Participants to analyze sentiments to beginwith, and we

Participants in the BLGNLP2017 shared task wereinvited to either build sentiment analysis systems(as a Builder team) or break them, by compilinglinguistically motivated test cases that resultin false predictions (as a Breaker team). A data setof movie reviews was provided as the domain forparticipating systems and as a source for generatingbreaking test cases. As a Breaker team, ourgoal was to construct minimal pairs consisting ofa review from the source data set, and a modifiedversion of the review that would be used to evaluatethe robustness or sensitivity of the participatingsystems predictions. The modified version of eachreview could either preserve the sentiment of theoriginal review, or reverse it.

Movie reviews from Rotten Tomatoes are a goodsource for comments full of sentiment, as the informalsetting provides for humor, pathos, wildcomparisons, sarcasm, artistic expressions and thelike. Hence, it was probably not an easy task forthe Builder systems to analyze sentiments to beginwith, and we tried to make it even harder. Basedon the sentiment analysis of our linguistic examplesit seems like there are several ways to trickthe Builder systems. In our own judgments of the provided items, wefollowed a positive/negative sentiment dichotomy, which was not always straightforward given thecomplexity of the data. However, even if a neutralsentiment option was included (as found inthe predictions of some of the Builder system) itwould not have accounted for the whole variation, as some items could have multiple plausible interpretations, affecting their perceived valence. Thus, it is important to bear in mind that the judgmentsprovided by us might not always coincide withthose of other people.

We begin this paper by describing the generalrationale we had employed in creating our testcases. We then present some examples of sentencesthat broke the Builder systems and discussthe nature of the errors, and the main difficultiesin analyzing sentiment. In addition, we discussthe linguistic processes that take place in inferringsentiment from the various examples.