L114: Coursework I (ticked): An exploration of the senses of "line"

Simone Teufel

22/10/2014

Consider the corpus examples for the noun "line" on the L114 course web-page (file line.html). The file contains 150 sentences containing the target word, extracted at random from the British National Corpus. Your task is to explore the senses of "line" on the basis of the corpus evidence given. In particular, the aim of this exercise is to explore the varying difference between word senses, by drilling down to some pairs of senses and showing how one could explore the differences empirically in an actual example.

- How many different senses of "line" are illustrated in the examples, in your opinion? Remember from the lecture that word senses can be defined at many different levels of granulary. There is no hard and fast right answer.

- To start answering this question, build sense clusters from the sentences together which in your opinion illustrate the same sense. You can do this bottom-up (by merging senses), or top-down (by splitting senses)

- Ignore compound words ("line thickness"; "tram line") and accidental verb occurrences. The reason for this is that it is difficult to distinguish senses in a compound context, due to the additional relation between premodifier and head noun. This is particularly so if the target word is the premodifier. Also ignore idioms ("in line").

- Cluster as many sentences as you think are necessary for determining how many senses there are. You may not need to cluster all sentences, but you can if you want to.

- After clustering the sentences, check that different clusters really carry different senses (no merging possible), and that each cluster really only represents one sense (no splitting possible).

- Write a definition for each sense, and select one representative sentence from the list for each sense (Sentence IDs are enough).

- For several interesting pairs of senses of your choice, demonstrate that the senses really are different, using appropriate ambiguity tests. Use of

different tests is encouraged. Demonstration of test application is a core aspect of the task.

- A particular note about Zeugma: this is a test for rather far-distant senses; it's quite difficult to make this test work, almost like inventing a joke with very strict constraints on what you can do (similar to writing poetry).

- NOTE: Do not simply list senses found in an existing sense inventory (e.g., WordNet). Your task is to arrive at an empirical observation purely from the data. You can however, include a post-hoc comparison of your set of senses with existing dictionaries.

- Submit a report that includes your list of senses with definitions. Indicate how you arrived at the list, including how many sentences you clustered.

- Summary: your report should consist of 3 parts:

    - List of senses with definitions
    - Description of how you arrived at them
    - Tests between pairs of senses – your choice how many pairs, your choice how many tests (aim for each kinds of tests at least once).

- Maximum length of report: 3 pages A4.

- Submission deadline: Tuesday 28 October, 12noon