

# 11-719 Computational Models of Discourse Analysis

## Instructor

Dr. Carolyn P. Rosé (cprose@cs.cmu.edu)

**Office hours:** Students are encouraged to request meetings with the instructor as needed

**Units:** 12 (PhD/Master's)

**Readings and On-Line Discussions :** All required readings will be posted as pdfs to blackboard. The following book is recommended but not required for the course:

James Paul Gee (2011). *An Introduction to Discourse Analysis: Theory and Method, Third Edition*, New York: Routledge

Students are expected to do the readings and post a response to discussion questions on-line to the weekly blogs in Blackboard by 8pm the night prior to each class meeting.

**Prerequisites:** students should be reasonably strong programmers and have taken or audited at least one machine learning course

**Class Meets:** Monday and Wednesday, 10:30am-11:50am, Gates Hillman Center 4102

## Course Description

Discourse analysis is the area of linguistics that focuses on the structure of language above the clause level. It is interesting both in the complexity of structures that operate at that level and in the insights it offers about how personality, relationships, and community identification are revealed through patterns of language use. A resurgence of interest in topics related to modeling language at the discourse level is in evidence at recent language technologies conferences. This course is designed to help students get up to speed with foundational linguistic work in the area of discourse analysis, and to use these concepts to challenge the state-of-the-art in language technologies for problems that have a strong connection with those concepts, such as dialogue act tagging, sentiment analysis, and bias detection.

This is meant to be a hands on and intensely interactive course with a heavy computational modeling component. The course is structured around 3 week units, each of which have a hands on assignment.

## Course Procedures and Grading Criteria

Most assignments will involve hands on computational work. You are welcome to use any tools you know and have access to. One way of approaching assignments with computational components is to program plugins for the LightSIDE text mining tool bench (<http://ankara.lti.cs.cmu.edu/side/download.html>). Plugins with either be novel feature extractors, classification algorithms, or meta-classifiers.

Grades will be assigned as follows:

20% for each of four assignments  
10% for class participation  
10% final project write up (Due by May 11)

## **Project/Assignments**

Each of the 4 assignments contributes directly to the course project, which will be a team project. The project should involve 3 (or at most 4) students each.

1. Preliminary: Pick one of the 4 data sets by January 25
  - a. *Example Data Sets*: Street Gangs Forum, Coursera MOOC discussions, Medical Support Forum, Wikipedia Talk Pages, GitHub Issue Corpus
2. Assignment 1 Part A: Model interpretation, due before class on February 1
3. Assignment 1 Part B: Analysis by hand, revisit model interpretation, due before class on February 10
4. Assignment 2: Preliminary analysis of project data set, due before class on February 29
5. Assignment 3: Project Proposal with preliminary results, due before class on March 28
6. Assignment 4: Final analysis, schedule meeting between team and instructor for the week of May 2
7. Final paper due May 11

## **Semester Reading Schedule and Deadlines**

### **UNIT 1: Foundations: Analysis of Conversational Strategy (Jan 11, 2016)**

#### **Week 1 Lecture 1 (Jan 11):** Course Intro

#### **Week 1 Lecture 2 (Jan 13):** Defining Social Goods and Achievement of Social Goals

- Gee, J. (2011). *An Introduction to Discourse Analysis: Theory and Method*, Routledge, Excerpts
- Gayo-Avello, D., Metaxas, P., Mustafaraj, E. (2011). Limits of Electoral Predictions Using Twitter, in *Proceedings of the Fifth International AAI Conference on Weblogs and Social Media*

#### **Week 2 Lectures 2 (Jan 20):** Modeling Achievement of Social Goals

- Vlad Niculae, Srijan Kumar, Jordan Boyd-Graber, Cristian Danescu-Niculescu-Mizil (2015). Linguistic Harbingers of Betrayal: A Case Study on an Online Strategy Game. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, China
- Caroline Langlet Chloe Clavel (2015). Improving social relationships in face-to-face human-agent interactions: when the agent wants to know user's likes and dislikes. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, China

#### **Week 3 Lecture 1 (Jan 25):** Simple Graphical Models of Conversational Structure

- Review LDA if you are not familiar with it:
  - Griffiths, T., Steyvers, M., Tenenbaum, J. (2007). Topics in Semantic Representation, *Psychological Review* 114(2), pp 211-244.
- Paul, M. (2012). Mixed Membership Markov Models for Unsupervised Conversation Modeling, in *Proceedings of Empirical Methods in Natural Language Processing*

- Roberts, M., Stewert, B., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S., Albertson, B., & Rand, D. (2014). Structural Topic Models for Open-Ended Survey Responses, *American Journal of Political Science* 58(4), pp 1064-1082. [Note that the code and data are publically available]
- DEADLINE: Project dataset Sign-Up (using Google Form sent by email)

**Week 3 Lecture 2 (Jan 27):** Studying Discussion Practices in Social Media

- Jones, R., Chik, A., Hafner, C. (2015). Discourse and Digital Practices: Doing Discourse Analysis in the Digital Age, Routledge (Chapter 4)
- Bender, E., Morgan, J., Oxley, M., Zachary, M., Hutchinson, B., Marin, A., Zhang, B., Ostendorf, M. (2011). Annotating Social Acts: Authority Claims and Alignment Moves in Wikipedia Talk Pages, *Proceedings of the HLT-NAACL 2011 Workshop on Language and Social Media*.
- Maren, A., Ostendorf, M. (2011). Detecting Forum Authority Claims in Online Discussions, *Proceedings of the HLT-NAACL 2011 Workshop on Language and Social Media*.

**UNIT 2: Power, Politeness and Intimacy (Feb 1, 2016)**

**Week 4 Lecture 1 (Feb 1):** Power, Gender and Style in Blogs

- Mooney et al. (2011). Language, Society, and Power, 3<sup>rd</sup> Edition, Routledge (Chapter 8)
- Page, R. (2012). *Stories and Social Media: Identities in Interaction*, Routledge (Chapter 3)
- Gianfortoni, P., Adamson, D., & Rosé, C. P. (2011). Modeling Stylistic Variation in Social Media with Stretchy Patterns, in *Proceedings of First Workshop on Algorithms and Resources for Modeling of Dialects and Language Varieties*
- DEADLINE: Assignment 1 Part A

**Week 4 Lecture 2 (Feb 3):** Theoretical Framing of Politeness Theory

- Penelope Brown and Stephen C Levinson. 1987. Politeness: some universals in language usage. Cambridge University Press. (excerpts)

**Week 5 Lecture 1 (Feb 8):** Modeling Politeness

- Danscu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., & Potts, C. (2013). A computational approach to politeness with application to social factors, *Proceedings of the 51<sup>st</sup> Annual Meeting of the Association for Computational Linguistics*, pp352-361.

**Week 5 Lecture 2 (Feb 10 – LP Guest Lecture):** Emotion and Intimacy

Reading TBA

- DEADLINE: Assignment 1 Part B

**Week 6 Lecture 1 (Feb 15):** Modeling Intimacy and Self-Disclosure

- Bak, J., Lin, C., & Oh, A. (2014). Self-disclosure topic model for classifying and analyzing Twitter conversations. Proceedings of the International Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), Doha, Qatar, 2014

**Week 6 Lecture 2 (Feb 17):** Power Differences at Work and in the Courtroom

- Danscu-Niculescu, C., Lee, L., Pang, B., Kleinberg, J. (2012). Echoes of power: Language effects and power differences in social interaction, in *Proceedings of WWW '12*
- Prabhakaran, V., John, A., & Seligmann, D. (2013). Who Had the Upper Hand? Ranking Participants of Interactions Based on Their Relative Power. *Proceedings of the International Joint Conference on Natural Language Processing*, pages 365–373.

**UNIT 3: Perspective and Meaning Making (Feb 22)**

**Week 7 Lecture 1 (Feb 22)**

- Richards, K. (2006). *Language and Professional Identity: Aspects of Collaborative Interaction*, Palgrave, Chapter 3
- Martin, J. & White, P. (2005). *The Language of Evaluation: Appraisal in English*, Palgrave, Chapter 3.

**Week 7 Lecture 2 (Feb 24):** Accommodation and Transactivity: From Theory to Computation

- Jain, M., McDonogh, J., Gweon, G., Raj, B., Rosé, C. P. (2012). An Unsupervised Dynamic Bayesian Network Approach to Measuring Speech Style Accommodation, *EACL 2012 Proceedings of the 13th Conference of the European Association for Computational Linguistics*, Avignon, France, April 23-27, 2012, pp787-797.
- Gweon, G., Jain, M., Mc Donough, J., Raj, B., Rosé, C. P. (2013). Measuring Prevalence of Other-Oriented Transactive Contributions Using an Automated Measure of Speech Style Accommodation, *International Journal of Computer Supported Collaborative Learning* 8(2), pp 245-265.

**Week 8 Lecture 1 (Feb 29):** Modeling Argumentation and Perspective

- Fairclough, I., Fairclough, N. (2012). *Political Discourse Analysis: A Method for Advanced Studies*, Routledge (Chapter 3).
- Oren Tsur Dan Calacci David Lazer A (2015). Frame of Mind: Using Statistical Models for Detection of Framing and Agenda Setting Campaigns. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, China
- DEADLINE: Assignment 2

**Week 8 Lecture 2 (Mar 2):** Metaphor and Meaning Making

- Shutova, E., Teufel, S., & Korhonen, A. (2013). Statistical Metaphor Processing, *Computational Linguistics* 39(2), pp 301-353

SPRING BREAK!!!

**Week 9 Lecture 1 (Mar 14):** Similarity and Difference

- Veale, T. & Li, G. (2013). Creating Similarity: Lateral Thinking for Vertical Similarity Judgments, *Proceedings of the 51<sup>st</sup> Annual Meeting of the Association for Computational Linguistics*

- Riloff, E., Qadir, A., Surve, P., de Silva, L., Gilbert, N., Huang, R. (2013). Sarcasm as Contrast between a Positive Sentiment and Negative Situation, *Proceedings of Empirical Methods in Natural Language Processing*
- Wallace, B., Choe, D., and Charniak, E. (2015). Sparse, Contextually Informed Models for Irony Detection: Exploiting User Communities, Entities and Sentiment, *Proceedings of Empirical Methods in Natural Language Processing*

**Week 9 Lecture 2 (Mar 16): Metaphor and Conversation**

- Jang, H., Piergallini, M., Wen, M., Rosé, C. P. (2014). Conversational Metaphors in Use: Exploring the Contrast between Technical and Every Day Notions of Metaphor, in *Proceedings of the Second Workshop on Metaphor in NLP (Meta4NLP)*, Association for Computational Linguistics.
- Jang, H., Moon, S., Jo, Y., Rosé, C. (2015). Metaphor Detection in Discourse, *Proceedings of SIGDIAL, 2015*

**Week 10: Project work week, full proposal due following Monday**

**UNIT 4: Legitimacy and Identity: Interactions “on the Street” (Mar 25, 2015)**

**Week 11 Lecture 1 (Mar 28): Sociolinguistic Perspective on Style: What Linguistic Resources Buy You**

- Myers-Scotton, C (1993). *Social Motivations for Codeswitching: Evidence from Africa*, Clarendon Press (excerpts)
- Eckert, P. & Rickford, J. (2001). *Style and sociolinguistic variation*. Cambridge University Press (excerpts)
- DEADLINE: Proposal and preliminary results

**Week 11 Lecture 2 (Mar 30) Negotiating Legitimacy in Interaction**

- Adams, K. & Winter, A. (1997). Gang graffiti as a discourse genre, *Journal of Sociolinguistics* 1/3. Pp 337-360.
- Garot, R. (2007). “Where You From!?”: Gang Identity as Performance, *Journal of Contemporary Ethnography*, 36, pp 50-84.

**Week 12 Lecture 1 (Apr 4) Computational Analysis of Gang Language**

- Piergallini, M., Gadde, P., Dogruoz, S., Rosé, C. P. (2014). Modeling the Use of Graffiti Style Features to Signal Social Relations within a Multi-Domain Learning Paradigm, *Proceedings of the European Chapter of the Association for Computational Linguistics*
- Hassan, A., Abu-Jbara, A., Radev, D. (2012). Detecting Subgroups in Online Discussions by Modeling Positive and Negative Relations among Participants, in *Proceedings of Empirical Methods in Natural Language Processing*
- Dian Yu, Yulia Tyshchuk, Heng Ji, William Wallace (2015). Detecting Deceptive Groups Using Conversations and Network Analysis. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, China

**Week 12 Lecture 2 (Apr 6) Impersonation and Multiple Identities**

- Schwartz, R., Tsur, O., Rappoport, A., Koppel, M. (2013). Authorship Attribution of Micro-Messages, *Proceedings of Empirical Methods in Natural Language Processing*
- Nguyen, V., Boyd-Graber, J., Resnik, P., Miler, K. (2015). Tea Party in the House: A Hierarchical Ideal Point Topic Model and Its Application to Republican Legislators in the 112th Congress, *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, China

**UNIT 5: Social Positioning through Narrative (April 11, 2015)**

**Week 13 Lecture 1 (Apr 11) Positioning events and characters**

- Page, R. (2012). *Stories and Social Media: Identities and Interaction*, Routledge (Chapter 4)
- Bamman, D., O'Connor, B., and Smith, N. (2013). Learning latent personas of film characters, *Proceedings of the 51<sup>st</sup> Annual Meeting of the Association for Computational Linguistics*, pp352-361.

**Week 13 Lecture 2 (Apr 13) Positioning the Author among Colleagues**

- Hyland, K. (2004). *Disciplinary Discourses: Social Interactions in Academic Writing*, The University of Michigan Press (excerpts)
- Perenz, P. & Webber, B. (2011). Stable Classification of Text Genres, *Computational Linguistics* 37(2), pp385-393

**Week 14 Lecture 2 (Apr 18): Analysis of Point of View**

- Recasens, M., Danescu-Niculescu-Mizil, C., Jurafsky, D. (2013). Linguistic Models for Analyzing and Detecting Biased Language, *Proceedings of the 51<sup>st</sup> Annual Meeting of the Association for Computational Linguistics*
- Sridhar, D., Foulds, J., Huang, B., Getoor, L., Walker, M. (2015). Joint Models of Disagreement and Stance in Online Debate. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, China

**Week 14 Lecture 2 (Apr 20) Positioning the Narrator**

- Booth, W. (1983). *The Rhetoric of Fiction*, The University of Chicago Press (excerpts)

**Week 15 Lecture 1 Reflecting situational factors through personal expression**

- Miaomiao Wen, Zeyu Zheng, Hyeju Jang, Guang Xiang, and Carolyn Rosé. Extracting Events with Informal Temporal References in Personal Histories in Online Communities. ACL'13, 2013
- Hyeju Jang, Miaomiao Wen, Carolyn Rosé, "Effects of Situational Factors on Metaphor Detection in an Online Discussion Forum", 3rd Workshop on Metaphor in NLP at NAACL, 2015.

**Week 15 Lecture 2: Group work day, work on final project**

- DEADLINE: Assignment 4, Each team must set up an in person meeting with the instructor during the week of May 2 to report on in progress results for the final project and get feedback. Final Paper due on May 11, 2016