

Week 1 – Corpus & Analysis

Lyubomira Dimitrova

January 2018

1 Source

www.chitanka.info - a free online library for books in Bulgarian (original works and translations).

2 Statistics

# Documents	# Tokens	# Types
21	1923379	98337

3 Graph

Counts of the 200 most frequent words

