

Contents

1	Word forms and their frequencies	1
1.1	Structure of entries and frequency data	2
1.2	Data in electronic form	3
1.3	Conception of a dictionary of word forms	3
1.4	On the definition of a word	4
2	Corpus	7
2.1	Data basis and extent of data collection	7
2.2	Data pre-processing	7
2.3	Proportions of old and new spellings	8
3	Statistical data on the word list	11
3.1	Character statistics	11
3.2	Vowels and consonants	13
3.3	Word length	14
3.4	Word structure	15
3.5	Zipf's law	17
3.6	Text coverage by the top-N most frequent words	17
3.7	Growth in word length	18
3.8	Number of syllables	20
3.9	Number of letter bigrams and trigrams	21
3.10	Summary of statistical data	22
4	References	23
5	Most frequent words orderd by rank	25
6	Alphabetical frequency list	37