

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 AND 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 14 classes (excluding the first class, the midterm, and the final exam) for a maximum of **28 points** (each unexcused absence reduces your score by 2 points)
 - submit 4 written homework assignments for a maximum of 5 points each, totaling a maximum of **20 points** (see the syllabus for due dates; each late day reduces your score by 1 point)
 - submit an article summary for a maximum of **5 points** (each late-day reduces your score by 1 point); you can pick any corpus-linguistic paper 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*; 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 **27 April in class**.
 - pass the midterm exam for a maximum of **20 points**
 - pass the final exam for a maximum of **27 points**

Both the midterm and the final exam are closed notes and closed books.

MAKEUP ASSIGNMENTS AND EXTRA CREDIT

If you are not satisfied with your performance in either the exams or attendance, you can contact me anytime to discuss a makeup or extra credit assignment. All makeup and extra credit assignments are due the last day of class (4 May).

INFORMATION ON DROPPING CLASSES

<<http://essc.unt.edu/registrar/schedule/scheduleclass.html>>

CONTACT

Office hours: by appointment

Web: <<http://www.lingtechcomm.unt.edu/~swulff/teaching/Ling5320.html>>

Email: Stefanie.Wulff@unt.edu

ODA STATEMENT

The University of North Texas is on record as being committed to both the spirit and letter of federal equal opportunity legislation; reference Public Law 92-112 – The Rehabilitation Act of 1973 as amended. With the passage of new federal legislation entitled Americans with Disabilities Act (ADA), pursuant to section 504 of the Rehabilitation Act, there is renewed focus on providing this population with the same opportunities enjoyed by all citizens.

THE BASICS

- 1** Introduction
19 Jan COCA online interface

CORPUS METHODS: ROUGH AND READY**HOMEWORK ASSIGNMENT 1 DUE TODAY**

- 2** Introduction to AntConc
26 Jan Frequency lists
Collocations

HOMEWORK ASSIGNMENT 2 DUE TODAY

- 3** Introduction to KfNgram
2 Feb Patterns and collocations

HOW TO COMPILE A CORPUS**HOMEWORK ASSIGNMENT 3 DUE TODAY**

- 4** Compiling corpora / The web as corpus
9 Feb Corpus transcription and annotation

CORPUS METHODS: THE REAL DEAL

- 5** R: Functions, arguments, vectors
16 Feb [Prep reading: Gries \(2009: 72-84\)](#)

- 6** R: Simple text processing
23 Feb [Prep reading: Gries \(2009: 84-105\)](#)

- 7** **MIDTERM EXAM**
2 Mar

- 8** R: Advanced text processing
9 Mar [Prep reading: Gries \(2009: 65-72; 113-117\)](#)

- 9** **SPRING BREAK**
16 Mar

- 10** R: Simple frequency lists
23 Mar [Prep reading: Gries \(2009: 120-123\)](#)

- 11** **HOMEWORK ASSIGNMENT 4 DUE TODAY**
30 Mar Guest teacher: Stefan Th. Gries (UC Santa Barbara)

- 12** R: Cultural word frequency comparisons
6 Apr [Prep reading: Gries \(2009: 63-65; 131-137\)](#)

- 13** R: Simple concordances
13 Apr

- 14** R: Complex concordances
20 Apr [Prep reading: Gries \(2009: 143-147\)](#)

IT'S ALL DOWNHILL FROM HERE...

- 15** **ARTICLE SUMMARY DUE TODAY**
27 Apr Make a Wish!

- 16** Revision and course evaluation
4 May

- 17** **FINAL EXAM**
11 May

REQUIRED READINGS

The required textbook for this class is:

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

The book has been ordered and should be available at the UNT book store. It is also available at amazon.com: < http://www.amazon.com/Quantitative-Corpus-Linguistics-Introduction-ebook/dp/B001Y35H5A/ref=sr_1_3?ie=UTF8&s=books&qid=1259387249&sr=8-3>.

All other course materials will be provided in class.

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 Kilgarrieff. *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.