IES> REFLEX <ES> ABIORMAL
0 1 RESPIRATORY INFECTION UPPER RECENT HX
0 1 STUPOR LETHARGY OR SOMNOLENCE

Last part, options: H for Help; SPACE; B; or N

F. Misc. examples underscoring need for IL grammar /techniques

6 PLASMA CELL DYSCRASIA AND/OR DYSPROTEINEMIA 7.4.
> 7 LYMPHOPROLIFERATIVE DISEASE 7.5.
8 DISEASE INVOLVING PLATELETS 7.6.

Current hierarchy depth is 3
> 1 Go BACK TO TOP level of hierarchy
2 Go BACK ONE level of hierarchy
3 * LEUKEMIA ACUTE LYMPHOBLASTIC 7.5.1.
4 * LEUKEMIA CHRONIC LYMPHOCYTIC 7.5.2.
5 * HODGKINS DISEASE SYSTEMIC. 7.5.3.
6 * MALIGNANT LYMPHOMA 101 HODGKINS TYPE 7.5.4.
7 * LEUKEMIA HAIRY CELL 7.5.5.
8 * AIGIOIMMUIOBLASTIC LYMPHADEIOPATHY 7.5.6.

6 PLASMA CELL DYSCRASIA AND/OR DYSPROTEINEMIA 7.4.
7 LYMPHOPROLIFERATIVE DISEASE 7.5.

```
> 8   DISEASE INVOLVING PLATELETS 7.6.

        Current hierarchy depth is 3
  1 Go BACK TO TOP level of hierarchy
  2 Go BACK ONE level of hierarchy
> 3   THROMBOCYTOPENIC PURPURA 7.6.1.
  4   * THROMBOCYTHEMIA IDIOPATHIC 7.6.2.

        Current hierarchy depth is 4
  1 Go BACK TO TOP level of hierarchy
  2 Go BACK ONE level of hierarchy
> 3   THROMBOCYTOPENIC PURPURA DUE TO INCREASED PLATELET UTILIZATION

        Current hierarchy depth is 5
> 1 Go BACK TO TOP level of hierarchy
  2 Go BACK ONE level of hierarchy
  3   * THROMBOCYTOPENIC PURPURA ACUTE 7.6.1.1.1.
  4   * THROMBOCYTOPENIC PURPURA CHRONIC 7.6.1.1.2.
  5   * THROMBOTIC THROMBOCYTOPENIC PURPURA 7.6.1.1.3.

  6   PULMONARY DISEASE 4.
> 7   DISEASE OF GASTROINTESTINAL SYSTEM 5.
  8   CARDIOVASCULAR DISEASE 6.

        Current hierarchy depth is 2
  1 Go BACK TO TOP level of hierarchy
  2 Go BACK ONE level of hierarchy
  3   ESOPHAGEAL DISEASE 5.1.
  4   GASTRODUODENAL DISEASE 5.2.
> 5   DISEASE OF SMALL INTESTINE 5.3.
  6   DISEASE OF LARGE INTESTINE 5.4.
  7   PANCREATIC DISEASE 5.5.

        Current hierarchy depth is 3
> 1 Go BACK TO TOP level of hierarchy
  2 Go BACK ONE level of hierarchy
  3   VASCULAR DISEASE OF SMALL INTESTINE 5.3.1.
  4   * CROHNS DISEASE OF SMALL INTESTINE 5.3.2.
  5   SMALL INTESTINAL NEOPLASM 5.3.3.
  6   * SMALL BOWEL OBSTRUCTION 5.3.4.
  7   * MALABSORPTION 5.3.5.
  8   INTESTINAL INFESTATION 5.3.6.
  9   * PANCREATIC CHOLERA 5.3.7.
  10  * EOSINOPHILIC GASTROENTERITIS <MUCOSAL FORM> 5.3.8.
```

11 * EOSIIOPHILIC GASTROEITERITIS <MUSCULAR F0RM> 5.3.9.
12 * EOSIIOPHILIC GASTROEITERITIS <SEROSAL F0RK> 5.3.10.
13 * LACTOSE IITOLERAICE 5.3.11.
14 * GASTROIITESTIIAL AMYLOIDOSIS 5.3.12.
15 * CELIAC SPRUE 5.3.13.
16 GASTROEITERITIS ACUTE 5.3.14.
17 * PROTEII-LQSIIG EITEROPATHY 5.3.15.