NLTK Bot

ChatScript comes with a bot used to analyze documents called NLTK. You build it using :build nltk. It's still in its infancy.

You start by selecting what you want to do:

tally rootpos – read documents and list all words in it and their frequencies by root and part of speech. The sentence *I have loved animals since I was a boy* will show
2 = I~pro
1 = have~aux
1 = animal~n
1 = since~con
1 = be~v
1 = a~det
1 = boy~n
though not necessarily in that order. You can then sort the output log file to find most common words. Note that *I have loved love* will list *love~n* and *love~v* separately since their part of speech usage was different.

tally root – read documents and list all words in it and their frequencies by root. No part of speech information is provided so *I have loved love* will display 2 = love.

tally raw – read documents and list all words in it and their frequencies as given. Swims and swim are two separate listings.

By default it displays all words. If you add *min 2* after some command, then any word displayed must be found at least twice.

sentiment – read incoming sentences or documents and display the sentence with a summary of + and – affect at start and pluss and minuses after words that have affect.

reset – turn off existing mode, in preparation for changing to a new mode.

For document reading, you would name the file or directory to read like this:

:document REGRESS/moby_dick.txt - to read a file relative to top level ChatScript
:document READER/BOOKS/ - to read some directory relative to top level ChatScript

Tallying on a big document can take a moment, so the system outputs a . every 1000 sentences it reads.