

---

# Ambiguity in the UMLS Metathesaurus

2004 Edition

**Alan R. Aronson and Sonya E. Shooshan**

July 16, 2004

## 1. Introduction

The UMLS<sup>®</sup> Metathesaurus<sup>®</sup> 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 included as another string for the concept involved.<sup>1</sup> 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’. The purpose of this report is to examine ambiguity in the 2004AA release of the Metathesaurus in the context of its effect on natural language processing (NLP) applications.

The number of explicitly ambiguous Metathesaurus terms has grown over the years. Table 1 shows several counts that indicate that growth in broad terms. Percentage changes are computed

	<b>2000</b>	<b>2001</b>	<b>2002</b>	<b>2003</b>	<b>2004</b>
<b>Strings with an ambiguity designator</b>	9,416	12,840 (+36%)	13,837 (+8%)	16,438 (+19%)	21,295 (+30%)
<b>Concepts with one or more ambiguity</b>	7,409	9,637 (+30%)	10,328 (+7%)	12,397 (+20%)	16,775 (+35%)
<b>Concepts with one or more non-suppressible ambiguity</b>	6,501	8,074 (+24%)	8,754 (+8%)	10,416 (+19%)	12,387 (+19%)
<b>Cases of ambiguity</b>	4,361	5,571 (+28%)	6,014 (+8%)	7,204 (+20%)	10,018 (+39%)
<b>Cases of non-suppressible ambiguity</b>	4,139	5,311 (+28%)	5,752 (+8%)	6,824 (+19%)	9,521 (+40%)

Table 1. Measures of ambiguity in the UMLS Metathesaurus

---

1. This year there are 138 exceptions to this rule.

relative to the previous year. Some examples will clarify the meaning of the counts in the table. There are 23 Metathesaurus strings ‘Protocols <n>’ for n ranging from 1 to 23; these strings occur in 23 distinct concepts and represent a single case of ambiguity. Some concepts contain more than one ambiguous string, i.e. they are ambiguous in more than one (not necessarily distinct) way. In fact, the concept ‘Optic Nerve Glioma, Childhood’ (C0278653) has 37 ambiguous strings, more than any other concept. A more manageable example of a concept with multiple ambiguities is ‘Mastodynia’ (C0024902) which has the following six ambiguous strings:

- Mastodynia <1>
- Breast pain <1>
- Pain breast <1>
- Mastalgia <2>
- Ache breast <1>
- Soreness breast <1>

All of these strings are ambiguous with the same concept ‘BREAST PAIN FEMALE’ (C0553713); so in this case, there is really only one ambiguity expressed six different ways. Finally, separate counts are given in the last line of the table after restricting focus to those Metathesaurus strings which are not suppressible synonyms. Since suppressible synonyms are generally unhelpful for NLP purposes, these counts give a more accurate view of Metathesaurus ambiguity.

The information in Table 1 shows that the amount of ambiguity in the Metathesaurus increased in 2004 significantly over the previous year and seems to be more than double that of 2000. Note, however, that the values for the different measures are not consistent. For example, the 39% increase in cases of ambiguity reported in 2004 is somewhat misleading because many of these cases are actually of degree 1 and hence not true cases at all. Ignoring the degree 1 cases, there are only 9,190 cases in 2004, a 28% increase over the corresponding number of cases, 7,171, in 2003. Even more interesting, ignoring degree 1 cases for non-suppressible ambiguity, the 40% increase becomes a 6% decrease (5,952 cases in 2004 instead of 9,521 vs. 6,330 cases in 2003 instead of 6,824). This instability in the numbers probably reflects the effect that marking strings suppressible has on removing false ambiguity.

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 occurs.<sup>1</sup> For example ‘Other <n>’ has degree 54. Table 2 contains the distribution of ambiguities in the Metathesaurus according to 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.

Ignoring suppressible synonyms produces the more realistic distribution shown in Table 3. Most of the ambiguity of degree greater than eight has disappeared, and all of that would disappear if appropriate strings were marked as suppressible. For example, ‘Other <n>’ now has degree two. One of these cases, ‘Other’ (C0205394), is legitimate; and the other, ‘Other location of complaint’

---

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.

Degree of ambiguity	2000 cases	2001 cases	2002 cases	2003 cases	2004 cases
56	0	0(0%)	0(0%)	1(-)	0(-100%)
55	0	1(-)	1(0%)	0(-100%)	0(0%)
54	1	0(-100%)	0(0%)	0(0%)	1(-)
23	1	1(0%)	1(0%)	1(0%)	1(0%)
18	1	1(0%)	2(+100%)	2(0%)	2(0%)
17	0	59(-)	58(-2%)	58(0%)	58(0%)
16	2	2(0%)	2(0%)	2(0%)	2(0%)
8	3	3(0%)	3(0%)	3(0%)	3(0%)
7	2	2(0%)	5(+150%)	6(+20%)	7(+17%)
6	2	4(+100%)	8(+100%)	8(0%)	7(-13%)
5	9	22(+144%)	21(-5%)	24(+14%)	22(-8%)
4	63	55(-13%)	59(+7%)	79(+34%)	76(-4%)
3	373	418(+12%)	472(+13%)	594(+26%)	600(+1%)
2	3,534	4,368(+24%)	4,641(+6%)	5,541(+19%)	7,250(+31%)
1	370	635(+72%)	741(+17%)	885(+19%)	1,989(+125%)
<b>Total</b>	<b>4,361</b>	<b>5,571(+28%)</b>	<b>6,014(+8%)</b>	<b>7,204(+20%)</b>	<b>10,018(+39%)</b>

Table 2. Metathesaurus ambiguity distribution by degree

Degree of ambiguity	2000 cases	2001 cases	2002 cases	2003 cases	2004 cases
56	0	0(0%)	0(0%)	0(0%)	0(0%)
55	0	0(0%)	0(0%)	0(0%)	0(0%)
54	0	0(0%)	0(0%)	0(0%)	0(0%)
23	1	1(0%)	1(0%)	1(0%)	1(0%)
18	1	1(0%)	1(0%)	1(0%)	1(0%)
17	0	1(-)	1(0%)	0(-100%)	0(0%)
16	1	0(-100%)	0(0%)	1(-)	1(0%)
8	3	3(0%)	3(0%)	3(0%)	3(0%)
7	2	2(0%)	3(+50%)	4(+33%)	2(-50%)
6	1	2(+100%)	2(0%)	1(-50%)	1(0%)
5	6	7(+17%)	10(+43%)	11(+10%)	7(-36%)
4	34	37(+9%)	46(+24%)	44(-4%)	42(-5%)
3	303	385(+27%)	448(+16%)	473(+6%)	416(-12%)
2	3,435	4,511(+31%)	4,878(+8%)	4,935(+1%)	4,309(-13%)
1	352	361(+3%)	359(-1%)	1,350(+276%)	4,738(+251%)
<b>Total</b>	<b>4,139</b>	<b>5,311(+28%)</b>	<b>5,752(+8%)</b>	<b>6,824(+19%)</b>	<b>9,521(+40%)</b>

Table 3. Metathesaurus ambiguity distribution after removing suppressibles

(C0220886), should be marked as suppressible. Suppressible synonyms are ignored for the remainder of this report.

Section 2 of this report describes only the most notable cases of ambiguity in the Metathesaurus, i.e., the cases of degree five 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. And Section 3 provides some conclusions derived from this study.

## 2. 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 five 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*.

### 2.1 ‘Protocols <n>’ (degree 23) <no change from last year>

Except for ‘Protocols <2>’ with preferred name ‘Protocols documentation’, ‘Protocols <n>’ are generalizations of their preferred names and should be marked as suppressible. Their concepts are

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

## 2.2 ‘Patient Education Plans <n>’ (degree 18) <no change from last year>

All 18 strings ‘Patient Education Plans <n>’ are generalizations and should be suppressed. Their concepts are

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

## 2.3 ‘Assessment <n>’ (degree 16) <no change from last year>

Similarly, all 17 strings ‘Assessment <n>’ are generalizations and should be suppressed. Note that the degree of ambiguity here is 16 because ‘Assessment <5>’ and ‘Assessment <17>’ both belong to the same concept, ‘Assessment: Cognition’. The concepts involved in this ambiguity are

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

**17. C0870300: Assessment: Cognition (Assessment <17>)****2.4 ‘cde genotype <n>’ (degree 8) <no change from last year>**

The 8 strings ‘cde genotype <n>’ differ only in the case of ‘c’, ‘d’ or ‘e’. They seem to be legitimate concepts from RCD99. In a case sensitive environment, there is no ambiguity; the presence of strings with ambiguity designators, although not harmful, is not necessary. When case is ignored, the ambiguity exists and the additional strings are warranted (albeit possibly confusing because they appear with mixed case). Thus, the ambiguous forms are correctly not suppressed.

**2.5 ‘Driver injured in collision with other and unspecified motor vehicles in nontraffic accident <n>’ (degree 8) <no change from last year>**

The 8 strings ‘Driver injured ... <n>’ occur in the Read Codes and ICD-10 with some variation in the wording. The concepts for these strings are

1. **C0476822: Driver pedal cycle injured in collision with other and unspecified motor vehicles in nontraffic accident (<1>)**
2. **C0476905: Motorcycle driver injured in collision with other and unspecified motor vehicles in nontraffic accident (<2>)**
3. **C0476973: Car driver injured in collision with other and unspecified motor vehicles in nontraffic accident (<4>)**
4. **C0477136: Bus driver injured in collision with other and unspecified motor vehicles in nontraffic accident (<7>)**
5. **C0496239: Driver of three-wheeled motor vehicle injured in collision with other and unspecified motor vehicles in nontraffic accident (<3>)**
6. **C0496408: Pick-up truck or van driver injured in collision with other and unspecified motor vehicles in nontraffic accident (<5>)**
7. **C0496481: Heavy transport vehicle driver injured in collision with other and unspecified motor vehicles in nontraffic accident (<6>)**
8. **C0596026: Driver injured in collision with other and unspecified motor vehicles in nontraffic accident (<8>)**

where, for ease of reading, the strings ‘Driver injured... <n>’ are shown simply as ‘<n>’. Except for the last case, the ambiguous strings are generalizations and should be suppressed.

**2.6 ‘Driver injured in collision with other and unspecified motor vehicles in traffic accident <n>’ (degree 8) <no change from last year>**

This case is analogous to the previous one (section 2.5); all but the last should be suppressed. The relevant concepts (with abbreviated ambiguity designator strings as before) are

1. **C0476826: Pedal cycle driver injured in collision with other and unspecified motor vehicles in traffic accident (<1>)**
2. **C0476909: Motor cycle driver injured in collision with other and unspecified motor vehicles in traffic accident (<2>)**
3. **C0476977: Car driver injured in collision with other and unspecified motor vehicles in traffic accident (<4>)**
4. **C0477140: Bus driver injured in collision with other and unspecified motor vehicles in traffic accident (<7>)**

5. **C0496241: Three-wheeled motor vehicle driver injured in collision with other and unspecified motor vehicles in traffic accident (<3>)**
6. **C0496410: Pick-up truck or van driver injured in collision with other and unspecified motor vehicles in traffic accident (<5>)**
7. **C0496483: Heavy transport driver injured in collision with other and unspecified motor vehicles in traffic accident (<6>)**
8. C0596027: Driver injured in collision with other and unspecified motor vehicles in traffic accident (<8>)

### **2.7 'Passenger injured in collision with other and unspecified motor vehicles in nontraffic accident <n>' (degree 7) <no change from last year>**

This case is analogous to the one in section 2.5 except that all no strings should remain unsuppressed. The relevant concepts are

1. **C0476823: Pedal cycle passenger injured in collision with other and unspecified motor vehicles in nontraffic accident (<1>)**
2. **C0476906: Motor cycle passenger injured in collision with other and unspecified motor vehicles in nontraffic accident (<2>)**
3. **C0496240: Three-wheeled motor vehicle passenger injured in collision with other and unspecified motor vehicles in nontraffic accident (<3>)**
4. **C0496324: Passenger of car injured in collision with other and unspecified motor vehicles in nontraffic accident (<4>)**
5. **C0496409: Pick-up truck or van passenger injured in collision with other and unspecified motor vehicles in nontraffic accident (<5>)**
6. **C0496482: Heavy transport vehicle passenger injured in collision with other and unspecified motor vehicles in nontraffic accident (<6>)**
7. **C0496486: Bus passenger injured in collision with other and unspecified motor vehicles in nontraffic accident (<7>)**

### **2.8 'Passenger injured in collision with other and unspecified motor vehicles in traffic accident <n>' (degree 7) <no change from last year>**

This case is analogous to the previous one; all ambiguous strings (except for sense <3>) should be suppressed. The relevant concepts are

1. **C0476827: Pedal cycle passenger injured in collision with other and unspecified motor vehicles in traffic accident (<1>)**
2. **C0476910: Motor cycle passenger injured in collision with other and unspecified motor vehicles in traffic accident (<2>)**
3. C0496242: Passenger injured in collision with other and unspecified motor vehicles in traffic accident (<3>)
4. **C0496326: Passenger of car injured in collision with other and unspecified motor vehicles in traffic accident (<4>)**
5. **C0496411: Passenger of pick-up truck or van vehicle injured in collision with other and unspecified motor vehicles in traffic accident (<5>)**
6. **C0496484: Passenger of heavy transport vehicle injured in collision with other and unspecified motor vehicles in traffic accident (<6>)**
7. **C0496488: Passenger of bus injured in collision with other and unspecified motor vehicles in traffic accident (<7>)**

**2.9 ‘Cold <n>’ (degree 6) <no change from last year>**

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

1. C0009264: cold temperature (Cold <1>)
2. C0009443: Common Cold (Cold <2>)
3. **C0010412: Cold Therapy (Cold <4>) [consider suppressing]**
4. C0024117: Chronic Obstructive Airway Disease (COLD <3>) (which has string ‘Chronic Obstructive Lung Disease’)
5. C0234192: Cold Sensation (Cold <5>)
6. C0719425: Cold brand of chlorpheniramine-phenylpropanolamine (Cold <6>)

Senses <3>, <4> and <6> are currently suppressed by MetaMap for mainly practical reasons. The only sense that might be marked as suppressible in the Metathesaurus is sense <4>, ‘Cold Therapy’, because it is extremely rare to find the word *cold* by itself meaning ‘Cold Therapy’.

**2.10 ‘[SO] Premolar tooth <n>’ (degree 5) <no change from last year>**

All ambiguous strings are generalizations and should be suppressed.

1. **C0005373: Bicuspid ([SO] Premolar tooth <1>)**
2. **C0447259: Permanent upper right first premolar tooth ([SO] Premolar tooth <2>)**
3. **C0447265: Permanent upper left second premolar tooth ([SO] Premolar tooth <3>)**
4. **C0447266: Permanent upper left first premolar tooth ([SO] Premolar tooth <4>)**
5. **C0447299: Permanent lower right second premolar tooth ([SO] Premolar tooth <5>)**

**2.11 ‘Buttercup <n>’ (degree 5)**

This seems to be legitimate ambiguity

1. C0330264: Ranunculus (Buttercup <3>)
2. C0697205: Ranunculus bulbosus (Buttercup <2>)
3. C0939924: Ranunculus bulbosus, Homeopathic preparation (Buttercup <4>)
4. C1176287: Cimicifuga racemosa, flower essences (buttercup <6>)
5. C1176288: Ranunculus acris, flower essences (Buttercup <5>)

**2.12 ‘CAM <n>’ (degree 5) <no change from last year>**

Suppress ambiguous forms (MetaMap only) because they are either abbreviatory or, in the case of ‘CAM brand of Ephedrine Hydrochloride’, a brand name.

1. C0007578: Cell Adhesion Molecules (CAM <1>)
2. C0178551: chorioallantoic membrane (CAM <3>)
3. C0678112: CAM brand of Ephedrine Hydrochloride (CAM <2>)
4. C0713465: Cam, topical lotion (Cam <4>)
5. C0936077: Complementary therapies (CAM <5>)

**2.13 ‘CD <n>’ (degree 5) <no change from last year>**

This is legitimate ambiguity (although it is not clear why ‘CP protocol’ is related to ‘CD’). However, the ambiguous forms are suppressed in MetaMap because they are abbreviations.



1. C0006632: Cadmium (Cd <3>)
2. C0056447: CP protocol (CD <2>)
3. C0079141: Compact discs (CD <4>)
4. C0332140: Diagnosis, clinical (CD <1>)
5. C0700300: candela (cd <5>)

#### 2.14 ‘Cowslip <n>’ (degree 5)

This seems to be legitimate ambiguity

1. C0330269: Cowslip (Cowslip <1>)
2. C0772340: Cowslip extract (COWSLIP <2>)
3. C0886348: Primula veris preparation (Cowslip <3>)
4. C1167628: Caltha palustris preparation (Cowslip <5>)
5. C1253907: Primula veris (cowslip <4>)

#### 2.15 ‘Dandelion <n>’ (degree 5)

This seems to be legitimate ambiguity

1. C0440015: Dandelion Extract (DANDELION <2>)
2. C0877851: Taraxacum officinale (Dandelion <1>)
3. C0939896: Taraxacum officinale, Homeopathic preparation (Dandelion <5>)
4. C1093500: Taraxacum (Dandelion <3>)
5. C1256220: Taraxacum officinale, flower essence (Dandelion <4>)

#### 2.16 ‘Premolar tooth <n>’ (degree 5) <no change from last year>

Like ‘[SO] Premolar tooth <n>’, all ambiguous strings are generalizations and should be suppressed.

1. **C0005373: Bicuspid (Premolar tooth <1>)**
2. **C0447259: Permanent upper right first premolar tooth (Premolar tooth <2>)**
3. **C0447265: Permanent upper left second premolar tooth (Premolar tooth <3>)**
4. **C0447266: Permanent upper left first premolar tooth (Premolar tooth <4>)**
5. **C0447299: Permanent lower right second premolar tooth (Premolar tooth <5>)**

#### 2.17 ‘Silver <n>’ (degree 5)

Since the measurement of a substance and that substance are not the same thing, the measurement sense could be suppressed.

1. C0037125: Silver (SILVER <1>)
2. C0037125: Silver (silver <2>)
3. **C0202473: Silver measurement (Silver <4>) [consider suppressing]**
4. C0947651: Silver preparation (SILVER <3>)
5. C1161330: Argentum metallicum, silver, Homeopathic preparation (Silver <5>)

### 3. Conclusions

Some concepts have ambiguous forms which should be marked as suppressible, where *ambiguous forms* means both the string with an ambiguity designator and one or more strings without it. Note that some concepts such as ‘Assessment: Cognition’ have more than one ambiguous form with ambiguity designator (‘Assessment <5>’ and ‘Assessment <17>’ in this case). 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’. 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’ 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, 22 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 ‘Other complications of procedures NEC in ICD10’, contain meta information. In this case it is the name of the vocabulary, ICD10 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.

One limitation of this study is that there was no follow-up of ambiguous strings discovered while examining the strings for a given case. For example, the set of all strings for the concepts containing a 'sound measurement <n>' string include the ambiguous strings 'Ultrasonography <n>', 'Echography <n>', 'Ultrasound <n>', 'Echotomography <n>' and 'ultrasound scanning <n>'. If the ambiguous forms for 'sound measurement <n>' are to be marked as suppressible, how does that affect the other ambiguous strings? Each of them is part of a cluster of concepts, possibly different from the one for 'sound measurement <n>'. Although it is probably not necessary to explore the new concept clusters, it is essential to examine the original cluster for additional ambiguous forms to suppress.