Terttu Nevalainen, Jukka Tyrkkö, Matti Rissanen
University of Helsinki, Department of English
VARIENG Research Unit

Corpus Resource Database (CoRD)

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

The *Research Unit for Variation, Contacts, and Change in English* (VARIENG) at the Universities of Helsinki and Jyväskylä is launching an open-access service for the benefit of both corpus linguists and students. This Corpus Resource Database is intended to grow into a significant focused source of information on corpora.

The impetus behind CoRD was the realization that, although link lists and portals have been and are maintained by individual corpus linguists, they rarely offer extensive and standardised information on the corpora.
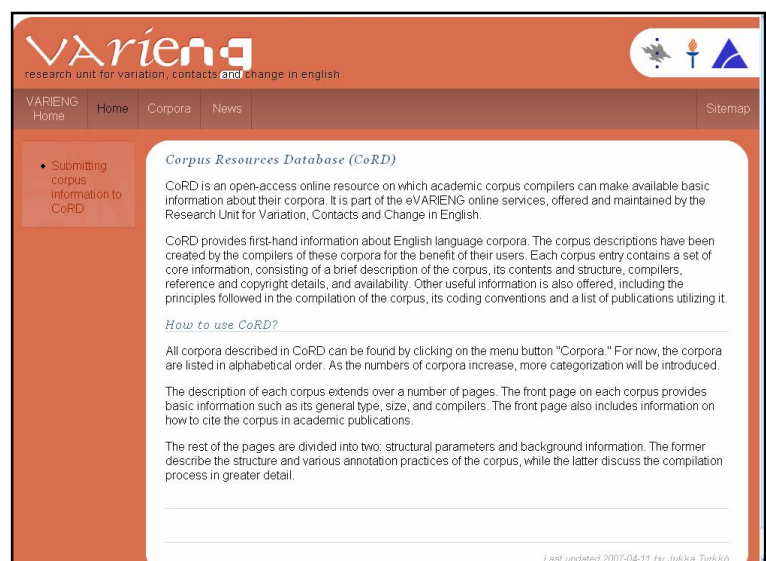
CoRD will consist of corpus descriptions submitted by compilers themselves. For each corpus, some basic information will be presented, along with detailed descriptions ranging from annotation and stratification to transcription practices. We also hope to include information on the unique compilation process of each corpus. What were the reasons behind the selection of the texts? How were the annotation practices decided on, how was annotation performed, and what problems were encountered? To provide a practical demonstration of how the source material has been turned into corpus material, we hope to include multimedia samples of original texts and audio clips alongside extracts from the corpus. We also welcome submission of photographs and other background material.

Another feature of CoRD is that each entry includes a bibliography of published work that utilises the corpus in question. At present, it is difficult for researchers, and especially for students, to discover whether research has been carried out on a topic using a particular corpus. Over time, the CoRD bibliography may grow into a quick reference guide for how published corpora have been used by researchers, and thus supplement the earlier bibliographies compiled by Bengt Altenberg and others.
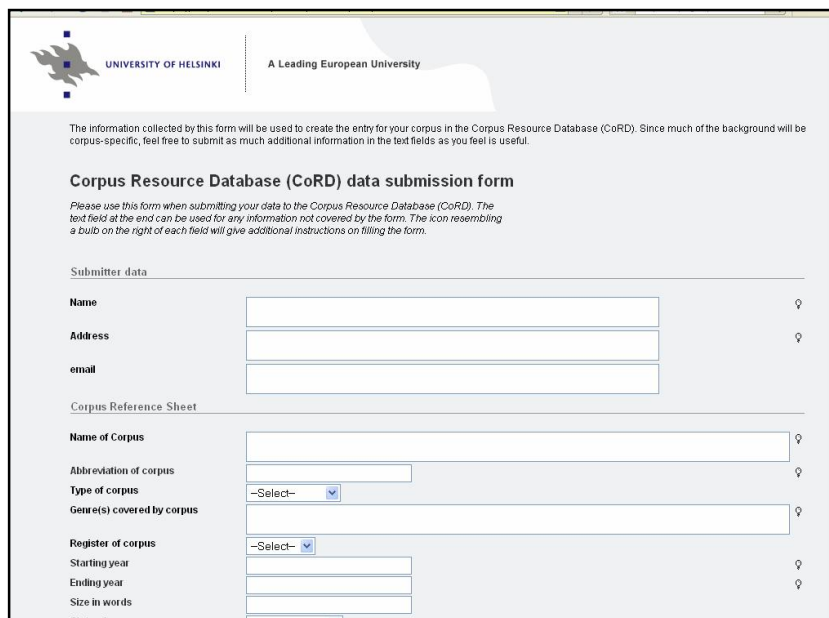
What's in CoRD now?

Since we're just getting started, not too much. The first corpus to be featured on CoRD is the *Helsinki Corpus* (if you don't know what that is, you can check CoRD!). We are currently preparing descriptions of other corpora compiled at the VARIENG Research Unit, such as the Corpus of Early English Correspondence (CEEC) and the Corpus of Early English Medical Writing (CEEM), so more is definitely on the way. Which brings us to...

How can I get involved?

Let's face it, nobody knows more about the compilation of a particular corpus than the people who actually did it. Therefore, it makes sense to have those same people write a description for CoRD. In this regard, CoRD can hopefully serve as an online historiography of corpus compilation, as well as as a useful resource for those planning the compilation of new corpora.

Submitting a corpus description to CoRD has been made very easy. On the CoRD website, you will find an electronic form. Fill it in – not necessarily all in one go – and click "submit". Upon receiving the data, we will edit it for the website and upload it there. Naturally, you are free to revise, add to, or delete any or all of the description. You are also welcome to add links to your own website with more in-depth information, to an online repository from which the corpus can be obtained or, in the case of a commercially published corpus, to the publishers' website.

I'm not a corpus compiler, but I'd still like to help. What can I do?

The easiest way is to submit bibliographical entries on research that makes use of a particular corpus. Feedback is submitted using another electronic form, also to be found on the website. If you have used a corpus in your published research or in your teaching, why not send that information in. Corpus compilers would love to know where the results of their hard work are being used, and fellow researchers will benefit from knowing about earlier work on the same topic – and can add a reference to your work in their articles!

If you have colleagues who compile corpora, you can tell them about CoRD. Remember, submitting information to CoRD doesn't cost you anything, nor do we profit from it. CoRD adheres to the principle of open access, and is provided as a service to the corpus community.

Finally, use CoRD! All information published on CoRD has been submitted by the compilers themselves, so it is reliable information which you can use both in your own work and in your teaching. In addition, please let us know of any changes you'd like to see or ideas which might improve the website.

Contact: jukka.tyrkko@helsinki.fi