

Linguistics 300: Words, networks, and complex systems

- **Time and Place:** MW 9 to 10:20, Cresap 101
- **Instructor:** Janet Pierrehumbert
2016 Sheridan Rm. 301
jbp@northwestern.edu
- **Prerequisites:** The course is intended equally for students who are interested understanding language as a system, and for students who are interested in using computational linguistic methods to study human social and cognitive dynamics. The recommended preparation is: *LING 334-0 20 Introduction to Computational Linguistics* or *LING 330-0 Research Methods in Linguistics* or a solid general background in computer science and/or mathematical methods in the social sciences. Contact Prof. Pierrehumbert by email for permission to enroll.
- **Objectives:**
 - Learn how networks are used to represent relations amongst words, amongst concepts, and amongst people.
 - Learn some tools for analyzing networks and their behaviors in complex systems.
 - Conceive and carry out an original research project on language and networks.
- **Grading:** Grades will be determined based on performance on two response papers 10% each, three labs (15% each), a final project (25%), and class participation.
- **Course description:** People know a lot of words, and combining words to communicate with each other is one of their most characteristic cognitive and social activities. One result is all the text contained in the estimated 12 billion web pages posted on the Internet as of the end of 2008. This course covers recent tools and results in network theory for understanding how words are related to each other. It uses the same tools to explore the relationships amongst documents and people that can be inferred by analyzing patterns of word use. Specific topics will include.
 1. **Spam:** How to generate spam, and why generating good spam is hard.
 2. **Ontologies:** Network-based approaches to relations amongst words, in psycholinguistics and AI.
 3. **Search engines:** Page ranking algorithms and document retrieval.
 4. **Social networks:** Convergence and differential of language communities on the internet and elsewhere.

Key concepts also find application in other complex systems domains such as epidemiology, finance, and systems biology. A brief overview of these parallels will be presented. Participants are encouraged to investigate applications compatible with their own interests, particularly when undertaking the final project

The course will be taught through lectures and discussion with laboratory exercises and a final project. For the laboratory exercises, students will explore the behavior of the Python code provided, and then modify the code to address additional questions. Previous knowledge of Python is not assumed, though background in some other programming or scripting language will be helpful. For the final project, students will read current literature on words, networks, and complex systems and design original followup studies (of approximately the scope of one laboratory exercise). Final projects will in general be done in teams that bring multiple sources of expertise to the problem. Individual projects by students with their own research program will also be accepted.

Course calendar: **L** denotes lectures, and **J** denotes jobs.

Date	Objectives
L: Mar 30- Apr 1	Introduction to meaningless language: Word sequences, ngrams, and spam.
L: Apr 6-15	Meaningful language: lexical relations and ontologies.
J: Apr 6	Post Lab 1 (Spam generator).
J: Apr 13	Lab 1 due. Ontology response paper assigned.
L: Apr 20-27	Structure of the internet. Issues in page ranking and page validity. Basic principles of the Google Page Ranking algorithm.
J: Apr 22	Ontology response paper due. Post Lab 2 (Explore the behavior of a mini page ranker).
L: Apr 29 - May 11	Social networks as communication networks. Measures of social cohesion. Information flow in social networks. Imitation and convergence.
J: Apr 29	Lab 2 due.
J: May 6	Form groups and brainstorm final projects.
J: May 11	Post Lab 3 (explore a small model of social influence).
L: May 13	Applications of course concepts to other domains.
J: May 18	Lab 3 due. Assign project-related response paper.
L: May 18-20	Discussion of Lab 3. Review of main course themes, with applications to projects.
L: May 25-27	Presentations of projects in progress: obtain feedback on your project.
J: May 25	Project-related response paper due.
J: June 1-5	WCAS Reading Week
J: June 12	Final project writeups due.

- There will be extended office hours during reading week.
- **Attendance:** Students are expected to attend class. Excessive absences will result in a substantial grade penalty.
- **Students with Disabilities:** If you have special needs, please contact the instructors as early as possible, and no later than the first week of class.

- **Academic Integrity at Northwestern:** You must comply with the Northwestern Academic Integrity guidelines; in particular, you should always acknowledge any external research. We will communicate more specific expectations for each assignment.
- **Homeworks and Labs:** You should feel free to work with other students on both the homeworks and the labs. However, each student must show their mastery of the material by writing their own lab solutions.

References

- [1] Tim Berners-Lee, James Hendler and Ora Lassila. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American Magazine*, May 17, 2001.
- [2] George A. Miller. WordNet: A Lexical Database for English. *Communications of the ACM*. 38(11): 39-41.
- [3] A. E. Motter, A. P. S. de Moura, Y.-C. Lai, and P. Dasgupta, Topology of the conceptual network of language. *Phys. Rev. E Rapid Communications*, 65, 065102.
- [4] Rebecca S. Wills. Google's PageRank: the math behind the search engine. *Math. Intelligencer*, 28(4):6-11, 2006.
- [5] Matthew J. Salganik, Peter Sheridan Dodds, and Duncan J. Watts. Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science* 311(5762), 854-856.
- [6] Dr. Seuss. The Sneetches. In *The Sneetches and Other Stories*, pages 2-25. Random House Inc., 1961.
- [7] and other selected articles. Tutorial material on network theory and formal language theory will also be provided.