

# Cristina M. Aggazzotti

✉ caggazz1@jhu.edu | 🏠 caggazzotti.github.io | 📧 caggazzotti | 🌐 c-aggazzotti

## Education

---

### Harvard University

PH.D. IN LINGUISTICS  
M.A. IN LINGUISTICS

Cambridge, MA

2019  
2017

### University College London

M.Sc. IN COGNITIVE AND DECISION SCIENCES (Distinction)

London, UK

2013

### University of Southern California

B.A. IN APPLIED AND COMPUTATIONAL MATHEMATICS  
B.A. IN LINGUISTICS  
Magna Cum Laude; Renaissance Scholar

Los Angeles, CA

2011

## Research and Work Experience

---

### Assistant Research Scientist

HUMAN LANGUAGE TECHNOLOGY CENTER OF EXCELLENCE, JOHNS HOPKINS UNIVERSITY

Baltimore, MD

Oct. 2024 – Present

- Develop methods for authorship attribution (who wrote something) and speaker attribution from speech transcripts (who said something), examining how someone's language use can reveal or protect (anonymization) their identity
- Work toward robust and explainable attribution models by exploring model representations, biases, and performance

### Postdoctoral Fellow in Computer Science

HUMAN LANGUAGE TECHNOLOGY CENTER OF EXCELLENCE, JOHNS HOPKINS UNIVERSITY

Baltimore, MD

Aug. 2023 – Oct. 2024

- Established a new benchmark and state-of-the-art performance for speaker attribution—identifying a speaker based on the content of their speech (using syntax/semantics/lexicon)—as a supplement to speaker recognition methods, which analyze the speech signal (using acoustics/phonetics)
- Identified the need for speaker anonymization methods to obfuscate both the speech signal and speech content
- Resulted in a first author ACL journal publication and an acceptance for oral presentation at ACL 2024

### Postdoctoral Fellow in Linguistics

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

Remote

May 2022 – July 2023

- Developed and tested stylometric and neural authorship attribution models across genres, domains, and dataset sizes
- Identified forensic linguistic data needs and began developing standards for multidisciplinary data use
- Resulted in 3 forensic science and forensic linguistics conference presentations and 2 papers (in preparation)

### Academic Editor and Proofreader

FREELANCER

Remote

Feb. 2017 – July 2023

- Edited and proofread journal papers, books, job/grant applications, and dissertations for non-native English speakers

### Visiting Research Scholar (Summer Camp for Applied Language Exploration)

HUMAN LANGUAGE TECHNOLOGY CENTER OF EXCELLENCE, JOHNS HOPKINS UNIVERSITY

Baltimore, MD

Jun. 2022 – Aug. 2022

- Experimented with schemas masking training data to explain neural author models' reliance on style versus content
- Developed a linguistic and stylometric feature model for determining authorship of various text types
- Resulted in a first author ACL journal publication and an acceptance for oral presentation at EMNLP 2023

### Research Contractor

INSTITUTE FOR LINGUISTIC EVIDENCE

Remote

2016 – 2018; Oct. 2021 – Apr. 2022

- Compared discriminant analysis (SPSS) and LSTM (Python) approaches to classifying suicide notes and last words
- Analyzed linguistic evidence across authors and text genres; researched statistical methods for presentation in court
- Resulted in 4 forensic science conference oral presentations and leading a half-day conference workshop

### Research Scientist and Customer Experience Engineer

POINTMAIL, INC. (STARTUP)

Cambridge, MA

Feb. 2018 – Jan. 2020

- Investigated ways to improve phrase- and sentence-level text prediction based on individual language usage and style
- Led bi-weekly team-wide research meetings and helped guide future research directions and product decisions
- Built rapport with customers, onboarded users, resolved customer issues, and elicited & evaluated customer feedback

## Graduate Researcher

HARVARD UNIVERSITY

Cambridge, MA

Aug. 2015 – May 2019

- Created a unified analysis of reflexives and reciprocals, implemented in synchronous tree adjoining grammar (a syntax-semantics hierarchical tree pairing system), and provided a novel extension of the formalism to capture morphology
- Developed a syntactic and semantic annotation scheme for reflexives using the Corpus of Contemporary American English by creating a hybrid classification system based on previous reflexives research and attested data
- Mentored 7 undergrads in annotation schemes, machine learning of anaphor data, and testing my model using OCaml

## Graduate Researcher

UNIVERSITY COLLEGE LONDON

London, UK

Sep. 2012 – Sep. 2013

- Created a formalized crime script of cash-in-transit robbery combining influence diagrams, causal models, Bayesian networks, and scripts that could ultimately be used by the police to identify effective prevention strategies

## Publications and Presentations

---

- **STEB: Style Text Embedding Benchmark.**  
Rafael Rivera Soto, Anna Wegmann, and Cristina Aggazzotti.  
Under review (2026).
- **A survey on representing linguistic style: Challenges and opportunities.**  
Anna Wegmann, Cristina Aggazzotti, Rafael Rivera Soto, and Dong Nguyen.  
Under review (2026).
- **A stylometric analysis of speaker attribution from speech transcripts.**  
Cristina Aggazzotti and Elizabeth Allyn Smith.  
In revision (2026).
- **The semiotics of ground truth: Avoiding misunderstanding in the presentation of computational forensic linguistics as scientific evidence.**  
Elizabeth Allyn Smith and Cristina Aggazzotti.  
In revision (2026).
- **CommonLID: Re-evaluating state-of-the-art language identification performance on web data.**  
Pedro Ortiz Suarez, Laurie Burchell, Catherine Arnett, ..., Cristina Aggazzotti, ..., Kenton Murray, and Sarah Luger  
*Proceedings of the Association for Computational Linguistics* (2026).
- **Content anonymization for privacy in long-form audio.**  
Cristina Aggazzotti, Ashi Garg, Zexin Cai, and Nicholas Andrews.  
*Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (2026). Oral presentation.
- **The impact of automatic speech transcription on speaker attribution.**  
Cristina Aggazzotti, Matthew Wiesner, Elizabeth Allyn Smith, and Nicholas Andrews.  
*Transactions of the Association for Computational Linguistics* 13, pp. 1578-1596 (2025). Presented at EMNLP 2025.
- **The relationship between linguistic accommodation and speaker attribution.**  
Cristina Aggazzotti and Elizabeth Allyn Smith.  
*5<sup>th</sup> Focus Conference on Forensic Linguistics of the International Language and Law Association (ILLA)* (2025).
- **Speaker attribution of speech transcripts: A stylometric approach.**  
Cristina Aggazzotti, Elizabeth Allyn Smith, and Nicholas Andrews.  
*17<sup>th</sup> Biennial Conference of the International Association for Forensic and Legal Linguistics (IAFL)* (2025).
- **Weaknesses of artificial intelligence proxies of topic for forensic linguistic applications.**  
Cristina Aggazzotti and Elizabeth Allyn Smith.  
*Proceedings of the American Academy of Forensic Sciences Annual Conference*, vol. 31, p. 505 (2025).
- **Can authorship attribution models distinguish speakers in speech transcripts?**  
Cristina Aggazzotti, Nicholas Andrews, and Elizabeth Allyn Smith.  
*Transactions of the Association for Computational Linguistics* 12, pp. 875–891 (2024). Presented at ACL 2024.
- **Can authorship representation learning capture stylistic features?**  
Andrew Wang\*, Cristina Aggazzotti\*, Rebecca Kotula, Rafael Rivera Soto, Marcus Bishop, and Nicholas Andrews.  
*Transactions of the Association for Computational Linguistics* 11, pp. 1416–1431 (2023). Presented at EMNLP 2023.
- **Testing authorship analysis on spoken language transcripts: Establishing a first benchmark.**  
Cristina Aggazzotti, Nicholas Andrews, and Elizabeth Allyn Smith.  
*4<sup>th</sup> Focus Conference on Forensic Linguistics of the International Language and Law Association (ILLA)* (2023).

- **Authorship attribution across right-wing manifestos.**  
Cristina Aggazzotti and Elizabeth Allyn Smith.  
*16<sup>th</sup> Biennial Conference of the International Association for Forensic and Legal Linguistics (IAFL)* (2023).
- **Disambiguating ‘ground truth’ for forensic applications in artificial intelligence.**  
Cristina Aggazzotti and Elizabeth Allyn Smith.  
*Proceedings of the American Academy of Forensic Sciences Annual Conference*, vol. 29, p. 380 (2023).
- **A comparison of two computational approaches to analyzing last words.**  
Cristina Aggazzotti.  
*Proceedings of the American Academy of Forensic Sciences Annual Conference*, vol. 28, p. 483 (2022).
- **A quantitative comparison of last words to related text genres.**  
Cristina Aggazzotti.  
*Proceedings of the American Academy of Forensic Sciences Annual Conference*, vol. 28, p. 486 (2022).
- **A quantitative assessment of last words using Suicide Note Assessment REsearch (SNARE).**  
Cristina Aggazzotti.  
*Proceedings of the American Academy of Forensic Sciences Annual Conference*, vol. 27, p. 370 (2021).
- **Reflexives and reciprocals in synchronous tree adjoining grammar.**  
Cristina Aggazzotti and Stuart M. Shieber.  
*Proceedings of the 13<sup>th</sup> Int’l Workshop on Tree Adjoining Grammars and Related Formalisms*, pp. 31–42 (2017).
- **Communication strategies to minimize bias and strengthen scientific foundations in forensic science.**  
Cristina Aggazzotti, Katherine Ramsland, Carole E. Chaski, Elizabeth Allyn Smith, and Roderick T. Kennedy.  
*Proceedings of the American Academy of Forensic Sciences Annual Conference*, vol. 23, p. 38 (2017).
- **Confirmation bias and metalinguistic awareness.**  
Carole E. Chaski, Elizabeth Allyn Smith, and Cristina Aggazzotti.  
*Proceedings of the American Academy of Forensic Sciences Annual Conference*, vol. 22, p. 902 (2016).

## Teaching and Mentoring Experience

---

### Prison Tutor and Justice-involved Youth Mentor

*Philadelphia, PA*

THE PETEY GREENE PROGRAM (SCI CHESTER; PHILLYCONNECT)

*Jan. 2025 – Present*

- Tutor incarcerated students for the GED or their associate’s degree; academically mentor system-impacted youth

### Research Mentor

*Remote*

POLYGENCE

*Sep. 2022 – Sep. 2023*

- Mentored high school students through their first research project, resulting in papers and symposium presentations
- Advised high school students on academics and possible career paths based on their interests and goals

### Math/SAT Tutor and College Coach

*Remote*

CARNEGIE PREP, CAMBRIDGE COACHING, INVICTUS

*Nov. 2021 – Aug. 2023*

- Devised lesson plans and SAT test prep guides; tutored for the SAT and for math from elementary to college level
- Guided students through application essay writing and college admissions from start to finish

### Teaching Fellow

*Cambridge, MA*

HARVARD UNIVERSITY

*Aug. 2016 – May 2018*

- Assisted in teaching two intro linguistics classes (Distinction in Teaching Award for each) and an intro syntax class
- Created curricula for and taught as the primary instructor tutorials on computational linguistics and forensic linguistics

### Prison Tutor and Careers Class Assistant

*Boston, MA*

SUFFOLK COUNTY HOUSE OF CORRECTIONS; BOSTON COMM. CORRECTIONS CENTER

*Feb. 2015 – May 2016*

- Taught English to Spanish-, Portuguese-, and Mandarin-speaking inmates; tutored men and women on probation/parole for the GED and assisted in a careers class creating resumes and helping with job applications

### Tutor and College Consultant

*Newbury Park, CA*

SELF-EMPLOYED

*Sep. 2013 – Aug. 2014*

- Tutored 30 students for the SAT, ACT, and all high school subjects; >300pt avg increase on SAT, 3pt avg increase on ACT
- Counseled 5 students through college selection, applications, resumes, and essays with successful admissions for all

## Grants and Awards

---

July 2025	<b>Demonstration Project Award</b> , Johns Hopkins University's Data Science and AI Institute
Aug. 2018	<b>Research grant</b> , Harvard Linguistics Department
April 2018	<b>Research grant</b> , Harvard Institute for Quantitative Social Science
Sep. 2017	<b>Travel grant</b> , Harvard Institute for Quantitative Social Science
2016, 2017	<b>Conference grant</b> , Harvard Linguistics Department
2016, 2017	<b>Distinction in Teaching Award</b> , Bok Center for Teaching and Learning

## Professional Memberships

---

<b>Society of Woman Geographers</b>	2025 – Present
<b>NIST's OSAC for Forensic Science</b> , Speaker Recognition	2024 – Present
<b>Association for Computational Linguistics</b>	2024 – Present
<b>International Association for Forensic and Legal Linguistics</b>	2022 – Present
<b>International Language and Law Association</b>	2022 – Present
<b>American Academy of Forensic Sciences</b> , Trainee Affiliate Member, Engineering & Applied Sciences	2022 – 2023

## Reviewing

---

<b>Books</b>	Research Methods in Applied Linguistics
<b>Journals</b>	Scientific Data (Nature), Cognitive Computation, Frontiers in Communication
<b>Conferences</b>	ICASSP

## Skills

---

<b>Computer</b>	Python, GitHub, $\LaTeX$ , MS Excel; R, MATLAB, SPSS, OCaml (prior experience)
<b>Linguistics</b>	Computational linguistics, syntax, semantics, corpus annotation and statistics, tree adjoining grammar
<b>Language</b>	English (native), Italian (proficient), Spanish (intermediate), French (reading knowledge)
<b>Soft skills</b>	Adaptable, good communicator, detail-oriented, problem solver, team player, quick learner, enthusiastic

## Mentoring and Leadership

---

<b>Board Member</b> , Society of Woman Geographers	2026 – Present
<b>Faculty Advisor</b> , JHU North American Computational Linguistics Open Competition Committee	2025 – Present
<b>Mentor for high school research projects</b> , Polygence	2022 – 2023
<b>Mentor for incoming graduate students</b> , Harvard Linguistics Department	2015 – 2018
<b>Assistant Head Tutor</b> , Harvard Linguistics Department	2017 – 2018
<b>Language Universals Colloquium Co-organizer</b> , Harvard Linguistics Department	2015 – 2017
<b>Captain</b> , Harvard Boxing Team	2016 – 2017

## Invited Talks

---

- “Authorship Attribution” *Los Angeles, CA*  
*Guest lecture in Linguistics 412: Language and the Law*, University of Southern California *Nov. 2024, Apr. 2022*
- Research and Graduate School Panel *Cambridge, MA*  
*Panelist for BLISS/SHARP programs*, Harvard University *July 2020, July 2018*
- Linguists in Industry Panel *Cambridge, MA*  
*Panelist*, Harvard University Linguistics Department *Feb. 2019*
- “Forensic Linguistics” *Cambridge, MA*  
*Guest lectures in Linguistics 83 and Linguistics 101*, Harvard University *Dec. 2018, Apr. 2017, Apr. 2016*
- “When words deceive you—Finding clues to justice through linguistics” *Albuquerque, NM*  
*Speaker in the Continuing Legal Education program*, New Mexico State Bar *Oct. 2017*

## Interests

---

- Traveling (76 countries, 47 U.S. states), sailing, running, learning new skills/hobbies, sports, music, knitting/crocheting