**Adjectives**

<http://aclweb.org/anthology/C00-1044>

* Hatzivassiloglou and McKeown (1997) presented a method for automatically assigning a + or orientation label to adjectives known to have some semantic orientation.
* Bruce and Wiebe (2000) performed a statistical analysis of the assigned classifications, finding that adjectives are statistically significantly and positively correlated with subjective sentences in the corpus on the basis of the log-likelihood ratio test statistic G2
* The probability of a sentence being subjective, simply given that there is at least one adjective in the sentence, is 56%, even though there are more objective than subjective sentences in the corpus. In addition, Bruce and Wiebe identified a type of adjective that is indicative of subjective sentences: those Quirk et al. (1985) term dynamic, which “denote qualities that are thought to be subject to control by the possessor” (p. 434). Examples are “kind” and “careful”. Bruce and Wiebe manually applied syntactic tests to identify dynamic adjectives in half of the corpus mentioned above. We include such adjectives in the analysis below, to assess whether additional lexical semantic features associated with subjectivity help improve predictability.
* Wiebe et al. (1999) developed an automatic system to perform subjectivity tagging. In 10-fold cross validation experiments applied to the corpus described above, a probabilistic classifier obtained an average accuracy on subjectivity tagging of 72.17%, more than 20 percentage points higher than the baseline accuracy obtained by always choosing the more frequent class. A binary feature is included for each of the following: the presence in the sentence of a pronoun, an adjective, a cardinal number, a modal other than will, and an adverb other than not. They also included a binary feature representing whether or not the sentence begins a new paragraph. Finally, a feature was included representing co-occurrence of word tokens and punctuation marks with the subjective and objective classification. An analysis of the system showed that the adjective feature was important to realizing the improvements over the baseline accuracy.

The data was coded according to part of speech and the number of adjectives for each article was extracted. This was done using the packages rdrpostagger for English, and Chinese.misc for Chinese. Graphs were created depicting the amount of adjectives in total, as well as the adjectives/total words in each article. This shows what percent of the words were adjectives. Articles with more adjectives are hypothesized to display more sentiment and opinions. The word count trend of the articles is also shown.

Scatterplots were made with date on the x axis and tokens on the y axis. This was done for each source - China Daily, People’s Daily, bullpiano, and xiakedao. The scatterplot was overlaid with a loess regression line implemented with the geom\_smooth function using the package ggplot2.