

Contents

1	Word forms and their frequencies	1
1.1	Previous frequency dictionaries for Hungarian	1
1.2	Structure of entries and frequency data	4
1.3	Data in electronic form	5
1.4	Conception of a dictionary of word forms	6
1.5	On the definition of a word	8
2	Corpus	11
2.1	Data basis and extent of data collection	11
2.2	Corpus timeline	12
2.3	Data pre-processing	13
3	Statistical data on the word list	15
3.1	Comparison criteria	15
3.2	Character statistics	16
3.3	Vowels and consonants	18
3.4	Word length	19
3.5	Word structure	20
3.6	Zipf's law	23
3.7	Text coverage of the top-N most frequent words	23
3.8	Growth in word length	24
3.9	Number of syllables	27
3.10	Number of letter bigrams and trigrams	27
3.11	Summary of statistical data	29
4	References	31
5	Most frequent words ordered by rank	33
6	Alphabetical frequency list	45