Options for cleaning up data

[ ] For expressions that start with an apostrophe, concatenate that entry with the entry before it.(or maybe just delete them. They dont carry much emotional meaning)

[ ] Concatenate the "n't" tokens to the previous entry

[ ] Keep "/5" "/10" and take away the rest of the numbers. Possibly label dates with date token

[ ] Split words that have a period between them (period isnt deleted. Stays as entry)

Any punctuations delete other than periods and parentheses

Replace number entries with number token

Remove periods parentheses that are in multiple succession within an entry

Keep exclamation points / keep any full stop character

Observations

[ ] Real reviews seem to havemore parentheses based on our sample

<https://lost-contact.mit.edu/afs/cs.pitt.edu/projects/nltk/docs/tutorial/introduction/nochunks.html>