

Resources:

Access to Computational Tools for Discourse Research

Overview

This document describes resources available to support research on spoken and written discourse. The resources are mainly of three types:

- software for working with sound, including sound editing and phonetic analysis programs
- software for working with transcriptions, especially for making concordances
- web pages that provide access to lots more information about corpora, concordances, linguistic analysis software, etc.

Several of the programs mentioned below are introduced in this course, and are available for use in the Linguistics Lab. If you find the software presented in this course to be useful, you may wish to obtain a copy for your own use. Among tools of interest to students of spoken discourse, two kinds are of special interest: programs for editing or analyzing sound, and for creating concordances.

Sound Editing and Analysis

SoundForge 8.0 is a very useful program for working with sound. Most of its features which are of interest to students of spoken discourse are also found in **SoundForge Audio Studio 8.0**. The Audio Studio version is significantly cheaper, and nearly as good for most purposes of spoken discourse research. Note that both programs require Windows XP.

SoundForge 8.0 (about \$300) or SoundForge Audio Studio (about \$70) can be ordered from Sony Media Software at:

<http://mediasoftware.sonypictures.com/products/soundforgefamily.asp>

But the cheapest way to obtain either SoundForge program is through a purchase at an educational price (e.g. from www.CreationEngine.com). (For some educational software purchases there may be a minimum order of 10 copies, so it may be best to try to do make such purchases jointly as a class.)

(However, if you are on a strict budget, some of the useful functions can be found in the non-commercial program PRAAT – although, for the features where SoundForge excels, PRAAT may seem much slower to work with, and not as user-friendly.)

A number of programs for working with sound are available for free on the Internet. Some of the most useful of these are VoiceWalker, Transcriber, PRAAT, and the Fujisaki editor. The latter two are oriented to phoneticians, and may require some technical sophistication to get the most out of them.

VoiceWalker is useful for transcribing. Version 1.0a is recommended, though you may also wish to try version 2.0. VoiceWalker is available for free from the UC Santa Barbara Linguistics Department, on the

Du Bois: Resources—Access to Computational Tools

Web at:

<http://www.linguistics.ucsb.edu/resources/computing/download/download.htm>

PRAAT can be downloaded from www.praat.org.

The **Fujisaki editor** can be downloaded from:
www.tfh-berlin.de/~mixdorff/fujisaki_analysis.htm

Transcriber can be downloaded from www.etca.fr/CTA/gip/Projets/Transcriber/. Normally you should choose one of their “binary distributions,” which are available for Windows, Mac, Unix, and Linux.

Concordance-Making

MonoConc Pro (about \$85.00) can be ordered at <http://www.athel.com/>

In addition, there are two other concordance programs that you may wish to consider. **Wordsmith Tools** has a large user base and a good reputation. **Concordance** by R.J.C. Watt claims to work with Unicode. I have not tried the latter program, but it does allow you a 30-day free trial.

Wordsmith Tools is available at www.lexically.net/wordsmith/

Concordance by R.J.C. Watt is available at:
www.concordance.ukgateway.net/Concordance200Manual.htm#top

If you want to work with British English, one of the very best concordance tools available anywhere is the **BNCWeb**, available at: <http://homepage.mac.com/bncweb/home.html>. But this software requires that you first set up the **British National Corpus** (BNC) on a UNIX or Mac OS X computer. See: www.natcorp.ox.ac.uk/

To use the World Wide Web as a source of linguistic data – a giant, unruly corpus – check out:
<http://webcorp.org.uk/index.html>

Corpus Linguistics Resources on the Web

The following URL's and home pages have lots of material of interest for researchers wishing to use computers to work on corpora of spoken or written discourse. Information is available on numerous and diverse languages (though there is far more material on written language, and on English).

Corpora List
<http://helmer.aksis.uib.no/corpora/>

Texts, text centres, resources and programs on the Web
<http://helmer.aksis.uib.no/text.htm>

Jane Edwards

http://www.icsi.berkeley.edu/~edwards/corpus_resources.html

Michael Barlow

<http://www.ruf.rice.edu/~barlow/corpus.html>

TalkBank

<http://www.talkbank.org/>

Linguistic Data Consortium

<http://www ldc.upenn.edu/>

Language Archives

<http://www ldc.upenn.edu/exploration/archives.html>

Software, Tools, Frequency Lists, etc.

<http://lingo.lancs.ac.uk/devotedto/corpora/software.htm>

Consent, Ethics & Human Subjects

Human Subjects Committee (UCSB):

<http://research.ucsb.edu/connect/pro/prs1.shtml>

Human Subjects Training Module (UCSB):

<http://hstraining.orda.ucsb.edu/training/>

Informed Consent Guidelines (TalkBank):

<http://www.talkbank.org/share/consent.html>

Code of Ethics (TalkBank):

<http://www.talkbank.org/share/ethics.html>

“Can We Tape?” -- Legal aspects of audio recording (phone calls, conversations, etc.) for journalists:

<http://www.rcfp.org/taping/>

[rev. 18-Sep-2004]