New Master of Science in Computational Linguistics

Montclair State University

Our program

- New: First cohort in Fall 2019
- 32 credit hours
- Flexible (full-time/part-time)
- No GRE (for Fall 2019 and Spring 2020)
- The first master's program of its kind in New Jersey
- The program is unusual in that it gives **equal emphasis** to **linguistics and computer science skills**. This combination best **meets the demands of the robust job market** for computational linguists.

What do computational linguists do?

- CL is an interdisciplinary field concerned with the statistical and rule-based modeling of natural language from a computational perspective as well as the study of computational approaches to linguistic questions.

- Computational linguists work for high tech companies, creating and testing models for improving or developing new software in areas such as **speech recognition**, **machine translation**, **grammar checking**, **dictionary developing, text-to-speech synthesis, conversational agents,** and more.

- They also work in research groups at universities and government research labs.

Computational linguists work on...

- speech recognition and synthesis
- sentiment analysis
- automatic text summarization
- machine translation
- natural language generation
- natural language understanding
- question answering
- and many others

Where are computational linguists employed?

- Notable companies that regularly employ computational linguists include **Amazon,* Apple, Audible,* Facebook, Google,* Grammarly, IBM, Intel, Interactions,* Lingsoft, Lionbridge, Microsoft, Nuance,* Oracle**, as well as many smaller companies in the region.

  *Our students are working at these companies.

Some demos

- Google translate (voice support):

https://translate.google.com/?hl=en&tab=TT&authuser=0

- Text summarization, sentiment analysis, entity extraction:

https://developer.aylien.com/text-api-demo?text=&language=en&tab=classify-taxonomy

Application Requirements

1. Submitted Online Application
2. Official transcript from each college attended (other than Montclair State)
3. Statement of Purpose
4. Two Letters of Recommendation
5. For Fall 2019 and Spring 2020 the GRE requirement is waived.

Applying Is Easy

1. Go to montclair.edu/graduate/how-to-apply/ and Register for an account for the Fall 2019 semester.
2. Begin and submit your online application.
3. Save $60.00: the application fee will be waived for Fall 2019 automatically.
4. Once you have submitted the application, you'll be able to view and upload the specific documents that are required in order to complete your application.

Applicants with International Educational Background

Students holding degrees from non-US institutions must request a course-by-course transcript evaluation from a member of NACES.org or by EducatedChoices.com. TOEFL or IELTS is required for applicants to any program who earned their degree in a country where English is not the official language. The institutional code for test reporting in 2520.

MONTCLAIR STATE
UNIVERSITY

Contact Information:
The Graduate School
Montclair State University
1 Normal Avenue
Montclair, NJ 07043
graduate.school@montclair.edu
Phone: 973-655-5147
montclair.edu/graduate

Upon acceptance into the program...

- Students will be assigned a faculty program advisor (PA)
- In case a student is accepted into the program without the necessary background, he/she can complete additional courses to make up the gaps.

Our Labs on the Montclair Campus

- **The Natural Language Processing Lab** (located in CCIS 424) conducts research in resource-light morphology, deception detection, computational approaches to figurative language, misinformation and Internet censorship, text mining, machine learning and big data analytics.

- **The Experimental Linguistics Lab** (located in CS 389) houses research that involves data collection from subjects, including research on the prosody/semantic interface and second language learner research using electroencephalography (EEG) to track language processing in real-time.

Recent Research Talks (Brown Bags)

- April 11, 2019
  Chelsey Hill, Assistant Professor of Information Management and Business Analytics, Feliciano School of Business
  Brown Bag Talk: Complaint Mining in the Automotive Industry

- March 27, 2019
  Richard Sproat, Research Scientist, Google Research, New York
  Neural Models of Text Normalization for Speech Applications

- March 7, 2019
  Heng Ji, Rensselaer Polytechnic Institute
  Universal Information Extraction

- February 13, 2019
  Joe Tetreault, Grammarly
  It's a Matter of Style: Experiments in Style Detection and Transformation with Natural Language Processing

Internship and Externship Opportunities

- The Linguistics and Computer Science faculty are all actively engaged in research. There are always paid and unpaid opportunities to get a hands-on experience.

- Montclair State is located 12 miles west of New York City. There are many companies both in New Jersey and New York that search for skills in Natural Language Processing/Computational Linguistics.

- Our alumni who took courses in CL either had internships or obtained full-time employment **at Google**, **Amazon**, **Nuance**, **Merck**, **Interactions**, **Red Hat, Acquire Media, Berlitz, ETS, Ethnic Technologies, Veracode, CQ Fluency, Ogilvy Common Health**, among others.

- Many continued for their doctoral studies at **Columbia University, Stevens Institute of Technology, CUNY, Washington University, and Penn State University**.

Coursework

- 26 credits of required courses
- 6 credits of electives

Required courses (26 credits)

1. APLN 504 Syntax (3)
2. APLN 505 Semantics & Pragmatics (3)
3. APLN 506 Phonetics and Phonology (3)
4. APLN 550 Computational Linguistics (3)
5. APLN 553 Text Analysis Tools (3)
6. APLN 552 Current Issues in NLP (3)
7. CSIT 505 Python (4)
8. CSIT 598 Machine Learning (3)
9. APLN 605: project in NLP (1)

Electives (at least two)

- APLN 540 Linguistic Annotation (3)
- APLN 563 Quantitative Linguistics (3)
- APLN 570 The Structure of ASL   (3)
- APLN 580 Corpus Linguistics (3)
- APLN 582 Languages & Mobile Communication (3)
- APLN 590 Topics in Applied Linguistics (3)
- CSIT 528 Statistics for Data Sciences (3)
- CSIT 553 Exploratory Data Analysis & Visualization (3)
- CSIT 554 Big Data Analytics (3)
- CSIT 555 Database Systems (3)
- CSIT 558 Data Mining (3)
- CSIT 571 Computer Algorithms & Analysis (3)
- INFO 583 Introduction to Business Analytics (3)
- MATH 535 Linear Algebra I   (3)
- **Other courses might be taken with an approval of the Program Advisor.**

Prerequisites (can be taken concurrently)

- **Prerequisites for the students with a B.A. in Linguistics:**

CSIT 503 Data Structures (4) (in Python)
CSIT 504 Discrete Math in Computing (4)

- **Prerequisites for the students with a BS in Computer Science:**
  APLN 500 Language and Linguistics (3)

Students coming in with a background other than Linguistics or Computer Science will have to complete up to 15 credits of prerequisite classes. Students must maintain a minimum GPA of 3.0 in order to continue in the program.

Faculty

- **Constantine Coutras, Chair of Computer Science** *Computer and network security*

- **Anna Feldman, MS in CL Coordinator** *Computational morphology, computational approaches to linguistic creativity, language learner corpora, language of disinformation*

- **Eileen Fitzpatrick, Chair Of Linguistics** *Automatic detection of deceptive language, language learner corpora*

- **Jonathan Howell, Linguistics** *Prosody, laboratory phonology, formal semantics, speech processing*

- **Chris Leberknight, Computer Science** (Social Network Analysis, Social Computing and Information Economics; computational methods for modeling and understanding human behavior in social networks)

- **Jing Peng, Computer Science** (Machine Learning, NLP; reinforcement learning, classification, learning representations; figurative language, disinformation

- **Aparna Varde, Computer Science** (data mining, artificial intelligence, machine learning, environmental management, big data analysis)

Costs

- In-state tuition: $677.10 per credit
- Fees: $36.09 per credit

- **32 credits = $22,822**
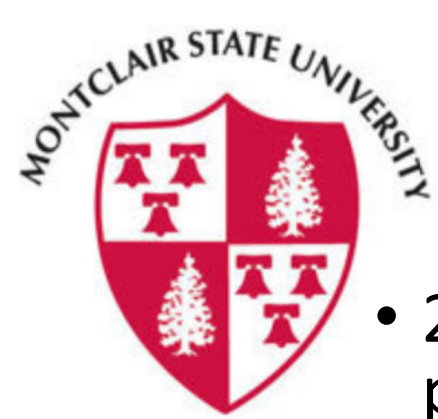
Research Grant Support

- 2017-2019 Howell, Jonathan. Establishing a ground truth for focus placement in naturally-occurring speech. NSF: Early-concept Grants for Exploratory Research. $105,894 (2 years).

- 2017-2020 Leberknight & Feldman. SaTC: CORE: Medium: Collaborative: A Linguistically-Informed Approach for Measuring and Circumventing Internet Censorship. $1M grant in collaboration with Princeton University.

- 2013-2017 Feldman & Peng. AIR: Automatic Idiom Recognition. National Science Foundation (NSF)(Computer and Information Sciences and Engineering) $176,514

- Jing Peng and Anna Feldman. DURIP. Dynamic Data-Driven Fusion and Scene Understanding for Real Time Queries over Live Streaming Video $145,170.

Student Support

- We have a limited number of GA positions available each year. They are competitive and usually cover a tuition or/and provide a stipend.

- Next fall we'll have 2-4 GA positions (1-2 through MSU and 1-2 through funded research).

- We also support travel costs for students with accepted papers.

Student Publications (1)

1. **Ng Kei Y.**, Feldman A., Peng J., and C. Leberknight **2019.** Neural Network Prediction of Censorable Language. In Proceedings of the 3rd Workshop on NLP and Computational Social Science (NLP+CSS) held in conjunction with NAACL 2019

2. **Ng Kei Y.**, Feldman A., Peng J. and C. Leberknight. **2018.** Linguistic Characteristics of Censorable Language on SinaWeibo. In Proceedings of The COLING 1st Natural Language Processing for Information Freedom workshop.

3. **Ng Kei Y.**, Feldman A., and C. Leberknight. **2018.** Detecting Censorable Content in Social Media: A Pilot Study. In Proceedings of Natural Language Processing for Social Media Analysis (NLP4SMA) 2018. 10th Hellenic Conference on Artificial Intelligence (SETN-2018).

4. **Joyce E., Goldeck M.,** Leberknight C., and Feldman A. **2018.** Apollo: A System for Tracking Internet Censorship. In Proceedings of the 13th Pre-ICIS Workshop on Information Security and Privacy

5. **Kateryna Kaplun**, Christopher Leberknight, Anna Feldman. **2018.** Controversy and Sentiment: An Exploratory Study. In Proceedings of Natural Language Processing for Social Media Analysis (NLP4SMA) 2018. 10th Hellenic Conference on Artificial Intelligence (SETN-2018)

6. **Kateryna Kaplun**, Christopher Leberknight, Anna Feldman. **2018.** A Comparison of Lexicons for Detecting Controversy. In Proceedings of the LREC 2018 Workshop: Natural Language Processing meets Journalism III, Miyazaki (Japan)

Student Publications (2)

7. **Katsiaryna Aharodnik**, Anna Feldman, Jing Peng. **2018.** Designing a Russian Idiom-Annotated Corpus. In Proceedings of the 11th edition of the Language Resources and Evaluation Conference, 7-12 May 2018, Miyazaki (Japan)

8. Jing Peng, **Katsiaryna Aharodnik**, Anna Feldman. **2018.** A Distributional Semantics Model for Idiom Detection: The Case of English and Russian. In Proceedings of the 10th International Conference on Agents and Artificial Intelligence (Special Session on Natural Language Processing in Artificial Intelligence - NLPinAI 2018

9. **Manali Pradhan**, Jing Peng, Anna Feldman, **Bianca Wright.** "Human or machines, it's all about context.," *Proceedings of Computational Linguistics and Intelligent Text Processing, CICLing*, 2017. **best paper award + best presentation award**

10. **Emily Olshefski.** 2015. Game-Changing Event Definition and Detection in an eSports Corpus. In Proceedings of the 3rd NAACL Workshop on EVENTS: Definition, Detection, Coreference, and Representation, Denver, Colorado.

11. Jing Peng, Anna Feldman, and **Hamza Jazmati**. 2015. Classifying Idiomatic and Literal Expressions Using Vector Space Representations. In Proceedings of RANLP, Hissar, Bulgaria, September 2015.

12. **Matthew Mulholland** and **Joanne Quinn**. "Suicidal Tendencies: The Automatic Classification of Suicidal and Non-Suicidal Lyricists Using NLP." IJCNLP 2013.

Student publications (3)

13. **Katsiaryna Aharodnik**, **Marco Chang**, Anna Feldman, and Jirka Hana. "Automatic identification of learners' language background based on their writing in Czech." IJCNLP 2013.

14. Jirka Hana, Anna Feldman, and **Katsiaryna Aharodnik**. 2011. A low-budget tagger for Old Czech. In Proceedings of the 5th ACL/HLT 2011 LaTeCH workshop.

15. **Amal Kaluarachchi**, Aparna Varde, Srikanta Bedathur, Gerhard Weikum, Jing Peng and Anna Feldman. 2010. Incorporating Terminology Evolution for Query Translation in Text Retrieval with Association Rules. In Proceedings of the 19th ACM/CIKM conference.

16. **Amal C. Kaluarachchige**, Aparna Varde, Jing Peng, and Anna Feldman. 2010. Intelligent Time-Aware Query Translation for Text Sources. In Proceedings of the 24th AAAI Conference.

17. **Laura Street, Rachel Silverstein, Nathan Michalov (and other students).** 2010. Like Finding a Needle in a Haystack: Annotating the American National Corpus for Idiomatic Expressions. In Proceedings of LREC 2010. **[BA/MA Fall 09 Class project].**

18. Jing Peng, Anna Feldman, and **Laura Street**. 2010 Computing Linear Discriminants for Idiomatic Sentence Detection. In *Research in Computing Science*.

19. **Hiroki Yamakawa**, Jing Peng, and Anna Feldman. 2010. Semantic Enrichment of Text Representation with Wikipedia for Text Classification. In Proceedings of SMC2010

Questions?