

Contents

1	Word forms and their frequencies	1
1.1	Previous works on Indonesian word frequency lists	1
1.2	Structure of entries and frequency data	2
1.3	Data in electronic form	3
1.4	Conception of a dictionary of word forms	4
1.5	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	Comparison criteria	13
3.2	Character statistics	14
3.3	Vowels and consonants	14
3.4	Word length	15
3.5	Word structure	16
3.6	Zipf's law	19
3.7	Text coverage of the top-N most frequent words	20
3.8	Growth in word length	21
3.9	Number of syllables and syllable length	22
3.10	Number of letter bigrams and trigrams	24
3.11	Summary of statistical data	25
4	References	27
5	Most frequent words ordered by rank	29
6	Alphabetical frequency list	41