

Contents

1	Word forms and their frequencies	1
1.1	The Georgian language and Georgian alphabet	1
1.2	Genetic classification, dialectal differentiation and distribution	2
1.3	Sociolinguistic situation	3
1.4	A short history of the Georgian lexicography	4
1.5	Structure of entries and frequency data	5
1.6	Data in electronic form	7
1.7	Conception of a dictionary of word forms	7
1.8	On the definition of a word	9
2	Corpus	11
2.1	Data basis and extent of data collection	11
2.2	Corpus timeline	12
2.3	Data pre-processing	12
3	Statistical data on the word list	15
3.1	Comparison criteria	15
3.2	Character statistics	16
3.3	Vowels and consonants	16
3.4	Word length	19
3.5	Word structure: Hyphenation	20
3.6	Zipf's law	21
3.7	Text coverage of the top-N most frequent words	22
3.8	Growth in word length	22
3.9	Number of syllables	24
3.10	Number of letter bigrams and trigrams	25
3.11	Summary of statistical data	27
4	References	29
5	Most frequent words ordered by rank	31
6	Alphabetical frequency list	43