I'm in agreement with the comment on your question that it's probably easiest (and fastest) to just provide all the variations you want to eliminate up front, unless you want to eliminate a lot more words than this but then it becomes more a question of spotting interesting ones than eliminating spurious ones.

1

You ask about speed, but you should be more concerned with accuracy. Both your suggestions will make a lot of mistakes, removing either too much or too little (for example, there are *a lot* of words that contain the substring "at"). I second the suggestion to look into the nltk module. In fact, one of the early examples in [the NLTK book](http://www.nltk.org/book) involves removing common words until the most common remaining ones reveal something about the genre. You'll get not only tools, but instruction on how to go about it.