Computational Linguistics 1 - Fall 2011
CMSC/LING723, LBSC 744
Homework 6 - Due 6 December 2011

Submission Guidelines

1. Print out any written portion of the assignment and submit in class on December 6, *as well as* emailing your writeup, in .pdf format, to `compling723.fall2011@gmail.com`.

2. Any code you write for this homework should also be submitted to `compling723.fall2011@gmail.com`.

3. Be sure include your full name and e-mail in your submission, and clearly indicate each problem number in your solution set.

Problem 1: Word Senses

(a) Pick three verbs of your choice. Google for them. Take ten snippets in which they appear (make sure that they appear as verbs: you might have to look at more than the top ten hits to get this). For each snippet, describe the sense in which this verb is being used. Which of these are the same sense (in your opinion) and which are not.

If youre a non-native speaker of English and would prefer to do this in your native language, that's fine. But please provide translations so that I can understand what you're talking about!

Problem 2: WordNet

(a) Write down, in English and without using WordNet or NLTK, between 5 and 10 different senses of the verb (not the noun) *break*. For example, here are two:

- **Sense**: break an object into pieces.
  Example: Edgar broke the vase.

- **Sense**: break a bone.
  Example: Mildred broke her wrist.

Try to do this without a dictionary if you can, but if you're not a native speaker of English, use a dictionary if you need to.

(b) Use WordNet (online, or using NLTK as Prof. Boyd-Graber showed us in class) to look up the verb senses for break. Which WordNet senses do your senses from part (a) match, if any? (One of your senses might match more than one WordNet sense, of course.) For example, Sense 1: matches WN senses 2,3,4,5. (Please list the synsets for each numbered sense, so as to avoid confusion in grading.)

(c) Do any of your senses group naturally into a class with common elements of meaning? How would you group them? (Use a hierarchy if that makes more sense.) Hint: You should examine the list of 5 to 10 senses in the context of the WordNet structure and determine whether there is a way to group these 5 to 10 senses into a smaller number of equivalence classes.

(d) Explore the WordNet hierarchy for one of the senses of the word break. List some of the related troponyms and hypernyms (including "sister" terms if available). Do these seem reasonable to you?

Problem 3: Semantic Roles

(Copied from SaLP, first edition, exercise 16.5) Assign the various verb arguments in the following WSJ examples to their appropriate thematic roles, using the set of roles in Figure 16.9:*

1. The intense heat buckled the highway about three feet.

2. He melted her reserve with a husky-voiced paean to her eyes.

3. But Mingo, a major Union Pacific shipping center in the 1890s, has melted away to little more than the grain elevator now.

Problem 4: Normalization

A number of weblog articles have been made available on the class list. Weblog data is hard for NLP because it is so different from news articles, which many of our systems are trained on. Choose one article, identify which you've chosen, and discuss how it's different from a typical news article (for example, from our Wall St Journal corpus).

(a) Does the weblog differ just in terms of vocabulary? Or does it also differ in the syntactic structure of the sentences (just eyeball the sentences, don't try to run your parser on them)?

(b) Try to edit your weblog article such that it looks more like a news article. Be creative, but also reasonable—imagine that we're going to build systems that can copy what you've done. Submit your edited article with your writeup, and include in the writeup a general description of what you changed, and why.

---

*The list is: Agent, Experiencer, Force, Theme, Result, Content, Instrument, Beneficiary, Source and Goal.