It was a dark and stormy night……

Identifying genre patterns in novels via NLP

Anne Nguyen
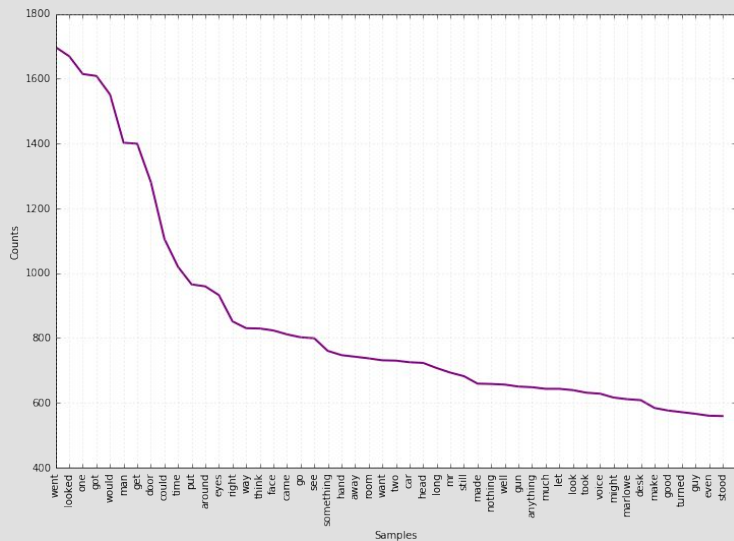5/31/2016

The Big Sleep (1939/1946)

Corpus: 7 Raymond
Chandler novels

Hardboiled/noir genre

Textual patterns

- 7 books: 560,000 words; 300,000 content; 15,500 unique
- girl: 454; gun: 698; cigarette: 480
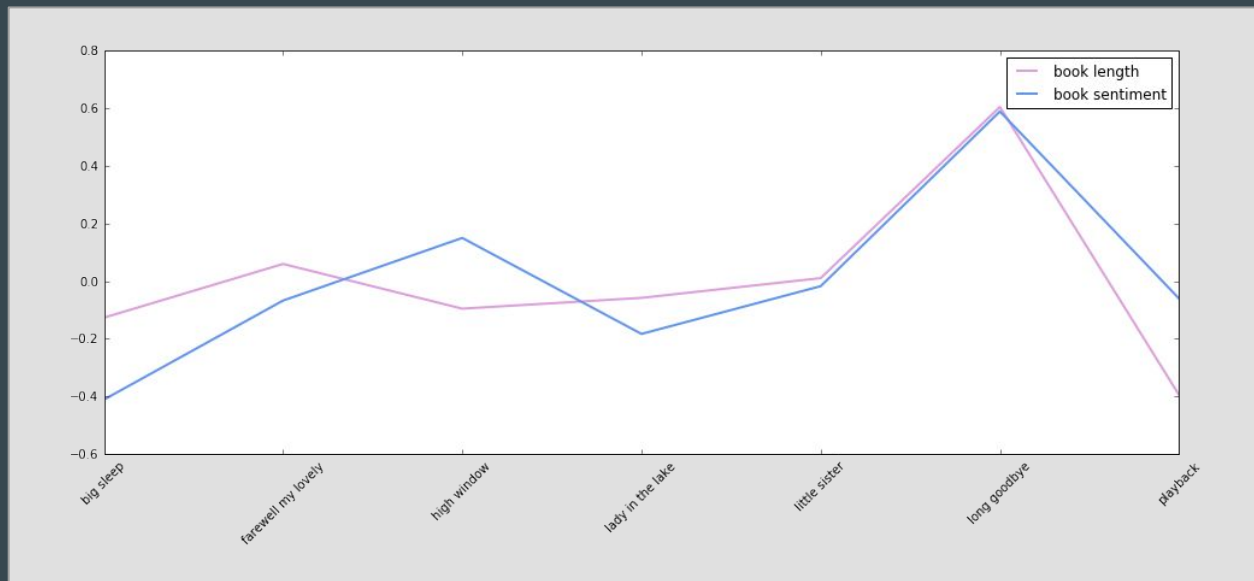- *The Big Sleep*: 12.3 words per sentence

Similes

platinumed hair shone like silver fruit bowl
fingers made movements like dying butterflies gun
85-cent dinner tasted like discarded mail bag
silvery tinkle sounded like putting away saucepans
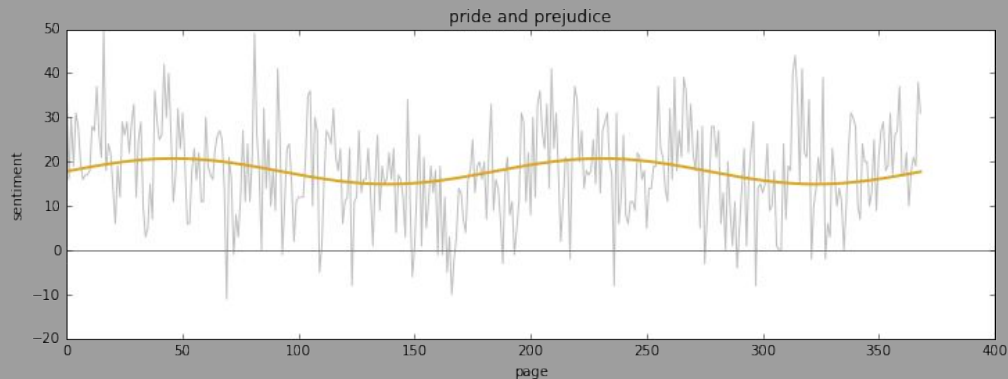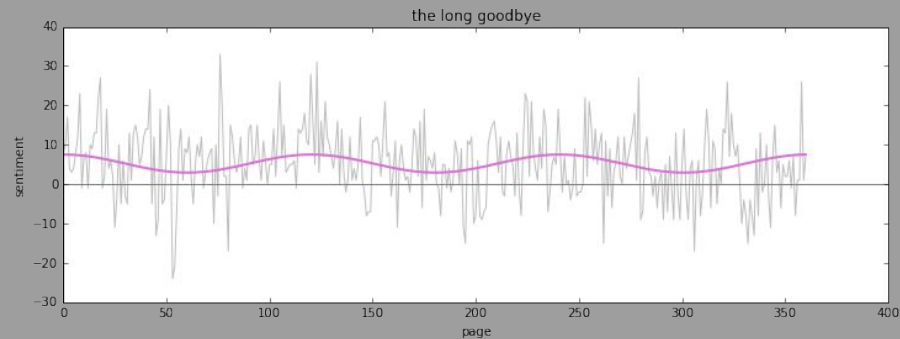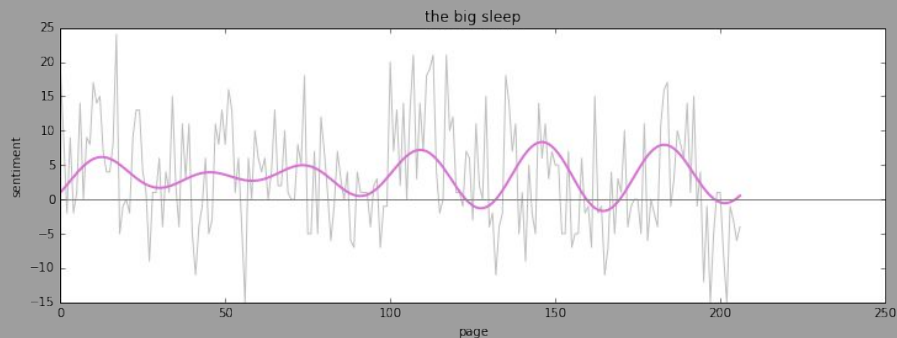snapped blackjack coiled like wounded snake went

Sentiment analysis

- Multiple lexicons compiled by Brian Foo (mainly NRC): 11k words, 40/60 pos/neg
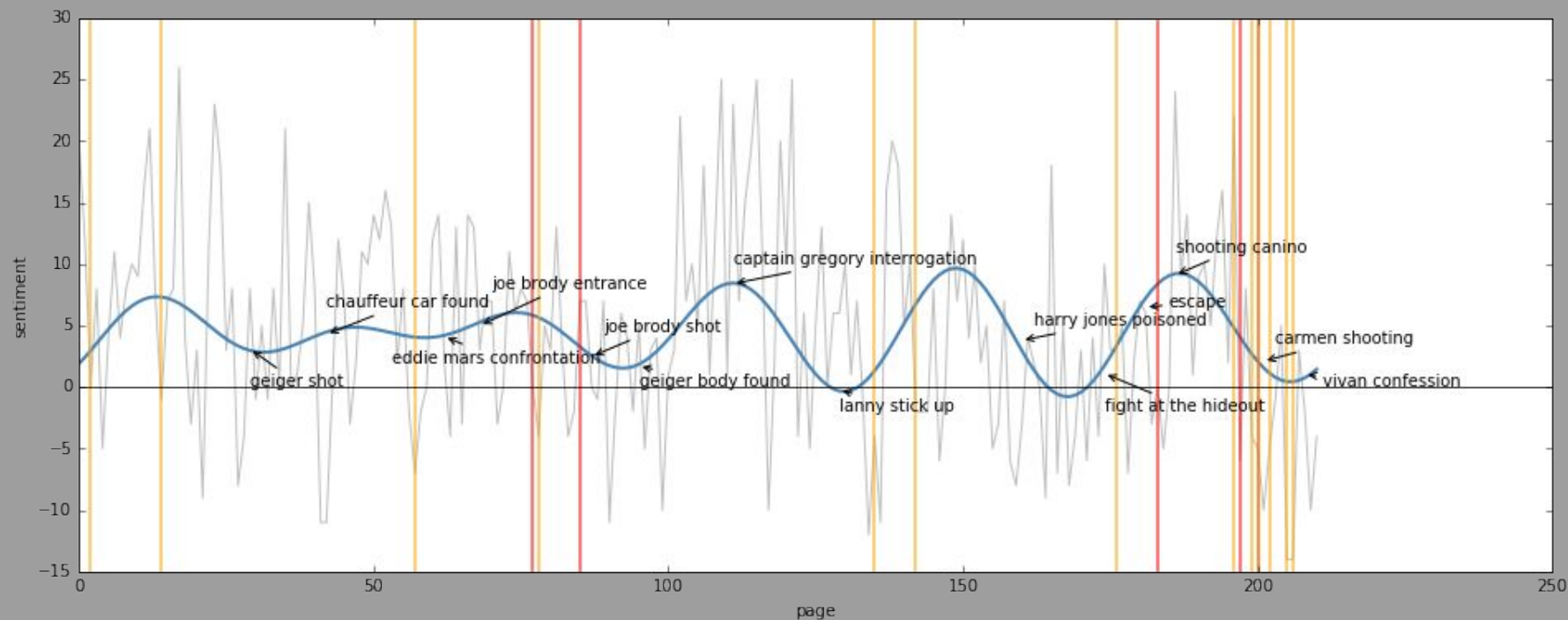- 6k out of 15k Chandler words tagged

Positive: ~14%
Negative: ~12%

Sentiment as plot structure

Words in lexicon: 15%
Positive: 9%
Negative: 4.5%
Words per sentence: 29

Sentiment in the Big Sleep

Topic analysis

By book: TFIDF/LDA, LSI, POS

degarmo + geiger + mitchell + kingsley + almor
geiger + degarmo + kingsley + lennox +regan

By sentence: TFIDF/LDA, LSI, NMF

door + eyes + came + still + stood + open + looking + mouth + light + house
see + anything + say + away + much + gun + french + turned + told + take + money
right + door + desk + hand + thought + stood + sat + open + chair + mouth + leaned + phone
reached + killed + softly + cigarette + smile + yeah + hands + glass

door eyes came face way open opened hand still head room car gun look away turned stood

Text generator

bit.ly/noirtext