This course is an introduction to computerized research methods, which are applied to large data bases of language used in natural communicative settings to supplement more traditional ways of linguistic analysis in all linguistic sub-disciplines.

COURSE CREDIT REQUIREMENTS

To meet the course credit requirements for this class, you are expected to
- attend class regularly
- come to class prepared, that is, do the required readings
- hand in a summary of a corpus-linguistic paper of your choice that appeared in a journal like, for instance, the *ICAME* journal, *Corpus Linguistics and Linguistic Theory*, *International Journal of Corpus Linguistics*, *Computers and the Humanities*, *Literary and Linguistic Computing*, *Corpora*, or *Computational Linguistics*, and that is not among the papers we discuss in class (cf. below); the length of this summary should be no shorter than 5 pages, double-spaced, 12 pt. font of running text (i.e., excluding a cover page, tables, figures, and the like). The summary is due 1 December in class.
- pass a midterm exam that covers the contents of the class to that date;
- pass a final exam that covers the contents of the complete class.

For both the midterm exam and the final exam, you are allowed to bring all course materials, readings, notes, etc.

GRADING

Your final grade will depend on the number of points you score. You can get the maximum of 100 points if you
- attend all classes for a maximum of 45 points (each unexcused absence reduces your score by 2.5 points)
- submit the article summary for a maximum of 15 points (each late-day reduces your score by 2.5 points)
- pass the midterm exam for a maximum of 15 points
- pass the final exam for a maximum of 25 points

Should you not be satisfied with your performance on either exam or the grade for the article summary, you can score a maximum of 10 additional points by handing in a special written assignment (feel free to talk to me about that anytime).

CONTACT

Office hours: Friday 2-3pm in SH 3523 or by appointment
Web: http:www.linguistics.ucsb.edu.faculty/swulff
Email: swulff@linguistics.ucsb.edu

THE BASICS

1	Sep 29	Introduction
2	Oct 1	Online interfaces I: MICASE online
3	Oct 6	Online interfaces II: BYU-CAE

HOW TO COMPILE A CORPUS

4 Oct 8 Compiling corpora / The web as corpus
Prep reading: Kennedy (1998: 13-87)

5 Oct 13 Corpus transcription and annotation
Prep reading: Atkins et al. (1992)

CORPUS METHODS: ROUGH AND READY

6 Oct 15 Frequency lists (Introduction to AntConc)
Prep reading: Leech and Fallon (1992)

7 Oct 20 Concordances
Prep reading: Hyland & Milton (1997)

8 Oct 22 Collocations
Prep reading: Biber et al. (1996: 21-54)

9 Oct 27 Patterns (Introduction to KfNgram)
Prep reading: Francis (1993)

10 Oct 29 Collostructions
Prep reading: Stefanowitsch & Gries (2003)

11 Nov 3 **MIDTERM EXAM**

CORPUS METHODS: THE R-EAL DEAL

12 Nov 5 R: Functions, arguments, vectors
Prep reading: Gries (2009: 72-84)

13 Nov 10 R: Simple text processing
Prep reading: Gries (2009: 84-105)

14 Nov 12 R: Advanced text processing
Prep reading: Gries (2009: 65-72; 113-117)

15 Nov 17 R: Simple frequency lists
Prep reading: Gries (2009: 120-123)

16 Nov 19 R: Cultural word frequency comparisons
Prep reading: Gries (2009: 63-65; 131-137)

17 Nov 24 R: Simple concordances

18 Nov 26 R: Complex concordances
Prep reading: Gries (2009: 143-147)

IT'S ALL DOWNHILL FROM HERE…

19 Dec 1 Make a Wish! **ARTICLE SUMMARY DUE TODAY**

20 Dec 3 Revision and course evaluation

Dec 12 12-3 pm: **FINAL EXAM**

REQUIRED READINGS

Atkins, Beryl T., Jeremy Clear and Nicholas Ostler. 1992. Corpus design criteria. *Literary and Linguistic Computing* 7.1:1-16.

Biber, Douglas, Susan Conrad and Randi Reppen. 1998. Corpus linguistics: investigating language structure and use. Cambridge: Cambridge University Press. [21-54]

Francis, Gill. 2004 [1993]. A corpus-driven approach to grammar: Principles, methods, and examples. In: Geoffrey Sampson and Diana McCarthy (eds.). *Corpus linguistics. Readings in a widening discipline*. London, New York: Continuum, 198-211.

Gries, Stefan Th. *Quantitative corpus linguistics with R: A practical introduction*. London, New York: Routledge, Taylor & Francis Group. [excerpts]

Hyland, Kenneth and John Milton. 2004 [1997]. Qualification and certainty in L1 and L2 students' writing. In: Geoffrey Sampson and Diana McCarthy (eds.). *Corpus linguistics. Readings in a widening discipline*. London, New York: Continuum, 371-386.

Kennedy, Graeme. 1998. *An introduction to corpus linguistics*. London, New York: Longman. [13-87]

Leech, Geoffrey and Roger Fallon. 2004 [1992]. Computer corpora – what do they tell us about culture? In: Geoffrey Sampson and Diana McCarthy (eds.). *Corpus linguistics. Readings in a widening discipline*. London, New York: Continuum, 160-173.

Stefanowitsch, Anatol and Stefan Th. Gries. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8.2:209-143.

SUGGESTIONS FOR FURTHER READINGS (NOT OBLIGATORY)

Baker, Paul, Andrew Hardie and Tony McEnery. 2006. *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.

Biber, D. & E. Finegan. 2004 [1987]. Historical drift in three English genres. In: Geoffrey Sampson & Diana McCarthy (eds.). *Corpus linguistics. Readings in a widening discipline*. London and New York: Continuum, 67-77.

Biber, D. 1993. The multi-dimensional approach to linguistic analyses of genre variation: An overview of methodology and findings. *Computers and the Humanities* 26: 331-345.

Biber, Douglas. 1990. Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing* 5.4:257-269.

Biber, Douglas. 1993. Representativeness in corpus design. *Literary and Linguistic Computing* 8.4:243-257.

Bybee, Joan. & Joanne Scheibman. 1999. The effect of usage on degrees of constituency: the reduction of *don't* in English. *Linguistics* 37.4:575-596.

Church, Kenneth W. and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* 16.1:22-9.

Diessel, Holger. 2005. Competing motivations for the ordering of main and adverbial clauses. *Linguistics* 43.3:449-470.

Diessel, Holger and Michael Tomasello. 2005. Particle placement in early child language: A multifactorial analysis. *Corpus Linguistics and Linguistic Theory* 1.1:89-112.

Gries, Stefan Th. 2001. A corpus-linguistic analysis of *-ic* and *-ical* adjectives. *ICAME Journal* 25:65-108.

Gries, Stefan Th. 2005. Null-hypothesis significance testing of word frequencies: a follow-up on Kilgarriff. *Corpus Linguistics and Linguistic Theory* 1.2:277-294.

Hanks, Patrick. 2004 [1986]. Typicality and meaning potentials. In: Geoffrey Sampson & Diana McCarthy (eds.). *Corpus linguistics. Readings in a widening discipline*. London and New York: Continuum, 58-66.

Hunston, Susan and Gill Francis. 1999. *Pattern grammar: […]*. Amsterdam, Philadelphia: John Benjamins.

Hunston, Susan. 2002. *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Ihalainen, Ossi. 2004 [1991]. A point of verb syntax in south-western British English: An analysis of a dialect continuum. In: Geoffrey Sampson & Diana McCarthy (eds.). *Corpus linguistics. Readings in a widening discipline*. London and New York: Continuum, 113-121.

Kaunisto, Mark. 1999. *Electric/electrical* and *classic/classical*: variation between the suffixes *-ic* and *-ical*. *English Studies* 4:343-370.

Kilgarriff, Adam and Gregory Grefenstette. 2003. Introduction to the special issue on the web as corpus. *Computational Linguistics* 29.3:333-347.

Kilgarriff, Adam. 2001. Comparing corpora. *International Journal of Corpus Linguistics* 6.1:1-37.

McEnery, Anthony, Richard Ziao and Yukio Tono. 2006. *Corpus-based language studies: an advanced resource book*. London, New York: Routledge.

McEnery, Tony and Andrew Wilson. 2001. *Corpus linguistics*. 2nd ed. Edinburgh: Edinburgh University Press.

Meyer, Charles F. 2002. *English corpus linguistics: an introduction*. Cambridge: Cambridge University Press.

Meyer, Charles F. and Hongyin Tao. 2005. Response to Newmeyer's 'Grammar is grammar and usage is usage'. *Language* 81.1:226-228.

Newmeyer, Frederick J. 2003. Grammar is grammar and usage is usage. *Language* 79.4:682-707.

Oakes, Michael. 1998. *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.

Partington, Alan. 1998. *Patterns and meanings: […]*. Amsterdam, Philadelphia: John Benjamins.

Renouf, Antoinette. 1987. Corpus development. In: John M. Sinclair (ed.). *Looking up*. London: Collins, p. 1-22.

Sinclair, John M. 2003. *Reading concordances*. London, New York: Pearson Education.

Stefanowitsch, Anatol. 2005. New York, Dayton (Ohio), and the Raw Frequency Fallacy. *Corpus Linguistics and Linguistic Theory* 1.2:295-301.

Teubert, Wolfgang and Anna Čermáková. 2007. *Corpus linguistics: a short introduction*. London, New York: Continuum.

Wiechmann, Daniel and Stefan Fuhs. 2006. Concordancing software. *Corpus Linguistics and Linguistic Theory* 2.1:107-127.

Wulff, Stefanie. 2003. A multifactorial corpus analysis of adjective order in English. *International Journal of Corpus Linguistics* 8.2:245-282.