

Ling. 5381
Computers and Natural Language: Corpus Linguistics
Fall 2005

Proposed Schedule [Updated: Dec. 7, 2005]

Background: Corpus Use and Famous Corpora

1. Mon. Aug. 22 Introduction to the class; Introduction to the lab
What is computational linguistics? What is corpus linguistics?

Wed. Aug. 24 The web as corpus; B&P Ch. 1 "Introducing Corpora and Corpus
Analysis Tools" (9-24).
2. Mon. Aug. 29 Fillmore, Charles J. (1992). "Corpus linguistics" or "Computer-
aided armchair linguistics."; "Henry Kucera (1991)

Wed. Aug. 31 Kennedy (1998). "Design and Development of Corpora."
(13-70);

Mukherjee, Joybrato. (2004). "Review Article: State of the
Art in CL"

Exercise 1 Due (Web Queries)
3. Mon. Sept. 5 **Labor Day Holiday—no classes**

Compiling a Corpus; Inputting and Annotating Texts

- Wed. Sept. 7 B&P (2002). Ch. 2 "Introducing LSP." (25-41).
Clear (1992). "Corpus Sampling." (21-31).
Demo on using the Scanner
4. Mon. Sept. 12 Crystal (2010). "Languages on the Web." (216-223).
Bowker (2002). "Capturing Data in Electronic Form." (22-42).
Demo on doing OCR

Wed. Sept. 14 B&P (2002). Ch. 3. "Designing a Special Purpose Corpus." (45-57)
B&P (2002). Ch. 4. "Compiling a Special Purpose Corpus." (58-74).
5. Mon. Sept. 19 B&P (2002). Ch. 5 "Markup and Annotation." (75-91).
Leech (1997). "Grammatical Tagging." (19-33)
Leech (1997) Appendix III, (256-260).

Exercise 2 Due (Scanning and Using OCR)

Wed. Sept. 21 Garside, Fligelstone, and Botley (1997). "Discourse Annotation:
Anaphoric Relations in Corpora." (66-84).

Tools to Use on a Corpus: Concordance Software and Indexing Software

6. Mon. Sept. 26 B&P (2002) Ch. 7 “Introduction to Basic Corpus Processing Tools”
Exercise 3 Due (Tagging Using Different Tagsets)
- Wed. Sept. 28 Workshop on paper topics
7. Mon. Oct. 3 WordSmith Tools demo
- Wed. Oct. 5 Work with WordSmith
See manual selections (available in the lab)
8. Mon. Oct. 10 B&P (2002) Ch. 6 “Bilingual and Multilingual corpora: Preprocessing, Alignment and Exploitation”
- Wed. Oct. 12 B&P (2002) Ch. 8 “Building Useful Glossaries”
Exercise 4 Due (Using WordSmith with Untagged Texts)
9. Mon. Oct. 17 B&P (2002) Ch. 9 “Term Extraction”
- Wed. Oct. 19 McEnery, Tony, Jean-Marc Langé, Michael Oakes, and Jean Véronis. (1997). (CA Ch. 15) “The exploitation of multilingual annotated corpora for term extraction” Pp. 220-230.
10. Mon. Oct. 24 B&P (2002) Ch. 10 “Using LSP Corpora as a Writing Guide”
- Wed. Oct. 26 B&P (2002) Ch. 11 “Using LSP Corpora as a Translation Resource”
11. Mon. Oct. 31 B&P (2002) Ch. 12 “Other Applications and Future Directions”
Exercise 5 Due (Concordancing with Tagged Texts)
- Wed. Nov. 2 Granger (1998). “The Computer Learner Corpus: A Testbed for Electronic EFL Tools.” Pp. 175-188.
12. Mon. Nov. 7 Biber, Conrad, and Reppen (1998). “Language Acquisition and Development.” Pp. 172-201.
[Dr. Stvan is out for jury duty—go over discussion questions]
- Wed. Nov. 9 Stvan (2005). “Inferring New Vocabulary Using Online Texts”
13. Mon. Nov. 14 Biber, Conrad, and Reppen (1998). “Historical and Stylistic Investigations.” Pp. 203-229 + Pp. 252-253.
- Wed. Nov. 16 Biber (1992). “Using Computer-based Text Corpora to Analyze

the Referential Strategies of Spoken and Written Texts.” Pp. 213-255.

14. Mon. Nov. 21 Stvan (in press). “Diachronic Change in The Discourse Markers Why and Say in American English”
Vocabulary Quiz
- Wed. Nov. 23 3 presentations
15. Mon. Nov. 28 3 presentations
- Wed. Nov. 30 Wind-up and evaluations
2 presentations

**EXAM Thurs. Dec. 8
WEEK (4:30-6:30 p.m.)**

**Final 6 in-class presentations of final projects
Final Paper on your Research is Due**

Additional Dates to Note

Wed. Sept. 7 Census Date
Fri. Nov. 11 Last day to Drop a Course
Wed. Dec. 14 Grades available via SAM (817) 272-2726
OR
via the web (<https://www4.uta.edu/StudentDataWeb/SLLogin.jsp>)