

# EMILY CHENG

emilyshana.cheng@upf.edu

## EDUCATION

---

- Universitat Pompeu Fabra** September 2022-2027  
PhD Candidate in Linguistics  
Representational complexity of language in biological and artificial brains  
Supervisor: Marco Baroni
- Massachusetts Institute of Technology** January 2022  
Master of Engineering in Computer Science (2022) GPA: 4.7/5.0  
Supervisors: Boris Katz and Andrei Barbu  
Bachelor of Science in Computer Science and Engineering (2020)  
Bachelor of Science in Mathematics (2020)

## EXPERIENCE

---

- Rycolab at ETH Zurich** Winter 2025-6  
*Visiting researcher with Ryan Cotterell* Zurich, CH
- Theoretical bounds on what we can decode from the internal representations of language models.
- Apple Machine Learning Research** Summer 2025  
*MLR Intern* Cambridge, UK
- Developed PAC bounds for reachable and controllable set estimation for generative models.
- MIT CSAIL** Summer 2024  
*Visiting researcher* Cambridge, MA
- Dimensionality of sEEG high-gamma power and feature richness during speech processing.
- Universitat Pompeu Fabra** Fall 2022-  
*Doctoral thesis* Barcelona
- Representational geometry of language in neural language models and the brain.
- Ecole Normale Supérieure, CNRS** Spring-Fall 2022  
*Visiting researcher with Thierry Poibeau* Paris
- Explore relationship between linguistic compositionality and ease of imitation learning.
- MIT CSAIL Infolab** Fall 2020 - Fall 2021  
*Master's Research* Cambridge, MA
- Thesis: *Understanding Symbolic Communication*
  - Supervised by Boris Katz and Andrei Barbu.
  - Characterized the transition from sub-symbolic to symbolic communication between human players, and later machine players via a communication game.

## PUBLICATIONS

---

1. **Emily Cheng**, Chris Wang, Andrei Barbu, Marco Baroni, Greta Tuckute. Effective dimensionality signifies feature abstraction during speech processing in the brain. *Submitted to Brain and Language. Abstract at CCN 2026.*
2. **Emily Cheng**, Aditya Vaidya, and Richard Antonello. Abstraction Induces the Brain Alignment of Language and Speech Models. *In Proceedings of ICML 2026.*  
Workshop version in UniReps@NeurIPS 2024. **Awarded best abstract.** (Oral)

3. **Emily Cheng**, Diego Doimo, Corentin Kervadec, Iuri Macocco, Jade Yu, Alessandro Laio, Marco Baroni. Emergence of a High-Dimensional Abstraction Phase in Language Transformers. *In Proceedings of ICLR 2025*. (Poster)
4. Marco Baroni, **Emily Cheng**, Iria de Dios Flores, and Francesca Franzon. Tracing the complexity profiles of different linguistic phenomena through the intrinsic dimension of LLM representations. *Submitted to EMNLP 2026*.
5. Jin Hwa Lee\*, Thomas Jiralerspong\*, Jade Yu, Yoshua Bengio, **Emily Cheng**. Geometric signatures of compositionality across a language model’s lifetime. *In Proceedings of ACL 2025*. (Oral, SAC award)
6. **Emily Cheng**, Corentin Kervadec, Marco Baroni. Bridging Information-Theoretic and Geometric Compression in Language Models. *In Proceedings of EMNLP 2023*. (Oral)
7. **Emily Cheng**, Carmen Amo Alonso, Federico Danieli, Arno Blaas, Luca Zappella, Pau Rodriguez, Xavier Suau-Cuadros. GenCtrl: A Formal Controllability Toolkit for Generative Models. *In Proceedings of ICLR 2026*.
8. **Emily Cheng** and Carmen Amo Alonso. Linearly Controlled Language Generation with Performative Guarantees. *TMLR 2026*.  
Workshop version in MINT@NeurIPS 2024.
9. **Emily Cheng**, Mathieu Rita, Thierry Poibeau. On the Correspondence between Compositionality and Imitation in Emergent Neural Communication. *In Findings of ACL 2023*.
10. **Emily Cheng**, Yen-Ling Kuo, Josefina Correa, Ignacio Cases, Boris Katz, and Andrei Barbu. Quantifying the Emergence of Symbolic Communication. *In Proceedings of CogSci 2022*.

## INVITED TALKS

---

### **Geometric Signatures of Compositionality Across a Language Model’s Lifetime**

*Oral at ACL, Interpretability track*

August 2025 *Vienna*

### **Evidence from fMRI supports a two-phase abstraction process in language models.**

*Talk at RycoLab Seminar @ ETH Zurich*

November 2024 *Zürich*

*Talk at Grammar and Cognition Group @ Universitat Pompeu Fabra*

October 2024 *Barcelona*

*Oral presentation at UniReps@NeurIPS*

December 2024 *Vancouver*

### **High-Dimensional Abstraction Phase in Language Transformers**

*Talk at Area Science Park*

June 2024 *Trieste, Italy*

*Talk at Infolab@MIT CSAIL group meeting*

June 2024 *Cambridge, MA*

*Talk at Johns Hopkins Center for Speech and Language Processing*

February 2025 *Baltimore*

*Talk at Pioneer Center for AI*

February 2025 *Copenhagen*

### **Bridging Information-Theoretic & Geometric Compression in LMs**

*Oral presentation at Deep Learning Barcelona Symposium*

December 2024 *Barcelona*

*Talk at Zaslavsky lab (NYU Psychology)*

August 2024 *NYC*

*Oral presentation at EMNLP*

December 2023 *Singapore*

*Talk at Swiss AI Lab (IDSIA)*

November 2023 *Lugano*

*Talk at Evolution in Language (EviL) Seminar*

September 2023 *Online*

### **Principles of Functional and Semantic Efficiency in Grammars**

*Talk at Evolution in Language (EviL) Seminar*

December 2024 *Online*

## AWARDS & GRANTS

---

### **ICML, NeurIPS top reviewer**

2025, 2026

*Free registration to ICML 2026 and NeurIPS 2025*

### **ACL Senior Area Chair Paper Award**

2025

*SAC Highlights paper award, < 1.5% of accepted papers*

*Vienna, AT*

### **NeurIPS UniReps Workshop Best Abstract**

2024

*Free registration to NeurIPS 2024*

*Vancouver, CA*

<b>UPF Department of Linguistics “Estada” Grant, € 3000</b> <i>Competitive grant for research stay with MIT CSAIL</i>	2024 Barcelona, ES
<b>UPF Department of Linguistics Travel Grant, € 2000</b> <i>Competitive grant for travel to EMNLP conference</i>	2023 Barcelona, ES
<b>Brains, Minds, Machines Travel Grant</b> <i>Fully paid BMM Summer Course, ≈ 10% acceptance</i>	2023 Woods Hole, MA

## TEACHING & SUPERVISION

---

<b>UPF Master’s thesis: Wanqiong Zhou</b> <i>Thesis co-supervision with Thomas Brochhagen</i>	Spring 2026 Barcelona, ES
Predictors of an iconic to symbolic transition in written historical Chinese.	
<b>UPF Computational Semantics</b> <i>Graduate Teaching Assistant</i>	Fall 2025, 2026 Barcelona, ES
Held recitations, created lab materials, and supervised course projects. Was rated 9.5/10 by students!	
<b>Brains Minds and Machines Summer Course</b> <i>Teaching Assistant</i>	Summer 2024, 2026 Woods Hole, MA
Gave tutorials and supervised course projects.	
<b>Madagascar ML Summer School</b> <i>Guest Lecture</i>	Winter 2022
Introduction to Statistical Learning Theory.	
<b>MIT 6.031 Software Construction</b> <i>Graduate Teaching Assistant</i>	Fall 2020 Cambridge, MA
Held lab hours, graded assignments for students in intermediate Java software class.	
<b>MIT-France Global Teaching Labs</b> <i>Instructor</i>	January 2020 Grenoble, France
Taught middle, high, and preparatory school students for STEM outreach program.	
<b>MIT Math Learning Center</b> <i>Teaching Assistant</i>	Fall 2018 - Spring 2019 Cambridge, MA
Office hours for Differential Equations, Linear Algebra, Probability, Physics and Calculus.	
<b>MIT Math Department</b> <i>Grader</i>	Fall 2017, Fall 2018 Cambridge, MA
Probability and Random Variables (18.600) and Statistics (18.650).	

## ACTIVITIES & OUTREACH

---

<b>Conference and workshop organization</b> <i>UniReps 2026 organizer; CoNLL 2025 Publication Chair; COLT Symposium 2025;</i>
<b>Reviewer</b> <i>ACL, ICML 2026; NeurIPS, ICLR 2025; LanGame@NeurIPS workshop, UniReps@NeurIPS workshop, NeurIPS, ARR, ICLR 2024; EMNLP 2023; NeurIPS 2022; NeurIPS, ICLR 2021</i>

## OTHER INDUSTRY EXPERIENCE

---

<b>Two Sigma Investments</b> <i>Quant Research Intern: News Team</i>	Summer 2019 New York, NY
Built alpha models in Python and Groovy to forecast equity and options returns with news data.	

**Virtu Financial**  
*Quant Research Intern*

January 2019  
New York, NY

Developed cross-asset market impact models in Python for cash equities execution.

**Goldman Sachs**  
*Securities Quant Research Intern: Equities Flow Vol, FICC SMM Execution Services*

Summer 2018  
New York, NY

- Developed alpha models in Python to forecast realized volatility for single stock options
- Built an order fill model for trading simulation in Java

## COURSEWORK

---

**Brains Minds and Machines Summer Course**  
Project: Intrinsic Dimensionality of Brain Responses to Language

Summer 2023  
Woods Hole, MA

**Institute of Language, Communication, and the Brain Summer School**

Summer 2022  
Marseille

### Official Coursework

6.867 Machine Learning (G)	6.860 Statistical Learning Theory (G)	6.337 Numerical Methods (G)
6.435 Bayesian Inference (G)	6.031 Software Construction	18.615 Stochastic Processes (G)
6.864 NLP (G)	6.046 Design & Analysis of Algorithms	6.436 Probability Theory (G)
6.884 Sensorimotor Learning (G)	24.933 Semantics & Pragmatics (G)	9.190 Comp. Linguistics (G)

## SKILLS

---

<b>Computer Languages</b>	Python, Java, C/C++, Julia
<b>Software &amp; Tools</b>	Pandas/Numpy/Scipy, PyTorch, Git, Linux

## LANGUAGES

---

English (native), Mandarin (fluent), French (C1-2), Spanish (C1), German (A1)