

Contents

1	Word forms and their frequencies	1
1.1	Previous frequency lists and dictionaries for Esperanto	1
1.2	Structure of entries and frequency data	2
1.3	Data in electronic form	4
1.4	Conception of a dictionary of word forms	4
1.5	On the definition of a word	6
1.6	Productivity and writing variants	7
2	Corpus	13
2.1	Data basis and extent of data collection	13
2.2	Corpus timeline	14
2.3	Data pre-processing	14
3	Statistical data on the word list	17
3.1	Comparison criteria	17
3.2	Character statistics	18
3.3	Vowels and consonants	18
3.4	Word length	21
3.5	Word structure	22
3.6	Zipf's law	24
3.7	Text coverage of the top-N most frequent words	25
3.8	Growth in word length	25
3.9	Number of syllables	29
3.10	Number of letter bigrams and trigrams	30
3.11	Summary of statistical data	31
4	References	33
5	Most frequent words ordered by rank	35
6	Alphabetical frequency list	47