Linguistic Methods

7. Video-and audio recording, Transcription and Coding
8. Corpus Linguistics

Readings
Allwood, J. Lingvistisk Metod: kap 7, 8a,b, 9, 10.

Available anuals:
1. GTS manual
See: http://www.ling.gu.se/projekt/tal/index.cgi?PAGE=3
2. Göteborg Coding Manual
See: Allwood, Jens (Ed.) (2001). Dialog Coding – Function and Grammar. Göteborg Coding Schemas. *Gothenburg Papers in Theoretical Linguistics*, 85. Göteborg University, Department of Linguistics. (Can be ordered from Ritva Aydi <ritva@ling.gu.se>
and
http://www.ling.gu.se/projekt/tal/index.cgi?PAGE=3

1. Introduction

Contents
Kinds of corpora
What is corpus linguistics?

Readings:
Allwood: Lingvistisk Metod, Chapter 7

Assignments
1. Find a corpus site on the web.
Visit it and explore it.
Write a short description of it giving theURL address:
- List its characteristics – written or spoken, size, sampling basis, conditions on access and use.
2. What is corpus linguistics? Give a brief description or definition.

Some hints
There is a website devoted to corpora http://devoted.to/corpora/
It has out-of-date links but also many which are still active.
Other web sites with links are:
http://www.athel.com/corpus_linguistics.html
http://www.ims.uni-stuttgart.de/projekte/corplex/
http://www.talkbank.org/

2. Corpora of spoken and written language

Contents
A. Written language
- sampling
- choice of material
- structure of corpus
B. Spoken language
1. Sampling
- What should you record?
2. Recording
- How should you record (audio – video)
- Archiving audio/video files
3. Transcribing
- Background information
- Transcription standard
- Reliability
C. Other relevant cultural and linguistic information
- What can/should be included?
- How should it be included?

Readings
Allwood: Lingvistisk metod 8a, 8b

Assignments
1. Think of a study where you would want to use a corpus. What would be your purpose?
What would be your criteria for what to include in the corpus?
2. Describe some of the problems you face in doing audio-video recordings for a spoken
language corpus.
3. Describe some of the types of contextual information you might want to include in creating
a data base or archive of (i) written and (ii) spoken language.

3. Transcribing and coding

Contents
A. Transcribing
- Background information
- Transcription standard
- Reliability
- Organizing the corpus for searching
B. Getting acquainted with one example of a transcription standard, GTS
C. Coding
- Some examples
- Reliability

Readings
Allwood: Lingvistisk metod, Chapters 9, 10.
GTS manual
Allwood (2001) Göteborg Coding Manual

Assignments
1. What choices can you make regarding the types of transcription (formats, standards) to be used in creating a spoken language corpus.
2. Describe three typical characteristics of spoken language and suggest a way to capture these characteristics in transcription.
3. Describe a procedure for creating reliability in a spoken language corpus
4. Make a short recording and present a 1 page transcription using GTS annotation
6. Code the transcription using one of the coding schemas in Allwood et al. (2001) Göteborg Coding Manual.

4. Use of corpora in language studies

Contents
Areas of linguistics and use of corpora

Speech	Sociolinguistics	Psycholinguistics
Lexicon	Stylistics	Cultural studies
Grammar	Teaching of language	Social psychology
Semantics	Historical linguistics	
Pragmatics	Dialectology	

Readings
Allwood: Lingv Metod, Chapter 7

Assignments
1. Select 3 areas of linguistics and give examples of how corpora have been used.