

CoRD

Corpus Resource Database

<http://www.helsinki.fi/varieng/CoRD/>

CoRD is an open access service launched by the Research Unit for Variation, Contacts, and Change in English (VARIENG) in 2007 to serve the corpus linguistic community in making information on available corpora easy to reach. CoRD is based on the notion of creative commons. All information on individual corpora featured in CoRD has been submitted by corpus compilers themselves. Ongoing projects can submit an entry early on and update it as work progresses.

Some key features of corpus descriptions on CoRD include:

Reference information

A common problem in corpus linguistic scholarship is deciding on the correct reference line when citing a particular corpus. CoRD provides the recommended reference line for each corpus, written by the compilers themselves.

A quick source of information

Through CoRD, users of corpora are one mouse click away from checking the basic data of a corpus such as its wordcount or that of its subcorpora, the annotation scheme, or perhaps the file format used in the distribution version.

Giving credit where credit is due

One of the shortcomings of conventional citation formats is that often only the leaders of research projects are mentioned by name. All the others, many of whom may have contributed extensive amounts of time and labour, are relegated to a footnote. CoRD allows corpus compilers to acknowledge all team members. Today, funding bodies are also increasingly interested in public acknowledgment of their contributions to research. Through CoRD, corpus compilation projects can credit sources of funding in as much detail as necessary.

Availability information

Some corpora are made available through open access academic repositories, while others are obtainable by subscription, membership in an organization, or directly from the compilers. CoRD tells you who to contact.

Teaching

Teachers of corpus linguistics can use CoRD in a number of different ways. In addition to referring students to CoRD for basic information on corpora, CoRD entries can serve as introductory material to the process of corpus compilation.

Recording the living history of corpus linguistics

Although a relatively young approach, corpus linguistics has in actual fact been around for quite some time. Over the years, corpus compilation methods have undergone considerable changes. CoRD is an online resource for recording and sharing not only hard facts, but also the stories and experiences of the compilers. Photographs, audio recordings and video clips can be used to liven up the presentation even further.

Submitting corpus information to CoRD

Basic submissions can be accomplished simply by filling in our electronic submission form. We collect the information and manually transform it into template compliant HTML, usually within a few days. Naturally, any of the information can be updated, removed or appended whenever necessary.

The *submission form* (left) is used for submitting corpus information. The predetermined fields ensure that the basic information is obtained for all featured corpora.

A separate *feedback form* is available for the purpose of informing us of any mistakes or omissions in corpus description. The form can also be used to submit your articles to the bibliographies of corpora used as source material.

If the description of a particular corpus requires extensive customizing of the basic structure, or if you'd like to include multimedia content, simply send us an e-mail and we'll be happy to work with you.

What's on CoRD right now?

Most of the major corpora compiled in Helsinki are now described on CoRD, as well as a small number of corpora submitted by colleagues. All of the Helsinki Corpora are described at length with charts, maps, photographs. A few article length discussions of compilation practices and structural issues are also included, as well as audio samples of dialect material and video presentations of each corpus. Bibliographies list research articles by Helsinki-based scholars making use of our own corpora.

Varieng
research unit for variation, contact and change in English

Helsinki Corpus of British English Dialects

The Helsinki Corpus of British English Dialects (HD) is a collection of orthographically transcribed audio recorded speech, mainly from East Anglia and the South-West, with a minor collection from Lancashire. The recordings were made in the 1970s and the 1980s by Finnish postgraduates.

The aim of the corpus is to put sociolinguistics, discourse analysis and other material for non-linguistic, multi-habit and history.

Project leaders: Ossi Ihalahti onwards.

Fieldwork supervisors: In the Orton (Leeds), Ossi Ihalahti.

Size: Altogether 187 files consist.

Time periods: 1970s and 1980s onwards.

Status: First stage completed ongoing, with plans of extending.

Corpus data: The primary data.

Funding: Finnish Culture Sig.

Language: English, rural (Cambridge, Lancashire).

Reference line and Copyright: The Helsinki Corpus of British English Dialects. All the material consists of copyright for the material. For

Region (urban)	Files	Words
CAMBRIDGESHIRE	26	53 (1.4%)
DEVON	20	22 (0.6%)
ISLE OF ELY	10	52 (1.6%)
SOMERSET	15	24 (0.7%)
SUFFOLK	15	27 (0.8%)
Total	86	176

Fieldworker's early transcription notes

At the time of the fieldwork in the 1970s and 1980s most of the field notes were written and in the mid 70s or early 80s they were typed. The notes were written by the fieldworkers who collected the data. The notes are in the form of a diary. The notes are in the form of a diary. The notes are in the form of a diary.

Audio player transcript:

S.W. from Birmingham.
Recorded in 1988.
Age: 45, female.
Occupation: cleaning instructor.
Description: S.W. talks about working with looms.
An 'I' takes the well article one way, but I don't go back again. So I'm to go on to take a conveyor belt, or that's why they've so many of them because there's one going across Stanger Street and then another so they get the axes back up again.

The CoRD entry for the *Helsinki Corpus of British English Dialects* was designed by Anna-Liisa Vasko and Simo Ahava, with Kirsti Peitsara.

To contact

Visit <http://www.helsinki.fi/varieng/CoRD/> or write an e-mail to jukka.tyrkko@helsinki.fi