DAVID ROTHWELL, MD, RICHARD WHEELER, MD, & NGÔ THANH NHÀN, Ph.D.
MEDICAL LANGUAGE PROCESSING, LLC
NEW YORK, NEW YORK

AN XML MEDICAL KNOWLEDGE LEXICON

句

Volume 1
A – I

NEW YORK
2005

DAVID ROTHWELL, MD, RICHARD WHEELER, MD & NGÔ THANH NHÀN, Ph.D.
MEDICAL LANGUAGE PROCESSING, LLC
NEW YORK, NEW YORK

AN XML MEDICAL KNOWLEDGE LEXICON

句

Volume 1
A – I

NEW YORK
2005

Printed in the United States of America

An XML Medical Knowledge Lexicon
by David Rothwell, MD, Richard Wheeler, MD, & Ngô Thanh Nhàn, Ph.D.

Volume 1 (A–I) 576 pp. 8½" x 11".

Includes index:

1. English medical knowledge lexicon. 2. natural language processing. 3. medical language processing. 4. medical term classification. 5. word class co-occurrence. 6. eXtensible Markup Language (XML). 7. Structured Health Markup Language (SHML).

句

Medical Language Processing, LLC
172 East 4th Street – Suite 3B
New York, New York 10009
Tel. (212) 420-1586
Fax: (212) 529-2891
Email: medical_language_processing@yahoo.com

Library of Congress Cataloging in Publication Data

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.

句

TABLE OF CONTENTS

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

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

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.

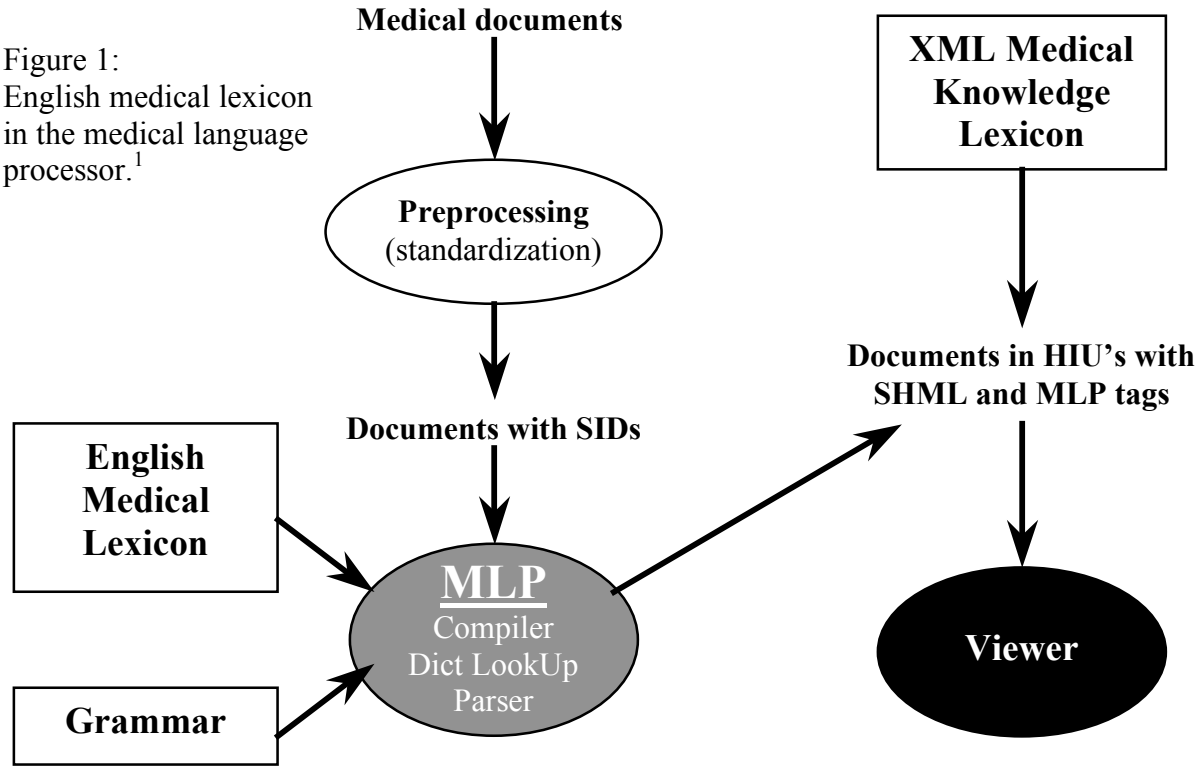

Figure 1:
English medical lexicon in the medical language processor.[1]

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

[1] 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	
Description	**Tag**
Diagnosis	dx
Diagnostic process	dx-prcss
Infectious diagnostic process	dx-prcss_infect
Immunologic diagnostic process	dx-prcss_imm
Neoplastic diagnostic process	dx-prcss_neopl
Diagnosis group	dx-kind
Neurologic disease	dx-kind_neuro
Migraine	dx- kind_neuro_migr
Reactive Airway Disease	dx-kind_d-k-resp_r-a-d
Asthma	dx-kind_d-k-resp_r-a-d

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 >
</dx>

Asthma

<dx>
<dx-prcss_imm_all><dx-kind_d-k-resp_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-k-resp_r-a-d ></dx-prcss_imm_all>
</dx>

Pneumonia

<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
</b-r_tk_thx_int-thor></a-s_resp_l-r_lng>
</or_mc>
</dx-kind_d-k-resp_pneum></dx-prcss_infect>
</dx>

Pneumonia, right lower lobe

<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>
</b-r_tk_thx_int-thor></a-s_resp_l-r_lng>
</or_mc>
</dx-kind_d-k-resp_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			
Term	**MLP Class**	**Part of Speech**	**SHML Tag**
antecede	H-TMLOC	TV	tm_tm-loc
on admission	H-TMLOC	D	tm_tm-loc
initially	H-TMBEG	D	tm_beg
emergent	H-TMBEG	ADJ	tm_beg
discontinue	H-TMEND	TV	tm_end
end-stage	H-TMEND	N	tm_end
unrelenting	H-TMDUR	ADJ	tm_dur
yearly	H-TMREP	D	tm_rep
after	H-TMPREP	P	tm_tm-prp
will *	FUT	W	tm_tense
09/30/2005 *		DT	tm_tm-exact

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

Terms expressing negation			
Term	**MLP Class**	**Part of Speech**	**SHML Tag**
deny	H-NEG	TV	md_ng
excepting	H-NEG	P	md_ng
exclude	H-NEG	V	md_ng
never	H-NEG	D	md_ng
not able	H-NEG	ADJ	md_ng
nothing	H-NEG	PRO	md_ng
rejected	H-NEG	VEN	md_ng
without	H-NEG	P	md_ng

Terms expressing uncertainty			
Term	**MLP Class**	**Part of Speech**	**SHML Tag**
allegedly	H-MODAL	D	md_modal
assume	H-MODAL	TV	md_modal
assumption	H-MODAL	N	md_modal
conceivably	H-MODAL	D	md_modal
doubtful	H-MODAL	ADJ	md_modal
hypothesis	H-MODAL	N	md_modal
hypothesize	H-MODAL	TV	md_modal
hypothetical	H-MODAL	ADJ	md_modal

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 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)
            - <lns> (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)
<vsc> (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> (Intracerbral 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)
<il> (Ileum)
<l-i> (Large Intestine)
<ccm> (Cecum)
<appdx> (Appendix)
<cln> (Colon)
<rctm> (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> (Interstitium)
<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)
<ml> (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)

<cort> (Cortex)
<medu> (Medulla)
<thy> (Thyroid)
<prthy> (Parathyroid)
<pan> (Endocrine Pancreas)
<crt-bdy> (Carotid Body)
<intg> (Integumentary System)
<skn> (Skin)
<appdg> (Appendage)
<hr> (Hair)
<nls> (Nail)
<sub> (Subcutaneous Tissue)

 (Breast)
<nppl> (Nipple)
<mss> (Musculoskeletal)
<msl> (Muscle)
<bn> (Bone)
<crt> (Cartilage)
<jnt> (Joint)
<dsc> (Intervertebral Disc)
<mnsc> (Meniscus)
<tn> (Tendon)
<lig> (Ligament)
<brs> (Bursa)
<fsc> (Fascia)
<sft> (Soft Tissue)
<interstit> (Interstitium)
<hm> (Hematopoietic System)
<r-i> (Reticuloendothelial/Immune System)
<lym-tis> (Lymphoid Tissue)
<wbc> (WBC)
<spleen> (Spleen)
<cg> (Coagulation System)
<rbc> (RBC)
<plt> (Platelet)
<marrow> (Bone Marrow)
<plc> (Products of Conception)
<elmic> (Ultrastructural anatomy)

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)

<itrl> (Infratentorial Region)
<b-o-s> (Base of the Skull)
<ps-fs> (Posterior Fossa)
<strl> (Supratentorial Region)
<pt-fs> (Pituitary Fossa)
<md-fs> (Middle Fossa)
<at-fs> (Anterior Fossa)
<jw> (Jaw)
<chn> (Chin)
<bcl> (Buccal Region)
<orl> (Oral Region)
<nsl> (Nasal Region)
<zyg> (Zygomatic Region)
<infr> (Infraorbital Region)
<supr> (Supraorbital Region)
<fc> (Face)
<frh> (Forehead)
<sbtl> (Suboccipital Region)
<octl> (Occipital Region)
<aur> (Auricular Region)
<tprl> (Temporal Region)
<prtl> (Parietal Region)
<frtl> (Frontal Region)
<sclp> (Scalp)
<orb-reg> (Orbital Region)
<mastoid> (Mastoid Region)
<parotid> (Parotid Region)
<nk> (Neck)
<a-n> (Anterior Neck)
<p-n> (Posterior Neck)
<supra> (Supraclavicular Region)
<spn> (Spine)
<c-s> (Cervical Spine)
<t-s> (Thoracic Spine)
<l-s> (Lumbar Spine)
<sacr-s> (Sacral Spine)
<cocc-s> (Coccygeal Spine)
<thor-lumb> (Thoracolumbar Spine)
<lumb-sac> (Lumbosacral Spine)
<tk> (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)
<fgr> (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/Malig)
<metast> (Metastatic)
<inflam> (Inflammatory Process)
<degen> (Degenerative Process)
<mort> (Mortality)
<ischem> (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)
<glauc> (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)
            <aminoquin> (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)

<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 acti)
<antiadr-agt-cent> (antiadrenergic agents, centrally acting)
<antianginal> (antianginal agents)
<antiarrrh> (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)

<vhtiiircp-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)

<gi-stim> (GI stimulants)
<htwo-antag> (H2 antagonists)
<laxat> (laxatives)
<misc-gi-agt> (miscellaneous GI agents)
<prot-pmp-inh> (proton pump inhibitors)
<aminosal> (5-aminosalicylates)
<ch-cl-hrm> (hormones)
<adrcort-ster> (adrenal cortical steroids)
<antidiab> (antidiabetic agents)
<sulfonurea> (sulfonylureas)
<non-sulfonurea> (non-sulfonylureas)
<insulin> (insulin)
<a-glusidase-inh> (alpha-glucosidase inhibitors)
<thzolidin> (thiazolidinediones)
<misc-hrm> (miscellaneous hormones)
<sex-hrm> (sex hormones)
<oral-contracept> (oral contraceptives)
<androg-anab-ster> (androgens and anabolic steroids)
<estrog> (estrogens)
<gonadtrp> (gonadotropins)
<progest> (progestins)
<sex-hrm-comb> (sex hormone combinations)
<misc-sex-hrm> (miscellaneous sex hormones)
<gonad-rel-hrm> (gonadotropin releasing hormones)
<thy-drg> (thyroid drugs)
<biphos> (bisphosphonates)
<misc-agt> (miscellaneous agents)
<antidote> (antidotes)
<chel-agt> (chelating agents)
<chol-musc-stim> (cholinergic muscle stimulants)
<loc-inject-anes> (local injectable anesthetics)
<misc-uncat-agt> (miscellaneous uncategorized agents)
<psoralen> (psoralens)
<radiocon-agt> (radiocontrast agents)
<gu-tr-agt> (genitourinary tract agents)
<urin-spasm> (urinary antispasmodics)
<urin-ph-mod> (urinary pH modifiers)
<misc-gu-agt> (miscellaneous genitourinary tract agents)
<illicit-dr> (illicit (street); drugs)
<antirheum> (antirheumatics)
<antigout> (antigout agents)
<impot-agt> (impotence agents)
<antipsor> (antipsoriatics)
<nutrit-prod> (nutritional products)
<iron-prod> (iron products)
<mineral-elect> (minerals and electrolytes)

<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)
<antihist> (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 antipsoriatics)
<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-antihist-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)
<nsl-prep> (nasal preparations)
<nsl-lub-irrig> (nasal lubricants and irrigations)
<nsl-ster> (nasal steroids)
<nsl-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)
<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)
<panel> (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)
<torch> (TORCH)
<thyr-panel> (Thyroid Panel)
<thyr-panel-tsh> (Thyroid Panel with TSH)
<blood-gas> (Blood Gases)
<trans> (Transfusion)
<micro> (Microbiology)
<tiss-type> (Tissue Typing)
<evo-supp> (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)
<dialysis> (Dialysis)
<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. **<rsk-fct>** (Risk Factor)
27. **<cntra>** (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. **<cmplic>** (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)

<chg-same> (Change-Same)

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. **<or>** (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 Enviroment)
    <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)
    - <vhv> (Verb Have)
    - <vtst> (Vtest)
    - <vbe> (Verb Be)
    - <bcnn> (Beconn)
    - <brp> (Be Representative)
    - <vdo> (Verb Do)
57. **<prnn>** (Pronouns)
58. **<evnt>** (Event)
59. **<trnsp>** (Transparent Term)
60. **<epnym>** (Eponym)
61. **<null>** (Null)
62. **<p-p>** (Parse Phrase)
63. **<pt-sp>** (Specimen from Patient)
64. **<q>** (Question)

TABLE 2
PARTS OF SPEECH
MAJOR LEXICAL CATEGORIES

Category	Description	Note
ADJ	Adjective	
CS0	Subordinate conjunction in SUB0	
CS1	Subordinate conjunction in SUB1	
CS2	Subordinate conjunction in SUB2	
CS3	Subordinate conjunction in SUB3	
CS4	Subordinate conjunction in SUB4	*not used in MLP*
CS5	Subordinate conjunction in SUB5	
CS6	Subordinate conjunction in SUB6	
CS7	Subordinate conjunction in SUB7	*not used in MLP*
CS8	Subordinate conjunction in SUB8	*not used in MLP*
CS9	Subordinate conjunction in SUB9	
D	Adverb	
DP	Adverbial Particle in DP object strings	
DS	Dose string, treated as a unit	
DT	Date in numerical form dd/mm/yy	
N	Noun	
NS	Possessive Noun	
NULL	Parser “pass”: No word required; Adjunct	
NULLC	Parser “pass”: No word required; Conjunct	
NULLFRAG	Parser “pass”: No word required; Fragment	
NULLN	Parser “pass”: No word required; Noun	*not used in MLP*
NULLOBJ	Parser “pass”: No word required; Object	
NULLRECIP	Parser “pass”: No word required; Reciprocals	*not used in MLP*
NULLWH	Parser “pass”: No word required; WH-strings	
P	Preposition	
PRO	Pronoun	
Q	Quantity	
T	Article	
TV	Tensed Verb	
V	Untensed Verb (infinitive)	
VEN	Past Participle	
VING	Present Participle	
W	Auxiliary (*will*, *can*, *may*, *should*, etc.)	

TABLE 3
LIST OF SUBORDINATE CONJUNCTION STRINGS

	DESCRIPTION	EXAMPLE
SUB0	Subordinate conjunction + Object of *'be'*	*when conscious of my observation*
SUB1	Subordinate conjunction + Assertion	*because her blood pressure has been well controlled*
SUB1-PHRASE	Subordinate conjunction phrase	He is responsive *in that* he would open his eyes
SUB2	Subordinate conjunction + VENPASS (passive verb with its passive object)	*since last documented*
SUB3	Subordinate conjunction + VINGO (V*ing* form of verb + Object)	*since being on Prednisone*
SUB4	Subordinate conjunction + VING string (either VINGOFN or NSVINGO)	
SUB5	Subordinate conjunction + SVINGO (Subject + V*ing* form of verb + Object)	*without anyone noticing it*
SUB6	Subordinate conjunction + Subject + Object of *'be'*	*with noone on the premises*
SUB7	Subordinate conjunction + SVEN	
SUB8	Subordinate conjunction (*as*) + partial inverted Assertion	*as were the other medications*
SUB9	Subordinate conjunction + infinitive verb + Object	*in order to break this cycle; in order to gain control*
SUB11	Time phrase (TM-PHRASE) + Assertion	*during which time* he remained unconscious
SUB12	*'should'* + tenseless Assertion	instruct to call *should she have any fevers*
SUB13	SUB1-PHRASE + Assertion	He is responsive *in that he would open his eyes*

TABLE 4
MLP SYNTACTIC MEDICAL SUBLANGUAGE CLASSES

	MLP CLASS	DESCRIPTION
PATIENT AREA		
H-PT	words referring to patient	*Patient complained of diarrhea.*
H-PTAREA	anatomical area	*Lesion is located at the base of the lung.*
H-PTDESCR	occupation, nationality, traits	*Patient is affable. Patient is a carpenter.*
H-PTFUNC	physiological function	*Appetite poor. Patient has a full range of motion of left shoulder.*
H-PTLOC	anatomical location relation	*Fracture radiates from left shoulder.*
H-PTMEAS	anatomical measure	*Weight 127 lbs. Normal cardiac size.*
H-PTPALP	palpated body part	*No liver.*
H-PTPART	body part	*Patient has metacarpal stiffness.*
H-PTSPEC	specimen type	*Right knee shavings.*
H-PTVERB	verb with patient subject	*Patient suffered a myocardial infarction.*
TEST/EXAM AREA		
H-OBSERVE	verbs of observation	*No ataxia was detected on finger-to-nose testing.*
H-TXCLIN	clinical exam, action	*Palpation of abdomen revealed tenderness. Kernig's sign positive.*
H-TXPROC	examination procedure	*MRI shows a grade 1 spinal listhesis.*
H-TXSPEC	test of specimen	*Blood chemistries showed electrolytes normal.*
H-TXVAR	test variable	*Acetone is normal. WBC 53.*
TREATMENT AREA		
H-DEVMED	medical devices	*Pt has a pacemaker. Pt uses a walker.*
H-INST	hospitals, clinics, physicians, staff	*Pt seen in ER. Pt will see a cardiologist.*
H-TTCOMP	complementary treatments	*Improved on bedrest.*
H-TTGEN	general medical management	*Pt returned to hospital. Discharge of patient from ZZZ hospital.*
H-TTMED	treatment by medication	*Doctor prescribed ampicillin.*
H-TTMODE	mode of administration	*intramuscular*
H-TTSURG	surgical procedure	*Excision of cataract in left eye.*
TIME		
H-POST	support class for prefix morpheme *post-*	*Postprandial discomfort.*
H-TMBEG	beginning	*Treatment was begun. Fever developed.*
H-TMEND	termination	*Fever disappeared. Medication was discontinued.*
H-TMLOC	location in time	*During the previous week…*
H-TMDUR	duration	*persistent cough. Episode of meningitis.*
H-TMPREP	time preposition	*After last week… During admission…*
H-TMREP	repetition	*Headache recurring. Second admission.*

FINDINGS AREA		
H-AMT	amount or degree	*Full range of motion. Severe pain.*
H-CHANGE	indication of change	*alterations in management.*
H-CHANGE-LESS	decrease	*Ampicillin tapered from 500 mg to 300 mg qid.*
H-CHANGE-MORE	increase	*a rise in his Creatinine to 2.3.* *weight gain*
H-CHANGE-SAME	no change	*Cardiac silhouette remains within normal limits. Neurologic diagnoses are static.*
H-DIAG	diagnosis	*Patient has glaucoma.*
H-INDIC	disease indicator word	*Pt complained of tenderness in joints.*
H-NORMAL	non-problematical	*Temperature normal. Growth and development – within normal limits.*
H-ORG	organism	*CSF shows H. Influenzae type B.*
H-RESP	patient response	*His blood gas improved on 35% oxygen.*
H-TXRES	test/exam result word	*Positive Kernig's. Gram stain negative.*
H-DESCR	descriptive information	*A punctuate dark lesion.*
H-DIET	name of foods, dietary prescriptions	*Allergic to tomatoes.*
EVIDENTIAL AREA		
H-NEG	negation of finding	*Patient refused medication. Patient never had headaches.*
H-MODAL	uncertainty of finding; event has not yet occurred	*Rule out meningitis. Doctor advised surgery.*
CONNECTIVE AREA		
H-BECONN	classifier verb	*Treatment consisted of ampicillin.*
H-CONN	connects 2 information units	*Fever caused the headache.*
H-SHOW	connects test and result	*Echo revealed apical akinesis.*
H-VTEST	lab or measurement verbs	*HCT ranged from 24 to 35.*
OTHER SEMANTIC CLASSES		
H-AGE	life stage	*adult, adolescence, years old, age*
VBE	verb '*be*'	
BEREP	*be* replacer verbs	*She becomes drowsy.*
VDO	verb *'do'*	
VHAVE	verb *'have'*	
H-FAMILY	family member or friend	*Brother also has sickle cell disease.*
H-GEOGR	geographical names	*Pt currently lives in San Francisco.*
H-ETHNIC	racial or ethnic background	*white male, African American female*
H-RECORD	patient record terms	*Review of health records showed PPD of 1972 was also positive.*
H-TESTVIEW	angles or positions of imaging	*lateral, axial, PA*
H-TRANSP	general terms that get classes from its modifiers	*a tonic, clonic type of convulsion.* *The list of medications...* *The state of the disease...*

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.

MLP Lexicon Classes		**XMK Lexicon Tags**	
MLP Class	***Part of Speech***	***Primary***	***Support***
	DS [dose string]	med-ds	
	DT [date]	tm	tm_exact
	T [English article]	null	
H-BECONN	V, TV, VEN, VING, P, ADJ, N:SI, N:PL	li	li_bcnn
BEREP	V, TV, VEN, VING	li	li_brp
H-CHANGE	*various parts of speech*	md	md_chg
H-CHANGE-LESS	*various parts of speech*	md	md_chg-less
H-CHANGE-MORE	*various parts of speech*	md	md_chg-more
H-CHANGE-SAME	*various parts of speech*	md	md_chg-same
H-CHEM	*various parts of speech*	ch	
H-CONN	*various parts of speech*	li	li_cnn
H-DEVMED	ADJ, D, N, N:SI, N:PL	dev	dev_dev-med
H-ETHNIC	ADJ, N:SI, N:PL	dem	dem_race-eth
H-GEOGR	ADJ, N:SI, N:PL	env	env_geo
H-MODAL	*various parts of speech*	md	md_modal
H-NEG	*various parts of speech*	md	md_ng
H-OBSERVE	V, TV, VEN, VING, ADJ, N:SI, N:PL	li	li_obs
H-PTVERB	V, TV, VEN, VING, N:SI, N:PL	li	li_ptv
H-RESP	*various parts of speech*	tr-rsp	
H-SHOW	V, TV, VEN, VING, P, ADJ, N:SI, N:PL	li	li_shw
H-TMBEG	*various parts of speech*	tm	tm_beg
H-TMDUR	*various parts of speech*	tm	tm_dur

H-TMEND	*various parts of speech*	tm	tm_end
H-TMLOC	*various parts of speech*	tm	tm_tm-loc
H-TMPREP	P	tm	tm_prp
class other than H-TMREP	P	li	li_prp
H-TMREP	*various parts of speech*	tm	tm_rep
H-VTENSE [PAST]	GRAM-NODE	tense	tm_tense_past
H-VTENSE [PRESENT]	GRAM-NODE	tense	tm_tense_present
H-VTENSE [PERF]	GRAM-NODE	tense	tm_tense_perf
H-VTENSE [PROG]	GRAM-NODE	tense	tm_tense_prog
H-VTENSE [FUT]	GRAM-NODE	tense	tm_tense_future
H-VTENSE [FUT-IMP]	GRAM-NODE	tense	tm_tense_fut-imp
NTIME1	N, N:SI, N:PL, ADJ	tm	tm_tm-loc
NTIME2	N, N:SI, N:PL	tm	tm_tm-loc
NUNIT	N, N:SI, N:PL, TV, VING, ADJ	unit	
QNUMBER	Q	num	
VBE	V, TV, VEN, VING	li	li_vbe
VDO	W	li	li_vdo
VHAVE	V, TV, VEN, VING	li	li_vhv