

# Contents

<b>1 Word forms and their frequencies</b>	<b>1</b>
1.1 Previous frequency lists and dictionaries of Czech . . . . .	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 . . . . .	4
<b>2 Corpus</b>	<b>7</b>
2.1 Data basis and extent of data collection . . . . .	7
2.2 Corpus timeline . . . . .	8
2.3 Data pre-processing . . . . .	8
<b>3 Statistical data on the word list</b>	<b>11</b>
3.1 Comparison criteria . . . . .	11
3.2 Character statistics . . . . .	12
3.3 Vowels and consonants . . . . .	12
3.4 Word length . . . . .	15
3.5 Word structure . . . . .	16
3.6 Zipf's law . . . . .	18
3.7 Text coverage of the top-N most frequent words . . . . .	18
3.8 Growth in word length . . . . .	19
3.9 Number of syllables . . . . .	22
3.10 Number of letter bigrams and trigrams . . . . .	23
3.11 Summary of statistical data . . . . .	25
<b>4 References</b>	<b>27</b>
<b>5 Most frequent words ordered by rank</b>	<b>29</b>
<b>6 Alphabetical frequency list</b>	<b>41</b>