

---

# Ambiguity in the UMLS<sup>®</sup> Metathesaurus<sup>®</sup>

2000 Edition

**Alan R. Aronson**

February 24, 2000

## 1. Introduction

The number of explicitly ambiguous Metathesaurus terms in the 2000 edition has grown somewhat over the 1999 edition. Table 1 shows several counts that indicate that growth in broad terms.

	<b>1999</b>	<b>2000</b>	<b>% change</b>
<b>Strings with an ambiguity designator</b>	7,912	9,416	+19%
<b>Concepts with one or more ambiguity</b>	6,565	7,409	+13%
<b>Cases of ambiguity</b>	3,669	4,361	+19%

**Table 1. Measures of ambiguity in the UMLS Metathesaurus**

Some examples will clarify the meaning of the counts. There are 54 Metathesaurus strings ‘Other <n>’ for n ranging from 1 to 54; these strings occur in 54 distinct concepts but represent a single case of ambiguity. Some concepts contain more than one ambiguous string. In fact, the concept ‘Optic Nerve Glioma, Childhood’ has 37 ambiguous strings. A more manageable example of a concept with multiple ambiguities is ‘Arthrogyrosis’ which has the following six ambiguous strings:

- Amyoplasia congenita <1>
- Congenital Arthromyodysplasia <1>
- Congenital multiple arthrogyrosis <1>
- myodystrophia fetalis deformans <2>
- myodystrophia foetalis deformans <2>
- pterygium universale <2>

All but the last string are ambiguous with strings belonging to the concept ‘Amyoplasia congenita disruptive sequence’. The concept containing ‘pterygium universale <1>’, however, is ‘Multiple pterygium syndrome’.

The information in Table 1 shows that the amount of ambiguity in the Metathesaurus increased almost 20% from 1999 to 2000 and that several concepts involved in the new ambiguities have more than one ambiguity. Examining the cases of ambiguity more closely, consider the *degree* of ambiguity, i.e., the number of ways a string is ambiguous. ‘Other <n>’, for example, has degree 54. Table 2 contains the distribution of ambiguities in the Metathesaurus according to degree.

Degree of ambiguity	1999 cases	2000 cases	% change
54	1	1	-
23	1	1	-
18	1	1	-
16	2	2	-
8	3	3	-
7	2	2	-
6	1	2	+100%
5	9	9	-
4	52	71	+37%
3	321	403	+26%
2	3,249	3,835	+18%
1	27	31	+15%
<b>Total</b>	<b>3,669</b>	<b>4,361</b>	<b>+19%</b>

**Table 2. Metathesaurus ambiguity distribution by degree**

Note that an ambiguity of degree one is not actually an ambiguity. ‘Abbreviations <1>’, for example, is not ambiguous since there are no other ‘Abbreviations <n>’ strings in the Metathesaurus. The most important thing to note about the table is that changes in ambiguity are focused on ambiguities of degree four or less and mainly of degree two. At least for 2000 there is no proliferation of the problematic ambiguity of high degree.

The purpose of this study is to examine the growth in ambiguity to be able to characterize its effect on text processing applications. Section 2 reviews the highly ambiguous cases discovered in the 1999 edition of this report. Section 3 examines the effect of removing suppressible synonyms from the study. Section 4 focuses on the cases of ambiguity of lesser degree. And Section 5 contains a summary of the results and recommendations for handling ambiguity in the Metathesaurus.

## 2. Review of the Problem

The UMLS Metathesaurus denotes explicit ambiguity by appending an ambiguity designator, a number in angle brackets, to the end of an ambiguous string. Because strings with ambiguity designators require significant special processing, the original string (with possible case changes) is always included as another string for the concept involved. Thus each concept with a string ‘aaa

<n>' also has string 'aaa' (ignoring case differences). But now any application which gains access to the Metathesaurus textually, cannot tell the difference among the concepts of an ambiguous cluster 'aaa <n>' each of which is represented by 'aaa'.

English Metathesaurus strings with the largest degree of ambiguity are listed below in decreasing order. Bold text indicates the difference between the 1999 and 2000 Metathesaurus. The differences for ambiguities of degree 5 or more are that 'Cold <n>' and 'Stomach <n>' each got an additional ambiguity. They are described in detail below.

- 54 'Other <n>'
- 23 'Protocols <n>'
- 18 'Patient Education Plans <n>'
- 16 'Assessment <n>'
- 16 'Limited function/disability <n>'
- 8 'Driver injured in collision with other and unspecified motor vehicles in nontraffic accident <n>'
- 8 'Driver injured in collision with other and unspecified motor vehicles in traffic accident <n>'
- 8 'cde genotype <n>'
- 7 'Passenger injured in collision with other and unspecified motor vehicles in nontraffic accident <n>'
- 7 'Passenger injured in collision with other and unspecified motor vehicles in traffic accident <n>'
- 6 'Other injuries <n>'
- **6** '**Cold <n>**' (was of degree 5)
- 2 'FEELING COLD <n>' (just <4> and <5>)
- 5 'Injuries <n>'
- 5 'Premolar tooth <n>'
- **5** '**Stomach <n>**' (was of degree 4)
- 5 'Tourniquet on <n>'
- 5 '[SO] Premolar tooth <n>'
- 5 'adjustment <n>'
- 5 'cd <n>'
- 5 'prostate <n>'
- 3 'sound measurement <n>' (just <3>, <4> and <5>)
- 5 'urethra <n>'
- 4 'Cl <n>'
- 4 'conjunctiva <n>'
- 4 'cornea <n>'
- 2 'Dihydropyridine <n>' (just <3> and <4>)
- 4 'ET - Esotropia <n>'
- 4 'HRF <n>'
- 4 'U <n>'
- 4 'lupus <n>'
- 4 'ms <n>'
- ...

## 2.1 ‘Cold <n>’

The concepts associated with the six senses of ‘Cold’ are

- ‘cold temperature’
- ‘Common Cold’
- ‘Cold Therapy’
- ‘Chronic Obstructive Airway Disease’ (which has strings ‘COLD’ and ‘Chronic Obstructive Lung Disease’)
- ‘Cold Sensation’
- **‘Cold brand of chlorpheniramine-phenylpropanolamine’**

The new sense is shown in bold and comes from the Multum MediSource Lexicon (MMSL99). In 1999 I argued that all five senses of cold were legitimate. COLD was considered a valid sense because the Metathesaurus does not yet distinguish case. Research on word sense disambiguation has made me rethink my position. I now believe that ‘cold temperature’, ‘Common Cold’ and ‘Cold Sensation’ are legitimate senses of cold. ‘Cold Therapy’ is not a legitimate sense of cold. The word *cold* does not convey enough of the meaning of ‘Cold Therapy’. Equivalently, it is difficult to construct a reasonable sentence containing the word *cold* without the word *therapy* but with the ‘Cold Therapy’ sense. COLD is also not legitimate and can be treated properly by any system which recognizes case differences. Finally, the new sense, ‘Cold brand of chlorpheniramine-phenylpropanolamine’ is not a legitimate sense for the same reason that ‘Cold Therapy’ is not.

## 2.2 ‘Stomach <n>’

The concepts associated with the five senses of ‘Stomach’ are

- ‘Stomach’
- ‘Benign neoplasm of stomach’
- ‘Carcinoma in situ of stomach’
- ‘Neoplasm of uncertain or unknown behavior of stomach’
- ‘Stomach problem’

Only the first concept, ‘Stomach’ is a legitimate sense of the word *stomach*. Appropriately, the ‘Stomach <n>’ and ‘Stomach’ strings associated with the remaining concepts are marked as suppressible synonyms.

# 3. Removing Suppressible Synonyms

Each Metathesaurus string is assigned a term status: P for preferred terms, S for synonyms, and s for suppressible synonyms. Suppressible synonyms are shortened forms of other terms (see sections 2.1 and 2.2 above for examples) and are best ignored for text processing of the Metathesaurus. The 2000 edition of the Metathesaurus has 1,018 suppressible synonyms, 9% more than the 932 such strings in the 1999 edition. Removing suppressible synonyms from the Meta-

thesaurus produces the ambiguity distribution shown in Table 3. There are now 4,139 cases of

Degree of ambiguity	Cases before removal	Cases after removal
54	1	0
23	1	1
18	1	1
16	2	1
8	3	3
7	2	2
6	2	1
5	9	6
4	71	34
3	403	303
2	3,835	3,435
1	31	352
<b>Total</b>	<b>4,361</b>	<b>4,139</b>

**Table 3. Ambiguity distribution before and after removal of suppressible synonyms**

ambiguity, 222 fewer (5% less) than before. While this is not a large difference numerically, it does represent a significant reduction in the high ambiguity cases. The list of cases after removal of suppressible synonyms is given below where differences from the previous list are shown in bold.

- **5** ‘Other <n>’ (previously 54)
- 23 ‘Protocols <n>’
- 18 ‘Patient Education Plans <n>’
- 16 ‘Assessment <n>’
- **0** ‘Limited function/disability <n>’ (previously 16)
- 8 ‘Driver injured in collision with other and unspecified motor vehicles in nontraffic accident <n>’
- 8 ‘Driver injured in collision with other and unspecified motor vehicles in traffic accident <n>’
- 8 ‘cde genotype <n>’
- 7 ‘Passenger injured in collision with other and unspecified motor vehicles in nontraffic accident <n>’
- 7 ‘Passenger injured in collision with other and unspecified motor vehicles in traffic accident <n>’
- **0** ‘Other injuries <n>’ (previously 6)
- 6 ‘Cold <n>’
- 2 ‘FEELING COLD <n>’ (just <4> and <5>)
- **2** ‘Injuries <n>’ (previously 5)
- 5 ‘Premolar tooth <n>’

- 1 **‘Stomach <n>’** (previously 5)
- 5 ‘Tourniquet on <n>’
- 5 ‘[SO] Premolar tooth <n>’
- 5 ‘adjustment <n>’
- 5 ‘cd <n>’
- 2 **‘prostate <n>’** (previously 5)
- 3 ‘sound measurement <n>’ (just <3>, <4> and <5>)
- 2 **‘urethra <n>’** (previously 5)
- 4 ‘Cl <n>’
- 2 **‘conjunctiva <n>’** (previously 4)
- 2 **‘cornea <n>’** (previously 4)
- 2 ‘Dihydropyridine <n>’ (just <3> and <4>)
- 4 ‘ET - Esotropia <n>’
- 4 ‘HRF <n>’
- 4 ‘U <n>’
- 4 ‘lupus <n>’
- 4 ‘ms <n>’
- ...

The five remaining senses of ‘Other <n>’ are

- ‘Other’ (<1>)
- ‘Other location of complaint’ (<2>)
- ‘Other activities involving preparation of a routine non-injectable drug product’ (<52>)
- ‘Other activities involving preparation of compounded non-parenteral medications’ (<53>)
- ‘Other activities involving preparation of compounded parenteral medications’ (<54>)

All but the first sense need to be suppressed.

The two remaining senses of ‘Injuries <n>’ are

- ‘Physical trauma’ (<1>)
- ‘Injury inflicted to the body by an external force’ (<2>)

Although the senses which were clearly wrong have been removed, the remaining senses seem to be synonymous (and the first sense is a MeSH subheading).

The two remaining senses of ‘prostate <n>’ are

- ‘Prostate’ (<1>)
- ‘Prostatic Diseases’ (<5>)

The second sense needs to be suppressed since a prostatic disease is not a prostate just as a stomach problem is not a stomach. Similar analyses apply to ‘urethra <n>’, ‘conjunctiva <n>’ and ‘cornea <n>’ each of which have an erroneous disease sense in addition to the correct body part sense.

Finally, all of the senses of the following high degree cases also need to be suppressed

- 23 ‘Protocols <n>’
- 18 ‘Patient Education Plans <n>’
- 16 ‘Assessment <n>’

- 8 ‘Driver injured in collision with other and unspecified motor vehicles in nontraffic accident <n>’
- 8 ‘Driver injured in collision with other and unspecified motor vehicles in traffic accident <n>’
- 8 ‘cde genotype <n>’
- 7 ‘Passenger injured in collision with other and unspecified motor vehicles in nontraffic accident <n>’
- 7 ‘Passenger injured in collision with other and unspecified motor vehicles in traffic accident <n>’
- 5 ‘Premolar tooth <n>’
- 5 ‘Tourniquet on <n>’
- 5 ‘[SO] Premolar tooth <n>’

## 4. Ambiguity of Lesser Degree

This section explores all 34 cases of ambiguity of degree 4 and 50 randomly selected cases of ambiguity of degree 3 or 2. (As Table 3 indicates, there are 303 cases of degree 3 and 3,435 cases of degree 2.)

### 4.1 Ambiguity of degree 4

#### 4.1.1 ‘AD <n>’

- ‘Alzheimer’s Disease’
- ‘Dactinomycin’ (which has string ‘Actinomycin D’)
- ‘Anterodorsal nucleus of thalamus’
- ‘Admitting diagnosis’

AD is an acronym in each case; suppress them.

#### 4.1.2 ‘Aspiration <n>’

- ‘Breathing’
- ‘Endotracheal aspiration’
- ‘Aspiration-action’
- ‘Pulmonary aspiration’

The two senses defined in Dorland seem to be represented by ‘Breathing’ and ‘Aspiration-action’. The other two senses may be too specific and should probably be suppressed.

#### 4.1.3 ‘Blood Pressure <n>’

- ‘Blood Pressure’ (senses <1> and <3>; Organism Function)
- ‘Blood pressure determination’ (Diagnostic Procedure)
- ‘Arterial pressure’ (Laboratory or Test Result)

It seems like there should only be two senses, the organism function (‘Blood Pressure’) and the diagnostic procedure/lab result (the other two concepts). This would argue for suppressing one of ‘Blood pressure determination’ or ‘Arterial pressure’, but it is not clear which one to choose.

## 4.1.4 'CAM &lt;n&gt;'

- 'Cell Adhesion Molecules'
- 'chorioallantoic membrane'
- 'CAM brand of Ephedrine Hydrochloride'
- 'Cam, topical lotion'

The first two senses are acronyms and should be suppressed. The third sense should probably be suppressed as being too specific. And the fourth sense might be legitimate.

## 4.1.5 'Cl &lt;n&gt;'

- 'Chlorine' ('Cl <4>')
- 'Cycloleucine' ('CL <2>')
- 'centiliter' ('cL <1>')
- 'Chloride Ion' ('CL <3>')

Although most of the senses can be distinguished because of case differences, I'm inclined to suppress them all because they are acronyms/abbreviations.

## 4.1.6 'COPE &lt;n&gt;'

- 'cisplatin/cyclophosphamide/etoposide/vincristine'
- 'COPE <1>' (which has only one other string, 'COPE')
- 'Cope brand of aspirin-caffeine'
- 'COPE <3>' (which also has only one other string, 'COPE')

'COPE <1>' and 'COPE <3>' appear to have incomplete representations in the Metathesaurus, and the third sense is another brand. Thus everything except the first sense could be suppressed.

## 4.1.7 'CVP &lt;n&gt;'

- 'cyclophosphamide/prednisone/vincristine'
- 'Measurement of central venous pressure'
- 'cisplatin/cyclophosphamide/etoposide'
- 'Central Venous Pressure-Biological function'

The third sense seems odd; but otherwise, they can all be suppressed as acronyms.

## 4.1.8 'Dandruff &lt;n&gt;'

- 'Seborrheic dermatitis of scalp' (Disease or Syndrome)
- 'Scurfiness of scalp' (Finding)
- 'Pityriasis simplex' (Disease or Syndrome)
- 'Dandruff brand of pyrithione zinc' (Pharmacologic Substance)

The brand name should be suppressed. Everything else seems reasonable.

## 4.1.9 'Dressing &lt;n&gt;'

- 'Sterile coverings' (Medical Device)
- 'Clothing assistance' (Health Care Activity)
- 'Dressing of skin or wound' (Therapeutic or Preventive Procedure)
- 'Dressing self-care' (Finding)

All senses seem to be legitimate.

## 4.1.10 'ET - Esotropia &lt;n&gt;'

- 'Esotropia'
- 'Manifest alternating convergent squint'
- 'Intermittent convergent squint'
- 'Incomitant esotropia'

All senses except the first seem to be findings associated with the first sense and could be suppressed.

## 4.1.11 'Evaluation &lt;n&gt;'

- 'Health evaluation' (<2>)
- 'Evaluation' (<1> and <3>)
- 'Pulmonary evaluation' (<4>)

The third sense is too specific and should be suppressed.

## 4.1.12 'Fire &lt;n&gt;'

- 'Accident caused by unspecified fire' (Injury or Poisoning)
- 'Fire - disasters' (Phenomenon or Process)
- 'Fire - physical phenomenon' (Phenomenon or Process)
- 'Fire as a heat source' (Human-caused Phenomenon or Process)

The first sense is a result of a fire, not a fire, itself; it should be suppressed.

## 4.1.13 'Grafts &lt;n&gt;'

- 'transplantation' (<2> and <3>; Functional Concept)
- 'Graft material' (<4>; Biomedical or Dental Material)
- 'Homologous Grafts' (<1>; Body Part, Organ, or Organ Component)

The second sense is used in a graft but is not, itself, a graft. It should be suppressed. The third sense may or may not be too specific. See also 'Graft <n>' which is three ways ambiguous involving the same concepts.

## 4.1.14 'HRF &lt;n&gt;'

- 'Hypothalamic Releasing Factor'
- 'Gonadorelin' (which has string 'FSH-Releasing Hormone')
- 'homologous restriction factor'
- 'HRF <3>'

The fourth sense is not completely defined, but all of the senses can be suppressed as acronyms.

## 4.1.15 'Lupus &lt;n&gt;'

- 'Lupus Vulgaris'
- 'Lupus Erythematosus, Discoid'
- 'Lupus Erythematosus, Systemic'
- 'Lupus Erythematosus'

As in the previous edition, the four senses seem to be legitimate although the second and third seem to be specific types of sense four and as such could be suppressed.

## 4.1.16 'Mole &lt;n&gt;'

- 'Mole the mammal'
- 'Nevus, Pigmented'
- 'Benign melanocytic nevus of skin'
- 'Mole, unit of measurement'

'Nevus, Pigmented' is a kind of 'Benign melanocytic nevus of skin' and could be suppressed.

## 4.1.17 'Morphology &lt;n&gt;'

- 'Morphology, NOS' (Finding)
- 'morphological' (Biomedical Occupation or Discipline)
- 'morphology <4>' (which has strings 'physical form' and 'physical shape'; Natural Phenomenon or Process)
- 'Science of Morphology' (Biomedical Occupation or Discipline)

Based on the semantic types of the sense concepts, the second and fourth senses seem to be the same thing; one of them should be suppressed. In addition, the difference between the Finding and Natural Phenomenon or Process senses is small but probably legitimate.

## 4.1.18 'Mosaic &lt;n&gt;'

- 'Mosaicism' (Organism Attribute)
- 'Embryonic Mosaic' (Cell or Molecular Dysfunction)
- 'Spatial Mosaic' (Spatial Concept)
- 'Mosaic - computer software' (Intellectual Product)

All senses seem legitimate although the first two are closely related.

## 4.1.19 'MS &lt;n&gt;'

- 'Mitral Valve Stenosis' ('ms <4>')!
- 'Multiple Sclerosis' ('MS <3>')
- 'Morphine Sulfate' ('MS <2>')
- 'millisecond' ('ms <1>')

The first three senses are acronyms and should be suppressed. The last sense, 'millisecond', is an example of an ubiquitous acronym (or abbreviation), that is one which almost always appears in text without definition. Ubiquitous acronyms can be left unsuppressed, but they should also be explicitly accumulated into a list for text processing purposes.

## 4.1.20 'Nutrition &lt;n&gt;'

- 'Science of nutrition' (Biomedical Occupation or Discipline)
- 'Nutritional status' (Organism Attribute)
- 'Nutrition outcomes' (Intellectual Product)
- 'Feeding and dietary regimes' (Therapeutic or Preventive Procedure)

Although senses two and three seem very close in meaning, all senses are probably reasonable.

## 4.1.21 'Orbital Prosthesis &lt;n&gt;'

- 'Ocular Prosthesis' (<1>)
- 'Orbital prosthesis implantation' (<2> and <4>)
- 'External orbital prosthesis' (<3>)

The second and third senses should be suppressed; the second is a procedure involving an orbital prosthesis, and the third is too specific.

#### 4.1.22 'P <n>'

- 'Phosphorus'
- 'Properdin'
- 'upper case pea'
- 'lower case pea'

The first two cases are acronyms/abbreviations, and the last two are self referential. They should all be suppressed.

#### 4.1.23 'Peripheral Neuroectodermal Tumor <n>'

- 'Neuroectodermal Tumor, Peripheral'
- 'Extraosseous Ewings sarcoma-primitive neuroepithelial tumor'
- 'Ewings sarcoma'
- 'Ewings sarcoma-primitive neuroectodermal tumor (PNET)'

All senses but the first are more specific than the first and should be suppressed.

#### 4.1.24 'PNET <n>'

- 'Neuroectodermal Tumor, Primitive'
- 'Extraosseous Ewings sarcoma-primitive neuroepithelial tumor'
- 'Ewings sarcoma'
- 'Ewings sarcoma-primitive neuroectodermal tumor (PNET)'

This case is identical to the previous one; suppress all senses except the first.

#### 4.1.25 'Pressure <n>'

- 'Pressure- physical agent' (Phenomenon or Process)
- 'Baresthesia' (Organ or Tissue Function)
- 'Pressure' (Functional Concept)
- 'Pressure - action' (Therapeutic or Preventive Procedure)

All senses seem reasonable.

#### 4.1.26 primitive neuroectodermal tumor

- 'Neuroectodermal Tumor, Primitive'
- 'Extraosseous Ewings sarcoma-primitive neuroepithelial tumor'
- 'Ewings sarcoma'
- 'Ewings sarcoma-primitive neuroectodermal tumor (PNET)'

Except for replacing the word *peripheral* with *primitive*, this is exactly the same as example

4.1.23. All senses except the first should be suppressed.

#### 4.1.27 'Relapse <n>'

- 'Relapse'
- 'Relapsing course'
- 'Relapse phase'
- 'Cancer Relapse'

All senses except for the first are too specific and should be suppressed.

## 4.1.28 ‘Screening &lt;n&gt;’

- ‘Screening for cancer’ (Health Care Activity)
- ‘Screening procedure’ (Diagnostic Procedure)
- ‘Aspects of disease screening’ (Intellectual Product)
- ‘Screening for unspecified condition’ (Finding)

It may be that the fourth sense should have semantic type Diagnostic Procedure in which case it seems synonymous with sense two. Meanwhile, it could be suppressed. Senses one and three are too specific and should be suppressed.

## 4.1.29 ‘Sensitivity &lt;n&gt;’

- ‘Statistical sensitivity’ (Quantitative Concept)
- ‘Personality Sensitivity’ (Mental Process)
- ‘Antimicrobial susceptibility’ (Laboratory or Test Result)
- ‘Sensitivity <4>’ (Finding)

Each sense seems reasonable although ‘Sensitivity <4>’ is not completely defined. (I was surprised to find that it is a child of ‘Social Interaction Skills’ rather than having to do with being sensitive to touch.) Note, too, that psychological senses of words having other senses (such as ‘Personality Sensitivity’) have often been suppressed.

## 4.1.30 ‘Sperm Count &lt;n&gt;’

- ‘Sperm Count Procedure’ (Laboratory Procedure)
- ‘Sperm number’ (Quantitative Concept)
- ‘Encounter due to sperm count’ (Finding)
- ‘Sperm number observed’ (Laboratory or Test Result)

There are too many senses here. Perhaps the fourth sense, which seems to be part of sense one, could be suppressed. In addition, the third sense seems to be part of the more general notion of semen analysis and could also be suppressed.

## 4.1.31 ‘TEM &lt;n&gt;’

- ‘Triethylenemelamine’
- ‘Transmission Electron Microscopes’
- ‘Transmissible mink encephalopathy’
- ‘Transmission Electron Microscopy’

All senses are acronyms/abbreviations and should be suppressed. Also, the third sense may be erroneous; it has no synonyms with the proper ordering of the acronym letters.

## 4.1.32 ‘Translocation &lt;n&gt;’

- ‘Chromosomal translocation’ (Genetic Function)
- ‘Cellular translocation’ (Cell Function)
- ‘Intracellular translocation’ (Cell Function)
- ‘Protein translocation’ (Cell Function)

Senses two through four come only from CSP98 and seem redundant. Perhaps senses three and four could be suppressed.

## 4.1.33 ‘U &lt;n&gt;’

- ‘Uranium’ (<3> and <4>)

- ‘Lower case you’ (<1>)
- ‘Unit’ (<2>)

All senses are abbreviatory or self referential and should be suppressed.

#### 4.1.34 ‘XT - Exotropia <n>’

- ‘Manifest divergent squint’
- ‘Manifest alternating divergent squint’
- ‘Intermittent divergent squint’
- ‘Incomitant exotropia’

All senses are findings of ‘Exotropia’ which is not present in this case; they should all be suppressed. Note the similarity with example 4.1.10 except for the presence of the underlying concept.

## 4.2 Ambiguity of degree 3 and 2

The first 4 cases below have ambiguity degree 3 and the remaining 46 have degree 2.

### 4.2.1 ‘Energy Conservation <n>’

- ‘Physiological energy management’ (Therapeutic or Preventive Procedure)
- ‘Thermodynamic energy conservation’ (Natural Phenomenon or Process)
- ‘Energy Conservation <2>’ (Finding)

All senses seem appropriate.

### 4.2.2 ‘MAC <n>’

- ‘Complement Membrane Attack Complex’
- ‘MacAndrew Alcoholism Scale’
- ‘Raincoat’

All senses are abbreviatory and should be suppressed.

### 4.2.3 ‘PTC <n>’

- ‘Factor IX’ (which has string ‘Plasma Thromboplastin Component’)
- ‘Percutaneous transhepatic cholangiography’
- ‘Oncogene, RET’ (which has string ‘ret proto-oncogene (multiple endocrine neoplasia MEN2A, MEN2B and medullary thyroid carcinoma 1, Hirschsprung disease)’) )

All senses seem to be abbreviatory and should be suppressed.

### 4.2.4 ‘Therapy <n>’

- ‘Care involving unspecified rehabilitation procedure’ (Finding)
- ‘therapy <1>’ (Functional Concept)
- ‘Therapeutic procedure’ (Therapeutic or Preventive Procedure)

Sense one may be subsumed by sense three and could be suppressed. Sense two is a MeSH sub-heading. Its definition is *Used with diseases for therapeutic interventions except drug therapy, diet therapy, radiotherapy, and surgery, for which specific subheadings exist. The concept is also used for articles and books dealing with multiple therapies.* As with all subheadings, it may be best to suppress because of its specialized use.

## 4.2.5 17beta-hydroxysteroid dehydrogenase

- ‘3 (or 17)-beta-hydroxysteroid dehydrogenase’ (<1> and <2>)

There is really only one sense.

## 4.2.6 ‘22G &lt;n&gt;’

- ‘22 gauge’
- ‘22 grams’

Both senses are abbreviatory and should be suppressed.

## 4.2.7 ‘alpha-L-iduronidase deficiency &lt;n&gt;’

- ‘Mucopolysaccharidosis I’ (which has synonym ‘Hurler-Pfaundler syndrome’; Disease or Syndrome)
- ‘Hurler-Scheie Syndrome’ (Disease or Syndrome)

The second sense is narrower than the first and so could be suppressed.

## 4.2.8 ‘Aminocaproic Acid &lt;n&gt;’

- ‘6-Aminocaproic Acid’ (Amino Acid, Peptide, or Protein and Pharmacologic Substance)
- ‘Aminocaproic Acids’ (Amino Acid, Peptide, or Protein and Pharmacologic Substance)

This is a common type of ambiguity in the Metathesaurus in which a term represents both a group of something and a specific, usually paradigmatic instance of that group. For text processing purposes, it would help to suppress the group sense. Then the choice between the senses reduces to determining whether the form of the word *acid* is singular or plural.

## 4.2.9 ‘BL19 &lt;n&gt;’

- ‘BL-19 (epsilon-biotinamidocaproyl-beta-Ala-beta-Ala-lisinopril)’
- ‘BL19 Bladder Acupuncture Point’

Both senses are abbreviatory and should be suppressed.

## 4.2.10 ‘Blood group antibody C’

- ‘Blood group antibody C’ (Amino Acid, Peptide, or Protein and Immunologic Factor)
- ‘Blood group antibody c’ (Amino Acid, Peptide, or Protein and Immunologic Factor)

These two senses differ only by case. Since they have synonyms ‘Blood group antibody Rh2’ and ‘Blood group antibody Rh4’, respectively, they appear to represent legitimate ambiguity (assuming case insensitivity).

## 4.2.11 ‘Body Length &lt;n&gt;’

- ‘Body Height’ (Organism Attribute and Quantitative Concept)
- ‘Length of body’ (Organism Attribute)

Oddly enough, ‘Length of body’ is a child of ‘Body Height’ in RCD99. Perhaps the senses are synonymous, but if not I would favor suppressing ‘Body Height’.

## 4.2.12 ‘BSE &lt;n&gt;’

- ‘Breast Self-Examination’
- ‘Encephalopathy, Bovine Spongiform’

Both senses are abbreviatory and should be suppressed.

## 4.2.13 'Comb &lt;n&gt;'

- 'bleomycin/cyclophosphamide/semustine/vincristine protocol' (with string 'COMB')
- 'Comb animal structure'

The first sense is abbreviatory (somehow) and should be suppressed.

## 4.2.14 'Context &lt;n&gt;'

- 'Context <1>' (Finding; child of 'Attribute' in RCD99)
- 'context <2>' (Functional Concept; with string 'setting' and child of 'social psychology' in AOD95)

The second sense is quite specific and could be suppressed.

## 4.2.15 'Cork &lt;n&gt;'

- 'Cork'
- 'Cork County, Ireland'

The senses are legitimate.

## 4.2.16 'Crabs - pubic lice &lt;n&gt;'

- 'Infestation by Phthirus pubis' (Disease or Syndrome)
- 'Phthirus pubis' (Invertebrate)

The senses seem legitimate.

## 4.2.17 'Degree &lt;n&gt;'

- 'Degree or extent'
- 'Academic degree'

The senses are legitimate.

## 4.2.18 'Delusional Disorders &lt;n&gt;'

- 'Delusional disorder'
- 'Paranoid Schizophrenia'

The second sense may be too specific and could be suppressed.

## 4.2.19 'Duration &lt;n&gt;'

- 'Duration'
- 'Duration brand of oxymetazoline'

The second sense should be suppressed due to its specific, contextual nature.

## 4.2.20 'Dysfunction &lt;n&gt;'

- 'physiopathology' (Functional Concept)
- 'Functional disorder' (Pathologic Function)

The first sense is a MeSH subheading and can be suppressed.

## 4.2.21 'Economics &lt;n&gt;'

- 'Economics' (Quantitative Concept and Occupation or Discipline)
- 'Economic' (Quantitative Concept and Functional Concept; with MSH2000 string 'economics')

The second sense is a MeSH subheading. In addition, this is another case where number makes a difference. For both reasons, the second sense should be suppressed.

#### 4.2.22 'Esophageal Reflux <n>'

- 'Gastroesophageal reflux disease' (Disease or Syndrome)
- 'Esophageal reflux observation' (Finding)

It seems unlikely that the text *esophageal reflux* would refer to the second sense, a finding. Thus the second sense could be suppressed.

#### 4.2.23 'Ganglion <n>'

- 'Ganglia' (Body Part, Organ, or Organ Component)
- 'Benign cystic mucinous tumour' (Acquired Abnormality and Neoplastic Process)

The senses seem legitimate.

#### 4.2.24 'Glucosuria <n>'

- 'Glycosuria' (<1> and <2>)

There is no ambiguity here.

#### 4.2.25 'Glutamate <n>'

- 'Glutamates' (Amino Acid, Peptide, or Protein)
- 'Glutamate' (Amino Acid, Peptide, or Protein and Biologically Active Substance)

Suppress 'Glutamates' and let text processing determine which sense is correct.

#### 4.2.26 'GPA <n>'

- 'GPA <1>' (Amino Acid, Peptide, or Protein and Immunologic Factor)
- 'Grade point average' (Intellectual Product)

Both senses are abbreviatory and should be suppressed. (Note that the first sense has only one other string, 'GPA'. It occurs in both LNC10M and PDQ99, but is not defined very well.)

#### 4.2.27 'Homeopathy <n>'

- 'Homeopathy' (Biomedical Occupation or Discipline)
- 'Homeopathic therapy' (Therapeutic or Preventive Procedure)

Both senses seem legitimate.

#### 4.2.28 'Induced Abortion <n>'

- 'Abortion, Induced' (Therapeutic or Preventive Procedure)
- 'Diagnosis of induced abortion' (Finding)

The two senses are not completely delineated in the Metathesaurus; nevertheless, they appear legitimate.

#### 4.2.29 'Influenza Virus <n>'

- 'Orthomyxoviridae' (Virus)
- 'Influenzavirus, NOS' (Virus)

One definition for the first sense is *A family of RNA viruses causing influenza and other diseases. There are three recognized genera: INFLUENZAVIRUS A, B; INFLUENZAVIRUS C; and*

*THOGOTO-LIKE VIRUSES*. Thus it may be too general to mean *influenza virus*. It could probably be suppressed without doing any harm.

4.2.30 'Inner Ear <n>'

- 'Labyrinth'
- 'Diseases of inner ear'

The second sense should be suppressed.

4.2.31 'Knee (right) <n>'

- 'Assessment of passive right knee joint movement'
- 'Assessment of active right knee joint movement'

Both senses should be suppressed; they are essentially abbreviatory.

4.2.32 'Lice, Head <n>'

- 'Pediculus capitis infestation'
- 'Head Lice'

Both senses seem legitimate.

4.2.33 'MAGE-3 <n>'

- 'MAGE-3.1' (Pharmacologic Substance)
- 'MAGE-3 gene' (Gene or Genome)

These senses seem to be distinct and hence legitimate.

4.2.34 'Manic <n>'

- 'Manic' (Functional Concept)
- 'Manic mood' (Finding)

Both senses seem legitimate.

4.2.35 'Mediation <n>'

- 'Negotiating' (Social Behavior)
- 'mediation <2>' (with just one string, 'mediation' from AOD95') (Occupational Activity)

Again, both senses are legitimate.

4.2.36 'Metanephrine <n>'

- 'Metanephrine' (Organic Chemical and Neuroreactive Substance or Biogenic Amine)
- 'Metanephrine measurement' (Laboratory Procedure)

This is the standard ambiguity between a substance and a laboratory procedure measuring it; both senses seem legitimate.

4.2.37 'Myelopathy <n>'

- 'Bone Marrow Diseases'
- 'Spinal Cord Diseases'

These senses correspond to the two definitions of myelopathy in Dorland's and are therefore legitimate.

## 4.2.38 'Other maltreatment syndromes &lt;n&gt;'

- 'Other maltreatment syndromes - cause' (Injury or Poisoning)
- 'Other maltreatment syndromes - disorder' (Mental or Behavioral Dysfunction)

These senses are confusing (and difficult to track down in the Metathesaurus). They could both be suppressed.

## 4.2.39 'Placidyl &lt;n&gt;'

- 'Ethchlorvynol measurement'
- 'Placidyl'

This is the substance/measurement ambiguity again.

## 4.2.40 'Pregnenolone &lt;n&gt;'

- 'Assay for pregnenolone'
- 'Pregnenolone'

And again.

## 4.2.41 'Rapid plasma reagin &lt;n&gt;'

- 'Rapid plasma reagin test procedure'
- 'Rapid plasma reagin'

And yet again.

## 4.2.42 'S ANTIGEN:ARBITRARY CONCENTRATION:POINT IN TIME:ERYTHROCYTES^BLOOD PRODUCT UNIT:ORDINAL:RED BLOOD CELL AGGLUTINATION &lt;n&gt;'

- 'Upper case ess ANTIGEN:ARBITRARY SUBSTANCE CONCENTRATION:POINT IN TIME:ERYTHROCYTES^BLOOD PRODUCT UNIT:SEMI-QUANTITATIVE:RBC AGGLUTINATION'
- 'Lower case ess ANTIGEN:ARBITRARY SUBSTANCE CONCENTRATION:POINT IN TIME:ERYTHROCYTES^BLOOD PRODUCT UNIT:SEMI-QUANTITATIVE:RBC AGGLUTINATION'

Since the strings will never occur in normal text, it makes no difference what is done here.

## 4.2.43 'Sedation &lt;n&gt;'

- 'Sedation procedure' (Therapeutic or Preventive Procedure)
- 'Sedation' (Sign or Symptom)

Both senses are legitimate.

## 4.2.44 'Separated from cohabitee &lt;n&gt;'

- 'Separated from cohabitee' (Social Behavior and Human-caused Phenomenon or Process)
- 'Separated from cohabitee <1>' (Spatial Concept)

I believe that sense two is a child of sense one in RCD99. It could be suppressed as being only part of the more general social behavior.

## 4.2.45 'SHBG &lt;n&gt;'

- 'Sex Hormone-Binding Globulin'

- ‘Somatotropin binding globulin’

Both senses are abbreviatory and should be suppressed.

#### 4.2.46 ‘Shigella sonnei <n>’

- ‘Shigella sonnei bacterium’
- ‘Disease caused by Shigella sonnei’

Both senses seem legitimate. (Note how this differs from ‘Prostate’ vs. ‘Prostatic Diseases’.)

#### 4.2.47 ‘Sludge <n>’

- ‘Sludge <1>’ (Environmental Effect of Humans)
- ‘Sludge, NOS’ (Body Substance)

Both senses seem legitimate.

#### 4.2.48 ‘Thyroid stimulating immunoglobulins <n>’

- ‘Thyroid stimulating immunoglobulins’
- ‘Thyroid stimulating immunoglobulins assay’

This is the substance/measurement ambiguity again; both senses are legitimate.

#### 4.2.49 ‘Vascular resistance <n>’

- ‘Vascular resistance’
- ‘Vascular resistance measurement’

Again the substance/measurement ambiguity.

#### 4.2.50 ‘Weight <n>’

- ‘Body Weight’ (Organism Attribute and Quantitative Concept)
- ‘Weight’ (Quantitative Concept)

Both senses are legitimate.

## 5. Summary and Conclusions

As the Metathesaurus grows, so does the number of ambiguous concepts within it. But the ambiguities of large degree, all of which are false ambiguities, are being made harmless largely by being marked as suppressible. Indeed, if the recommendations made earlier are followed, the greatest degree of remaining ambiguity will be of degree 4.

The analysis described here reveals some classes of possible ambiguity commonly occurring in the Metathesaurus:

- Body part/disease ambiguity. This class of false ambiguity is exemplified by ‘Prostate’ and ‘Prostatic Diseases’. It arises from terms which require context within their vocabulary in order to be properly understood.
- Abbreviatory ambiguity. This is another, large class of ambiguity caused by distinct concepts having the same acronyms (or abbreviations). An example from above is that ‘Mitral Valve Stenosis’, ‘Multiple Sclerosis’, ‘Morphine Sulfate’ and ‘millisecond’ all have abbreviation ‘MS’ or ‘ms’. Although this class represents true ambiguity in a strict sense, it is better to disal-

low it in many text processing situations, especially those in which authors define the abbreviations they use. Furthermore, because of the large number of expansions for most acronyms, it might make sense to disallow such ambiguity universally.

- Substance/measurement ambiguity. This class, exemplified by ‘Thyroid stimulating immunoglobulins’ and ‘Thyroid stimulating immunoglobulins assay’, represents true ambiguity, especially in patient records but also in the biomedical literature.

Trying to decide what level of ambiguity is acceptable in some textual application is analogous to determining what granularity of meaning is appropriate in a knowledge source such as the Metathesaurus. Different levels are appropriate for different purposes. For this reason, it remains to be decided where the recommendations in this report are implemented. Some of them, including the cases of ambiguity of large degree, should be handled by marking more Metathesaurus strings as suppressible. Others, including abbreviatory ambiguity, may best be handled outside the Metathesaurus.