

Contents

1	Word forms and their frequencies	1
1.1	Previous frequency lists and dictionaries for Zulu	1
1.2	Structure of entries and frequency data	4
1.3	Data in electronic form	6
1.4	Conception of a dictionary of word forms	6
1.5	On the definition of a word	7
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	18
3.5	Word structure	19
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 and syllable length	25
3.10	Number of letter bigrams and trigrams	26
3.11	Summary of statistical data	27
4	References	29
5	Most frequent words ordered by rank	33
6	Alphabetical frequency list	45