

EMILY CHENG

emilyshana.cheng@upf.edu

EDUCATION

Universitat Pompeu Fabra

PhD Candidate in Linguistics

Linguistic complexity and its representation in biological and artificial brains

Supervisor: Marco Baroni

September 2022-

Massachusetts Institute of Technology

Master of Engineering in Computer Science (2021)

Supervisors: Boris Katz and Andrei Barbu

Bachelor of Science in Computer Science and Engineering (2020)

Bachelor of Science in Mathematics (2020)

December 2021

GPA: 4.7/5.0

PUBLICATIONS

1. **Emily Cheng**, Carmen Amo Alonso, Marco Baroni. Linearly Controlled Language Generation with Performative Guarantees. *Under review*.
2. **Emily Cheng**, Diego Doimo, Corentin Kervadec, Iuri Macocco, Jade Yu, Alessandro Laio, Marco Baroni. Emergence of a High-Dimensional Abstraction Phase in Language Transformers. *Under review*.
3. **Emily Cheng*** and Richard Antonello*. Evidence from fMRI supports a two-phase abstraction process in language models. *Under review*.
4. Thomas Jiralerspong*, Jin Hwa Lee*, Jade Yu, Yoshua Bengio, **Emily Cheng**. Geometric signatures of compositionality across a language model's lifetime. *Under review*.
5. **Emily Cheng** and Francesca Franzon. Distributional Universals for Grammatical Features. *In prep*.
6. **Emily Cheng**, Corentin Kervadec, Marco Baroni. Bridging Information-Theoretic and Geometric Compression in Language Models. *In Proceedings of EMNLP 2023*.
7. **Emily Cheng**, Mathieu Rita, Thierry Poibeau. On the Correspondence between Compositionality and Imitation in Emergent Neural Communication. *In Findings of ACL 2023*.
8. **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*.
9. **Emily Cheng**, Yen-Ling Kuo, Ignacio Cases, Boris Katz, and Andrei Barbu. Towards Modeling the Emergence of Symbolic Communication. *In Proceedings of the ICRA-2021 Social Intelligence Workshop*.

TALKS & PRESENTATIONS

High-Dimensional Abstraction Phase in Language Transformers

Talk at Area Science Park

Talk at Infolab@MIT CSAIL group meeting

June 2024 Trieste, Italy

June 2024 Cambridge, MA

Bridging Information-Theoretic & Geometric Compression in LMs

Talk at Zaslavsky lab (NYU Psychology)

Oral presentation at EMNLP

Talk at Swiss AI Lab (IDSIA)

Talk at Evolution in Language (EviL) Seminar

August 2024 NYC

December 2023 Singapore

November 2023 Lugano

September 2023 Online

Interplay of functional & semantic aspects in shaping inflectional morphology: a case study on Romance languages

Poster at Crosslinguistic Perspectives in Linguistics Conference (X-PPL)

November 2023

Zürich

EXPERIENCE

Intrinsic Dimensionality & Representational Compression in Neural Language Models

Doctoral thesis

Fall 2022-
Barcelona

- Explore relationship between linguistic structure, information-theoretic compression, and intrinsic dimensionality in neural language models.

Compositionality and Imitation Learning in Artificial Emergent Language

Visiting researcher at ENS, CNRS

Spring-Fall 2022
Paris

- Supervised by Thierry Poibeau. Supported by Paris AI Research Institute.
- Explore relationship between compositionality of emergent languages and ease of imitation learning.

Emergent Symbolic Communication in Humans and Machines

Master's Research: MIT Infolab

Fall 2020 - Fall 2021
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.

AWARDS

Travel Grant

Brains, Minds, and Machines Summer Course

2023
Woods Hole, MA

Fulbright France Open Research Grant Semifinalist

Meta-learning in low-resource multilingual generalization

2021
Paris, France

In collaboration with LATTICE at CNRS and École Normale Supérieure.

TEACHING

Brains Minds and Machines Summer Course

Teaching Assistant

Summer 2024
Woods Hole, MA

6.031 Software Construction

Graduate Teaching Assistant

Fall 2020
Cambridge, MA

MIT-France Global Teaching Labs

Instructor

January 2020
Grenoble, France

- Taught middle, high, and preparatory school students for STEM outreach program.

MIT Math Learning Center

Teaching Assistant

Fall 2018 - Spring 2019
Cambridge, MA

- Office hours for Differential Equations, Linear Algebra, Probability, Physics and Calculus.

MIT Math Department

Grader

Fall 2017, Fall 2018
Cambridge, MA

- Probability and Random Variables (18.600) and Statistics (18.650).

ACTIVITIES & OUTREACH

Reviewer

NeurIPS, ACL Rolling Review, ICLR 2024; EMNLP 2023; NeurIPS 2022; NeurIPS, ICLR 2021

INDUSTRY EXPERIENCE

Palantir Technologies

Software Engineering Intern

Summer 2020
Remote

Two Sigma Investments
Quant Research Intern: News Team

Summer 2019
New York, NY

Virtu Financial
Algo Quant Research Intern

January 2019
New York, NY

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

Summer 2018
New York, NY

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 Natural Language Processing (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

Computer Languages	Python, Java, C/C++
Software & Tools	Pandas/Numpy/Scipy, PyTorch, Git, Linux

LANGUAGES

English (native), Mandarin (fluent), French (C1), Spanish (C1)