Ambiguity in the UMLS Metathesaurus

2008 Edition

Sonya E. Shooshan and Alan R. Aronson

August 1, 2008

1. Introduction

The UMLS® Metathesaurus® contains a significant amount of ambiguity. For example, the string "Cold" (or "cold" or "COLD") occurs in six distinct concepts with six distinct meanings. The purpose of this report is to examine ambiguity in the 2008AA release of the Metathesaurus in the context of its effect on natural language processing (NLP) applications.

Until the 2004AC release of the UMLS Knowledge Sources, ambiguity was denoted explicitly by appending an ambiguity designator, a number in angle brackets, to the end of an ambiguous string. Thus the ambiguity for "cold" was denoted by 'Cold <1>', 'Cold <2>', 'COLD <3>', etc. Now ambiguity is computed by finding concepts with strings that differ only with respect to case.[1]

Table 1 shows that the degree of Metathesaurus ambiguity has grown over the years and was particularly explosive in 2005, partly due to the direct computation of ambiguity mentioned above.

	2003	**2004**	**2005**	**2006**	**2007**	**2008**
Strings with an ambiguity designator	16,438	21,295 (+30%)	N/A	N/A	N/A	N/A
Concepts with one or more ambiguity	12,397	16,775 (+35%)	36,133 (+115%)	44,591 (+23%)	48,820 (+9%)	61,473 (+26%)
Concepts with one or more non-suppressible ambiguity	10,416	12,387 (+19%)	33,513 (+171%)	40,977 (+22%)	43,499 (+6%)	55,168 (+27%)
Cases of ambiguity	7,204	10,018 (+39%)	22,218 (+122%)	27,599 (+24%)	29,415 (+7%)	40,574 (+38%)
Cases of non-suppressible ambiguity	6,824	9,521 (+40%)	20,996 (+121%)	25,290 (+20%)	26,084 (+3%)	36,266 (+39%)

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

More recently, ambiguity grew significantly in 2006 and 2008, less so in 2007. (In Table 1, percentage changes are computed relative to the previous year.)

1. Note that AMBIGSUI.RRF or AMBIG.SUI cannot be used for this purpose because they do not conflate case.

Examining the cases of ambiguity more closely, consider the *degree* of ambiguity, i.e., the number of ways a string is ambiguous or, equivalently, the number of concepts in which it (or one of its case variants) occurs.[1] For example “deprecated ^ wbc-acnc” has degree 124 in 2008 all of which are marked as suppressible; “other” has degree 89 (43 if suppressibles are ignored). Table 2 con-

Degree of ambiguity	2005 cases	2006 cases	2007 cases	2008 cases
124			1	1 (0%)
93			1	
92		1		
54	1			
89				1
39		1	1 (0%)	1 (0%)
36		1	1 (0%)	1 (0%)
24		1		
23	1		1	1 (0%)
20		1		1 (0%)
19		1	1 (0%)	
18	1	1 (0%)	2 (+100%)	2 (0%)
17	2			2 (0%)
16	1	2 (+100%)	1 (-50%)	1 (0%)
15		1	3 (+200%)	2 (-33%)
14		1		3 (+200%)
13		1	1 (0%)	3 (+200%)
12	1	1 (0%)	3 (+200%)	6 (+100%)
11		3	4 (+33%)	10 (+150%)
10		4	7 (+75%)	17 (+143%)
9	6	13 (+117%)	14 (+8%)	25 (+79%)
8	10	23 (+130%)	24 (+4%)	61 (+154%)
7	11	28 (+155%)	42 (+50%)	70 (+67%)
6	24	66 (+175%)	104 (+58%)	185 (78%)
5	54	158 (+193%)	195 (+23%)	404 (+107%)
4	208	452 (+117%)	562 (+24%)	996 (77%)
3	1,239	1,868 (+51%)	2,380 (+27%)	4,226 (+78%)
2	20,659	24,971 (+21%)	26,067 (+4%)	34,555 (+32%)
1				
Total	**22,218**	**27,599 (+24%)**	**29,415 (+7%)**	**40,574 (+38%)**

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

1. The computation of the degree of an ambiguity was corrected in 2002. As a result, there are some differences from previous editions of this report in the counts reported in the tables.

tains the distribution of ambiguities in the Metathesaurus according to degree. Note that an ambiguity of degree one is not actually an ambiguity. In 2004 and before, for example, ‘Abbreviations <1>’ is not ambiguous since there were no other ‘Abbreviations <n>’ strings in the Metathesaurus.

Ignoring suppressible synonyms produces the more realistic distribution shown in Table 3. Most of the ambiguity of higher degree has disappeared, and all of that would disappear if appropriate

Degree of ambiguity	2005 cases	2006 cases	2007 cases	2008 cases
43				1
41			1	
40		1		
39		1		
36		1	1 (0%)	1 (0%)
24		1		
23	1		1	1 (0%)
20		1		1 (0%)
19		1	1 (0%)	
18	1	1 (0%)	2 (+100%)	2 (0%)
17	2			
16				
15		1	1 (0%)	1 (0%)
14			1	4 (+300%)
13		1		1 (0%)
12	1	1 (0%)	3 (+200%)	6 (+100%)
11		1	2 (+100%)	7 (+250%)
10		4	6 (+50%)	16 (+167%)
9	5	9 (+80%)	12 (+33%)	22 (+83%)
8	8	16 (+100%)	19 (+19%)	40 (+110%)
7	5	16 (+220%)	25 (+56%)	60 (+140%)
6	7	39 (+457%)	87 (+123%)	142 (+63%)
5	31	123 (+297%)	160 (+30%)	306 (+91%)
4	156	360 (+131%)	481 (+34%)	899 (+87%)
3	1,000	1,586 (+59%)	2,076 (+31%)	3,857 (+86%)
2	19,779	23,126 (+17%)	23,205 (+0%)	30,899 (+33%)
1				
Total	**20,996**	**25,290 (+20%)**	**26,084 (+3%)**	**36,266 (+39%)**

Table 3. **Metathesaurus ambiguity distribution after removing suppressibles**

strings were marked as suppressible. Suppressible synonyms are ignored for the remainder of this report.

Section 2 of this report describes general classes of ambiguity found in the Metathesaurus. Finally, Section 3 of this report, an appendix, describes only the most notable cases of ambiguity in the Metathesaurus, i.e., the cases of degree 10 or more. The bulk of the cases are now reported automatically by the Migration Assistant, a tool developed generally for annotating ambiguity and specifically for the purpose of marking appropriate cases as suppressible.

2. Classes of Metathesaurus Ambiguity

Some concepts contain strings which should be marked as suppressible. Many of these strings are already marked suppressible for a given UMLS release; this report recommends further cases some of which are universally applicable and some of which are appropriate in more limited environments such as the natural language processing done by MetaMap.

The analysis in this and previous editions of this report reveals some classes of ambiguity commonly occurring in the Metathesaurus:

- **Contextual (or hierarchical) ambiguity.** This class of false ambiguity is exemplified by the string ‘prostate’ for ‘Prostatic Diseases’. (Many of these problems have been fixed by suppressing the misleading string for the concept; but the problems continue to reappear as the Metathesaurus grows.) It normally arises from terms which require context within their vocabulary (in this case, a disease hierarchy) in order to be properly understood. Contextual ambiguities can be classified according to their participants:
  - **Body part/disease ambiguity** exemplified by ‘Prostate’ and ‘Prostatic Diseases’
  - **Body part/procedure ambiguity** exemplified by ‘Stomach’ and ‘Procedures on the stomach’
  - **Pathology/procedure ambiguity** exemplified by ‘Pathology’ and ‘Pathology procedure’
  - **Medical device/procedure ambiguity** exemplified by ‘Prosthesis’ and ‘Prosthesis Implantation’
  - **Substance/therapy ambiguity** exemplified by ‘Anthracyclines’ and ‘prior anthracycline therapy’
  - **Substance/measurement ambiguity** exemplified by ‘Thyroid stimulating immunoglobulins (TSI)’ and ‘Thyroid stimulating immunoglobulins assay’
- **Generalization ambiguity.** This is also false ambiguity caused by grouping several concepts together using a more general term. For example, 23 concepts including ‘Protocols: Activities’ and ‘Protocols: Pre- or Intra- or Post-Procedure**’** are generalized to ‘Protocols’ which does seem to be a legitimate synonym of the concept ‘Protocols documentation’.
- **Meta ambiguity.** This new class of ambiguity, represented by strings such as ‘Stress fracture, NEC in ICD10_1998’, contain meta information. In this case it is the name of the vocabulary, ICD10_1998 in the example. As opposed to the first class of ambiguity above in which strings such as ‘Prostate’ meaning ‘Prostatic Diseases’ do not say enough about themselves, these strings say too much. It is true that the meaning of a string containing ‘NEC’, ‘not elsewhere classified’ or like phrase, depends upon its vocabulary, but such information is already available in the MSRO file (where it belongs). It is also true that such strings have different meanings and strictly speaking should be different concepts. But the practical result of such a representational scheme is to introduce an ambiguity that most users do not want or need to

resolve. (It is not even clear that those who might want to resolve the ambiguity can do so with the information available in the Metathesaurus.)

- **Abbreviation 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 disallow it in many text processing situations, especially those in which authors define the abbreviations they use. Unlike the other classes of ambiguity defined above, we do not recommend that this case be reflected in changes to the Metathesaurus. This kind of ambiguity will be suppressed for MetaMap processing only.

3. Appendix: Higher Degree Metathesaurus Ambiguity

Ambiguous English Metathesaurus strings are described in this section in decreasing order of degree of ambiguity. Only those cases of degree 10 or more are covered. See Migration Assistant reports for cases of ambiguity of lesser degree.

In all cases, suppressible synonyms are ignored as is done in Table 3. Ambiguous forms for concepts shown in bold should be marked as suppressible. Recommendations for cases which are not clear are introduced with the word *consider.* Ambiguous forms for concepts shown in italics should be marked as suppressible in MetaMap only.

3.1 "other" (degree 43)

Except for 'Other', the remaining cases should be suppressed because they mean something more specific than "other". The concepts involved are

1. C0205394| Other
2. **C0220886| Other location of complaint**
3. **C1271040| Other health professional**
4. **C1521979| Other Routes of Drug Administration**
5. **C1546380| Other - Event Reason**
6. **C1546725| Other Specimen Source Code**
7. **C1546836| Other - Special Program Code**
8. **C1546840| Other - Publicity Code**
9. **C1546902| Other - Diagnosis Classification**
10. **C1546930| Other - Report Source**
11. **C1547110| Other - Modality**
12. **C1547196| Other - Organization unit type**
13. **C1547233| Other - Triage Code**
14. **C1547241| Other - Newborn Code**
15. **C1547267| Other - Risk Management Incident Code**
16. **C1547272| Other - Incident Type Code**
17. **C1547281| Other - Production Class Code**
18. **C1547292| Other - Recreational Drug Use Code**
19. **C1547304| Other - Precaution Code**
20. **C1547309| Other - Patient Condition Code**

21. **C1547994| Other - Diagnostic Service Section ID**
22. **C1549063| Other - Notify Clergy Code**
23. **C1549104| Other - Administrative Gender**
24. **C1549110| Other - Marital Status**
25. **C1550146| Other - Substance Type**
26. **C1556042| Other - Relationship**
27. **C1556043| Other - Religion**
28. **C1556044| other - No Information**
29. **C1556045| Other - What subject filter**
30. **C1556046| Other - Employment Status**
31. **C1556048| Other - Contact Role**
32. **C1556049| Other - Mail Claim Party**
33. **C1556050| Other - Living Dependency**
34. **C1556051| Other - Event Consequence**
35. **C1556052| Other - Indirect exposure mechanism**
36. **C1556053| Other - Action Taken in Response to the Event**
37. **C1556054| Other - Status of Evaluation**
38. **C1556055| Other - Causality Observations**
39. **C1556056| Other - Job Status**
40. **C1556057| Other - Immunization Registry Status**
41. **C1561608| Other - Mode of Arrival**
42. **C1868670| Other Growth**
43. **C1996846| Other (qualifier in LNC)**

3.2 "unknown" (degree 36) <no change from last year>

Except for 'Unknown' (occurs twice), the remaining cases should be suppressed because they mean something more specific than "unknown". The concepts involved are

1. C0439673| Unknown
2. **C1521803| Unknown Route of Drug Administration**
3. **C1546837| Unknown - Special Program Code**
4. **C1546841| Unknown Publicity Code**
5. **C1547283| Unknown - Production Class Code**
6. **C1547294| Unknown - Recreational Drug Use Code**
7. **C1547306| Unknown - Precaution Code**
8. **C1547312| Unknown - Patient Condition Code**
9. **C1548340| Unknown - Allergy Severity**
10. **C1548502| Unknown - Vaccines administered**
11. **C1548543| Unknown - Living Will Code**
12. **C1548550| Unknown - Organ Donor Code**
13. **C1549064| Unknown - Notify Clergy Code**
14. **C1549105| Unknown - Administrative Gender**
15. **C1549115| Marital Status - Unknown**
16. **C1549625| Unknown - Ethnic Group**
17. **C1556120| Unknown - Religion**
18. **C1556121| Unknown - Event reason**
19. **C1556122| Unknown - Relationship**
20. **C1556123| Unknown - Employment Status**

21. **C1556124| Unknown - Living Arrangement**
22. **C1556125| Unknown - Transport Arranged**
23. **C1556126| Unknown - Escort Required**
24. **C1556127| Unknown - Patient Outcome**
25. **C1556128| Unknown - Job Status**
26. **C1556129| Unknown - Patient_s Relationship to Insured**
27. **C1556130| Unknown - CWE statuses**
28. **C1556131| Unknown - Container status**
29. **C1556132| Unknown - Immunization Registry Status**
30. **C1556133| Unknown - Expanded yes/no indicator**
31. **C1556134| Unknown - Event Expected**
32. **C1556135| Unknown - Patient Class**
33. **C1556136| Unknown - Living Dependency**
34. **C1556137| Unknown - Contact Role**
35. C1561529| Unknown
36. **C1609613| unknown - NullFlavor**

3.3 "protocols" (degree 23) <no change from last year>

Except for 'Protocols documentation', the remaining cases should be suppressed because they mean something more specific than "protocols". The concepts involved are

1. C0442711| Protocols documentation
2. **C0542547| Protocols: Activities**
3. **C0677556| Protocols: Pre- or Intra- or Post-Procedure**
4. **C0677557| Protocols: Urinary Elimination**
5. **C0677558| Protocols: Tissue Perfusion**
6. **C0677559| Protocols: Tissue Integrity**
7. **C0677560| Protocols: Sensation, Pain and Comfort**
8. **C0677561| Protocols: Self-Concept**
9. **C0677562| Protocols: Self-Care**
10. **C0677563| Protocols: Safety**
11. **C0677564| Protocols: Role Relationship**
12. **C0677565| Protocols: Respiration**
13. **C0677566| Protocols: Physical Regulation**
14. **C0677567| Protocols: Nutrition**
15. **C0677568| Protocols: Metabolism**
16. **C0677569| Protocols: Medications and Blood Products**
17. **C0677570| Protocols: Immunology**
18. **C0677571| Protocols: Health Behavior**
19. **C0677572| Protocols: Fluid and Electrolyte**
20. **C0677573| Protocols: Coping**
21. **C0677574| Protocols: Cognition**
22. **C0677575| Protocols: Circulation**
23. **C0677576| Protocols: Bowel Elimination**

3.4 "assessment" (degree 20)

Except for 'Evaluation', 'Evaluation procedure', and 'Assessed', the remaining cases should be suppressed because they are specific kinds of "assessment". The concepts involved in this ambiguity are

1. **C0028708| Nutrition Assessment**
2. **C0031809| Physical Examination**
3. C0220825| Evaluation
4. **C0542573| Assessment: Bowel Elimination**
5. **C0549068| Assessment: Circulation**
6. **C0549070| Assessment: Coping**
7. **C0549071| Assessment: Fluid and Electrolytes**
8. **C0549072| Assessment: Health Behavior**
9. **C0549073| Assessment: Medications and Blood Products**
10. **C0549074| Assessment: Metabolism**
11. **C0549075| Assessment: Respiration**
12. **C0549076| Assessment: Safety**
13. **C0549077| Assessment: Self-Care**
14. **C0549078| Assessment: Sensation, Pain and Comfort**
15. **C0549079| Assessment: Urinary Elimination**
16. **C0549080| Assessment: Pre- or Intra- or Post-Procedure**
17. **C0679207| Knowledge acquisition using a method of assessment**
18. **C0870300| Assessment: Cognition**
19. C1261322| Evaluation procedure
20. C1516048| Assessed

3.5 "ec 2.7.1.112" (degree 18) <no change from last year>

All Enzyme Commission (EC) numbers (strings beginning "ec <integer>.") are suppressed by MetaMap because they represent classes of enzymes and are consequently highly ambiguous.

1. *C0033681| Protein Tyrosine Kinase*
2. *C0065344| Lymphocyte Specific Protein Tyrosine Kinase p56(lck)*
3. *C0109317| EphB2 Receptor*
4. *C0117718| fibroblast growth factor receptor 3*
5. *C0138965| protein-tyrosine kinase c-src*
6. *C0169658| Janus kinase 1*
7. *C0169661| Janus kinase 2*
8. *C0290067| Platelet-Derived Growth Factor alpha Receptor*
9. *C0290068| Platelet-Derived Growth Factor beta Receptor*
10. *C0907648| Ephrin Receptor EphB1*
11. *C0915156| Ephrin Receptor EphA8*
12. *C1259418| MERTK protein, human*
13. *C1333408| EPHA4 protein, human*
14. *C1333409| EPHB3 protein, human*
15. *C1333410| EPHA2 protein, human*
16. *C1334392| LTK protein, human*
17. *C1370509| EPHA1 protein, human*
18. *C1504624| KDR protein, human*

3.6 "patient education plans" (degree 18) <no change from last year>

All eighteen cases should be suppressed because they are specific kinds of "patient education plans". Their concepts are

1. **C0549081| Patient Education Plans: Activities**
2. **C0549082| Patient Education Plans: Bowel Elimination**
3. **C0549083| Patient Education Plans: Circulation**
4. **C0549084| Patient Education Plans: Coping**
5. **C0549085| Patient Education Plans: Health Behavior**
6. **C0549086| Patient Education Plans: Immunology**
7. **C0549087| Patient Education Plans: Medications and Blood Products**
8. **C0549088| Patient Education Plans: Metabolism**
9. **C0549089| Patient Education Plans: Nutrition**
10. **C0549090| Patient Education Plans: Physical Regulation**
11. **C0549091| Patient Education Plans: Respiration**
12. **C0549092| Patient Education Plans: Role Relationship**
13. **C0549093| Patient Education Plans: Safety**
14. **C0549094| Patient Education Plans: Self-Care**
15. **C0549095| Patient Education Plans: Sensation, Pain and Comfort**
16. **C0549096| Patient Education Plans: Tissue Integrity**
17. **C0549097| Patient Education Plans: Urinary Elimination**
18. **C0549098| Patient Education Plans: Pre- or Intra- or Post-Procedure**

3.7 "emergency" (degree 15) <no change from last year>

Except for 'Emergency Situation' and 'Bale out', the remaining cases should be suppressed because they are specific kinds of "emergency". The concepts involved in this ambiguity are

1. C0013956| Emergency Situation
2. C0175673| Bale out
3. **C1546399| Encounter Admission Source - emergency**
4. **C1546844| Visit Priority Code - Emergency**
5. **C1547144| Specialty Type - Emergency**
6. **C1552231| Clinical Nurse Specialist - Emergency**
7. **C1553500| Act Code - emergency**
8. **C1555975| Registered Nurse - Emergency**
9. **C1561583| Patient Class - Emergency**
10. **C1561584| Certification patient type - Emergency**
11. **C1561585| Level of Care - Emergency**
12. **C1561586| Consent Bypass Reason - Emergency**
13. **C1561587| Referral category - Emergency**
14. **C1561588| Admission Type - Emergency**
15. **C1561589| Consent Non-Disclosure Reason - Emergency**

3.8 "1" (degree 14)

All numbers are suppressed by MetaMap because they are highly ambiguous.

1. *C0205447|One*
2. *C0227032| Maxillary right third molar*

3. *C0920321*| *Phase I Clinical Trials*
4. *C1314970*| *Laboratory Class*
5. *C1334882*| *NKX3-1 gene*
6. *C1418979*| *PSCD2 gene*
7. *C1424905*| *WAC gene*
8. *C1426071*| *SASH1 gene*
9. *C1427384*|*DEADC1 gene*
10. *C1513302*| *Mild Adverse Event*
11. *C1539859*| *ST3GAL1 gene*
12. *C1579809*| *Pack size 1*
13. *C1823121*| *SPTLC3 gene*
14. *C1825098*| *FAAH2 gene*

3.9 "cap" (degree 14)

Except for 'Caps', 'Cap Device Component, and 'Syringe Caps', the remaining cases should be suppressed (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0006935*| *capsule (pharmacologic)*
2. C0179586| Caps
3. *C0278651*| *cyclophosphamide/doxorubicin/prednisone protocol*
4. *C0280547*| *cisplatin/cyclophosphamide/doxorubicin protocol*
5. *C1416891*| *LNPEP gene*
6. *C1418551*| *SERPINB6 gene*
7. *C1419093*| *PTPLA gene*
8. *C1422073*| *BRD4 gene*
9. *C1422760*| *SORBS1 gene*
10. *C1426630*| *CAP1 gene*
11. C1706092| Cap Device Component
12. C1657858| Syringe Caps
13. *C1706433*| *Capsule Dosing Unit*
14. *C1855179*| *CATARACT, ANTERIOR POLAR*

3.10 "none" (degree 14) <no change from last year>

Except for 'None', the remaining cases should be suppressed because they are specific kinds of "none". The concepts involved in this ambiguity are

1. C0549184| None
2. **C1546509| none - TableRules**
3. **C1547191| none - ResponseLevel**
4. **C1550083| None - EntityCode**
5. **C1550437| None - Sequencing**
6. **C1551387| None - ContainerSeparator**
7. **C1553523| none - SubstanceAdminSubstitution**
8. **C1556146| None - Relationship**
9. **C1556147| None - Eligibility Source**
10. **C1556148| None - Action Taken in Response to the Event**
11. **C1556150| None - ObservationValue**
12. **C1556151| None - Language Proficiency**

13. **C1556152| None - Additive/Preservative**
14. **C1706277| None Device Component**

3.11 “pap” (degree 14)

Suppress ambiguous form(s) (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0030350| Papaverine*
2. *C0312402| Acid phosphatase isoenzyme, prostatic fraction*
3. *C1367456|ACPP gene*
4. *C1413944| DDEF1 gene*
5. *C1413945| DDEF2 gene*
6. *C1418410| MRPS30 gene*
7. *C1422804| PDAP1 gene*
8. *C1423108| PAPOLA gene*
9. *C1424700| TUSC2 gene*
10. *C1538823| REG3A gene*
11. *C1705529| ACPP wt Allele*
12. *C1705530| PAPOLA wt Allele*
13. *C1705531| TUSC2 wt Allele*
14. *C1970472| PULMONARY ALVEOLAR PROTEINOSIS, ACQUIRED*

3.12 “alp” (degree 13)

Suppress ambiguous form(s) (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0102159| alizarinprimeveroside*
2. *C0201850| Alkaline phosphatase measurement*
3. *C0663932| SLPI protein, human*
4. *C1366565| SLPI gene*
5. *C1366566| CCL27 gene*
6. *C1412624| ATHS gene*
7. *C1424288| ASRGL1 gene*
8. *C1427121| PDLIM3 gene*
9. *C1428783| ATRNL1 gene*
10. *C1531719| Atherogenic lipoprotein phenotype*
11. *C1705078| CCL27 wt Allele*
12. *C1706468| SLPI wt Allele*
13. *C1826354| NAT10 gene*

3.13 “active” (degree 12) <no change from last year>

Except for ‘Active’ and ‘Active brand of pseudoephedrine-triprolidine’, the remaining cases should be suppressed because they are specific kinds of “active”. Suppress ‘Active brand of pseudoephedrine-triprolidine’ (MetaMap only) because it is a brand name. The concepts involved in this ambiguity are

1. C0205177| Active

2. *C0718247| Active brand of pseudoephedrine-triprolidine*
3. **C1547419| ActStatus - active**
4. **C1553875| Concept Status - Active**
5. **C1561507| EditStatus - Active**
6. **C1561508| Managed Participation Status - active**
7. **C1561509| Role Status - active**
8. **C1561510| Entity Status - active**
9. **C1561511| Document Storage - active**
10. **C1561512| Document Storage Status - Active**
11. **C1561513| Immunization Registry Status - Active**
12. **C1706449| Active Control**

3.14 "ar" (degree 12)

Suppress ambiguous form(s) (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0003504| Aortic Valve Insufficiency*
2. *C0003761| Country of Argentina*
3. *C0003790| Arkansas*
4. *C0051755| Amphiregulin*
5. *C0332284| Arising in*
6. *C0559546| Adverse reactions*
7. *C1367578| AR gene*
8. *C1412322| AKR1B1 gene*
9. *C1447749| AR protein, human*
10. *C1514768| Recombinant Amphiregulin*
11. *C1704903| AREG wt Allele*
12. *C1705240| AR wt Allele*

3.15 "cd" (degree 12)

Suppress ambiguous form(s) (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0006632| Cadmium*
2. *C0007570| Celiac Disease*
3. *C0018553| Hamartoma Syndrome, Multiple*
4. *C0043444| Democratic Republic of the Congo*
5. *C0056447| CP protocol*
6. *C0079141| Compact discs*
7. *C0332140| Diagnosis, clinical*
8. *C0700300| candela*
9. *C1426202| CELIAC3 gene*
10. *C1426204| CELIAC2 gene*
11. *C1826449| NOD2 gene*
12. *C1955216| Clusters of differentiation*

3.16 "ec 2.7.1.-" (degree 12)

All Enzyme Commission (EC) numbers (strings beginning "ec <integer>.") are suppressed by MetaMap because they represent classes of enzymes and are consequently highly ambiguous.

1. *C0108836| CDC7 protein, human*
2. *C0108855| CDK2 protein, human*
3. *C0259367| PCTAIRE Protein Kinase 1*
4. *C0659150| CHEK1 protein, human*
5. *C0673406| GPRK7 protein, human*
6. *C1333180| Cyclin-Dependent Kinase 10*
7. *C1333735| GPRK2L protein, human*
8. *C1333738| G Protein-Coupled Receptor Kinase Family*
9. *C1337052| PAK6 protein, human*
10. *C1447440| CDK3 protein, human*
11. *C1744605| G-protein-coupled receptor kinase 5*
12. *C1744606| G-protein-coupled receptor kinase 6*

3.17 "not applicable" (degree 12) <no change from last year>

Except for 'not applicable', the remaining cases should be suppressed because they are specific kinds of "not applicable". The concepts involved in this ambiguity are

1. C1272460| not applicable
2. **C1546968| No Information - not applicable**
3. **C1547280| Production Class Code - Not Applicable**
4. **C1549103| Administrative Sex - Not applicable**
5. **C1609491| Patient Class - Not Applicable**
6. **C1610044| Derived specimen - Not Applicable**
7. **C1610595| Identity May Be Divulged - Not applicable**
8. **C1611147| CWE statuses - Not applicable**
9. **C1619691| Expanded yes/no indicator - not applicable**
10. **C1705112| Potency Not Applicable**
11. **C1705113| Dosage Form Not Applicable**
12. **C1705512| Route of Administration Not Applicable**

3.18 "ptc" (degree 12)

Suppress ambiguous form(s) (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0015491| Factor IX*
2. *C0203085| Percutaneous transhepatic cholangiography*
3. *C0238463| Papillary thyroid carcinoma*
4. *C0694890| RET gene*
5. *C1366464| F9 gene*
6. *C1419055| TAS2R38 gene*
7. *C1425774| CCDC6 gene*
8. *C1704885| RET wt Allele*
9. *C1705338| F9 wt Allele*
10. *C1705339|PTCH wt Allele*

11. *C1706229*| *CCDC6 wt Allele*
12. *C1826732*| *PTCH1 gene*

3.19 "ad" (degree 11)

Suppress ambiguous form(s) (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0002395*| *Alzheimer's Disease*
2. *C0002838*| *Andorra*
3. *C0010934*| *Dactinomycin*
4. *C0050841*| *dacarbazine/doxorubicin protocol*
5. *C0228318*| *Anterodorsal nucleus of thalamus*
6. *C0280573*| *cytarabine/daunorubicin protocol*
7. *C0332133*| *Admitting Diagnosis*
8. *C0547043*| *Up*
9. *C1630418*| *AD Substance*
10. *C1704642*| *Analysis Dataset Domain*
11. *C1706476*| *AD Term Type*

3.20 "cam" (degree 11)

Except for 'Cam, topical lotion' and 'CAM brand of Ephedrine Hydrochloride', the remaining cases should be suppressed (MetaMap only) because they are abbreviatory. Suppress 'CAM brand of Ephedrine Hydrochloride' (MetaMap only) because it is a brand name. The concepts involved in this ambiguity are

1. *C0007578*| *Cell Adhesion Molecules*
2. *C0054551*| *cyclophosphamide/doxorubicin/methotrexate protocol*
3. *C0178551*| *chorioallantoic membrane*
4. *C0678112*| *CAM brand of Ephedrine Hydrochloride*
5. C0713465| Cam, topical lotion
6. *C1148475*| *Complementary and alternative medicine*
7. *C1366910*| *Calmodulin 1*
8. *C1366911*| *Cerebral Cavernous Malformations 1*
9. *C1537503*| *KRIT1 gene*
10. *C1706432*| *KRIT1 wt Allele*
11. *C1861784*| *CEREBRAL CAVERNOUS MALFORMATIONS*

3.21 "cat" (degree 11)

Except for 'Felis catus', 'Felis silvestris', 'Genus Felis', 'Family Felidae', 'Subfamily Felinae', and 'Cat (antigen)', the remaining cases should be suppressed (MetaMap only) because they are abbreviatory. 'Family Felidae', 'Subfamily Felinae', and 'Cat (antigen)' should be suppressed because they are specific kinds of "cat". The concepts involved in this ambiguity are

1. C0007450| Felis catus
2. *C0008169*| *Chloramphenicol O-Acetyltransferase*
3. *C0040405*| *X-Ray Computed Tomography*
4. *C0280589*| *cytarabine/thioguanine*

5. **C0325089| Family Felidae**
6. C0325090| Felis silvestris
7. C0524517| Genus Felis
8. **C1270185| Subfamily Felinae**
9. *C1366498| Chloramphenicol Acetyl Transferase Gene*
10. *C1413138| CAT gene*
11. **C1963009| Cat (antigen)**

3.22 "ms" (degree 11)

Suppress ambiguous form(s) (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0025867| Metric System*
2. *C0026221| Mississippi (geographic location)*
3. *C0026269| Mitral Valve Stenosis*
4. *C0026514| Montserrat*
5. *C0026769| Multiple Sclerosis*
6. *C0439223| millisecond*
7. *C1417453| MTR gene*
8. *C1513009| Master of Science*
9. *C1552156| Supernumerary mandibular left primary canine*
10. *C1868685| MULTIPLE SCLEROSIS, SUSCEPTIBILITY TO*
11. *C1881819| Microbiology Susceptibility Domain*

3.23 "p14" (degree 11)

Suppress ambiguous form(s) (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0054505| Calgranulin B*
2. *C0292779| activated RNA polymerase II transcription cofactor 4*
3. *C0525037| CDKN2A gene*
4. *C1335798| S100A9 gene*
5. *C1423800| CTNNBL1 gene*
6. *C1428962| RPP14 gene*
7. *C1540306| CDK2AP2 gene*
8. *C1704874| S100A9 wt Allele*
9. *C1709390| SUB1 gene*
10. *C1835861| MAPBP-INTERACTING PROTEIN GENE*
11. *C1842980| SPLICING FACTOR 3B, 14-KD SUBUNIT GENE*

3.24 "patient" (degree 11)

Except for 'Patients', the remaining cases should be suppressed because they are specific kinds of "patient". The concepts involved in this ambiguity are

1. C0030705| Patients
2. **C1550655| Specimen Type - Patient**
3. **C1578478| Role Class - patient**

4. **C1578479| Role Code - Patient recipient**
5. **C1578480| Role Code - Patient specimen**
6. **C1578481| Mail Claim Party - Patient**
7. **C1578483| Report source - Patient**
8. **C1578484| Relationship modifier - Patient**
9. **C1578485| Specimen Source Codes - Patient**
10. **C1578486| Disabled Person Code - Patient**
11. **C1705908| Veterinary Patient**

3.25 "yes" (degree 11)

Except for 'YES1 gene', 'Yes (indicator)' and 'Yes - Yes/no indicator', the remaining cases should be suppressed because they are specific kinds of "Yes". Suppress 'YES1 gene' (MetaMap only) because it is abbreviatory. The concepts involved in this ambiguity are

1. *C0919479| YES1 gene*
2. **C1298907| Yes - Presence findings**
3. **C1546945| Yes - Event Seriousness**
4. **C1546947| Yes - Event Expected**
5. **C1546969| Yes - Identity May Be Divulged**
6. **C1548171| Yes - Release Information**
7. **C1549060| Yes - Expanded yes/no indicator**
8. **C1549065| Yes - Notify Clergy Code**
9. **C1549443| Yes - Assignment of Benefits**
10. C1549445| Yes - Yes/no indicator
11. C1705108| Yes (indicator)

3.26 "a" (degree 10)

All single letters are suppressed by MetaMap because they are highly ambiguous.

1. *C0227089|Deciduous maxillary right second molar tooth*
2. *C0348042| Blood group antigen A*
3. *C0439234| year*
4. *C0457243| Ampere*
5. *C1442985| Tumor staging descriptor a*
6. *C1442986| Abdominal lymph node tumor invasion status A (tumor staging)*
7. *C1522424| A Mouse*
8. *C1706280| Lower case Roman letter a*
9. *C1706281| Upper case Roman letter A*
10. *C1706282| Lymphoma staging symptom status A*

3.27 "ADENINE 30 MG / ANTICOAGULANT CITRATE DEXTROSE SOLUTION / CITRIC ACID 209 MG / DEXTROSE 1.78 GM / DEXTROSE 2.42 GM / MANNITOL 825 MG / RED CELL PRESERVATION SOLUTION / SODIUM CHLORIDE 990 MG / SODIUM CITRATE 1.84 GM / SODIUM HYDROXIDE / SODIUM PHOSPHATE, MONOBASIC, MONOHYDRATE 155 MG / WATER FOR

INJECTION, STERILE QS SOLUTION [ANTICOAGULANT]" (degree 10)

All cases should be suppressed because should be suppressed because they are specific kinds of "ADENINE 30 MG....". The concepts involved in this ambiguity are

1. **C1737879| ANTICOAGULANT CITRATE PHOSPHATE DEXTROSE SOLUTION WITH ADSOL 4R3335**
2. **C1738147| ANTICOAGULANT CITRATE PHOSPHATE DEXTROSE SOLUTION WITH ADSOL 4R3445**
3. **C1738688| ANTICOAGULANT CITRATE PHOSPHATE DEXTROSE SOLUTION WITH ADSOL 4R3475**
4. **C1738689| ANTICOAGULANT CITRATE PHOSPHATE DEXTROSE SOLUTION WITH ADSOL 4R3330**
5. **C1740111| ANTICOAGULANT CITRATE PHOSPHATE DEXTROSE SOLUTION WITH ADSOL 4R3463**
6. **C1742071| ANTICOAGULANT CITRATE PHOSPHATE DEXTROSE SOLUTION WITH ADSOL 4R3464**
7. **C1742339| ANTICOAGULANT CITRATE PHOSPHATE DEXTROSE SOLUTION WITH ADSOL 4R3467**
8. **C1743294| ANTICOAGULANT CITRATE PHOSPHATE DEXTROSE SOLUTION WITH ADSOL 4R3468**
9. **C1743295| ANTICOAGULANT CITRATE PHOSPHATE DEXTROSE SOLUTION WITH ADSOL 4R1488**
10. **C1966740| ANTICOAGULANT CITRATE PHOSPHATE DEXTROSE SOLUTION WITH ADSOL 4R3440**

3.28 "at3" (degree 10)

Except for 'Antithrombin III', the remaining cases should be suppressed because they are specific kinds of "Antithrombin III Deficiency". Suppress 'Antithrombin III' (MetaMap only) because it is abbreviatory. The concepts involved are

1. *C0003438| Antithrombin III*
2. **C1862776| Antithrombin III Deficiency PADUA 2**
3. **C1862777| Antithrombin III Defieincy ROMA [sic]**
4. **C1862778| Antithrombin III Deficiency TRENTO**
5. **C1862781| Antithrombin III Deficiency FONTAINBLEAU**
6. **C1862784| Antithrombin III Deficiency CLICHY**
7. **C1862786| Antithrombin III Deficiency Barcelona**
8. **C1862789| Antithrombin III Deficiency BARCELONA 2**
9. **C1862790| Antithrombin III Deficiency AVRANCHES**
10. **C1862797| Antithrombin III Deficiency Paris**

3.29 "bar" (degree 10)

Except for 'bar unit of measure', 'Taverns', 'Bar form', and 'External fixator bar', the remaining cases should be suppressed (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0001643| beta-2 Adrenergic Receptors*
2. C0441233| External fixator bar

3. C0687760| Taverns
4. C0993613| Bar form
5. *C1367657| ADRB2 Gene*
6. *C1417825| NR1H4 gene*
7. *C1425012| BFAR gene*
8. C1551065| bar unit of measure
9. *C1704463| ADRB2 wt Allele*
10. *C1704759| Bar Dosing Unit*

3.30 "car" (degree 10)

Except for 'Automobiles' and 'Car - Mode of Arrival Code', the remaining cases should be suppressed (MetaMap only) because they are abbreviatory. 'Car - Mode of Arrival Code' should be suppressed because it is a specific kind of "car". The concepts involved are

1. C0004381| Automobiles
2. *C0406810| Atrial myxoma with lentigines*
3. *C1166663| actomyosin contractile ring*
4. *C1413828| CXADR gene*
5. *C1417827| NR1I3 gene*
6. *C1420354| SPG7 gene*
7. **C1547285| Car - Mode of Arrival Code**
8. *C1622899| car <invertebrate>*
9. *C1706434| RFP2 wt Allele*
10. *C1858724| Caronte Gene*

3.31 "ec 2.7.1.37" (degree 10)

All Enzyme Commission (EC) numbers (strings beginning "ec <integer>.") are suppressed by MetaMap because they represent classes of enzymes and are consequently highly ambiguous.

1. *C0033640| PROTEIN KINASE*
2. *C0072402| Protein-Serine-Threonine Kinases*
3. *C0244987| glycogen synthase kinase 3 alpha*
4. *C0294209| LIM Domain Kinase 1*
5. *C0380146| activin receptor-like kinase 1*
6. *C0541150| 3-Phosphoinositide Dependent Protein Kinase-1*
7. *C1314894| Col4A3 protein, human*
8. *C1332856| Casein Kinase 2, Alpha 1 Polypeptide*
9. *C1447968| ACVR1 protein, human*
10. *C1880254| Death-Associated Protein Kinase 1 Protein*

3.32 "kit" (degree 10) <no change from last year>

Except for 'Kit device', 'Kit Component of Device', and 'Drug Kit', the remaining cases should be suppressed (MetaMap only) because they are abbreviatory. The concepts involved in this ambiguity are

1. *C0072470| Proto-Oncogene Protein c-kit*
2. C0812225| Kit device

3. *C0920288*| *C-KIT Gene*
4. *C1416655*| *KIT gene*
5. *C1553450*| *Kit Code*
6. *C1690540*| *Kit Dosing Unit*
7. *C1704742*| *Kit Dosage Form*
8. *C1704888*| *KIT wt Allele*
9. C1705212| Kit Component of Device
10. C1705213| Drug Kit

3.33 "m" (degree 10)

Suppress ambiguous form(s) (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0024554*| *Male gender*
2. *C0221134*| *Blood group antigen M*
3. *C0227102*| *Mandibular left primary canine tooth*
4. *C0439113*| *Upper case emm*
5. *C0439232*| *Minute of time*
6. *C0456533*| *M - Metastasis stages*
7. *C0475209*| *meter*
8. *C1706456*| *Roman numeral upper case emm*
9. *C1706457*| *lower case emm*|
10. *C1883310*| *One Thousand*

3.34 "mac" (degree 10)

Suppress ambiguous form(s) (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0009545*| *Complement Membrane Attack Complex*
2. *C0024403*| *Macao*
3. *C0026916*| *Mycobacterium avium-intracellulare Infection*
4. *C0065465*| *cyclophosphamide/dactinomycin/methotrexate protocol*
5. *C0083360*| *chlorambucil/dactinomycin/methotrexate protocol*
6. *C0279190*| *cyclophosphamide/doxorubicin/mitomycin protocol*
7. *C0451273*| *MacAndrew Alcoholism Scale*
8. *C0453947*| *Raincoat*
9. *C1167383*| *membrane attack complex location*
10. *C1416956*| *MARCKS gene*

3.35 "no" (degree 10) <no change from last year>

Except for 'Norway', 'no', and 'No - yes/no indicator', the remaining cases should be suppressed because they are specific kinds of "no". Suppress 'Norway' (MetaMap only) because it is abbreviatory. The concepts involved in this ambiguity are

1. *C0028423*| *Norway*
2. C1298908| no
3. **C1546943| No - Event Seriousness**

4. **C1546946| No - Event Expected**
5. **C1546967| No - Identity May Be Divulged**
6. **C1548170| No - Release Information**
7. **C1549056| No - Expanded yes/no indicator**
8. **C1549062| No - Notify Clergy Code**
9. **C1549442| No - Assignment of Benefits**
10. C1549444| No - yes/no indicator

3.36 "normal" (degree 10)

Except for 'Normal' and 'Normal assessment finding', the remaining cases should be suppressed because they are specific kinds of "normal". The concepts involved in this ambiguity are

1. C0205307| Normal
2. **C1550457| Normal Observation Interpretation**
3. **C1550469| normal Confidentiality**
4. **C1551394| normal Device Alert Level**
5. **C1553386| normal Act Status**
6. **C1553399| normal Managed Participation Status**
7. **C1553402| normal Role Status**
8. **C1553406| normal Entity Status**
9. **C1704701| Normality-Based Dosing Unit**
10. C1873497| Normal assessment finding

3.37 "p40" (degree 10)

Suppress ambiguous form(s) (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0050854| adjuvant P40*
2. *C0085424| Interleukin-9*
3. *C1367780| Laminin Receptor-1*
4. *C1412528| ARHGEF2 gene*
5. *C1416795| LANCL1 gene*
6. *C1419038| PSMD7 gene*
7. *C1456382| EBNA1BP2 gene*
8. *C1539696| RPSA gene*
9. *C1705231| RPSA wt Allele*
10. *C1826761| RABEPK gene*

3.38 "radiology" (degree 10)

Except for 'Radiology Speciality', 'Radiology studies', and 'Diagnostic radiologic examination', the remaining cases should be suppressed because they are specific kinds of "radiology". The concepts involved in this ambiguity are

1. C0034599| Radiology Specialty
2. C0807679| Radiology studies
3. **C1405978| Encounter due to radiological examination**
4. **C1548000| Radiology Section ID**

5. **C1548429| radiology referral type**
6. **C1552284| Radiology Podiatrist**
7. **C1555923| Radiology Chiropractor**
8. **C1608525| Radiology - NUCCProvider Codes**
9. **C1610162| Radiology - Clinic/Center - NUCCProviderCodes**
10. C1962945| Radiographic imaging procedure

3.39 "sports medicine" (degree 10) <no change from last year>

Except for 'sports medicine specialty', the remaining cases should be suppressed because they are specific kinds of "sports medicine". The concepts involved in this ambiguity are

1. C0038040| sports medicine specialty
2. **C1552285| Podiatrist - Sports Medicine**
3. **C1555741| Emergency Medicine - Sports Medicine**
4. **C1555748| Family Practice - Sports Medicine**
5. **C1555771| Internal Medicine - Sports Medicine**
6. **C1555800| Orthopedic Surgery - Sports Medicine**
7. **C1555844| Pediatrics - Sports Medicine**
8. **C1555849| Physical Medicine & Rehabilitation - Sports Medicine**
9. **C1555858| Preventive Medicine - Sports Medicine**
10. **C1555872| Psychiatry & Neurology - Sports Medicine**

3.40 "tr" (degree 10)

Suppress ambiguous form(s) (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0040961| Tricuspid Valve Insufficiency*
2. *C0041400| Country of Turkey*
3. *C0332121| Treatment required for*
4. *C1366448| TERC gene*
5. *C1366449| F2R gene*
6. *C1420775| TMEFF2 gene*
7. *C1425351| TXNRD2 gene*
8. *C1619635| CD71 antigen*
9. *C1705312| TERC wt Allele*
10. *C1705939| F2R wt Allele*

3.41 "ts" (degree 10)

Suppress ambiguous form(s) (MetaMap only) because they are abbreviatory. The concepts involved are

1. *C0040517|Gilles de la Tourette's syndr.*
2. C0040963| *Tricuspid Valve Stenosis*
3. *C0041341| Tuberose Sclerosis*
4. *C1366824| TYMS gene*
5. *C1420620| TBXAS1 gene*
6. *C1552162| Supernumerary mandibular right second primary molar*

7. *C1704618*| *Trial Summary Domain*
8. *C1705746*| *TYMS wt Allele*
9. *C1832916*| *TIMOTHY SYNDROME*
10. *C1868676*| *GROWTH CONTROL, Y-CHROMOSOME INFLUENCED*