

L113: Coursework I (ticked): An exploration of the senses of “line”

Simone Teufel

12/10/2012

Consider the corpus examples for the noun “line” on the L113 course webpage (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.

- How many different senses of “line” are illustrated in the examples?
- Start by clustering the sentence together which illustrate the same sense.
- Ignore compound words (“line thickness”; “tram line”) and accidental verb occurrences.
- 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).
- Submit a report that includes your list of senses with definitions. Indicate how you arrived at the list, including how many sentences you clustered.
- For each pair of similar senses, demonstrate that they are really different, using ambiguity tests of your choice (the more different ones, the better). This is a core aspect of the task.
- 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.
- Maximum length of report: 3 pages A4.
- Submission deadline: Friday 19 October, 5pm