

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	4
1.4	On the definition of a word	6
2	Corpus	9
2.1	Data basis and extent of data collection	9
2.2	Corpus timeline	10
2.3	Data pre-processing	10
3	Statistical data on the word list	13
3.1	Character statistics	13
3.2	Vowels and consonants	14
3.3	Word length	16
3.4	Word structure	17
3.5	Zipf's law	20
3.6	Text coverage by the top-N most frequent words	20
3.7	Growth in word length	21
3.8	Number of syllables	24
3.9	Number of letter bigrams and trigrams	25
3.10	Summary of statistical data	26
4	References	27
5	Most frequent words ordered by rank	29
6	Alphabetical frequency list	41