RICE SOCIAL SCIENCES
Department of Linguistics

what is computational linguistics & why is language difficult?

LING 430 Computational Linguistics Fall 2026
lecture 1a
Aug 25 2026

Siyu Liang
siyu.liang@rice.edu

agenda

- getting started
- introduction
- logistics

about me

- Dr. Siyu Liang
  - assistant professor at the department of linguistics
  - I work on: computational linguistics, speech processing, variation, low-resource & endangered languages, fieldwork, historical linguistics, etc.
  - PhD in Computational Linguistics, University of Washington
  - BA in Linguistics, Georgetown University

about me

- 📧 email: siyu.liang@rice.edu
- 🔗 website: https://siyuliang.github.io/
- 🏢 office: 205 Herring Hall
- 👨‍💻 office hour: Tuesdays, 12:30-2:30pm

about you!

- major
- year
- any languages you know/are studying/are interested in
- why LING 430 and what you hope to learn
- any concerns about the class

why is language hard?

activity

“The chef made her duck.”

1. think about what this sentence means
2. now think about a different thing it could mean
3. keep going, and share with people around you

activity

- **made**
  - cooked (made her duck confit)
  - created (made a duck for her)
  - caused to (made her duck under the beam)
  - transformed (turned her into a duck)
- **her** and **duck**
  - *her*: possessive (her duck) or object (made her … duck)
  - *duck*: noun (the bird) or verb (to dodge)
- **key idea:** ambiguity multiplies, and a few two-way choices produce many readings.

ambiguity

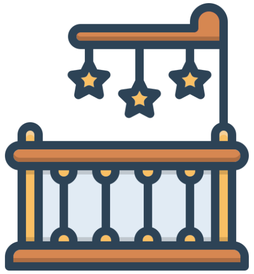
level	e.g.
phonetic	“Siri can recognize speech” ≈ “Siri can wreck a nice beach”
morphological	number = number, or numb + -er (“more numb”)?
syntactic	“Squad helps dog bite victim”
semantic	“She went to the bank”: river or money?
pragmatic	“Can you pass the salt?”: a question about ability?

- humans barely notice, but machines **must** decide, at every level and every word.

looking ahead

“The box was in the pen.”

- Yehoshua Bar-Hillel (1960) argued “no computer could ever translate this correctly”
  - choosing writing pen vs playpen requires world knowledge, not dictionaries
- he concluded fully automatic high-quality translation was impossible.
- at the end of the class, we will explain what a 2026 system does with this sentence, why it usually works, and what Bar-Hillel got right anyway.



what is computational linguistics?

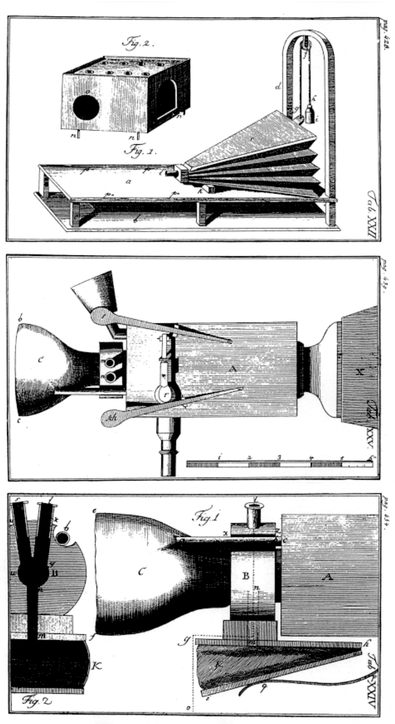
recap: *Is ACL Interdisciplinary?*

- "**Computational linguistics** is per definition interdisciplinary. **NLP**, however, may or may not be"
- engineering can succeed without imitating humans, "just as aeronautical engineering does not have to draw inspiration from birds."

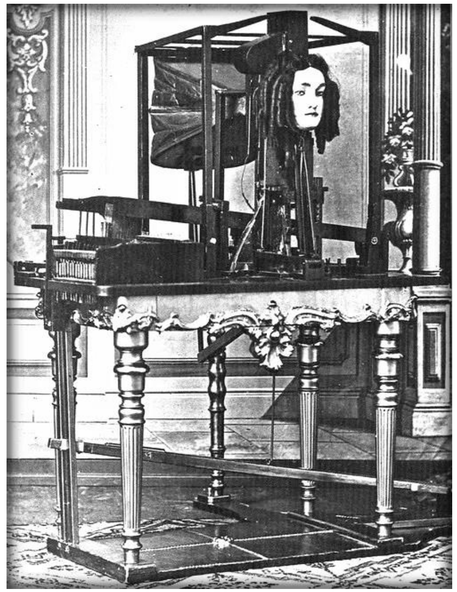
Ryan Cotterell

rationalism vs empiricism

examples from speech



Wolfgang von Kepelen's speaking machine, 1769



Joseph Faber's Euphonia. c. 1870

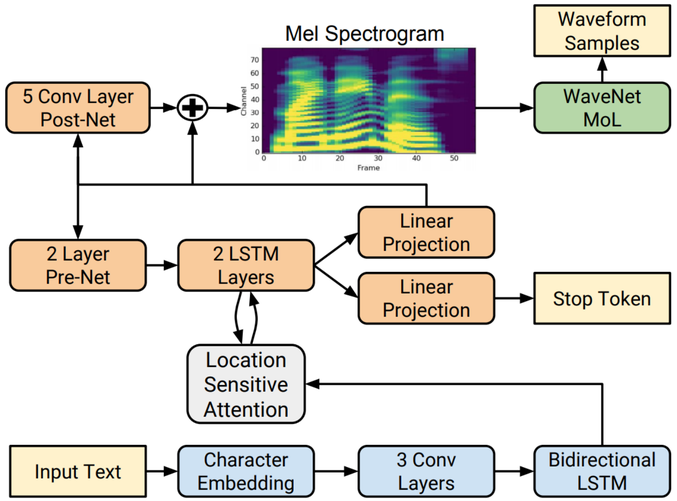
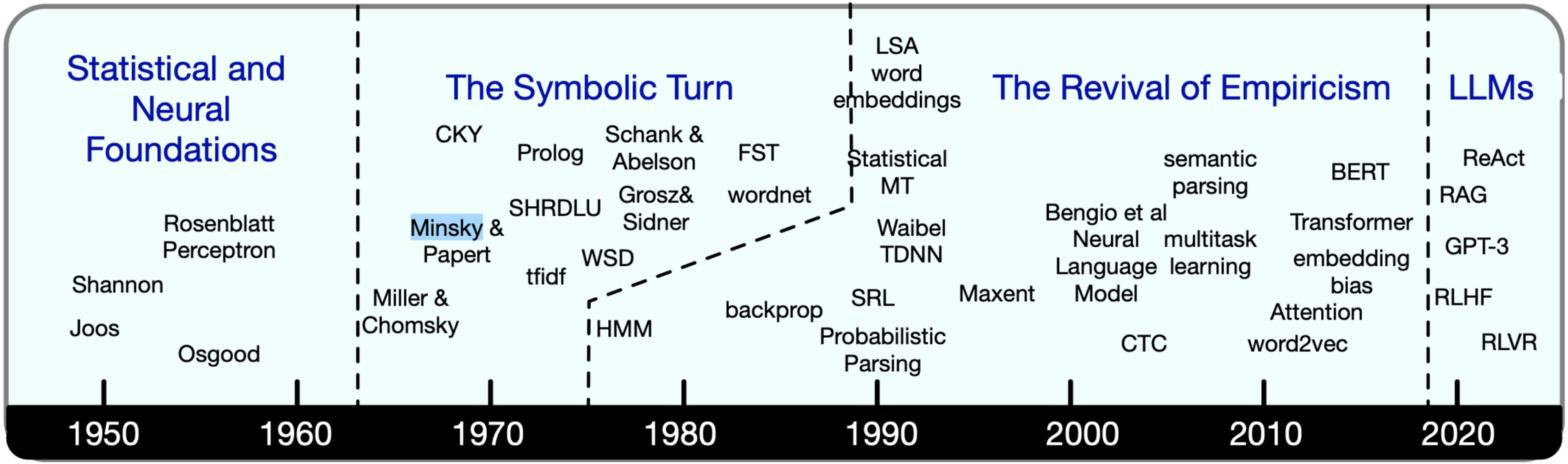


Fig. 1. Block diagram of the Tacotron 2 system architecture.

An RNN model, Shen et al. 2017

evolution of computational linguistics



slide from Jurafsky & Martin (SLP)

a tale of two names

computational linguistics vs NLP

computational linguistics

- using computation to answer **questions about language**
  - What are the rules? How is meaning built? How do children, and machines, learn it?

natural language processing (NLP)

- **engineering systems** that do useful things with language
  - translate, transcribe, search, summarize, converse.

- they are more intertwined than ever
  - we travel both worlds in this course
  - every tool we build makes a linguistic assumption

computational linguistics

another take

- ***theoretical*** computational linguistics
  - investigates the computability of human language and linguistic theories
  - formal language theory and automata theory
- ***applied*** computational linguistics
  - the focus is on building (working) models of human language
  - NLP, human language technology (HLT), artificial intelligence (AI)
  - often centers around building practical applications: spell checkers, machine translation systems, etc.
- distinct flavors: symbolic vs. statistical approaches vs. neural approach

slide from Janet & Na-Rae (Pitt)

so what is this class?

- this course is about (Python) coding, language processing, and some machine learning
- ***theoretical*** computational linguistics
  - formal language theory and automata theory
- ***applied*** computational linguistics
  - text and corpus processing
  - using NLTK as an NLP library
  - NLP applications:
    - spelling correction, POS tagger, morphological analyzer/synthesizer, machine translation, text classification, ASR

slide from Janet & Na-Rae (Pitt)

so what is this class?

learning goals

- to gain an overall and broad appreciation and understanding other the field
  - linguistic phenomena & NLP tasks/applications
- to develop robust Python skills and acquire basic ML concepts/knowledge
- a stepping stone for intermediate/advanced CL/NLP classes

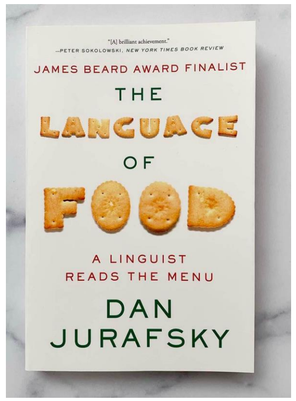
where linguistics is used

- annotation & guidelines: deciding what counts as a verb, an entity, a toxic comment
- evaluation & error analysis: how do models make mistakes and what they mean
- language documentation & low-resource NLP: most of the world's 7,000+ languages are underserved
- research: psycholinguistics, corpus linguistics, sociolinguistics all influence applications

linguistics at scale

fun things you can do with CL

- a study of ~900,000 Yelp reviews (Jurafsky et al. 2014, *Narrative framing of consumer sentiment in online restaurant reviews*)
  - 1-star reviews share the linguistic signature of **trauma narratives**
    - past tense (distancing), we/us (seeking solace), people-words
  - positive reviews focus on **sensory pleasure**
    - reads like “sex, drugs, and dessert”
    - “orgasmic pastry”, “the fries are like crack"



Dan Jurafsky

living with LLMs

the landscape in 2026

- large language models are great
  - they **write**, **translate**, and **transcribe** astonishingly well
  - most of this course's classic tasks now sit inside them
- they also **hallucinate**, are **overconfident**, are **sycophantic**, and **reproduce stereotypes** about essentially every group

living with LLMs

our stance

- both facts are course content
  - the components (tokenizers, probabilities, embeddings, attention)
  - the evaluation mindset that catches the failures
- neither hype nor dismissal
  - a computational linguist’s job in 2026 is to know exactly what these systems do with language, and to check
- more on learning in the age of LLM on Thursday

logistics

communication

- for course syllabus/slides/readings → https://siyuliang.github.io/teaching/ling430/fall26/
- for assignments/grades → Canvas
- for general questions → Canvas Discussion Boards
- for extensions/specifics about your own homework → email
- for anything you want to talk about → office hour or by appointment

before Thursday

- bring a laptop (for every class)
- read: Ma et al. 2025, “Not Everyone Wins with LLMs”. Skim for the main finding
- also: Cotterell’s short post “Is ACL Interdisciplinary?” if you haven’t already. Skim okay.
- HW 1 releases Thursday: setup and easy coding