Frequency List of Written Finnish Word Forms

1 BASIC INFORMATION

1.1 Lexicon type
Frequency list

1.2 Representation of the corpora (flat files, database, markup)
Flat file of Unix-style records

1.3 Character encoding
ISO-8859-1 (Latin-1)

2 ADMINISTRATIVE INFORMATION

2.1 Contact person
Name: Toni Suutari
E-mail: toni.suutari@kotus.fi

2.2 Copyright statement and information on IPR
Freely downloadable

3 TECHNICAL INFORMATION

3.1 Data structure of an entry
Three text files of different size (and an HTML page of the 5000 most frequent forms)

3.2 Lexicon size
1 339 787 (full list), 542 521 (frequency > 1), 362 514 words (frequency > 2)

4 CONTENT INFORMATION

4.1 The natural language(s) of the lexicon
Finnish

4.2 Entry type
Word form per line

4.3 Attributes
Word form with its rank, absolute frequency, and relative frequency in the Finnish Parole corpus

4.4 Coverage of the lexicon
The inflected vocabulary of the Finnish Parole corpus (17 million tokens)

4.5 Intended application of the corpus
NLP applications

4.6 Reliability (automatically/manually constructed)
Automatically constructed

5 **RELEVANT REFERENCES AND OTHER INFORMATION**

http://kaino.kotus.fi/sanat/taajuuslista/parole.php (In Finnish)