AN XML MEDICAL KNOWLEDGE LEXICON

Volume 1
A – I

NEW YORK
2005
AN
XML
MEDICAL
KNOWLEDGE
LEXICON

句
Volume 1
A – I

NEW YORK
2005
ACKNOWLEDGEMENTS

Many colleagues have contributed to the development of the XML Medical Knowledge Lexicon and the Delphi Knowledge dBMS, Delphi and PHP Viewers that use it. Naomi Sager has been central in coordinating this work with the English Medical Language Lexicon for Natural Language Processing. We wish to thank Ronald Tarrant, Nancy Wheeler and Jorge Roccatagliata for their significant contributions. We also wish to thank Anita Parmalee, Clara Hager, and Richard Wheeler’s family: Michele, Nathan, Ben and Shauna for their support and generosity.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to the Structured Health Markup Language (SHML)</td>
<td>vii – ix</td>
</tr>
<tr>
<td><strong>David Rothwell, MD</strong></td>
<td></td>
</tr>
<tr>
<td>Table 1: SHML Medical Tag Hierarchy</td>
<td>xiv-xxxii</td>
</tr>
<tr>
<td>Table 2: EML Parts of Speech—Major lexical categories</td>
<td>xxxiii</td>
</tr>
<tr>
<td>Table 3: EML List of Subordinate Conjunction Strings</td>
<td>xxxiv</td>
</tr>
<tr>
<td>Table 4: Syntactic Medical Sublanguage Classes</td>
<td>xxxv – xxxvi</td>
</tr>
<tr>
<td>Table 5: Correspondence between MLP and XML lexical attributes</td>
<td>xxxvii – xxxviii</td>
</tr>
<tr>
<td><strong>An XML Medical Knowledge Lexicon</strong></td>
<td></td>
</tr>
<tr>
<td>Notes on Use of this Lexicon</td>
<td>1</td>
</tr>
<tr>
<td>The Main section (67,642 entries including literals and coded name</td>
<td>2 – 1086</td>
</tr>
<tr>
<td>entries)</td>
<td></td>
</tr>
<tr>
<td>Credit</td>
<td>1087</td>
</tr>
</tbody>
</table>
INTRODUCTION
July 2005

The Structured Health Markup Language (SHML) consists of a set of tags and accompanying lexicon, constructed within the eXtensible Markup Language (XML) formalism, designed to capture the medical, administrative and biopsychosocial elements of a patient encounter. A markup language consists primarily of a set of labels (tags) developed within XML rules.

SHML consists of a collection of tags that capture and describe the entire content of a medical document across each medical domain. Tags are applied to each term found in a document. SHML tags describe both the traditional biological elements of a medical encounter as well as the psychosocial aspects, conformant with the biopsychosocial model of care. In addition, tags for many of the administrative elements of care are also included.

**SHML Tag Types**

<table>
<thead>
<tr>
<th>BIOLOGICAL</th>
<th>BIOPSYCHOSOCIAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anatomic structure</td>
<td>Activities (sports,...)</td>
</tr>
<tr>
<td>Body region</td>
<td>Medications: (Multum), med-class</td>
</tr>
<tr>
<td>Sign-symptom</td>
<td>Chemicals</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Time: freq, repetition, exact, begin, end</td>
</tr>
<tr>
<td>Dx-process</td>
<td>Links</td>
</tr>
<tr>
<td>Dx group by system</td>
<td>Modifiers: modal, negation, changes, amount, desc, s-q</td>
</tr>
<tr>
<td>Procedures</td>
<td>Person: kin, civil</td>
</tr>
<tr>
<td>Organisms</td>
<td>Health status (adl...)</td>
</tr>
<tr>
<td>Allergies</td>
<td>Demographic</td>
</tr>
<tr>
<td>Pt. social behavior</td>
<td>Socio-cultural</td>
</tr>
<tr>
<td>Health status (adl...)</td>
<td>Patient direction</td>
</tr>
<tr>
<td></td>
<td>Patient preference</td>
</tr>
<tr>
<td></td>
<td>Patient understanding</td>
</tr>
<tr>
<td></td>
<td>Relationships</td>
</tr>
<tr>
<td></td>
<td>Beliefs</td>
</tr>
<tr>
<td></td>
<td>Values</td>
</tr>
<tr>
<td></td>
<td>Living situation</td>
</tr>
</tbody>
</table>

The lexicon of SHML-tagged terms, the XML Medical Knowledge Lexicon, works in conjunction with the English Medical Lexicon for Natural Language Processing, the lexicon developed for natural language processing of clinical documents, used by the Medical Language Processor (MLP). The two lexicons are in concordance with one another, comprising, in effects, a single combined lexicon. This effort is a maturation and marriage of two developments, both aimed at improving accessibility to relevant patient data found in text. See Figure 1.
MLP in conjunction with SHML-tagging functions to transform the content of clinical documents into individual clinical facts, which are referred to as Health Information Units (HIU’s). SHML tagging assists in identifying and structuring the medically relevant content of documents. MLP provides linguistic and broad medical characterization of each term while SHML tagging provides more precise medical characterization of these terms.

A challenge and primary task in developing an Electronic Medical Record (EMR) is to provide both immediate and long range access to information in the clinical record. Since significant parts of current medical records consist of transcribed or written notes, access to this information demands Natural Language Processing (NLP) techniques, (and for our purposes MLP), to isolate and retrieve (i.e. unravel) the informational units from that text. In short, to transform all information captured, whether dictated or written, into retrievable clinical facts.

The goals of this MLP/SHML effort are to retrieve clinical information for display from all previous encounters in a succinct, user defined manner; to make available that information for subsequent data analysis; and to support clinical prompts-and-alerts software. With the use of Viewer software, displays of the HIU’s are integrated with and conformant to HL-7’s Clinical Document Architecture (CDA).

**Mission of the SHML**

Notes: SID (sentence identification number), SHML (Structured Health Markup Language), HIU (Health Information Unit), Viewer (an XML display of MLP processed documents with access to text via SHML tags).
- Define a granular representation of terms and phrases that within a given language (domain) unambiguously define clinical concepts
- Provide for an adequate representation of these terms and concepts in a simple and easily understood architecture
- Provide for discrete mapping to any other “nomenclature” and/or “code set”
- Utilize easily available, inexpensive and widely supported tools for authoring, maintenance and use
- Provide this as a non-proprietary standard under the auspices of a private not-for-profit entity

SHML tags are specifically designed to characterize and label medical knowledge. Tags act as an initial sort and retrieval device for the EMR; they get at the core of what was stated. They provide uniformity to data elements. SHML tagging, when used to its full potential, defines what is worth accessing, viewing, and counting in a medical document. Tags are views of data, not attributes or properties of the data. They are intended to be inclusive of all parts of the record. Tagging provides access to raw (original) data; they are not interpretive e.g. if BUN is elevated, it captures only this data element, not the possibility of “renal failure”. SHML tagging does not require clinical language to be forced into categories of a predetermined data model, i.e. Structured Data Entry menus. Tag classes are illustrated and brief examples are shown below:

| Anatomy         | <a-s>            |
| Body region     | <b-r>            |
| Organisms       | <or>             |
| Chemicals       | <chem>           |
| Meds            | <med>            |
| Diagnoses       | <dx>             |
| Procedures      | <pr>             |

### SHML Tag System

<table>
<thead>
<tr>
<th>Description</th>
<th>Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis</td>
<td>dx</td>
</tr>
<tr>
<td>Diagnostic process</td>
<td>dx-precss</td>
</tr>
<tr>
<td>Infectious diagnostic process</td>
<td>dx-precss_infect</td>
</tr>
<tr>
<td>Immunologic diagnostic process</td>
<td>dx-precss_imm</td>
</tr>
<tr>
<td>Neoplastic diagnostic process</td>
<td>dx-precss_neopl</td>
</tr>
<tr>
<td>Diagnosis group</td>
<td>dx-kind</td>
</tr>
<tr>
<td>Neurologic disease</td>
<td>dx-kind_neuro</td>
</tr>
<tr>
<td>Migraine</td>
<td>dx-kind_neuro_migr</td>
</tr>
<tr>
<td>Reactive Airway Disease</td>
<td>dx-kind_d-k-resp_r-a-d</td>
</tr>
<tr>
<td>Asthma</td>
<td>dx-kind_d-k-resp_r-a-d</td>
</tr>
</tbody>
</table>
**Congestive cardiomyopathy**

<dx>
  <dx-kind_cardiov_cardmy>
  <a-s_cv_hrt_myc>
  <b-r_tk_thx>
    Congestive cardiomyopathy
  </b-r_tk_thx>
  </a-s_cv_hrt_myc>
</dx-kind_cardiov_cardmy>

**Asthma**

<dx>
  <dx-prcss_imm_all><dx-kind_d-kresp_r-a-d>
  <a-s_resp_l-r><b-r_m-r>
    Asthma
  </b-r_m-r></a-s_resp_l-r>
  </dx-kind_d-kresp_r-a-d></dx-prcss_imm_all>
</dx>

**Pneumonia**

<dx>
  <dx-prcss_infect><dx-kind_d-kresp_pneum>
  <or_mc>
  <a-s_resp_l-r_lng><b-r_tk_thx_int-thor>
    Pneumonia
  </b-r_tk_thx_int-thor></a-s_resp_l-r_lng>
  </or_mc>
  </dx-kind_d-kresp_pneum></dx-prcss_infect>
</dx>

**Pneumonia, right lower lobe**

<dx>
  <dx-prcss_infect><dx-kind_d-kresp_pneum>
  <or_mc>
  <a-s_resp_l-r_lng><b-r_tk_thx_int-thor>
    Pneumonia
    <a-s_rsp_l-r_lng>
    <b-r_tk_thx_int-thor>
      right lower lobe
      </b-r_tk_thx_int-thor>
    </a-s_rsp_l-r_lng>
  </or_mc>
  </dx-kind_d-kresp_pneum></dx-prcss_infect>
</dx>
Pneumonia, right lower lobe, superior, due to Klebsiella.

<dx><dx-prcss_infect><dx-kind_d-k.resp_pneum><or_MC><a-s_resp_l_r_lng><b-r_tk_thx_int-thor>
Pneumonia,
<a-s_rsp_l_r_lng><b-r_tk_thx_int-thor>
right lower lobe
</b-r_tk_thx_int-thor><a-s_resp_l-r_lng>
,
<p-o>
superior
</p-o>
,
<li>
<li_CNN>
due to
</li_CNN>
</li>
<OR>
<or_MC_bct_gm-neg>
Klebsiella
</or_MC_bct_gm-neg>
</OR>
</b-r_tk_thx_int-thor><a-s_resp_l-r_lng>
</or_MC>
</dx-kind_d-k.resp_pneum></dx-prcss_infect>
</dx>

PRESENTATION FORMAT (one of many)

Diagnosis: Pneumonia
Location: RLL, superior
Organism: Klebsiella

The combined lexicon used to underlie and drive the MLP/SHML system is sorted by class and terms are labeled for their linguistic and clinical properties. Classes are shown in the accompanying tables. The lexicon includes all terms encountered in a document, nouns, verbs, and all classes of modifiers. Traditional terminologies are known to contain only a small percentage of the terms encountered in medical text; their focus has been on nouns. The MLP/SHML lexicon addresses and classes/tags each term for their linguistic and medical content.

Of great importance are the terms expressing uncertainty, negation and time. Each of these classes consists of several hundred terms. When present, they modify the data element to which they refer and are included in the appropriate HIU.
Ambiguity is a particularly difficult issue in medicine. Ambiguous language can be of several types. The first is intentional ambiguity that is to be intentionally uncertain. The second is non-intentional (can’t use the language properly). The third is within the language itself, e.g. homonyms (foot, depression). It has been estimated that up to 60% of statements found in medical records include terms expressing ambiguity. All must be recognized and accounted for to achieve an accurate rendition of a record. MLP/SHML does this.

### Terms expressing time

<table>
<thead>
<tr>
<th>Term</th>
<th>MLP Class</th>
<th>Part of Speech</th>
<th>SHML Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>antecede</td>
<td>H-TMLOC</td>
<td>TV</td>
<td>tm_tm-loc</td>
</tr>
<tr>
<td>on admission</td>
<td>H-TMLOC</td>
<td>D</td>
<td>tm_tm-loc</td>
</tr>
<tr>
<td>initially</td>
<td>H-TMBEG</td>
<td>D</td>
<td>tm_beg</td>
</tr>
<tr>
<td>emergent</td>
<td>H-TMBEG</td>
<td>ADJ</td>
<td>tm_beg</td>
</tr>
<tr>
<td>discontinue</td>
<td>H-TMEND</td>
<td>TV</td>
<td>tm_end</td>
</tr>
<tr>
<td>end-stage</td>
<td>H-TMEND</td>
<td>N</td>
<td>tm_end</td>
</tr>
<tr>
<td>unrelenting</td>
<td>H-TMDUR</td>
<td>ADJ</td>
<td>tm_dur</td>
</tr>
<tr>
<td>yearly</td>
<td>H-TMREP</td>
<td>D</td>
<td>tm_rep</td>
</tr>
<tr>
<td>after</td>
<td>H-TMPREP</td>
<td>P</td>
<td>tm_tm-prp</td>
</tr>
<tr>
<td>will *</td>
<td>FUT</td>
<td>W</td>
<td>tm_tense</td>
</tr>
<tr>
<td>09/30/2005 *</td>
<td>DT</td>
<td></td>
<td>tm_tm-exact</td>
</tr>
</tbody>
</table>

* FUT and DT are provided by the medical language processor.

### Terms expressing negation

<table>
<thead>
<tr>
<th>Term</th>
<th>MLP Class</th>
<th>Part of Speech</th>
<th>SHML Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>deny</td>
<td>H-NEG</td>
<td>TV</td>
<td>md_ng</td>
</tr>
<tr>
<td>expecting</td>
<td>H-NEG</td>
<td>P</td>
<td>md_ng</td>
</tr>
<tr>
<td>exclude</td>
<td>H-NEG</td>
<td>V</td>
<td>md_ng</td>
</tr>
<tr>
<td>never</td>
<td>H-NEG</td>
<td>D</td>
<td>md_ng</td>
</tr>
<tr>
<td>not able</td>
<td>H-NEG</td>
<td>ADJ</td>
<td>md_ng</td>
</tr>
<tr>
<td>nothing</td>
<td>H-NEG</td>
<td>PRO</td>
<td>md_ng</td>
</tr>
<tr>
<td>rejected</td>
<td>H-NEG</td>
<td>VEN</td>
<td>md_ng</td>
</tr>
<tr>
<td>without</td>
<td>H-NEG</td>
<td>P</td>
<td>md_ng</td>
</tr>
</tbody>
</table>

### Terms expressing uncertainty

<table>
<thead>
<tr>
<th>Term</th>
<th>MLP Class</th>
<th>Part of Speech</th>
<th>SHML Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>allegedly</td>
<td>H-MODAL</td>
<td>D</td>
<td>md_modal</td>
</tr>
<tr>
<td>assume</td>
<td>H-MODAL</td>
<td>TV</td>
<td>md_modal</td>
</tr>
<tr>
<td>assumption</td>
<td>H-MODAL</td>
<td>N</td>
<td>md_modal</td>
</tr>
<tr>
<td>conceivably</td>
<td>H-MODAL</td>
<td>D</td>
<td>md_modal</td>
</tr>
<tr>
<td>doubtful</td>
<td>H-MODAL</td>
<td>ADJ</td>
<td>md_modal</td>
</tr>
<tr>
<td>hypothesis</td>
<td>H-MODAL</td>
<td>N</td>
<td>md_modal</td>
</tr>
<tr>
<td>hypothesize</td>
<td>H-MODAL</td>
<td>TV</td>
<td>md_modal</td>
</tr>
<tr>
<td>hypothetical</td>
<td>H-MODAL</td>
<td>ADJ</td>
<td>md_modal</td>
</tr>
</tbody>
</table>
The current combined lexicon is derived from terms found in actual records and from publicly available sources. It is possible that the lexicon could be extended to include all terms found in the UMLS, available from NLM. Classing and tagging such a source could be a formidable, but worthwhile task. SHML tagging of this or other terminology sources would in effect make them more operational.

The enumeration of the current version of SHML tags is shown in the accompanying Table 1. Each term is given a primary class tag and when deemed useful additional support tags are added (see Notes on Use of this Lexicon).

Documents when processed by MLP/SHML technology are both human and machine readable. Tools needed for adoption of MLP/SHML are few, readily available and inexpensive. MLP/SHML markup provides a digital representation of a medical document. The software outlined above can store, display, process, transmit, search, and print each identified informational element.

David J. Rothwell, MD
July 15, 2005
<table>
<thead>
<tr>
<th>Medical Tag Hierarchy</th>
</tr>
</thead>
</table>

### TABLE 1

**MEDICAL TAG HIERARCHY**

by David Rothwell, MD, Richard Wheeler, MD, & Ngô Thanh Nhàn, Ph.D.

available from the *Health Language Center*

---

1. `<a-s>` (Anatomic System)
   - `<sys>` (Systemic)
   - `<multi-sys>` (Multi-System)
   - `<npsych>` (Neuropsychological System)
   - `<gn-sys>` (Genetic System)
     - `<chrom>` (Chromosome)
     - `<gene>` (Gene)
   - `<nr>` (Neurologic System)
     - `<cns>` (Central Nervous System)
       - `<brain>` (Brain)
       - `<mng>` (Meninges)
       - `<c-n>` (Cranial Nerves)
       - `<spinal-c>` (Spinal Cord)
     - `<p-n-s>` (Peripheral Nervous System)
     - `<a-n-s>` (Autonomic Nervous System)
     - `<e-n-s>` (Enteric Nervous System)
     - `<prg>` (Paraganglion)
   - `<eye>` (Eye)
     - `<eld>` (Eyelid)
     - `<cj>` (Conjunctiva)
   - `<glb>` (Eye Globe)
     - `<op-ds>` (Optic Disc)
     - `<rtn>` (Retina)
     - `<Ins>` (Lens)
     - `<ppl>` (Pupil)
     - `<uv-tr>` (Uveal Tract)
   - `<chd>` (Choroid)
   - `<cl-bd>` (Ciliary Body)
   - `<ir>` (Iris)
   - `<scl>` (Sclera)
   - `<crn>` (Cornea)
   - `<ps-ch>` (Posterior Chamber)
   - `<an-ch>` (Anterior Chamber)
   - `<lc-dc>` (Lacrimal Duct)
   - `<lc-gl>` (Lacrimal Gland)
   - `<ear>` (Ear)
     - `<i-e>` (Inner Ear)
     - `<m-e>` (Middle Ear)
     - `<e-e>` (External Ear)
   - `<dnt>` (Dental Structures)
<tth> (Teeth)
<gm> (Gums)
<resp> (Respiratory System)
<u-r> (Upper Respiratory Tract)
<nse> (Nose)
<sns> (Sinuses)
<nasoph> (Nasopharynx)
<l-r> (Lower Respiratory Tract)
<epgl> (Epiglottis)
<lrx> (Larynx)
<trch> (Trachea)
<brchi> (Bronchi)
<lng> (Lung)
<plr> (Pleura)
<dgm> (Diaphragm)
<cv> (Cardiovascular System)
<hrt> (Heart)
<myc> (Myocardium)
<atr> (Atrium)
<hrt-vnt> (Ventricle)
<hrt-vlv> (Valve)
<cnd> (Conduction System)
<prcm> (Pericardium)
<vscc> (Vasculature)
<art> (Artery)
<cor-a> (Coronary Arteries)
<pul-a> (Pulmonary Arteries)
<arta> (Aorta)
<crt-a> (Carotid Arteries)
<ic-a> (Intracerebral Arteries)
<vn> (Vein)
<ic-v> (Intracerebral Vein)
<d-s> (Dural Sinus)
<v-c> (Vena Cava)
<lym-sys> (Lymphatic System)
<gi> (GI System)
<omn> (Omentum)
<gi-tr> (GI Tract)
<u-gi> (Upper GI Tract)
<mth> (Mouth)
<phryx> (Pharynx)
<esph> (Esophagus)
<stm> (Stomach)
<tongue> (Tongue)
<l-gi> (Lower GI Tract)
<s-i> (Small Intestine)
<du> (Duodenum)
<jj> (Jejunum)
<i> (Ileum)
<l-i> (Large Intestine)
<ccm> (Cecum)
<appdx> (Appendix)
<cln> (Colon)
<rtm> (Rectum)
<anus> (Anus)
<gi-or> (Gastrointestinal Organs)
<s-gd> (Salivary Gland)
<lvr> (Liver)
<gb> (Gallbladder)
<b-d> (Bile Duct, Extrahepatic)
<ex-p> (Exocrine Pancreas)
<perit> (Peritoneum)
<gu> (Genitourinary System)
<urn> (Urinary System)
<up-urn> (Upper Urinary System)
<kd> (Kidney)
<glm> (Glomerulus)
<tub> (Tubule)
<inter> (Interstitial)
<urt> (Ureter)
<lw-urn> (Lower Urinary System)
<bldr> (Bladder)
<urth> (Urethra)
<gen-sys> (Genital System)
<gyn> (Female Genital System)
<vulva> (Vulva)
<vgn> (Vagina)
<cx> (Cervix)
<ut> (Uterus)
<ov> (Ovary)
<tb> (Fallopian Tube)
<mI> (Male Genital System)
<penis> (Penis)
<tst> (Testis)
<prost> (Prostate)
<scrot> (Scrotum)
<endo> (Endocrine)
<pit> (Pituitary)
<ant-l> (Anterior Lobe)
<pos-l> (Posterior Lobe)
<pnl> (Pineal)
<adr> (Adrenal)
2. <em-ft> (Embryo and Fetus)
3. <cell> (Cell)
4. <b-r> (Body Region)
   <bdy> (Body)
   <m-r> (Multi-region)
   <h-n> (Head and Neck)
   <hd> (Head)
   <cran> (Cranial Vault)
<itr> (Infratentorial Region)
  <b-o-s> (Base of the Skull)
  <ps-fs> (Posterior Fossa)
<br> (Supratentorial Region)
  <pt-fs> (Pituitary Fossa)
  <md-fs> (Middle Fossa)
  <at-fs> (Anterior Fossa)

<jw> (Jaw)
<chn> (Chin)
<bcl> (Buccal Region)
<orl> (Oral Region)
<ns> (Nasal Region)
<zyg> (Zygomatic Region)
<infr> (Infraorbital Region)
<supr> (Supraorbital Region)
<fc> (Face)
<brh> (Forehead)
<sbtl> (Suboccipital Region)
<oct> (Occipital Region)
<aur> (Auricular Region)
<tprl> (Temporal Region)
<prt> (Parietal Region)
<frt> (Frontal Region)
<br> (Scalp)
<orb-reg> (Orbital Region)
<mastoid> (Mastoid Region)
<parotid> (Parotid Region)

<nk> (Neck)
  <a-n> (Anterior Neck)
  <p-n> (Posterior Neck)
  <supra> (Supraclavicular Region)
<br> (Spine)
  <c-s> (Cervical Spine)
  <t-s> (Thoracic Spine)
  <l-s> (Lumbar Spine)
  <sac-s> (Sacral Spine)
  <cocc-s> (Coccygeal Spine)
  <thor-lumb> (Thoracolumbar Spine)
  <lumb-sac> (Lumbosacral Spine)
<br> (Trunk)
  <ax> (Axilla)
  <thx> (Thorax)
    <int-thor> (Intrathoracic)
      <mediast> (Mediastinum)
  <ch-wl> (Chest Wall)
    <prc> (Precordium)
<abd> (Abdomen)
<abd-wl> (Abdominal Wall)
<retroper> (Retroperitoneum)
<int-abd> (Intraabdominal Cavity)
<mes> (Mesentery)
<pel-cav> (Pelvic Cavity)
<pel> (Pelvis)
<bk> (Back)
<cv-j> (Costovertebral Junction)
<hip> (Hip)
<ing> (Inguinal Region)
<perin> (Perineum)
<butt> (Buttock)
<ex> (Extremity)
<u-e> (Upper Extremity)
<shl> (Shoulder)
<u-a> (Upper Arm)
<elb> (Elbow)
<frm> (Forearm)
<wr> (Wrist)
<hnd> (Hand)
<fg> (Finger)
<l-e> (Lower Extremity)
<fem> (Femoral Region)
<thg> (Thigh)
<kn> (Knee)
<l-l> (Lower Leg)
<ank> (Ankle)
<ft> (Foot)
<toe> (Toe)
<pop-fossa> (Popliteal Fossa)

5. <s-s> (Sign or Symptom)
6. <pe-sp> (Finding Specific to PE)
7. <dx> (Diagnosis)
8. <dx-prcss> (Known Process or Cause)
   <imm> (Immunologic Process)
   <all> (Allergic Process)
   <auto> (Autoimmune Process)
   <gen> (Genetic Cause)
   <met> (Metabolic Process)
   <infect> (Infectious Process)
   <inj> (Injury)
   <exp> (Exposure)
   <pois> (Poisoning)
   <nutr> (Nutritional Cause)
   <ab-negl> (Abuse/Neglect)
<neopl> (Neoplastic Process)
  <benign> (Benign)
  <in-situ> (In-Situ)
  <malig> (Malignant)
  <uncert-benign-malig> (Uncertain Benign/Malign)
  <metast> (Metastatic)
<i nflam> (Inflammatory Process)
<degen> (Degenerative Process)
<mort> (Mortality)
<i schem> (Ischemic Process)
<wnd-heal> (Wound Healing)
<atheroscl> (Atherosclerosis)
9. <dx-kind> (Diagnostic Groups)
  <neuro> (Neurological Groups)
    <demyel> (Demyelinating Disease)
    <paraly> (Paralytic Disease)
    <migr> (Migraine Disorder)
    <epilep> (Epilepsy)
    <extpyr> (Extrapyramidal Disease)
    <ataxia> (Ataxia)
    <myelop> (Myelopathy)
    <mot-neur> (Motor Neuron Disease)
  <neuropsy> (Neuropsychological Groups)
    <eat-dis> (Eating Disorder)
    <mood> (Mood Disorder)
    <demen> (Dementia)
    <schizo> (Schizophrenia)
    <psysex> (Psychosexual Disease)
    <sleep> (Sleep Disorder)
    <somato> (Somatoform Disorder)
    <thgt> (Thought Disorder)
  <d-k-eye> (Eye Groups)
    <refrac> (Refraction Disorder)
    <eye-mov> (Eye Movement Disorder)
    <glau> (Glaucoma)
  <d-k-ear> (Ear Groups)
    <hear> (Hearing Disorder)
    <vest> (Vestibular Disorder)
  <d-k-resp> (Respiratory Groups)
    <pneum> (Pneumonia)
    <hy-pneum> (Hypersensitivity Pneumonitis)
    <r-a-d> (Reactive Airway Disease)
    <emphy> (Emphysema)
    <int-lung-dis> (Interstitial Lung Disease)
    <prot> (Proteinosis)
  <cardiov> (Cardiovascular Groups)
<htn> (Hypertensive Disease)
<fail> (Heart Failure)
<rheum> (Rheumatic Disease)
<cardmy> (Cardiomyopathy)
<arrhy> (Arrhythmia)
<thromb> (Thrombotic Disease)
<hemorr> (Hemorrhagic Disease)
<gastro> (Gastrointestinal Groups)
<malab> (Malabsorption)
<hepat> (Hepatitis)
<cirr> (Cirrhotic Disease)
<urinary> (Urinary Tract Groups)
<nephro> (Nephrotic Disease)
<d-k-endo> (Endocrine Groups)
<hypthy> (Hyperthyroidism)
<hypothy> (Hypothyroidism)
<goiter> (Goiter)
<thyditis> (Thyroiditis)
<thytox> (Thyrotoxicosis)
<d-k-mss> (Musculoskeletal Groups)
<arthrit> (Arthritis)
<re-imm> (Reticulo-Immune Groups)
<imm-def> (Immune Deficiency)
<gammop> (Gammopathy)
<amylo> (Amyloid)
<hemato> (Hematologic Groups)
<anemia> (Anemia)
<ref-anem> (Refractory Anemia)
<hemo-anem> (Hemolytic Anemia)
<megalo> (Megaloblastic Anemia)
<aplast-anem> (Aplastic Anemia)
<hgbopath> (Hemoglobinopathy)
10. <v-s> (Vital Sign)
<bp> (Arterial Blood Pressure)
<pls-rate> (Pulse Rate)
<resp-rate> (Respiratory Rate)
<tmp> (Temperature)
<v-press> (Venous Pressure)
11. <bdy-ms> (Body Measurement)
<bdy-hgt> (Body Height or Length)
<wgt> (Weight)
<hd-cir> (Head Circumference)
<r-o-m> (Range of Motion)
12. <phy-fun> (Physiologic Function)
13. <sec> (Secretion)
14. <hlth-st> (Health Status)
<dis> (Disability)
15. <act> (Physical Activity)
   <adl> (ADL)
   <sprt> (Sports Activity)
16. <p-s-b> (Patient Behavior)
   <hyg> (Hygiene)
   <tob-u> (Tobacco Use)
   <alc-u> (Alcohol Use)
   <cff-u> (Caffeine Use)
   <d-u> (Illicit Drug Use)
   <cr-b> (Criminal Behavior)
   <sex> (Sexual Behavior)
   <b-c> (Birth Control)
17. <s-c> (Sociocultural Information)
   <rel> (Relationship)
   <bel> (Belief)
   <val> (Value)
   <liv-sit> (Living Situation)
   <trvl> (Travel)
18. <diet-in> (Dietary Information)
19. <food> (Food)
20. <med> (Medication)
21. <med-ds> (Medication Dose)
22. <med-rt> (Medication Route)
23. <med-cl> (Medication Class)
   <antiinf> (anti-infectives)
      <amebicide> (amebicides)
      <antihel> (anthelmintics)
      <antifung> (antifungals)
         <amphoter> (amphotericins)
      <azole-antifungal> (azole antifungals)
      <misc-antifungal> (miscellaneous antifungals)
      <antimalar> (antimalarial agents)
      <a aminoquin> (a aminoquinolones)
      <misc-antimalar> (miscellaneous antimalarials)
      <antitb> (antituberculosis agents)
      <aminosalic> (aminosalicylates)
      <nicot-acid-der> (nicotinic acid derivatives)
      <rifamyc-der> (rifamycin derivatives)
      <strep-deriv> (streptomyces derivatives)
      <misc-antitb> (miscellaneous antituberculosis agents)
      <antiviral> (antiviral agents)
      <protease-inh> (protease inhibitors)
      <nrti> (NRTIs)
      <misc-antiviral> (miscellaneous antivirals)
      <nnrti> (NNRTIs)
<adam-antiviral> (adamantane antivirals)
<pur-nucl> (purine nucleosides)
<carbap> (carbapenems)
<cephalo> (cephalosporins)
<i-gen-ceph> (first generation cephalosporins)
<ii-gen-ceph> (second generation cephalosporins)
<iii-gen-ceph> (third generation cephalosporins)
<iv-gen-ceph> (fourth generation cephalosporins)
<leprostat> (leprostatics)
<macrolid> (macrolides)
<misc-antibiot> (miscellaneous antibiotics)
<pcn> (penicillins)
<pcnase-resis-pcn> (penicillinase resistant penicillins)
<antipseudomon-pcn> (antipseudomonal penicillins)
<aminopen> (aminopenicillins)
<beta-lact-inh> (beta-lactamase inhibitors)
<natural-pcn> (natural penicillins)
<quinolone> (quinolones)
<sulfamide> (sulfonamides)
<tetracyc> (tetracyclines)
<urin-antiinf> (urinary anti-infectives)
<aminogly> (aminoglycosides)
<linco-der> (lincomycin derivatives)
<antilipid> (antihyperlipidemic agents)
<hmg-coa-reduc> (HMG-CoA reductase inhibitors)
<misc-antilipid> (miscellaneous antihyperlipidemic agents)
<fib-acid-der> (fibric acid derivatives)
<bile-acid-seq> (bile acid sequestrants)
<antineopl> (antineoplastics)
<alk-agt> (alkylating agents)
<antibiot-antineopl> (antibiotics/antineoplastics)
<antimetab> (antimetabolites)
<ch-cl-hrm-antineopl> (hormones/antineoplastics)
<misc-antineopl> (miscellaneous antineoplastics)
<mitot-inh> (mitotic inhibitors)
<radiopharm> (radiopharmaceuticals)
<med-biol> (biologicals)
<antitox-antiven> (antitoxins and antivenins)
<col-stim-fac> (colony stimulating factors)
<in-vivo-dx-biol> (in vivo diagnostic biologicals)
<recomb-hum-erythro> (recombinant human erythropoietins)
<immuno-agt> (immunologic agents)
<bact-vac> (bacterial vaccines)
<immunoglob> (immune globulins)
<toxoid> (toxoids)
<viral-vac> (viral vaccines)
<misc-biol> (miscellaneous biologicals)
<immunosupp-agt> (immunosuppressive agents)
<interferon> (interferons)
<monoclon-ab> (monoclonal antibodies)
<cv-agt> (cardiovascular agents)
<hypert-emerg-agt> (agents for hypertensive emergencies)
<angioten-conv-enz-> (angiotensin converting enzyme inhibitors)
<antiadr-agt-per> (antiadrenergic agents, peripherally acting)
<antiadr-agt-cent> (antiadrenergic agents, centrally acting)
<antianginal> (antianginal agents)
<antiarrhh> (antiarrhythmic agents)
<beta-adr-blk> (beta-adrenergic blocking agents)
<beta-blk-cv-sel> (cardioselective beta blockers)
<beta-blk-cv-nonsel> (non-cardioselective beta blockers)
<ca-chan-blk> (calcium channel blocking agents)
<diuret> (diuretics)
<loop-diur> (loop diuretics)
<k-sp-diur> (potassium-sparing diuretics)
<thia-diur> (thiazide diuretics)
<car-anh-inh> (carbonic anhydrase inhibitors)
<misc-diur> (miscellaneous diuretics)
<ionotrop-agt> (inotropic agents)
<misc-cv-agt> (miscellaneous cardiovascular agents)
<per-vasodil> (peripheral vasodilators)
<vasodil> (vasodilators)
<vasopr> (vasopressors)
<hyperten-comb> (antihypertensive combinations)
<angioten-ii-inh> (angiotensin II inhibitors)
<cns-agt> (central nervous system agents)
<analg> (analgesics)
<misc-analg> (miscellaneous analgesics)
<narc-analg> (narcotic analgesics)
<nsaid> (nonsteroidal anti-inflammatory agents)
<salic> (salicylates)
<analg-comb> (analgesic combinations)
<narc-comb> (narcotic analgesic combinations)
<antimigr> (antimigraine agents)
<cox-ii> (cox-2 inhibitors)
<anticonv> (anticonvulsants)
<hydant-anticonv> (hydantoin anticonvulsants)
<succ-anticonv> (succinimide anticonvulsants)
<barb-anticonv> (barbiturate anticonvulsants)
<oxazo-anticonv> (oxazolidinedione anticonvulsants)
<benzo-anticonv> (benzodiazepine anticonvulsants)
<misc-anticonv> (miscellaneous anticonvulsants)
<antiemet-antivert> (antiemetic/antivertigo agents)
<vhtiiirep-antg> (5HT3 receptor antagonists)
<pheno-antiemet> (phenothiazine antiemetics)
<antichol-antiemet> (anticholinergic antiemetics)
<misc-antiemet> (miscellaneous antiemetics)
<antipark> (antiparkinson agents)
<antichol-antipark> (anticholinergic antiparkinson agents)
<misc-antipark> (miscellaneous antiparkinson agents)
<dopa-antipark> (dopaminergic antiparkinsonism agents)
<gen-anes> (general anesthetics)
<musc-relax> (muscle relaxants)
<neuromusc-blk-agt> (neuromuscular blocking agents)
<skelmusc-relax> (skeletal muscle relaxants)
<skelmusc-relax-com> (skeletal muscle relaxant combinations)
<misc-cns-agt> (miscellaneous central nervous system age)
<anorex> (anorexiants)
<psy-agt> (psychotherapeutic agents)
<anx-sed-hyp> (anxiolytics, sedatives, and hypnotics)
<barb> (barbiturates)
<benzodia> (benzodiazepines)
<misc-anx-sed-hyp> (miscellaneous anxiolytics, sedatives and)
<cns-stim> (CNS stimulants)
<antidepr> (antidepressants)
<misc-antidepr> (miscellaneous antidepressants)
<ssri-anti-depr> (SSRI antidepressants)
<tricyc-antidepr> (tricyclic antidepressants)
<monamine-ox-inh> (monoamine oxidase inhibitors)
<antipsy> (antipsychotics)
<misc-antipsy> (miscellaneous antipsychotic agents)
<psy-comb> (psychotherapeutic combinations)
<pheno-antipsy> (phenothiazine antipsychotics)
<coag-mod> (coagulation modifiers)
<anticoag> (anticoagulants)
<hep> (heparins)
<coum-indand> (coumarins and indandiones)
<antiplt> (antiplatelet agents)
<plt-aggr-inh> (platelet aggregation inhibitors)
<glyprot-plt-inh> (glycoprotein platelet inhibitors)
<hep-antag> (heparin antagonists)
<misc-coag-mod> (miscellaneous coagulation modifiers)
<thrmlytic> (thrombolytics)
<gi-agt> (gastrointestinal agents)
<antacid> (antacids)
<antichol-antispas> (anticholinergics/antispasmodics)
<antidiarr> (antidiarrheals)
<dig-enz> (digestive enzymes)
<gall-stn-sol-agt> (gallstone solubilizing agents)
<oral-nutrit-supp> (oral nutritional supplements)
<vit> (Vitamin)
<vit-min-comb> (vitamin and mineral combinations)
<iv-nutrit-prod> (intravenous nutritional products)
<resp-agt> (respiratory agents)
<antihi> (antihistamines)
<antituss> (antitussives)
<brondil> (bronchodilators)
<methxanth> (methylxanthines)
<adr-brondil> (adrenergic bronchodilators)
<brondil-comb> (bronchodilator combinations)
<decong> (decongestants)
<expect> (expectorants)
<misc-resp-agt> (miscellaneous respiratory agents)
<resp-inhal-prod> (respiratory inhalant products)
<antiasthmat-comb> (antiasthmatic combinations)
<upper-resp-comb> (upper respiratory combinations)
<leukotr-mod> (leukotriene modifiers)
<lung-surf> (lung surfactants)
<tpcl-agt> (topical agents)
<anorect-prep> (anorectal preparations)
<antisept-antigerm> (antiseptic and germicides)
<derm-agt> (dermatological agents)
<top-antiinf> (topical anti-infectives)
<top-ster> (topical steroids)
<top-anes> (topical anesthetics)
<misc-top-agt> (miscellaneous topical agents)
<top-ster-antiinf> (topical steroids with anti-infectives)
<top-acne> (topical acne agents)
<top-antipsor> (topical antipsorials)
<top-emoll> (topical emollients)
<mouth-thr-prod> (mouth and throat products)
<ophth-prep> (ophthalmic preparations)
<ophth-antiinf> (ophthalmic anti-infectives)
<ophth-glauc-agt> (ophthalmic glaucoma agents)
<ophth-ster> (ophthalmic steroids)
<ophth-ster-antiinf> (ophthalmic steroids with anti-infectives)
<ophth-antiinflamm> (ophthalmic anti-inflammatory agents)
<ophth-lub-irrig> (ophthalmic lubricants and irrigations)
<misc-ophth-agt> (miscellaneous ophthalmic agents)
<ophth-antihi-dec> (ophthalmic antihistamines and decongesta)
<otic-prep> (otic preparations)
<otic-antiinf> (otic anti-infectives)
<otic-ster-antiinf> (otic steroids with anti-infectives)
<misc-otic-agt> (miscellaneous otic agents)
<vag-prep> (vaginal preparations)
<spermicide> (spermicides)
<vag-antiinf> (vaginal anti-infectives)
<misc-vag-agt> (miscellaneous vaginal agents)
<sterile-irrig-sol> (sterile irrigating solutions)
<nlsl-prep> (nasal preparations)
<nlsl-lub-irrig> (nasal lubricants and irrigations)
<nlsl-ster> (nasal steroids)
<nlsl-antihist-decon> (nasal antihistamines and decongestants)
<altmed> (alternative medicines)
<altmed-nut> (nutraceutical products)
<altmed-herb> (herbal products)

24. <pr> (Procedure)
<admit> (Admit)
<discharge> (Discharge)
<ref> (Referral)
<con> (Consult)
<adm> (Administrative)
<csg> (Counseling)
<p-e> (Physical Exam)
<lab> (Lab)
<lab-chem> (Lab Chemistry)
<enz> (Enzyme)
<bili> (Bilirubin)
<amino> (Amino Acid)
<l lipid> (Lipid)
<prot> (Proteinosis)
<carbo> (Carbohydrate)
<vit> (Vitamin)
<prgland> (Prostaglandin)
<horm> (Hormone)
<elect> (Electrolyte)
<neurotrans> (Neurotransmitters)
<lipopro> (Lipoprotein)
<radionuc> (Radionuclide)
<lab-immuno> (Immunology Lab)
<recept> (Receptor)
<cmplmt> (Complement)
<antigen> (Antigen)
<antib> (Antibody)
<mediator> (Mediator)
<genetic-test> (Genetic Testing)
<nuc-ac> (Nucleic Acid)
<cytogen-test> (Cytogenetic Test)
<hemat> (Hematology)
<porph> (Porphyrin)
<hgb-type> (Hemoglobin Type)
<coag> (Coagulation)
<branel> (Panel)
<metab-basic> (Metabolic Basic)
<metab-comp> (Metabolic Comprehensive)
<gen-hlth> (General Health)
<elect-panel> (Electrolyte Panel)
<ob-panel> (Obstetric Panel)
<hepatic-panel> (Hepatic Panel)
<hepatitis-panel> (Hepatitis Panel)
<arthit-panel> (Arthritis Panel)
<thyroid-panel> (Thyroid Panel)
<thyroid-panel-tsh> (Thyroid Panel with TSH)
<trans> (Transfusion)
<micro> (Microbiology)
<tiss-type> (Tissue Typing)
<evo-suppl> (Evocative/Suppression Test)
<tox> (Toxicology)
<toxin> (Toxin)
<tol-test> (Tolerance Test)
<cyto> (Cytology)
<urinal> (Urinalysis)
<im> (Imaging Procedure)
<n-m> (Nuclear Medicine)
<m-s> (Med/Surgical Procedure)
<endosc> (Endoscopy)
<ecg> (ECG)
<ecg-axis> (ECG Axis)
<ecg-int> (ECG Interval)
<ecg-wv> (ECG Wave)
<transplant> (Transplant)
<anes> (Anesthesia)
<p-m> (Physical Medicine)
<s-t> (Speech Therapy)
<c-m> (Complementary Therapy)
<plx> (Prophylactic Procedure)
<eval> (Evaluation)
<foll-up> (Follow up)
<mgt> (Management)
<rad-rx> (Radiation Therapy)
<transfer> (Transfer)
<clin-trial> (Clinical Trial)
25. <inv-pr> (Invasive Procedure)
<maj-pr> (Major Procedure)
<min-pr> (Minor Procedure)
26. <risk-fct> (Risk Factor)
27. <contra> (Contraindication)
28. <pt-dir> (Patient Directive)
29. <pt-pref-agree> (Patient Preference/Agreement)
30. <pt-ed> (Patient Education)
31. <trt-rec> (Treatment Recommendation)
32. <trt-opt> (Discussion of Options)
33. <pt-und> (Understanding of Illness/Therapy)
34. <gl-obj> (Goal/Objective of Therapy)
35. <cmp> (Compliance)
36. <response> (Response)
37. <complic> (Complication)
38. <tm> (Time)
   <dur> (Duration)
   <tm-loc> (Location)
      <life-stg> (Life Stage)
         <preg-ob> (Pregnancy/Obstetric)
         <peds> (Pediatric)
         <adult> (Adult)
         <birth-neonatal> (Birth/Neonatal)
         <congen> (Congenital)
      <urg> (Urgency)
   <rep> (Repetition)
   <beg> (Begin)
   <end> (End)
   <tm-prp> (Preposition)
   <tense> (Tense)
      <past> (Past)
      <present> (Present)
      <perf> (Perfect)
      <prog> (Progressive)
      <future> (Future)
      <fut-imp> (Future-Implied)
   <tm-exact> (Exact Time)
39. <md> (Modifier)
   <amt> (Amount)
   <chg> (Change)
   <st> (Status)
   <s-g> (Stage-grade)
   <des> (Description)
   <app> (Appearance)
   <ng> (Negation)
   <modal> (Modal)
   <chg-less> (Change-Less)
   <chg-more> (Change-More)
40. <p-o> (Physical Orientation)
   <laterality> (Laterality)
41. <per> (Person)
   <kin> (Kinship)
   <civ-rel> (Civil Relation)
42. <dem> (Demographic Information)
   <relig> (Religion)
   <lang> (Language)
   <mlty> (Military Service)
   <edu> (Education)
   <race-eth> (Race/Ethnicity)
43. <occ> (Occupation)
   <med-occ> (Medical Occupation)
44. <org> (Organism)
   <mc> (Microbe)
   <bct> (Bacteria)
   <gm-pos> (Gram Positive)
   <gm-neg> (Gram Negative)
   <vr> (Virus)
   <rck> (Rickettsia)
   <fgs> (Fungus)
   <par> (Parasite)
   <arthropod> (Arthropod)
   <pl> (Plant)
   <anim> (Animal)
45. <chem> (Chemical)
46. <dev> (Device)
   <dev-med> (Medical Device)
47. <env> (Environment)
   <env-med> (Medical Environment)
   <geo> (Geographic)
   <hz> (Hazard)
   <atm> (Atmosphere)
   <work> (Work)
48. <p-f> (Physical Force)
49. <nl> (Normal)
50. <grft> (Graft)
51. <srg-app> (Surgical Approach)
52. <unit> (Unit)
53. <num> (Number)
   <sys-press> (Systolic Pressure)
   <diast-press> (Diastolic Pressure)
   <end-dias-press> (End-Diastolic Pressure)
   <mean-press> (Mean Pressure)
54. <dct> (Document)
55. <rgn> (Regulation)
56. <li> (Link)
   <cnn> (Connector)
   <li-prp> (Preposition)
   <ptv> (Verb)
   <ttg> (Verb Ttgen)
   <shw> (Show)
   <obs> (Observe)
   <vkv> (Verb Have)
   <vtst> (Vtest)
   <vbe> (Verb Be)
   <bcnn> (Beconn)
   <brp> (Be Representative)
   <vdo> (Verb Do)
57. <prnn> (Pronouns)
58. <evnt> (Event)
59. <trns> (Transparent Term)
60. <epnym> (Eponym)
61. <null> (Null)
62. <p-p> (Parse Phrase)
63. <pt-sp> (Specimen from Patient)
64. <q> (Question)
<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>Adjective</td>
<td></td>
</tr>
<tr>
<td>CS0</td>
<td>Subordinate conjunction in SUB0</td>
<td></td>
</tr>
<tr>
<td>CS1</td>
<td>Subordinate conjunction in SUB1</td>
<td></td>
</tr>
<tr>
<td>CS2</td>
<td>Subordinate conjunction in SUB2</td>
<td></td>
</tr>
<tr>
<td>CS3</td>
<td>Subordinate conjunction in SUB3</td>
<td></td>
</tr>
<tr>
<td>CS4</td>
<td>Subordinate conjunction in SUB4</td>
<td>not used in MLP</td>
</tr>
<tr>
<td>CS5</td>
<td>Subordinate conjunction in SUB5</td>
<td></td>
</tr>
<tr>
<td>CS6</td>
<td>Subordinate conjunction in SUB6</td>
<td></td>
</tr>
<tr>
<td>CS7</td>
<td>Subordinate conjunction in SUB7</td>
<td>not used in MLP</td>
</tr>
<tr>
<td>CS8</td>
<td>Subordinate conjunction in SUB8</td>
<td>not used in MLP</td>
</tr>
<tr>
<td>CS9</td>
<td>Subordinate conjunction in SUB9</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Adverb</td>
<td></td>
</tr>
<tr>
<td>DP</td>
<td>Adverbal Particle in DP object strings</td>
<td></td>
</tr>
<tr>
<td>DS</td>
<td>Dose string, treated as a unit</td>
<td></td>
</tr>
<tr>
<td>DT</td>
<td>Date in numerical form dd/mm/yy</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>Noun</td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td>Possessive Noun</td>
<td></td>
</tr>
<tr>
<td>NULL</td>
<td>Parser “pass”: No word required; Adjunct</td>
<td></td>
</tr>
<tr>
<td>NULLC</td>
<td>Parser “pass”: No word required; Conjunct</td>
<td></td>
</tr>
<tr>
<td>NULLFRAG</td>
<td>Parser “pass”: No word required; Fragment</td>
<td></td>
</tr>
<tr>
<td>NULLN</td>
<td>Parser “pass”: No word required; Noun</td>
<td>not used in MLP</td>
</tr>
<tr>
<td>NULLOBJ</td>
<td>Parser “pass”: No word required; Object</td>
<td></td>
</tr>
<tr>
<td>NULLRECIP</td>
<td>Parser “pass”: No word required; Reciprocals</td>
<td>not used in MLP</td>
</tr>
<tr>
<td>NULLWH</td>
<td>Parser “pass”: No word required; WH-strings</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>Preposition</td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>Pronoun</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>Quantity</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>Article</td>
<td></td>
</tr>
<tr>
<td>TV</td>
<td>Tensed Verb</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>Untensed Verb (infinitive)</td>
<td></td>
</tr>
<tr>
<td>VEN</td>
<td>Past Participle</td>
<td></td>
</tr>
<tr>
<td>VING</td>
<td>Present Participle</td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>Auxiliary (will, can, may, should, etc.)</td>
<td></td>
</tr>
</tbody>
</table>
**TABLE 3**

**LIST OF SUBORDINATE CONJUNCTION STRINGS**

<table>
<thead>
<tr>
<th>DESCRIPTION</th>
<th>EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUB0</td>
<td>Subordinate conjunction + Object of ‘be’ when conscious of my observation</td>
</tr>
<tr>
<td>SUB1</td>
<td>Subordinate conjunction + Assertion because her blood pressure has been well controlled</td>
</tr>
<tr>
<td>SUB1-PHRASE</td>
<td>Subordinate conjunction phrase He is responsive in that he would open his eyes</td>
</tr>
<tr>
<td>SUB2</td>
<td>Subordinate conjunction + VENPASS (passive verb with its passive object) since last documented</td>
</tr>
<tr>
<td>SUB3</td>
<td>Subordinate conjunction + VINGO (V ingl form of verb + Object) since being on Prednisone</td>
</tr>
<tr>
<td>SUB4</td>
<td>Subordinate conjunction + VING string (either VINGOFN or NSVINGO)</td>
</tr>
<tr>
<td>SUB5</td>
<td>Subordinate conjunction + SVINGO (Subject + V ingl form of verb + Object) without anyone noticing it</td>
</tr>
<tr>
<td>SUB6</td>
<td>Subordinate conjunction + Subject + Object of ‘be’ with noone on the premises</td>
</tr>
<tr>
<td>SUB7</td>
<td>Subordinate conjunction + SVEN</td>
</tr>
<tr>
<td>SUB8</td>
<td>Subordinate conjunction (as) + partial inverted Assertion as were the other medications</td>
</tr>
<tr>
<td>SUB9</td>
<td>Subordinate conjunction + infinitive verb + Object in order to break this cycle; in order to gain control</td>
</tr>
<tr>
<td>SUB11</td>
<td>Time phrase (TM-PHRASE) + Assertion during which time he remained unconscious</td>
</tr>
<tr>
<td>SUB12</td>
<td>‘should’ + tenseless Assertion instruct to call should she have any fevers</td>
</tr>
<tr>
<td>SUB13</td>
<td>SUB1-PHRASE + Assertion He is responsive in that he would open his eyes</td>
</tr>
<tr>
<td>MLP CLASS</td>
<td>DESCRIPTION</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>PATIENT AREA</strong></td>
<td></td>
</tr>
<tr>
<td>H-PT</td>
<td>words referring to patient</td>
</tr>
<tr>
<td>H-PTAREA</td>
<td>anatomical area</td>
</tr>
<tr>
<td>H-PTDESCR</td>
<td>occupation, nationality, traits</td>
</tr>
<tr>
<td>H-PTFUNC</td>
<td>physiological function</td>
</tr>
<tr>
<td>H-PTLOC</td>
<td>anatomical location relation</td>
</tr>
<tr>
<td>H-PTMEAS</td>
<td>anatomical measure</td>
</tr>
<tr>
<td>H-PTPALP</td>
<td>palpated body part</td>
</tr>
<tr>
<td>H-PTPART</td>
<td>body part</td>
</tr>
<tr>
<td>H-PTSPEC</td>
<td>specimen type</td>
</tr>
<tr>
<td>H-PTVERB</td>
<td>verb with patient subject</td>
</tr>
<tr>
<td><strong>TEST/EXAM AREA</strong></td>
<td></td>
</tr>
<tr>
<td>H-OBSERVE</td>
<td>verbs of observation</td>
</tr>
<tr>
<td>H-TXCLIN</td>
<td>clinical exam, action</td>
</tr>
<tr>
<td>H-TXPROC</td>
<td>examination procedure</td>
</tr>
<tr>
<td>H-TXSPEC</td>
<td>test of specimen</td>
</tr>
<tr>
<td>H-TXVAR</td>
<td>test variable</td>
</tr>
<tr>
<td><strong>TREATMENT AREA</strong></td>
<td></td>
</tr>
<tr>
<td>H-DEVMED</td>
<td>medical devices</td>
</tr>
<tr>
<td>H-INST</td>
<td>hospitals, clinics, physicians, staff</td>
</tr>
<tr>
<td>H-TTCOMP</td>
<td>complementary treatments</td>
</tr>
<tr>
<td>H-TTGEN</td>
<td>general medical management</td>
</tr>
<tr>
<td>H-TTMED</td>
<td>treatment by medication</td>
</tr>
<tr>
<td>H-TTMODE</td>
<td>mode of administration</td>
</tr>
<tr>
<td>H-TTSURG</td>
<td>surgical procedure</td>
</tr>
<tr>
<td><strong>TIME</strong></td>
<td></td>
</tr>
<tr>
<td>H-POST</td>
<td>support class for prefix morpheme post-</td>
</tr>
<tr>
<td>H-TMBEG</td>
<td>beginning</td>
</tr>
<tr>
<td>H-TMEND</td>
<td>termination</td>
</tr>
<tr>
<td>H-TMLOC</td>
<td>location in time</td>
</tr>
<tr>
<td>H-TMDUR</td>
<td>duration</td>
</tr>
<tr>
<td>H-TMPREP</td>
<td>time preposition</td>
</tr>
</tbody>
</table>
### FINDINGS AREA

<table>
<thead>
<tr>
<th>H-AMT</th>
<th>amount or degree</th>
<th>Full range of motion. Severe pain.</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-CHANGE</td>
<td>indication of change</td>
<td>alterations in management.</td>
</tr>
<tr>
<td>H-CHANGE-LESS</td>
<td>decrease</td>
<td>Ampicillin tapered from 500 mg to 300 mg qid.</td>
</tr>
<tr>
<td>H-CHANGE-MORE</td>
<td>increase</td>
<td>a rise in his Creatinine to 2.3. weight gain</td>
</tr>
<tr>
<td>H-CHANGE-SAME</td>
<td>no change</td>
<td>Cardiac silhouette remains within normal limits. Neurologic diagnoses are static.</td>
</tr>
<tr>
<td>H-DIAG</td>
<td>diagnosis</td>
<td>Patient has glaucoma.</td>
</tr>
<tr>
<td>H-INDIC</td>
<td>disease indicator word</td>
<td>Pt complained of tenderness in joints.</td>
</tr>
<tr>
<td>H-ORG</td>
<td>organism</td>
<td>CSF shows H. Influenzae type B.</td>
</tr>
<tr>
<td>H-RESP</td>
<td>patient response</td>
<td>His blood gas improved on 35% oxygen.</td>
</tr>
<tr>
<td>H-TXRES</td>
<td>test/exam result word</td>
<td>Positive Kernig’s. Gram stain negative.</td>
</tr>
<tr>
<td>H-DESCR</td>
<td>descriptive information</td>
<td>A punctuate dark lesion.</td>
</tr>
<tr>
<td>H-DIET</td>
<td>name of foods, dietary prescriptions</td>
<td>Allergic to tomatoes.</td>
</tr>
</tbody>
</table>

### EVIDENTIAL AREA

<table>
<thead>
<tr>
<th>H-NEG</th>
<th>negation of finding</th>
<th>Patient refused medication. Patient never had headaches.</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-MODAL</td>
<td>uncertainty of finding; event has not yet occurred</td>
<td>Rule out meningitis. Doctor advised surgery.</td>
</tr>
</tbody>
</table>

### CONNECTIVE AREA

<table>
<thead>
<tr>
<th>H-BECOMN</th>
<th>classifier verb</th>
<th>Treatment consisted of ampicillin.</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-CONN</td>
<td>connects 2 information units</td>
<td>Fever caused the headache.</td>
</tr>
<tr>
<td>H-SHOW</td>
<td>connects test and result</td>
<td>Echo revealed apical akinesis.</td>
</tr>
<tr>
<td>H-VTEST</td>
<td>lab or measurement verbs</td>
<td>HCT ranged from 24 to 35.</td>
</tr>
</tbody>
</table>

### OTHER SEMANTIC CLASSES

<table>
<thead>
<tr>
<th>H-AGE</th>
<th>life stage</th>
<th>adult, adolescence, years old, age</th>
</tr>
</thead>
<tbody>
<tr>
<td>VBE</td>
<td>verb ‘be’</td>
<td></td>
</tr>
<tr>
<td>BEREP</td>
<td>be replacer verbs</td>
<td>She becomes drowsy.</td>
</tr>
<tr>
<td>VDO</td>
<td>verb ‘do’</td>
<td></td>
</tr>
<tr>
<td>VHAVE</td>
<td>verb ‘have’</td>
<td></td>
</tr>
<tr>
<td>H-FAMILY</td>
<td>family member or friend</td>
<td>Brother also has sickle cell disease.</td>
</tr>
<tr>
<td>H-GEOGR</td>
<td>geographical names</td>
<td>Pt currently lives in San Francisco.</td>
</tr>
<tr>
<td>H-ETHNIC</td>
<td>racial or ethnic background</td>
<td>white male, African American female</td>
</tr>
<tr>
<td>H-RECORD</td>
<td>patient record terms</td>
<td>Review of health records showed PPD of 1972 was also positive.</td>
</tr>
<tr>
<td>H-TESTVIEW</td>
<td>angles or positions of imaging</td>
<td>lateral, axial, PA</td>
</tr>
<tr>
<td>H-TRANSP</td>
<td>general terms that get classes from its modifiers</td>
<td>a tonic, clonic type of convulsion. The list of medications... The state of the disease...</td>
</tr>
</tbody>
</table>
TABLE 5
CORRESPONDENCE BETWEEN MLP AND XML LEXICAL ATTRIBUTES

The following table shows the correspondence of lexical attributes between the English Medical Lexicon for Natural Language Processing [EML] and the XML Medical Knowledge Lexicon [XMK]. The medical language processor [MLP] automatically generates lexical attributes, and their XMK lexical tags, for entries that are not included in the EML, such as:
- numbers and dates;
- grammatical markers, such as TENSE;
- dose phrase patterns which inherently contain specific numbers and units
- names of patient, family members, friends, physicians and staff, and institutions, which are incidental (and can be anonymized);
- geographic names (or addresses), which can be updated from a standard list; and
- organism and medication lists which can be updated from other sources.

Thus, during postprocessing, the MLP inserts XMK tags into parsed information format trees, using the correspondence table below to generate corresponding XMK lexical tags.

<table>
<thead>
<tr>
<th>MLP Lexicon Classes</th>
<th>Part of Speech</th>
<th>XMK Lexicon Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLP Class</td>
<td>Primary</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DS [dose string]</td>
<td>med-ds</td>
</tr>
<tr>
<td></td>
<td>DT [date]</td>
<td>tm</td>
</tr>
<tr>
<td></td>
<td>T [English article]</td>
<td>null</td>
</tr>
<tr>
<td></td>
<td>H-BEREP [v, TV, VEN, VING]</td>
<td>li</td>
</tr>
<tr>
<td></td>
<td>H-CHANGE [various parts of speech]</td>
<td>md</td>
</tr>
<tr>
<td></td>
<td>H-CHANGE-LESS [various parts of speech]</td>
<td>md</td>
</tr>
<tr>
<td></td>
<td>H-CHANGE-MORE [various parts of speech]</td>
<td>md</td>
</tr>
<tr>
<td></td>
<td>H-CHANGE-SAME [various parts of speech]</td>
<td>md</td>
</tr>
<tr>
<td></td>
<td>H-CHEM [various parts of speech]</td>
<td>ch</td>
</tr>
<tr>
<td></td>
<td>H-CONN [various parts of speech]</td>
<td>li</td>
</tr>
<tr>
<td></td>
<td>H-DEVMED [ADJ, D, N, N:SI, N:PL]</td>
<td>dev</td>
</tr>
<tr>
<td></td>
<td>H-MODAL [various parts of speech]</td>
<td>md</td>
</tr>
<tr>
<td></td>
<td>H-NEG [various parts of speech]</td>
<td>md</td>
</tr>
<tr>
<td></td>
<td>H-RESP [various parts of speech]</td>
<td>tr-rsp</td>
</tr>
<tr>
<td></td>
<td>H-TMBEG [various parts of speech]</td>
<td>tm</td>
</tr>
<tr>
<td></td>
<td>H-TMDUR [various parts of speech]</td>
<td>tm</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
<td>Node</td>
</tr>
<tr>
<td>------</td>
<td>----------------------</td>
<td>------</td>
</tr>
<tr>
<td>H-TMEND</td>
<td>various parts of speech</td>
<td>tm</td>
</tr>
<tr>
<td>H-TMLOC</td>
<td>various parts of speech</td>
<td>tm</td>
</tr>
<tr>
<td>H-TMPREP</td>
<td>P</td>
<td>tm</td>
</tr>
<tr>
<td><em>class other than H-TMREP</em></td>
<td>P</td>
<td>li</td>
</tr>
<tr>
<td>H-TMREP</td>
<td>various parts of speech</td>
<td>tm</td>
</tr>
<tr>
<td>H-VTENSE [PAST]</td>
<td>GRAM-NODE</td>
<td>tense</td>
</tr>
<tr>
<td>H-VTENSE [PRESENT]</td>
<td>GRAM-NODE</td>
<td>tense</td>
</tr>
<tr>
<td>H-VTENSE [PERF]</td>
<td>GRAM-NODE</td>
<td>tense</td>
</tr>
<tr>
<td>H-VTENSE [PROG]</td>
<td>GRAM-NODE</td>
<td>tense</td>
</tr>
<tr>
<td>H-VTENSE [FUT]</td>
<td>GRAM-NODE</td>
<td>tense</td>
</tr>
<tr>
<td>H-VTENSE [FUT-IMP]</td>
<td>GRAM-NODE</td>
<td>tense</td>
</tr>
<tr>
<td>NTIME1</td>
<td>N, N:SI, N:PL, ADJ</td>
<td>tm</td>
</tr>
<tr>
<td>NTIME2</td>
<td>N, N:SI, N:PL</td>
<td>tm</td>
</tr>
<tr>
<td>NUNIT</td>
<td>N, N:SI, N:PL, TV, VING, ADJ</td>
<td>unit</td>
</tr>
<tr>
<td>QNUMBER</td>
<td>Q</td>
<td>num</td>
</tr>
<tr>
<td>VBE</td>
<td>V, TV, VEN, VING</td>
<td>li</td>
</tr>
<tr>
<td>VDO</td>
<td>W</td>
<td>li</td>
</tr>
<tr>
<td>VHAVE</td>
<td>V, TV, VEN, VING</td>
<td>li</td>
</tr>
</tbody>
</table>