San José State University
Linguistics & Language Development
LING115, Corpus Linguistics, Fall 2011

Instructor: Hahn Koo
Office Location: Clark Hall 485
Telephone: (408)-924-7093
Email: hahn.koo@sjsu.edu
Office Hours: M 1600 – 1700
Class Days/Time: M 1800 – 2045
Classroom: Clark Hall 242
Prerequisites: LING101 or instructor’s consent

Faculty Web Page and MYSJSU Messaging
Copies of the course materials such as the syllabus, major assignment handouts, etc. may be found at http://www.sjsu.edu/faculty/hahn.koo/teaching/ling115/ling115.html. You are responsible for regularly checking the class website for announcements.

Course Description
Introduction to the use of large collections of computer-readable text (“corpora”) in linguistics. Focus on text processing techniques and quantitative data analysis.

Course Goals and Student Learning Objectives
This is an introductory course on corpus linguistics. We will examine various uses of corpora in linguistic research and learn the basic computational skills and quantitative methods necessary in carrying out a corpus investigation.

Course Content Learning Outcomes
Upon successful completion of this course, students will be able to:
LO1 Demonstrate understanding of various uses of corpora in linguistic research;
LO2 Demonstrate understanding of quantitative methods utilized in corpus-based linguistic research;
LO3 Use basic Unix/Linux shell commands and the Python programming language to implement scripts for corpus analysis.

**Recommended Texts/Readings**

For those of you who are new to programming, I recommend the following book:


We will read and discuss the following scholarly articles below:


**Classroom Protocol**

All students are expected to act civilly in classroom.

- Arrive in time for class.
- Be respectful to the instructor and your classmates.
- Turn your cell phones off.
- Do not use personal computers for activities not directly related to class.
- Any activities deemed disruptive will be met with appropriate disciplinary actions, including expulsion from the classroom.

**Dropping and Adding**

Students are responsible for understanding the policies and procedures about add/drops, academic renewal, etc. Information on add/drops are available at [http://info.sjsu.edu/web-dbgen/narr/soc-fall/rec-298.html](http://info.sjsu.edu/web-dbgen/narr/soc-fall/rec-298.html). Information about late drop is available at [http://www.sjsu.edu/sac/advising/latedrops/policy/](http://www.sjsu.edu/sac/advising/latedrops/policy/). Students should be aware of the current deadlines and penalties for adding and dropping classes.

**Assignments and Grading Policy**

Your final grade in LING115 will be based on your performance on the following:

**Lab assignments (6 points \( \times 10 = 60 \) points)**

There will be ten lab assignments, each of which is worth 6 points. They are designed to help you practice what you learned during the lab sessions.

**Paper review (20 points)**

Find a scholarly journal article whose author(s) use corpus-based methods to address a topic that you find interesting. The article can be about any field of linguistics including TESOL. To get started, you can search for articles with keywords that reflect your interests on [http://scholar.google.com](http://scholar.google.com). Or you can browse articles published in journals such as the *International Journal of Corpus Linguistics*, *Corpus Linguistics and Linguistic Theory*, and *ICAME Journal*.

You will give an **in-class presentation (10 pts)** and submit a **written paper review (10 pts)** for the article you choose. Regarding format, the presentation should be about 15–20 minutes including Q&A. You should either prepare powerpoint slides or hand-outs for your presentation. The paper review should be about 3–4 pages long – single-spaced, 12 pt font, and one-inch margin all around. As for the contents, make sure you address the following:

- What is the research question of the article?
- What corpus did the author(s) use?
- How was the data from corpus analyzed?
- What are the conclusion and implications of the study?
- Why did you find the article interesting? That is, what is your research interest and in what way is the article relevant?

You should give your presentation either on 11/28 or 12/5. You should submit a hard-copy of your review by 12/5.

**Final project (20 points)**

You will write a program in Python that automatically constructs a thesaurus from the Wall Street Journal corpus. Further instructions will be announced in class.
Based on the total score you will have cumulated by the end of semester, you will receive your letter grade as summarized in the following table:

<table>
<thead>
<tr>
<th>Score(s)</th>
<th>Letter Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>98 – 100</td>
<td>A+</td>
</tr>
<tr>
<td>90 – 97</td>
<td>A</td>
</tr>
<tr>
<td>81 – 89</td>
<td>A-</td>
</tr>
<tr>
<td>74 – 80</td>
<td>B+</td>
</tr>
<tr>
<td>67 – 73</td>
<td>B</td>
</tr>
<tr>
<td>61 – 66</td>
<td>B-</td>
</tr>
<tr>
<td>54 – 60</td>
<td>C+</td>
</tr>
<tr>
<td>47 – 53</td>
<td>C</td>
</tr>
<tr>
<td>41 – 46</td>
<td>C-</td>
</tr>
<tr>
<td>34 – 40</td>
<td>D+</td>
</tr>
<tr>
<td>27 – 33</td>
<td>D</td>
</tr>
<tr>
<td>21 – 26</td>
<td>D-</td>
</tr>
<tr>
<td>0 – 20</td>
<td>F</td>
</tr>
</tbody>
</table>

A passing grade for the course is C or above.

**University Policies**

**Academic integrity**

Students should know that the University’s Academic Integrity Policy is available at [http://www.sa.sjsu.edu/download/judicial_affairs/Academic_Integrity_Policy_S07-2.pdf](http://www.sa.sjsu.edu/download/judicial_affairs/Academic_Integrity_Policy_S07-2.pdf). Your own commitment to learning, as evidenced by your enrollment at San Jose State University and the University’s integrity policy, require you to be honest in all your academic course work. Faculty members are required to report all infractions to the office of Student Conduct and Ethical Development. The website for Student Conduct and Ethical Development is available at [http://www.sa.sjsu.edu/judicial_affairs/index.html](http://www.sa.sjsu.edu/judicial_affairs/index.html).

Instances of academic dishonesty will not be tolerated. Cheating on exams or plagiarism (presenting the work of another as your own, or the use of another person’s ideas without giving proper credit) will result in a failing grade and sanctions by the University. For this class, all assignments are to be completed by the individual student unless otherwise specified. If you would like to include in your assignment any material you have submitted, or plan to submit for another class, please note that SJSU’s Academic Policy F06-1 requires approval of instructors.

**Campus Policy in Compliance with the American Disabilities Act**

If you need course adaptations or accommodations because of a disability, or if you need to make special arrangements in case the building must be evacuated, please make an appointment with me as soon as possible, or see me during office hours. Presidential Directive 97-03 requires that students with disabilities requesting accommodations must register with the DRC (Disability Resource Center) to establish a record of their disability.
Student Technology Resources

Computer labs for student use are available in the Academic Success Center located on the 1st floor of Clark Hall and on the 2nd floor of the Student Union. Additional computer labs may be available in your department/college. Computers are also available in the Martin Luther King Library.

A wide variety of audio-visual equipment is available for student checkout from Media Services located in IRC 112. These items include digital and VHS camcorders, VHS and Beta video players, 16 mm, slide, overhead, DVD, CD, and audiotape players, sound systems, wireless microphones, projection screens and monitors.

Learning Assistance Resource Center

The Learning Assistance Resource Center (LARC) is located in Room 600 in the Student Services Center. It is designed to assist students in the development of their full academic potential and to motivate them to become self-directed learners. The center provides support services, such as skills assessment, individual or group tutorials, subject advising, learning assistance, summer academic preparation and basic skills development. The LARC website is located at http://www.sjsu.edu/larc/.

SJSU Writing Center

The SJSU Writing Center is located in Room 126 in Clark Hall. It is staffed by professional instructors and upper-division or graduate-level writing specialists from each of the seven SJSU colleges. Our writing specialists have met a rigorous GPA requirement, and they are well trained to assist all students at all levels within all disciplines to become better writers. The Writing Center website is located at http://www.sjsu.edu/writingcenter/about/staff/.

Peer Mentor Center

The Peer Mentor Center is located on the 1st floor of Clark Hall in the Academic Success Center. The Peer Mentor Center is staffed with Peer Mentors who excel in helping students manage university life, tackling problems that range from academic challenges to interpersonal struggles. On the road to graduation, Peer Mentors are navigators, offering “roadside assistance” to peers who feel a bit lost or simply need help mapping out the locations of campus resources. Peer Mentor services are free and available on a drop-in basis, no reservation required. Website of Peer Mentor Center is located at http://www.sjsu.edu/muse/peermentor/.
The schedule is subject to change with fair notice. Changes will be announced in class and on the class website.

Table 1 Course Schedule

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8/29</td>
<td>Discussion: Introduction to Corpus Linguistics&lt;br&gt;Lab: Introduction to Unix/Linux</td>
</tr>
<tr>
<td>2</td>
<td>9/5</td>
<td>NO CLASS (Labor Day)</td>
</tr>
<tr>
<td>3</td>
<td>9/12</td>
<td>Discussion: Basic statistics and probability&lt;br&gt;Lab: Basic shell commands; Vim text editor</td>
</tr>
<tr>
<td>4</td>
<td>9/19</td>
<td>Discussion: Biber (1987)&lt;br&gt;Lab: Regular expressions</td>
</tr>
<tr>
<td>5</td>
<td>9/26</td>
<td>Discussion: Tsui (2005)&lt;br&gt;Lab: TREEBANK and TGrep2</td>
</tr>
<tr>
<td>6</td>
<td>10/3</td>
<td>Discussion: Erman (2001)&lt;br&gt;Lab: Python</td>
</tr>
<tr>
<td>7</td>
<td>10/10</td>
<td>Discussion: Nevalainen (2000)&lt;br&gt;Lab: Python</td>
</tr>
<tr>
<td>8</td>
<td>10/17</td>
<td>Discussion: Hay and Bauer (2007)&lt;br&gt;Lab: Python</td>
</tr>
<tr>
<td>9</td>
<td>10/24</td>
<td>Discussion: Baayen and Renouf (1996)&lt;br&gt;Lab: Python</td>
</tr>
<tr>
<td>10</td>
<td>10/31</td>
<td>Discussion: Gries (2006)&lt;br&gt;Lab: Python</td>
</tr>
<tr>
<td>11</td>
<td>11/7</td>
<td>Discussion: Hindle and Rooth (1993)&lt;br&gt;Lab: Python</td>
</tr>
<tr>
<td>12</td>
<td>11/14</td>
<td>Discussion: Gale et al. (1992)&lt;br&gt;Lab: Python</td>
</tr>
<tr>
<td>14</td>
<td>11/28</td>
<td>Student presentation</td>
</tr>
<tr>
<td>15</td>
<td>12/5</td>
<td>Student presentation</td>
</tr>
<tr>
<td>Final</td>
<td>12/14</td>
<td>Final project due by 19:30</td>
</tr>
</tbody>
</table>