Annotator Guidelines
for the WSD Test Collection

Alan R. Aronson

July 11, 2000

These guidelines serve as recommendations for annotators as they disambiguate the instances of ambiguity constituting the Indexing Initiative's Word Sense Disambiguation Test Collection (WSD TC). The Test Collection is being built by considering 50 frequently occurring cases of ambiguity, where ambiguity is defined as the inability of MetaMap to choose among competing concepts for some text taken from a MEDLINE citation. For example, the 1999 UMLS Metathesaurus has five concepts corresponding to the text *cold*: cold temperature, the sensation of cold, the common cold, an acronym COLD, and cold therapy. Each of these concepts has the string "cold" (with possible case variation) as a synonym for the concept. For each ambiguity case, 100 instances of the ambiguity have been randomly selected from MEDLINE citations with entry month in 1998. The Test Collection will be constructed by several annotators each of whom will choose a sense (or "none of the above") for each instance of each case of ambiguity. The construction process will be accomplished using a flexible web interface developed by Jim Mork. A commonly understood set of guidelines for the annotation process will hopefully reduce inter-annotator disagreement.

Each ambiguity case has the following supporting material:

- the preferred concept name;
- the semantic types of the concept;
- definition(s) when available; and
- a hierarchy, when available, for the concept taken from one of the vocabularies providing one.

Note that the annotator can review the Metathesaurus information for a concept directly to obtain further information (especially additional hierarchies) about the concept. This motivates the first guideline:

1. **Make sure you fully understand the competing senses for a given case of ambiguity before annotating its 100 instances.**

While testing the TC interface, we were reminded that the Metathesaurus does not necessarily contain all relevant senses of a given word. Furthermore, annotating the examples from MEDLINE often affects one's a priori understanding of the senses. Thus:

2. **Review your understanding of the senses as required throughout the annotation process.**

Also,

3. **Don't hesitate to pass an instance that isn't clear to you; you can return to it later.**

and

4. **Take frequent breaks; this should increase both your accuracy and your enjoyment of the annotation process.**

Discussions held while selecting the 50 cases of ambiguity for inclusion in the TC made it clear that there is a tendency to broaden the task from selecting one from a number of possible senses to finding the best concept that represents the text. This changes the disambiguation task to an indexing task which is important but not relevant for the construction of the TC. So:

5. **Stay focused on selecting the best sense from among the senses provided rather than finding a better concept altogether.**

One of the most difficult aspects of deciding whether or not to include a given ambiguity case in the TC was reconciling our differing understanding of the senses. One of the biggest factors affecting our perspective was how strictly we viewed the Metathesaurus information (definitions, but even more so, hierarchies) defining each sense. Since the Metathesaurus is a growing body of knowledge, I think it best to take a relaxed view on how the current version of the Metathesaurus constrains the sense. Therefore,

6. **Take a broad perspective on the Metathesaurus information defining the various senses to be disambiguated; try to capture the spirit rather than the letter of the "law".**

Finally, we have found "Replacement Rule" (or "Susanne's Law") to be useful in making some decisions. The idea is to replace the ambiguous text with a full name for a given sense. If the sentence does not make sense after the replacement, then the sense probably is not correct. For example, replacing *radiation* in the sentence *The patient underwent radiation treatment for ...* with "Radiation therapy" produces *The patient underwent ratiation therapy treatment for ...* . This does not sound right, so the "Radiation therapy" sense is discarded in favor of the "Electromagnetic Energy" sense.

7. **Use the "Replacement Rule" to eliminate senses that cannot replace the text they represent.**