

Corpus of Early Modern Finnish

1 BASIC INFORMATION

1.1 Corpus composition

Written Finnish from the 19th century (mostly from the years 1810-1880)

1.2 Representation of the corpora (flat files, database, markup)

Browsable and searchable on the web

1.3 Character encoding

UTF-8

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 browsable on the web

3 TECHNICAL INFORMATION

3.1 Data structure of an entry

Browsable plain text

3.2 Corpora size (num. of tokens)

8 645 700 words

4 CONTENT INFORMATION

4.1 Type of the corpus (monolingual/multilingual, parallel/comparable, raw/annotated)

Monolingual, raw

4.2 The natural language(s) of the corpus

Finnish

4.3 Domain(s)/register(s) of the corpus

Published literature, periodicals, newspapers, dictionaries, among others, with a focus on the earliest and most important publications and a wide thematic coverage, and a preference for originally Finnish texts over translations.

4.4 Annotations in the corpus (if an annotated corpus)

No annotations

4.5 Intended application of the corpus

Research

4.6 Reliability of the annotations (automatically/manually assigned) – if any

No annotations

5 RELEVANT REFERENCES AND OTHER INFORMATION

http://kaino.kotus.fi/korpus/1800/meta/1800_coll_rdf.xml