Data Analysis 1
C. Hidey
February 11, 2015

Problem 1:
I hypothesize that when reviewers use contrasting sentiment (both a positive and negative adjective), the order in which they present the words is important. I expect that the conjunctions that express a contrast ('but', 'yet') tend to have the negative polarity word occur first, whereas conjunctions that express an equal relationship ('and', 'or', 'nor', 'versus') tend to have the positive polarity word first. The following counts from the data (where the adjectives have opposite polarity) bear this out (all the results follow a binomial distribution and are statistically significant except for 'yet', where there is not enough data).

Conjunction	Order 1 (Positive, Negative)	Order 2 (Negative, Positive)
but	1155	1826
yet	18	34
and	12157	7646
or	3924	910
nor	223	86
versus	387	5

Consider one case where we have both positive and negative sentiment: 'good versus evil' or 'good and evil'. We would not often say 'evil versus good' unless we want to explicitly put emphasis on 'evil,' as if we were siding with evil. We may intuitively put the positive polarity words first because we deem that side/view more important.

Problem 2:
From examining the data, there are a few features that stand out that might be useful. Assuming that the data is permitted to be used as is (with the identifying tokens removed), the following features seem to be indicative:

1. Capitalization. The FML site has well-written text (correct capitalization, spelling, formatting), which may be a site requirement. In contrast, IMMD contains many examples with sentences not capitalized at the beginning, the first-person pronoun not capitalized, or words in all caps for emphasis.

2. Cursing. The intuition is that people curse more when something bad happens and some of the examples indicate this (especially since FML has a curse word in the title). In contrast, none of the IMMD examples I examined had curse words.

3. Emoticons/Punctuation. Many of the IMMD examples contain smiley face emoticons and we would not expect to see that in FML. There is a lack of emoticons in FML that may be due to some kind of posting requirement as well. Some of the IMMD examples also contain additional exclamation points for emphasis, and the FML examples I believe would be less likely to have them.