USC Viterbi
School of Engineering

Course ID and Title: CSCI 699 History of Language and Computing
Units: 4
Term—Day—Time: Spring 2025 — TuTh — 4–5:50pm
Location: GFS 220

Instructor: Jesse Thomason
Office Hours: TBA
Location: TBA

Course Description

This course is designed for PhD students with an interest in understanding the bases and common assumptions in modern natural language processing research. We will study the history of thought and paradigms surrounding language and computing. We will read original texts as well as retrospectives and summary arguments from influential writers and researchers in recent history as well as those predating modern computation. Students will draw connections between historical perspectives and abstractions to modern day technological innovations and assumptions in natural language processing. Students will develop a rich understanding of the historical context of their own work in computing and language, and be better prepared to situate their research contributions in the long context of language processing.

Learning Objectives

By the end of this course, students will be able to:

- **O1**: Review historical perspectives from writers and actors in the history of natural language processing to communicate it effectively to colleagues.
- **O2**: Identify the linguistic assumptions made by an individual research contribution or common paradigm in natural language processing.
- **O3**: Develop an insight based on a historical perspective and demonstrate that insight on a problem faced by modern natural language processing.

Description of Assignments and How They Will Be Assessed

Roleplaying Paper Reading Seminars

Students will take on active Roles (detailed here) throughout the semester that serve as anchors for discussing reading materials in class. Roles include things like: presenting a summary of the work verbally to the class; acting as a "reviewer" of the paper as though it were a conference submission; an industry practitioner aiming to monetize aspects of the paper; a hacker trying to

reimplement the methods in the paper; being a teacher who scopes out interesting discussion questions to ask the group about the paper and leads that discussion; or a societal impact assessment officer who studies how the results in the paper could lead to help or harm for different groups of people. For your breakout and chosen Role, you will turn in a 2 page document based on your Role. If you're curious to get an idea in advance, the original ideas for these seminar breakouts are detailed here. Students will be graded on both their participation in the seminar discussion in the capacity of their Role and their page report for each session. Reports should be prepared in gDoc or LaTeX (templates here).

There are 13 planned discussion weeks throughout the semester, and students are expected to play an active Role in 10 discussion weeks. In other words, the 30% course grade for discussion and 20% course grade for Role reports are earned in 3% and 2% increments, respectively, per week. For weeks where students are not playing an active Role, they should in general still expect to attend and participate freely in the discussion.

Each seminar discussion week will involve reading historical material and reading modern research papers in NLP before class and discussion of this work during class. Reviewing historical perspectives from writers and actors in the history of natural language processing and communicating it effectively to colleagues develops skills for learning objective **O1**, while identifying the linguistic assumptions made by an individual modern research contribution in natural language processing develops skills for learning objective **O2**.

In-class Quizzes and Activities
Throughout the semester, some discussion days may include "project roundtables", activities in which you will work within your group and across other groups to refine your project ideas, hypotheses, approaches, and science communication plans. Additionally, the *Assessor* Role may create graded quizzes and activities to test knowledge of the reading material or engage more deeply with concepts covered therein. In expectation, there will be 10 or more such in-class activities, each worth 1%. Your aggregate score over the 10 or more graded activities will be the sum of your 10 highest scores across those activities.

Course Project
Throughout the semester, students will develop an insight based on a historical perspective and demonstrate that insight on a problem faced by modern natural language processing, developing and demonstrating skills for learning objective **O3**. The course project deliverables will include written midterm and final reports, as well as midterm and final oral presentations to the class. The course project can take the form of a research project demonstrating the implementation of the

developed insight and how it affects outcomes on the identified problem faced by modern NLP. Alternatively, the course project can take the form of a detailed literature review contextualizing an insight or paradigm the student identifies through historical literature; this review must survey material comprising 50% or more works beyond the reading materials covered in the course. Students should work in small teams of two.

Midterm Presentation Rubric (Total 05%):

- Answer the question: What are you trying to do? Articulate your objectives using absolutely no jargon. (01%)
- State a testable hypothesis; clearly identify the independent and dependent variables and the metric by which you will measure the dependent variables for comparison. (02%)
- Discuss preliminary results and findings and relate them to the stated hypothesis; cover any immediate next steps you plan to take. (01%)
- Remain under the time limit of 15 minutes for speaking, then answer audience questions for up to 3 minutes following. (01%)

Midterm Report Rubric (Total 10%):

- Conforms to the ACL style templates here with a 4 page limit excluding references. (01%)
- Contains an *Introduction* section that succinctly and clearly answers Heilmeier questions 1–4. (03%)
- Contains a *Related Work* section that indexes related literature in both the immediate *research* background and the intellectual/philosophical/linguistic *context* background and school(s) of thought at play in the proposed work. (02%)
- Contains a *Preliminary Methods* section outlining the high level hypothesis or claim and at least one testable hypothesis with dependent and independent variables identified alongside the metric by which the dependent variable will be measured. The hypothesis should be stated as a truth condition, and the associated statistical or other concrete evaluation test planned to reject or accept the hypothesis should be included. (02%)
- Contains a *Preliminary Results and Next Steps* section outlining any experiments conducted so far aligned to the stated hypotheses. Lays out weekly or biweekly milestones up to the final report date and an associated division of responsibilities among team members. (02%)

Final Presentation Rubric (Total 10%):

- Tell us about your project and key findings. You can shorten introductory content that the audience should remember or have absorbed from discussions and the midterm report. You want to deliver the core idea of what you wanted to achieve, how it goes beyond what was done before, why it matters, the hypotheses you wanted to evaluate, how you evaluated them experimentally, and the conclusions of your findings so far. Remain

under the time limit of 15 minutes for speaking, then answer audience questions for up to 3 minutes following.

Final Report Rubric (Total 15%):

- Conforms to the ACL style templates here with an 8 page limit. The page limit excludes: references, the ACL-required Limitations section, and optional Appendix. The Limitations and References sections post-page-limit are required. (01%)
- Contains an *Introduction* section that succinctly and clearly answers Heilmeier questions 1–4. (02%)
- Contains a *Related Work* section that indexes related literature in both the immediate *research* background and the intellectual/philosophical/linguistic *context* background and school(s) of thought at play in the submitted work. This section should contain updates since the midterm report. (03%)
- Contains a *Methods and Experiments* section outlining the high level hypothesis or claim and at least one testable hypothesis with dependent and independent variables identified alongside the metric by which the dependent variable will be measured. The hypothesis should be stated as a truth condition, and the associated statistical or other concrete evaluation test planned to reject or accept the hypothesis should be included. The experimental protocol for testing these hypotheses should be described clearly in this section. (04%)
- Contains a *Results and Conclusions* section outlining any experiments conducted aligned to the stated hypotheses and how these results support or contradict those hypotheses. Where applicable, appropriate statistical tests and corrections should be run to make quantified claims. (04%)
- Contains a *Future Work* section divided into two parts:
    - What *Future Work* could be built on this work, similar to what one would see in a published research paper. (01%)
    - What *Future Work* you, the team, are actually considering, if any. If you are not planning to continue the work in any way, state that. (00%; this section is just for planning)

Participation

In-class roleplaying paper reading seminar session discussions require in-class participation.

Grading Breakdown

Assessment Tool (assignments)	**% of Grade**
Reading Seminar Session Discussion Contributions	30%
Reading Seminar Session Written Role Reports	20%
In-Class Quizzes and Activities	10%
Course Project Midterm Presentation	05%
Course Project Midterm Report	10%

Assessment Tool (assignments)	% of Grade
Course Project Final Presentation	10%
Course Project Final Report	15%
TOTAL	**100%**

Grading Scale

A [93-
A- [90-93)
B+ [87-90)
B [83-86)
B- [80-83)
C+ [77-80)
C [73-76)
C- [70-73)
D+ [67-70)
D [63-66)
D- [60-63)
F [0-59)

Assignment Submission Policy

Written reports must be submitted electronically by 11:59pm on the due date.

Late Submission Policy

Late assignments will have their total grade reduced by 5% for every day late they are turned in. Each student will have **5 Late Day Tokens** to be used in **integer amounts** and distributed as the student sees fit. Any exception needs to be discussed within the first 2 weeks of the semester (no exception otherwise). Late Day Tokens *cannot* be redeemed for in-class quizzes, presentations, or discussions, but *can* be used for written deliverables. Late Day Tokens *cannot* be redeemed for any other purpose than removing 5% per-day penalties on late assignments. Late Day Tokens may be used by a subset of group members for project deliverables that are turned in late (e.g., members who do not utilize tokens would receive late day penalties, while those that redeem tokens would not; that is to say, Late Day Tokens are all *personal*, not group-level).

Attendance

Because attendance is needed for paper roleplaying seminar sessions, students should give advance notice of expected absences when possible. Each roleplaying session will be “staffed” by a subset of students taking on specific Roles. While most students will have a Role for every discussion, there will be slack built into the schedule to enable scheduled absences from students corresponding to class days on which they have no assigned seminar Role. Students who miss class unexpectedly can utilize Late Day Tokens for written

deliverables, and may be able to make up Role discussions by taking on a Role in an additional session later in the semester.

Use of Generative AI in this Course

Creating, analytical, and critical thinking skills are part of the learning outcomes of this course. All assignments should be prepared by the student working individually or in groups (Scribes and projects). Students may not have another person or program complete any portion of any assignment. Generative AI tools are trained, often without appropriate license, on text and images from folks whose intellectual property you do not own and cannot appropriately license or credit. Therefore, using AI generation tools is prohibited in this course unless explicitly marked as example outputs from such tools as part of an assessment or analysis of their behavior. *Note that all media generated by such tools is inherently plagiarized content*.

Course Schedule and Deliverables

Each class meeting will be run as a discussion group. Students will take on particular roles in that discussion to summarize reading material, connect ideas between sessions and schools of thought, and engage critically with the assumptions and presuppositions of work in language and computing as characterized by philosophy and cognitive theory of language. A 2 page report produced with respect to constraints and instructions for each role must be turned in by 11:59pm on the day of the discussion. Reading materials are subject to change at the instructor's discretion as the course progresses.

Week	Readings	Deliverables
The Rise and Fall and Rise of Machine Translation		
Week 1 Jan 14 & Jan 16	*Preparatory background [Read before Jan 14]:* Language. Mark Aronoff (2007), Scholarpedia, 2(5):3175. Language in Brief. American Speech-Language-Hearing Association. 11.8: Introduction to Linguistics. Paris, Ricardo, Raymond, & Johnson. *Further Course Resources [For your reference]:* Speech and Language Processing (3rd ed. draft). Dan Jurafsky and James H. Martin. 2023. ^ *A great reference textbook for all things NLP.* The Quest for Artificial Intelligence. Nils J. Nilsson. 2009. ^ *A primer on the state and history of AI in 2009, just before deep learning swept the field.* *Context [read before Jan 16]:* The Cryptological Origins of Machine Translation. Quinn DuPont. Amodern 2017.	(Weekly): Roleplaying session discussion contributions (Weekly): 2-page reports for each of your session Roles
Week 2 Jan 21 & Jan 23	*Research:* A Mathematical Theory of Communication. Claude E. Shannon. The Bell System Technical Journal Vol. 27 pp. 379–423, 623–656. 1948. *Context:* Translation. Warren Weaver. Rockefeller Foundation, Memorandum, 1949.	

Week 3 Jan 28 & Jan 30	*Context:* Joachim Becher Provides an Early Model for Machine Translation. History of Information; Circa 1661. Troyanskii of St. Petersburg Invents a Mechanical "Translating Machine". History of Information; Circa 1933. The first public demonstration of machine translation: the Georgetown-IBM system, 7th January 1954. John Hutchins. *Research:* Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, Jeffrey Dean. 2016.	Form project teams and sign up for midterm + final presentation slots. Optional, ungraded project pitches and outlines for feedback round 1.
Week 4 Feb 4 & Feb 6	*Context:* Leibniz: Logic. Wolfgang Lenzen. Internet Encyclopedia of Philosophy. The Universal Language: Chapter 3. Louis Couturat. The Logic of Leibniz (1901). ALPAC: the (in)famous report. John Hutchins. MT News International, no. 14. 1996. *Research:* A Framework of a Mechanical Translation Between Japanese and English by Analogy Principle. Makoto Nagao. Artificial and Human Intelligence, 1984. A Statistical Approach to Language Translation. P. Brown, J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, R. Mercer, and P. Roossin. COLING 1988.	Optional, ungraded project pitches and outlines for feedback round 2.
Meaning Making and Meaning Learning		
Week 5 Feb 11 & Feb 13	*Context:* The Kekulé Problem. Cormac McCarthy. Nautilus. 2017. A Linguist Responds to Cormac McCarthy. Julie Sedivy. Nautilus. 2017. Plato, Platonic Idealism, and Neo-Platonism. Tom Drake. Theory of Forms. Wikipedia summary. *Research:* A Structured Vector Space Model for Word Meaning in Context. Katrin Erk & Sebastian Padó. EMNLP 2008.	Optional, ungraded project pitches and outlines for feedback round 3.
Week 6 Feb 18 *No Class Feb 20*	*Research:* Expert Systems. Wayne Goddard. CpSc810 Notes Chapter 7. ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine. Joseph Weizenbaum. Computational Linguistics 1(9). 1966. *Context:* Research on Expert Systems. Bruce G. Buchanan. Report No. STAN-CS-81-837. 1981. Human and Computational Question Answering. Wendy Lehnert. Cognitive Science 1(1) pp. 47–73. 1977.	
Week 7 Feb 25 & Feb 27	Project Midterm Presentations	Project Midterm Report

Week 8 Mar 4 & Mar 6	*Context:* On sense and reference. Gottlob Frege. Zeitscheift für Philosophie and philosophische Kritik, 100 (1892), 25-50; as translated in Translations from the Philosophical Writings of Gottlob Frege (1952); as reprinted in A.W. Moore (ed.) Meaning and Reference. Oxford: Oxford University Press. Logical Syntax and Semantics: Their Linguistic Relevance. Noam Chomsky. Language 31(1) pp 36–45. 1955. *Research:* Distributed Representations of Words and Phrases and their Compositionality. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean. NeurIPS 2013. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. NAACL 2019.	
Spring Break Mar 11 & Mar 13		
Week 9 Mar 18 & Mar 20	*Context:* The Symbol Grounding Problem. Harnad, S. Physica D 42: 335-346 (1990). Meaning, Form and the Limits of Natural Language Processing. Oliver Dürr, Jan Segessenmann and Jan Juhani Steinmann. Philosophy, Theology and the Sciences Vol. 10 (2023). *Research:* Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. Emily M. Bender, Alexander Koller. ACL 2020.	
Week 10 Mar 25 & Mar 27	*Context:* General Semantics. David Lewis. Synthese, Vol. 22, No. 1/2, Semantics of Natural Language, II (Dec., 1970), pp. 18-67. *Research:* WordNet: A Lexical Database for English. George A. Miller. 1992. Building a Large Annotated Corpus of English: The Penn Treebank. Mitchell P. Marcus, Beatrice Santorini, Mary Ann Marcinkiewicz. Computational Linguistics Vol. 19(2). 1993. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J. Liu. JMLR 21(1) 2020.	
Computing and The Teleology of Language		
Week 11 Apr 1 & Apr 3	*Context:* Meaning is use: Wittgenstein on the limits of language. Tim Rayner. Philosophy for Change, 2014. Russel, Wittgenstein, and Moderate Realism. Dave Seng. The Socratic Dictum, 2023. Twenty-five years of information extraction. Ralph Grishman. Cambridge University Press, 2019. *Research:* Weakly Supervised Learning of Semantic Parsers for Mapping Instructions to Actions. Yoav Artzi & Luke Zettlemoyer. ACL 2013.	

Week 12 Apr 8 & Apr 10	*Context:* Logic and Conversation. H. P. Grice. Reprinted from Syntax and Semantics 3: Speech Arts, Cole et al., pp 41-58 (1975). *Research:* Scripts, plans, goals and understanding: Chapters 1-3. Schank, R. C., & Abelson, R. P. (1977)	
Week 13 Apr 15 & Apr 17	*Context:* Born This Way: Chomsky's Theory Explains Why We're So Good at Acquiring Language. Rebecca Joy. Healthline 2019. Evidence Rebuts Chomsky's Theory of Language Learning. Paul Ibbotson & Michael Tomasello. Scientific American 2016. *Research:* How Adults Understand What Young Children Say. Stephan C. Meylan, Ruthe Foushee, Nicole H. Wong, Elika Bergelson, Roger P. Levy. Nature Human Behavior 7(12). 2023.	
Week 14 Apr 22 & Apr 24	*Research:* Including Signed Languages in Natural Language Processing. Kayo Yin, Amit Moryossef, Julie Hochgesang, Yoav Goldberg, Malihe Alikhani. ACL 2021. YouTube-ASL: A Large-Scale, Open-Domain American Sign Language-English Parallel Corpus. David Uthus, Garrett Tanzer, Manfred Georg. 2023. *Context:* To Build Our Future, We Must Know Our Past: Contextualizing Paradigm Shifts in Natural Language Processing. Sireesh Gururaja, Amanda Bertsch, Clara Na, David Gray Widder, Emma Strubell. EMNLP 2023.	
Project Final Presentations and Reports		
Week 15 Apr 28 & May 1	Project Final Presentations	
FINAL	**May 8:** Final exam period 4:30pm–6:30pm	Project Final Report DUE on the University scheduled exam period end, **6:30pm May 8**

Statement on Academic Conduct and Support Systems

Academic Integrity:
The University of Southern California is a learning community committed to developing successful scholars and researchers dedicated to the pursuit of knowledge and the dissemination of ideas. Academic misconduct, which includes any act of dishonesty in the production or submission of academic work, comprises the integrity of the person who commits the act and can impugn the perceived integrity of the entire university community. It stands in opposition to the university's mission to research, educate, and contribute productively to our community and the world.

All students are expected to submit assignments that represent their own original work, and that have been prepared specifically for the course or section for which they have been submitted. You may not submit work written by others or "recycle" work prepared for other courses without obtaining written permission from the instructor(s).

For this class, unless specifically designated as a 'group project,' all assignments are expected to be completed individually.

Other violations of academic integrity include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), collusion, knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

The impact of academic dishonesty is far-reaching and is considered a serious offense against the university. All incidences of academic misconduct will be reported to the Office of Academic Integrity and could result in outcomes such as failure on the assignment, failure in the course, suspension, or even expulsion from the university.

For more information about academic integrity see the student handbook or the Office of Academic Integrity's website, and university policies on Research and Scholarship Misconduct.

Please ask your instructor if you are unsure what constitutes unauthorized assistance on an exam or assignment, or what information requires citation and/or attribution.

Course Content Distribution and Synchronous Session Recordings Policies
USC has policies that prohibit recording and distribution of any synchronous and asynchronous course content outside of the learning environment.

Recording a university class without the express permission of the instructor and announcement to the class, or unless conducted pursuant to an Office of Student Accessibility Services (OSAS) accommodation. Recording can inhibit free discussion in the future, and thus infringe on the academic freedom of other students as well as the instructor. (Living our Unifying Values: The USC Student Handbook, page 13).

Distribution or use of notes, recordings, exams, or other intellectual property, based on university classes or lectures without the express permission of the instructor for purposes other than individual or group study. This includes but is not limited to providing materials for distribution by services publishing course materials. This restriction on unauthorized use also applies to all information, which had been distributed to students or in any way had been displayed for use in relationship to the class, whether obtained in class, via email, on the internet, or via any other media. (Living our Unifying Values: The USC Student Handbook, page 13).

Students and Disability Accommodations:
USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

Support Systems:

Counseling and Mental Health - (213) 740-9355 – 24/7 on call
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

988 Suicide and Crisis Lifeline - 988 for both calls and text messages – 24/7 on call
The 988 Suicide and Crisis Lifeline (formerly known as the National Suicide Prevention Lifeline) provides free and confidential emotional support to people in

suicidal crisis or emotional distress 24 hours a day, 7 days a week, across the United States. The Lifeline is comprised of a national network of over 200 local crisis centers, combining custom local care and resources with national standards and best practices. The new, shorter phone number makes it easier for people to remember and access mental health crisis services (though the previous 1 (800) 273-8255 number will continue to function indefinitely) and represents a continued commitment to those in crisis.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL) – 24/7 on call
Free and confidential therapy services, workshops, and training for situations related to gender- and power-based harm (including sexual assault, intimate partner violence, and stalking).

Office for Equity, Equal Opportunity, and Title IX (EEO-TIX) - (213) 740-5086
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

The Office of Student Accessibility Services (OSAS) - (213) 740-0776
OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

USC Campus Support and Intervention - (213) 740-0411
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity, Equity and Inclusion - (213) 740-2101
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-1200 – 24/7 on call
Non-emergency assistance or information.

Office of the Ombuds - (213) 821-9556 (UPC) / (323-442-0382 (HSC)
A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

Occupational Therapy Faculty Practice - (323) 442-2850 or otfp@med.usc.edu
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.