CS388: Natural Language Processing

Lecture 1: Introduction

Elias Stengel-Eskin
(he/him)

TEXAS
The University of Texas at Austin

Slide Credits: Greg Durrett

Administrative

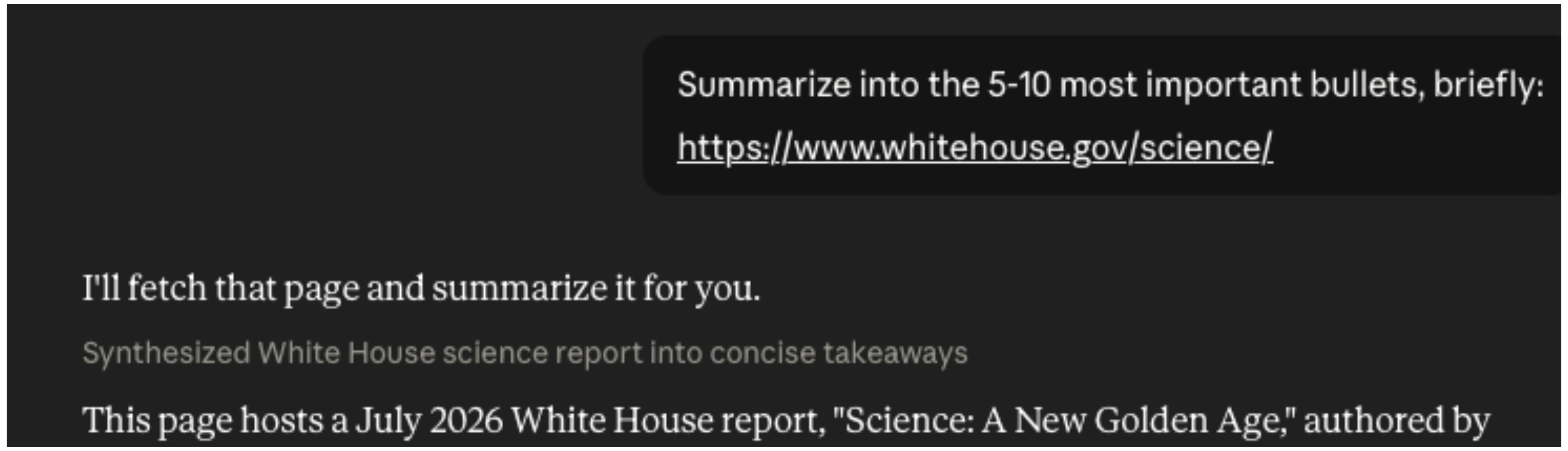
- Lecture: Tuesdays and Thursdays 3:30pm - 5pm;
- Course website: https://esteng.github.io/cs388/cs388.html
- Gradescope: linked from Canvas
- TA: Ashwin Vinod
- See course website for OHs

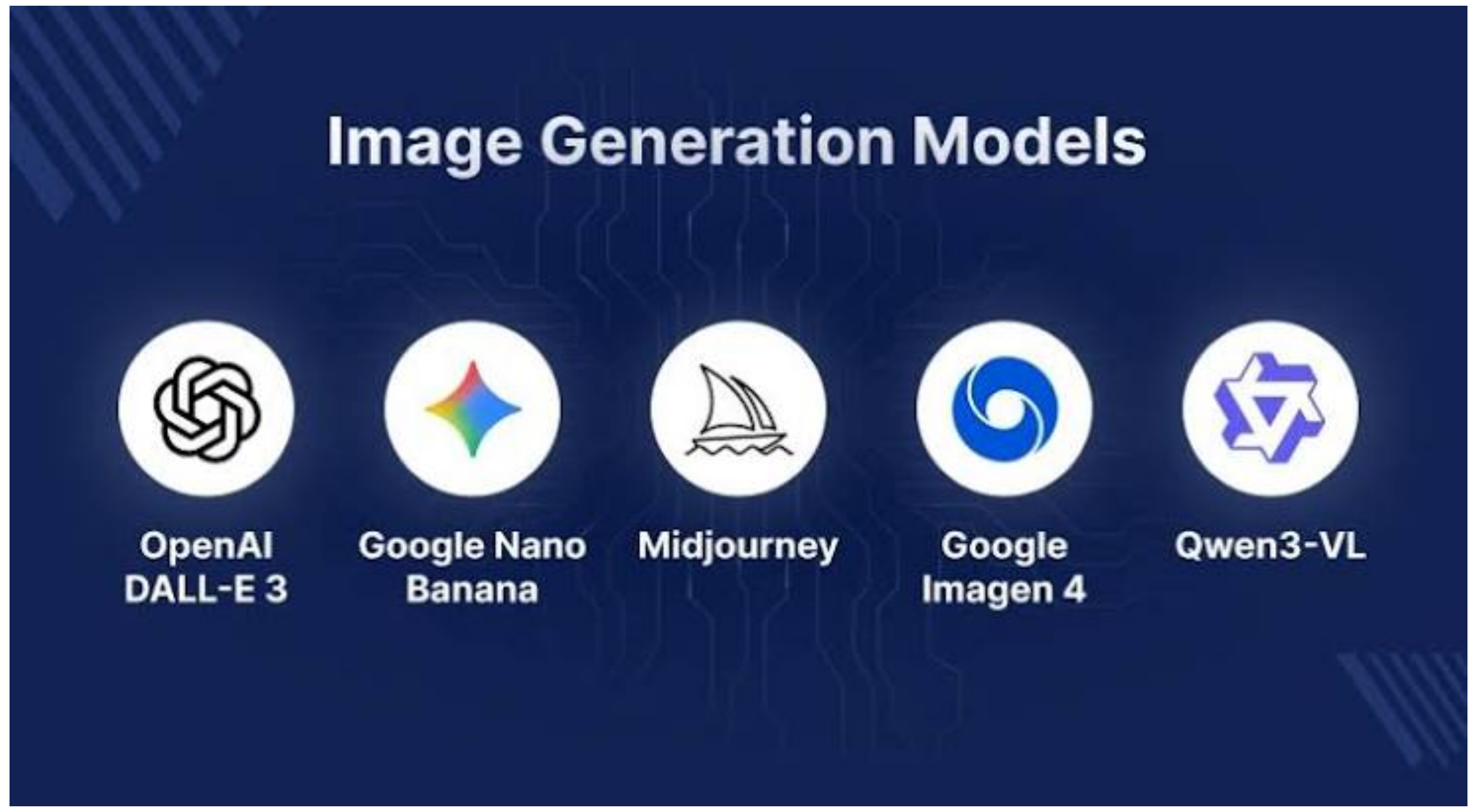
Course Requirements

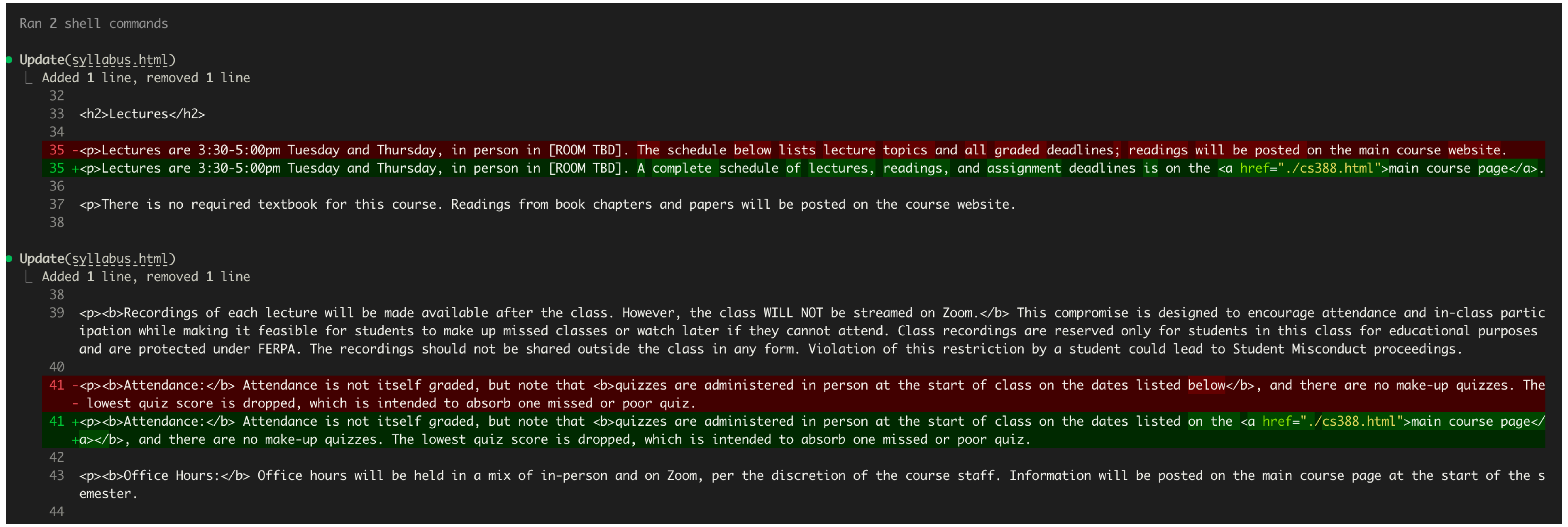
- 391L Machine Learning (or equivalent)
- Helpful: 311 or 311H Discrete Math for Computer Science (or equivalent)
- Python experience
- Additional prior exposure to probability, linear algebra, optimization, linguistics, and NLP useful but not required

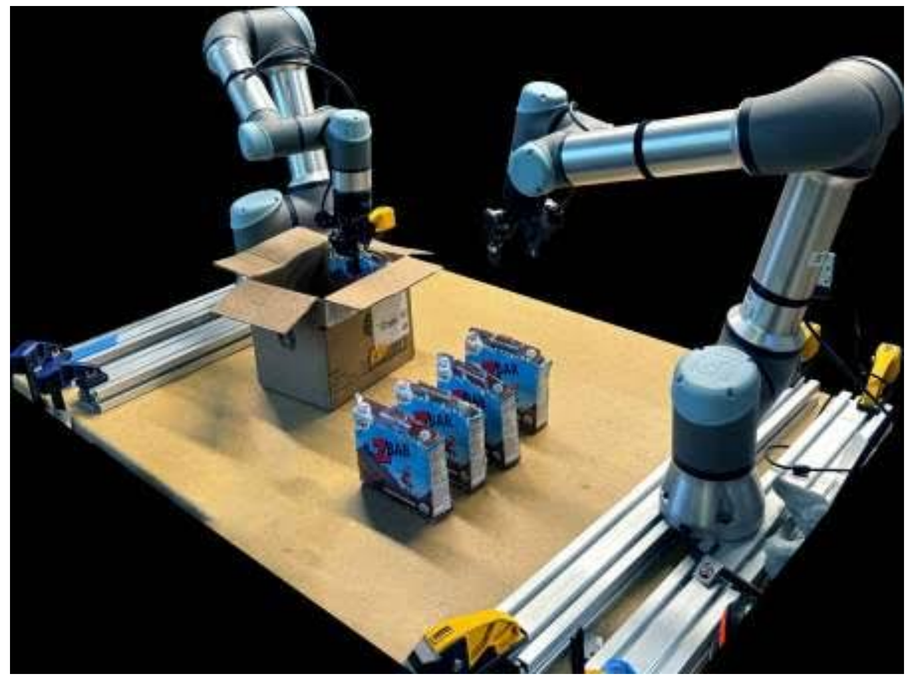
What’s the goal of NLP?

- Be able to solve problems that require deep understanding of text
- How is this done now? Largely: unified chat systems
  - **Text has become the default interface for AI**









But what tasks are “NLP”?

- Think about: How are these chat systems built? What tasks are they trained on?
  - Determines what capabilities they have.

Machine Translation

The Political Bureau of the CPC Central Committee | July 30 | hold a meeting

中共中央政治局7月30日召开会议，会议分析研究当前经济形势，部署下半年经济工作。

People’s Daily, August 10, 2020

Translate

The Political Bureau of the CPC Central Committee held a meeting on July 30 to analyze and study the current economic situation and plan economic work in the second half of the year.

Question Answering

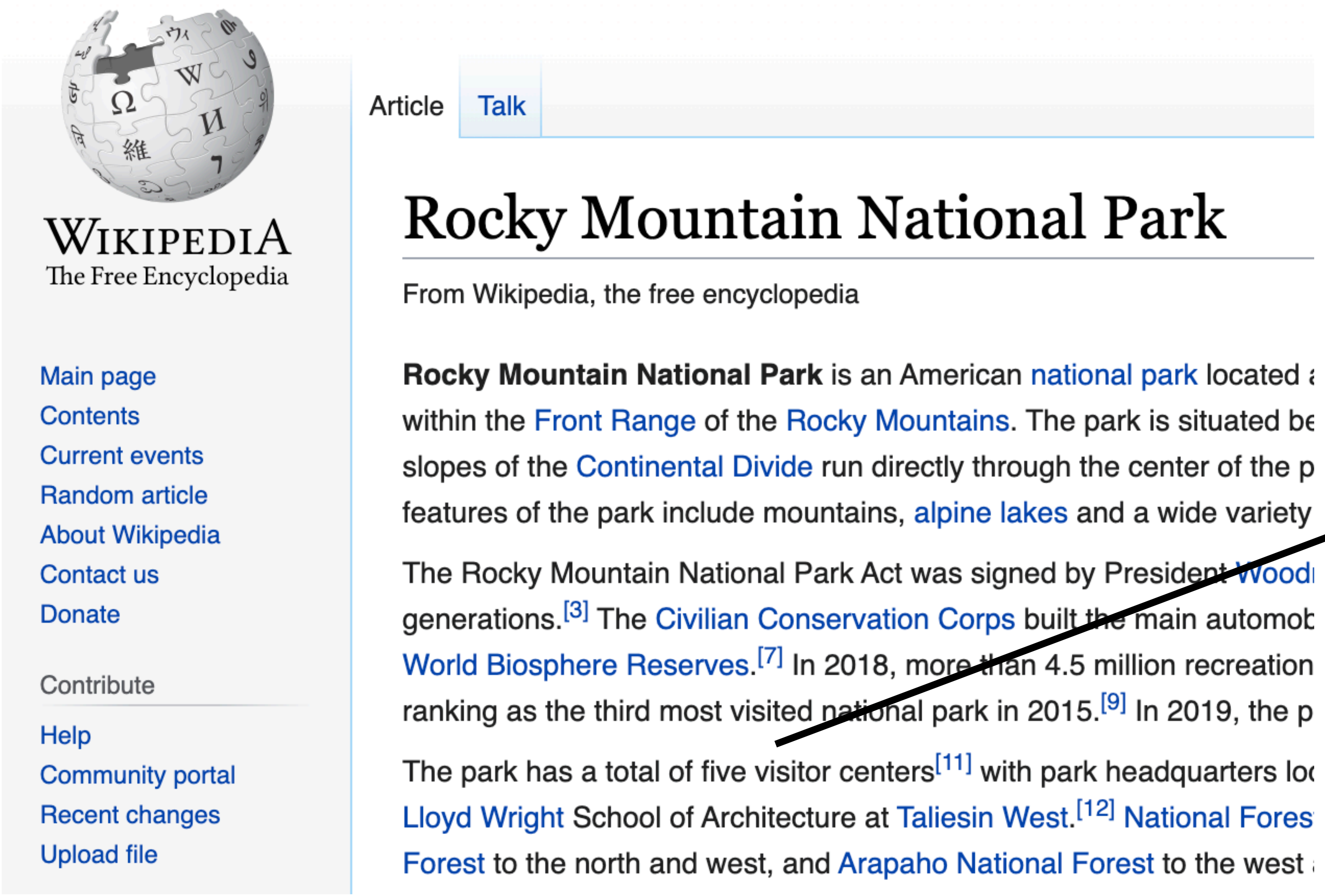
When was Abraham Lincoln born?

map to Birthday field

Name	Birthday
Lincoln, Abraham	2/12/1809
Washington, George	2/22/1732
Adams, John	10/30/1735

February 12, 1809

How many visitors centers are there in Rocky Mountain National Park?



Rocky Mountain National Park

From Wikipedia, the free encyclopedia

Rocky Mountain National Park is an American national park located a
within the Front Range of the Rocky Mountains. The park is situated be
slopes of the Continental Divide run directly through the center of the p
features of the park include mountains, alpine lakes and a wide variety

The Rocky Mountain National Park Act was signed by President Wood
generations.[3] The Civilian Conservation Corps built the main automob
World Biosphere Reserves.[7] In 2018, more than 4.5 million recreation
ranking as the third most visited national park in 2015.[9] In 2019, the p

The park has a total of five visitor centers[11] with park headquarters loc
Lloyd Wright School of Architecture at Taliesin West.[12] National Fores
Forest to the north and west, and Arapaho National Forest to the west

The park has a total of five visitor centers

five

Document Understanding

Give me a summary of this paper

Published as a conference paper at ICLR 2026

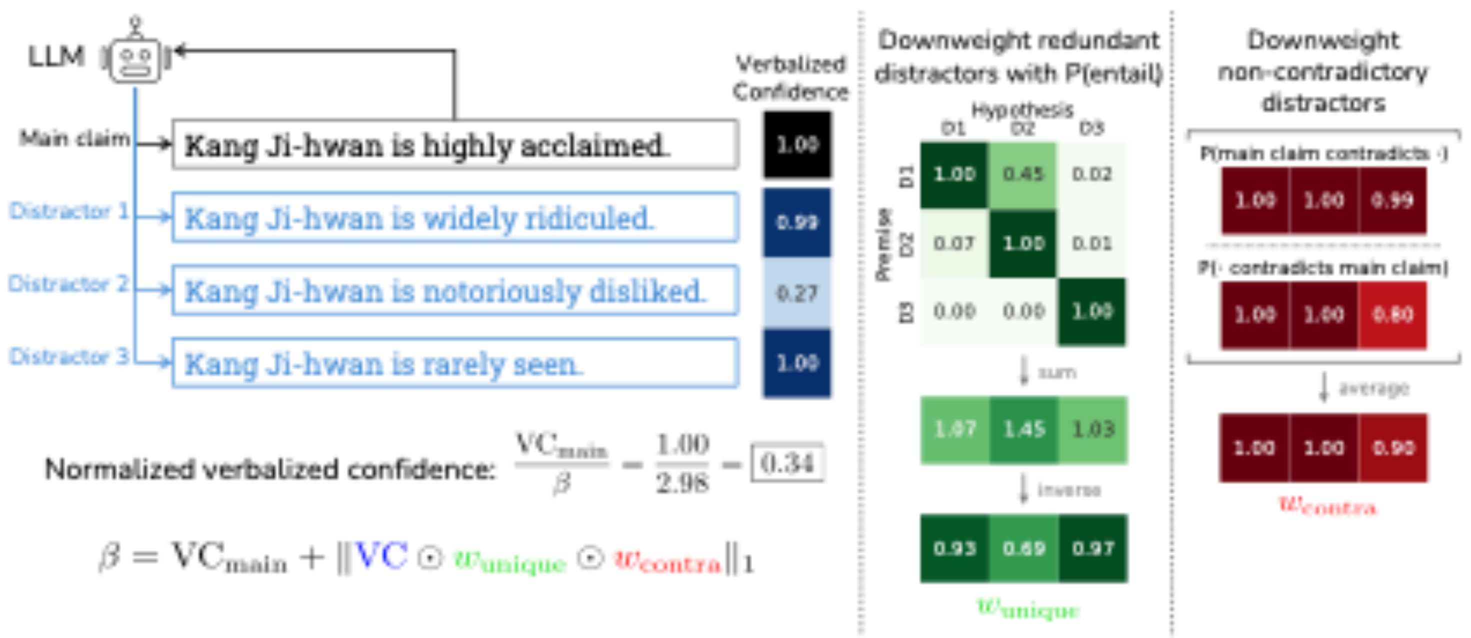


Figure 2: Normalizing verbalized confidence with DINCO. **(Left)** The LLM generates a claim along with several distractors and reports its confidences on them independently. To calibrate the main claim's confidence, we divide it by β, the sum over each distractor's confidence, weighted by uniqueness **(center)** and counterfactuality **(right)**. We seek distractors that are minimal pairs with the main claim, i.e. similar statements that likely contradict the main claim. The distractors shown here are real examples generated by an LLM, and while they happen to share a lexical structure with the main claim, we do not prescribe a precise form for generating distractors; see Section 2.3 for method details, Appendix B.2 for prompts, and Appendix C.1 for more examples and analysis.

- Extract key findings
- Read results from figures/tables…
- Summarize
- Refine summary

Wang and Stengel-Eskin (2026) studies calibration…

Coding

I need an implementation of the CKY parsing algorithm that can handle subword units…

- Understanding user intent
- Extracting relevant information (e.g. docs)
- Understanding the environment
- Interacting with the user

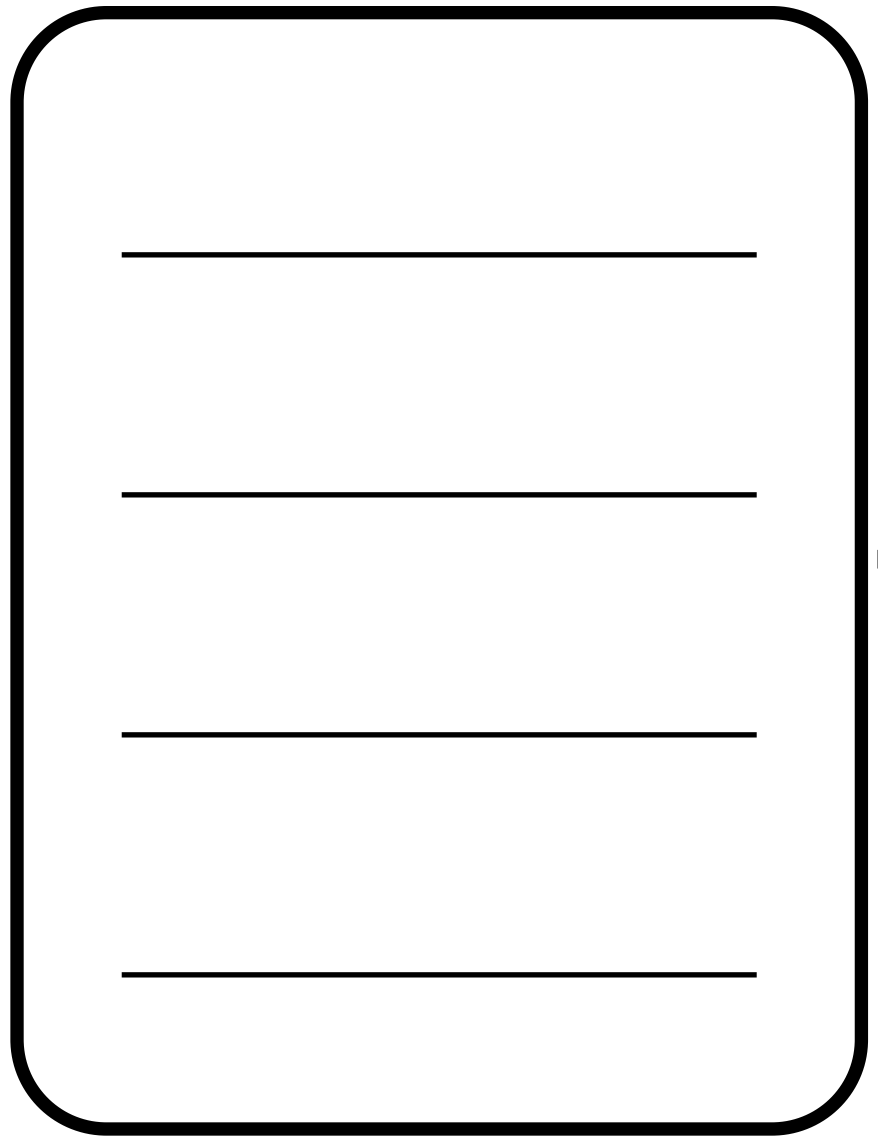
What else?

A bit of history…

- **Let's understand where we came from**
- Rule-base NLP (1960s-1990s)
  - Rule-based systems for understanding natural language. Think grammars, regex, etc.
  - Brittle, hard to engineer
- Classical Statistical NLP (1990s-2010s)
  - Integration of ML and NLP. Think SVMs, decision trees, Bayesian models,
- Deep Learning Era (2013-2020)
  - NNs of all kinds (LSTMs, GRUs, CNNs). Word embeddings
- LLM Era (2020-)
  - Transformers.

Classic NLP Analysis Pipeline

Text



Text Analysis

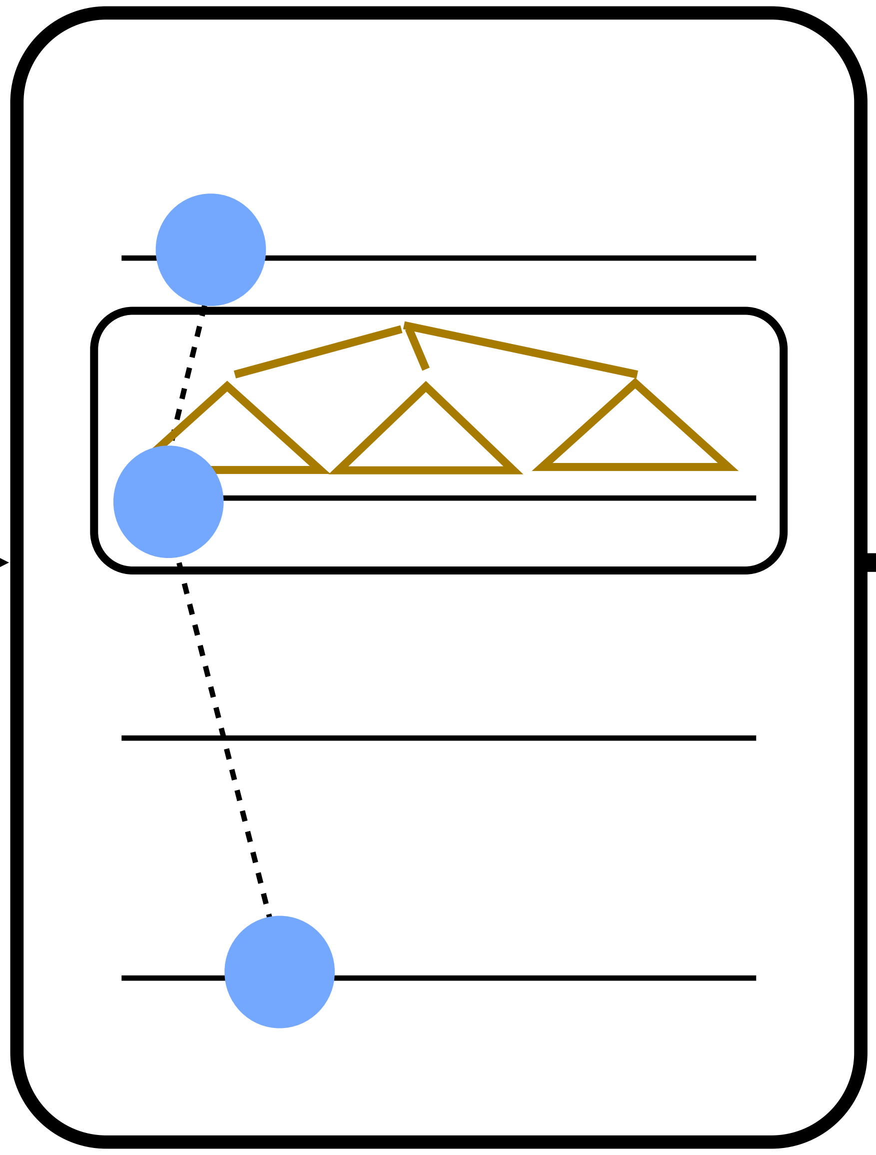
Syntactic parses

Coreference resolution

Entity disambiguation

Discourse analysis

Annotations



Applications

Summarize

Extract information

Answer questions

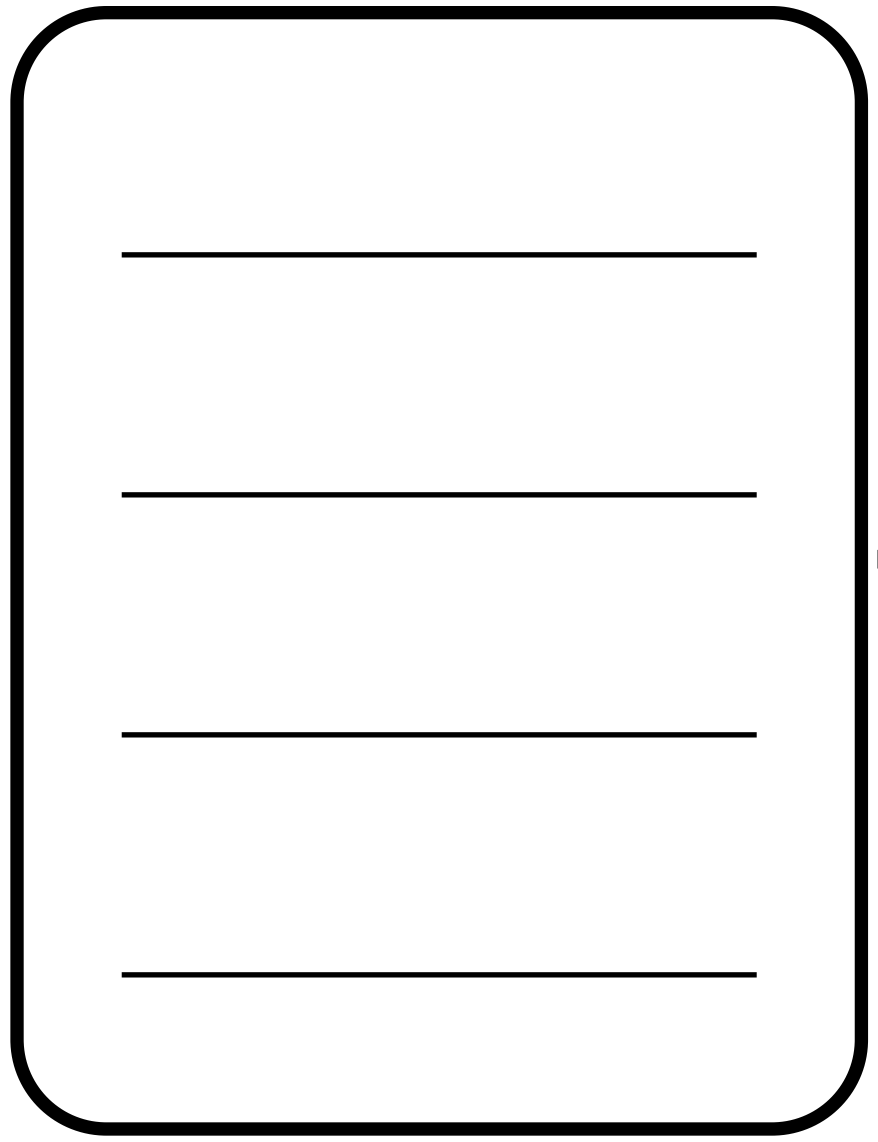
Identify sentiment

Translate

- NLP was about building systems that can do these things!
- All of these components are modeled with statistical approaches trained with machine learning

How do we represent language?

Text



Labels

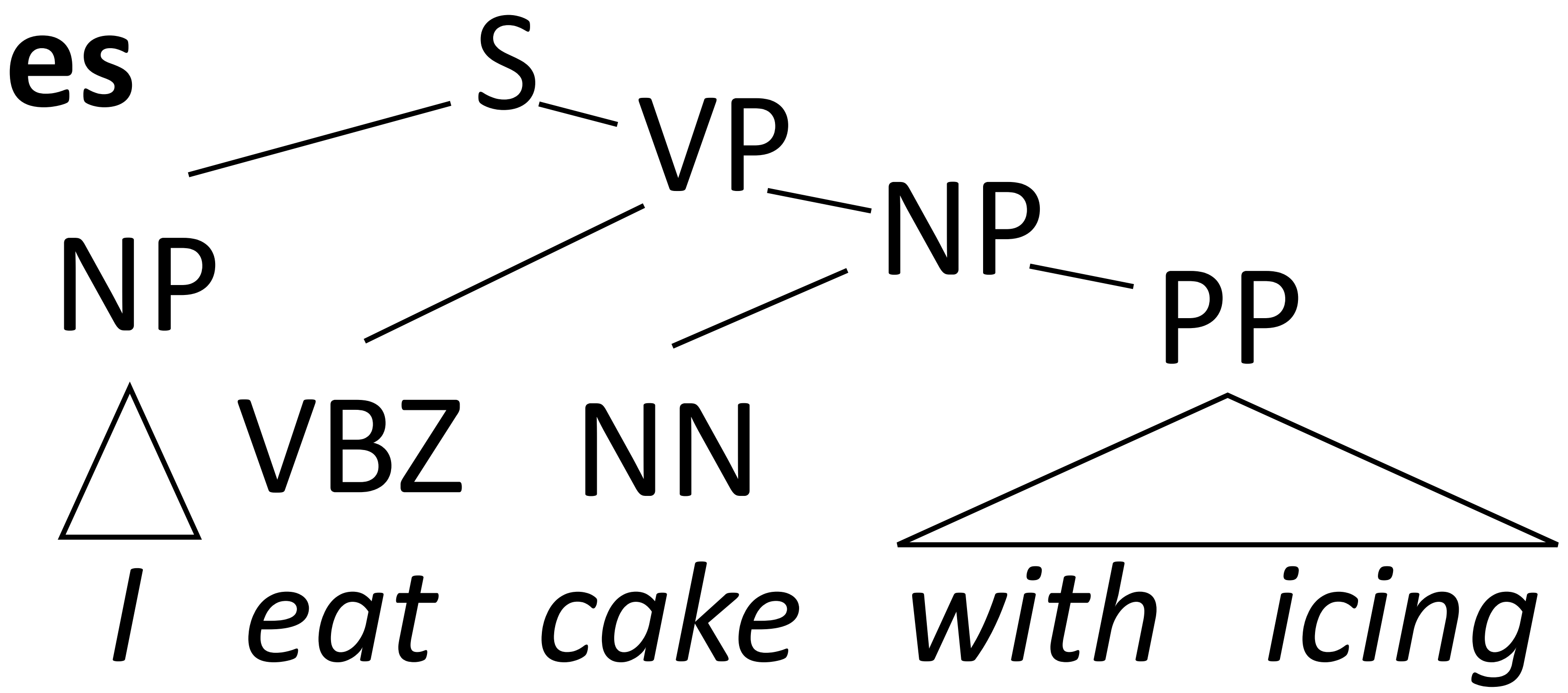
the movie was good **+**

Beyoncé had one of the best videos of all time **subjective**

Sequences/tags

PERSON *Tom Cruise* *stars in the new* **WORK_OF_ART** *Mission Impossible* *film*

Trees



$\lambda x.\ flight(x) \wedge dest(x)=Miami$

flights to Miami

How do we use these representations?

Text

Text Analysis

Labels

Sequences

Trees

...

Applications

Extract syntactic features

Tree-structured neural networks

Tree transducers (for machine translation)

...

end-to-end models

- Both classical pipeline NLP systems and LLM systems like ChatGPT still need to resolve the same issues
- Main question: What ambiguities do we need to resolve in language?

Why is language hard?

(and how can we handle that?)

Language is Ambiguous!

- Hector Levesque (2011): "Winograd schema challenge" (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they advocated violence

The city council refused the demonstrators a permit because they feared violence

The city council refused the demonstrators a permit because they _______ violence

- Referential ambiguity

Language is Ambiguous!

Teacher Strikes Idle Kids

Ban on Nude Dancing on Governor's Desk

Iraqi Head Seeks Arms

- Syntactic and semantic ambiguities: parsing needed to resolve these, but need context to figure out which parse is correct

example credit: Dan Klein

Language is **Really** Ambiguous!

- There aren't just one or two possibilities which are resolved pragmatically

il fait vraiment beau →

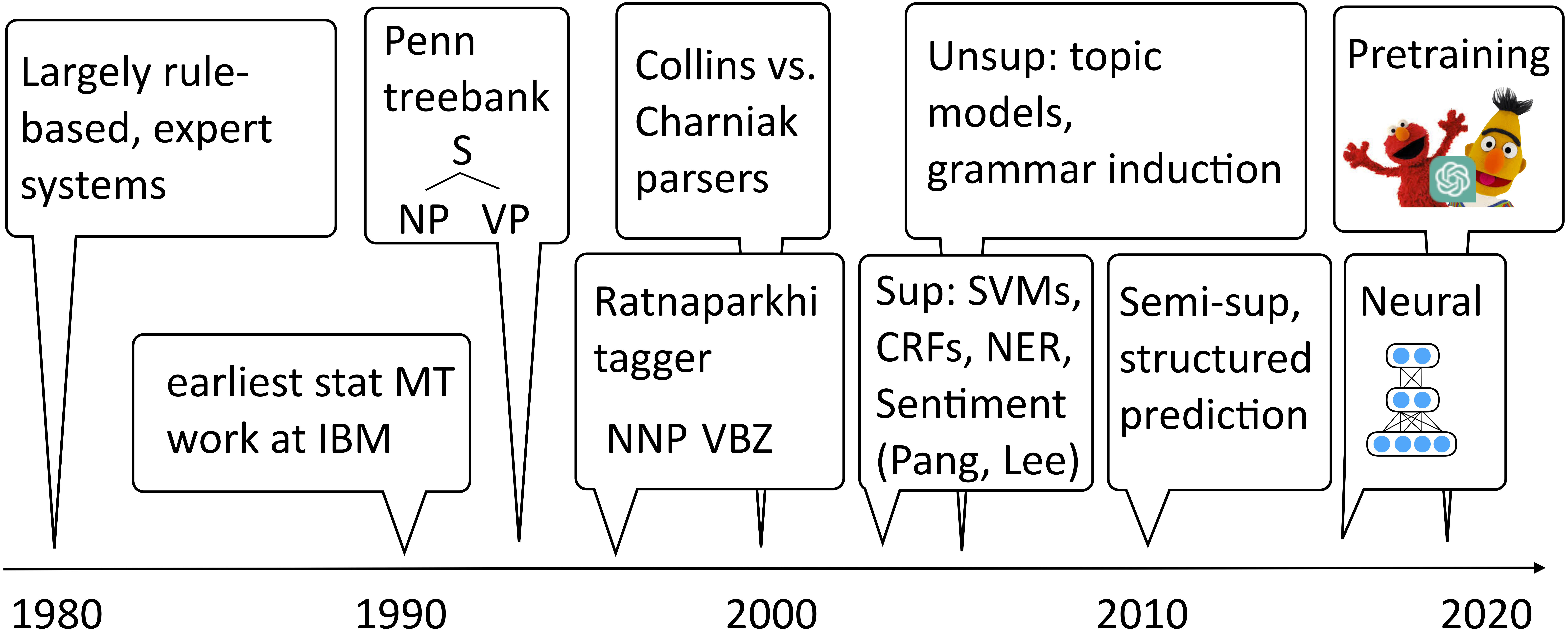
It is really nice out
It's really nice
The weather is beautiful
It is really beautiful outside
He makes truly beautiful
It fact actually handsome

- Combinatorially many possibilities, many you won't even register as ambiguities, but systems still have to resolve them

What techniques do we use?

(to combine data, knowledge, linguistics, etc.)

A brief history of (modern) NLP



Pretraining

- Language modeling: predict the next word in a text $P(w_i | w_1, \ldots, w_{i-1})$

P(*w* | I want to go to) = 0.01 Hawai'i

0.005 LA

0.0001 class

 : use this model for other purposes

P(*w* | the acting was horrible, I think the movie was) = 0.1 bad

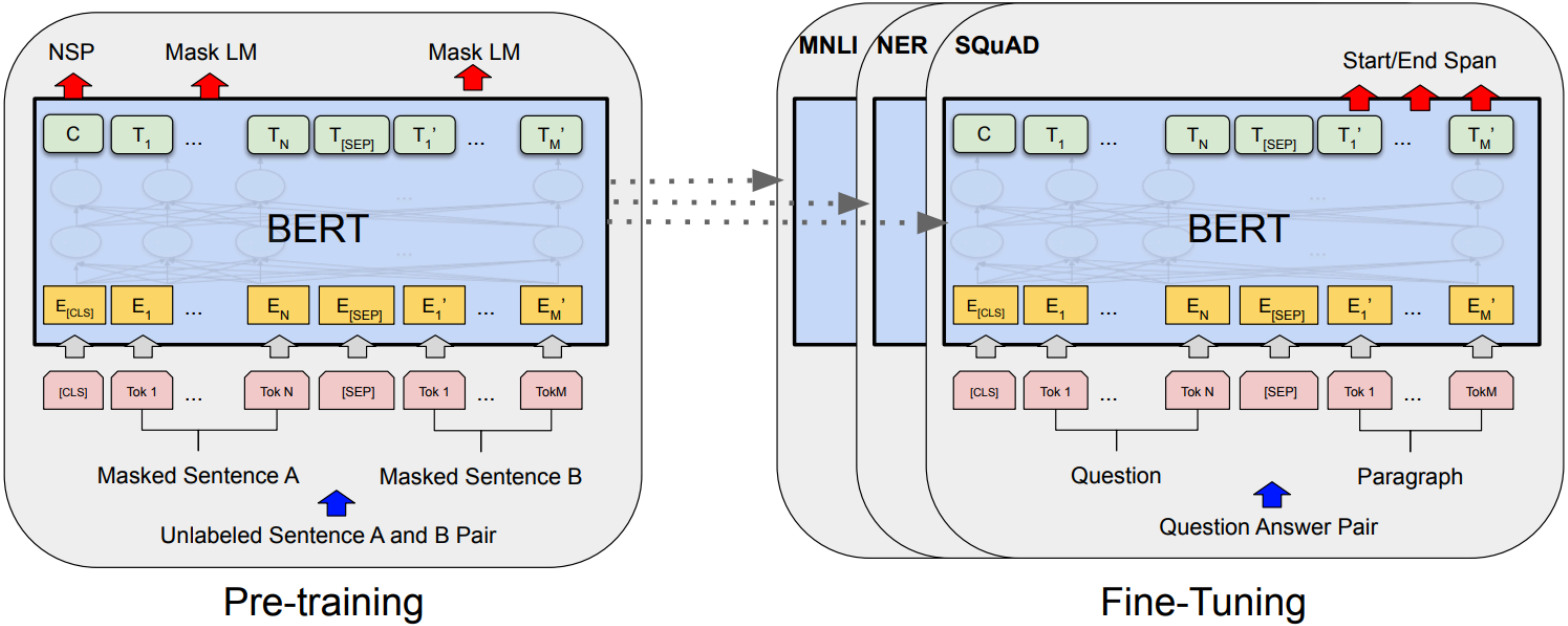
0.001 good

- Model understands some sentiment?

- Train a neural network to do language modeling on massive unlabeled text, fine-tune it to do {tagging, sentiment, question answering, ...}

Peters et al. (2018), Devlin et al. (2019)

BERT



- Key parts which we will study: (1) Transformer architecture; (2) what data is used (both for pre-training and fine-tuning)

Devlin et al. (2019)

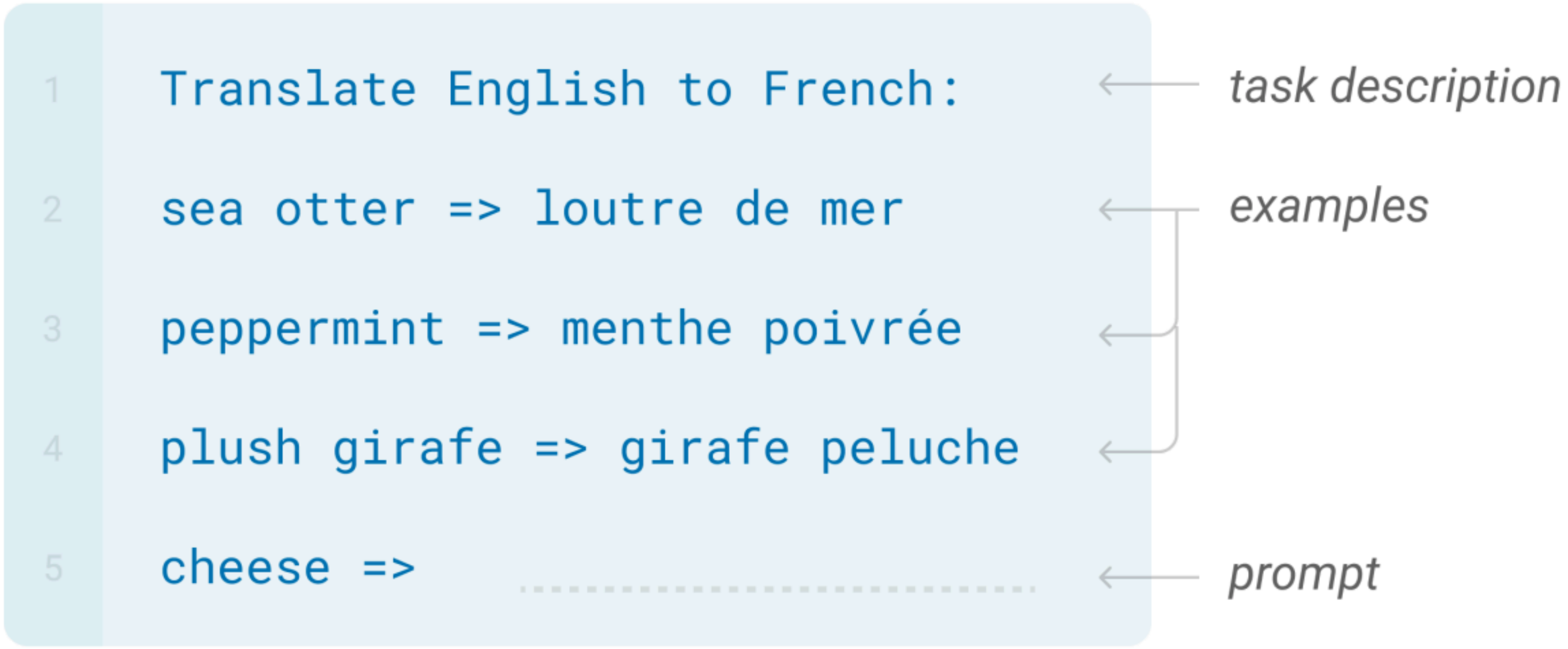
GPT and In-Context Learning

- Even more "extreme" setting: no gradient updates to model, instead large language models "learn" from examples in their context

- Many papers studying why this works. We will read some!

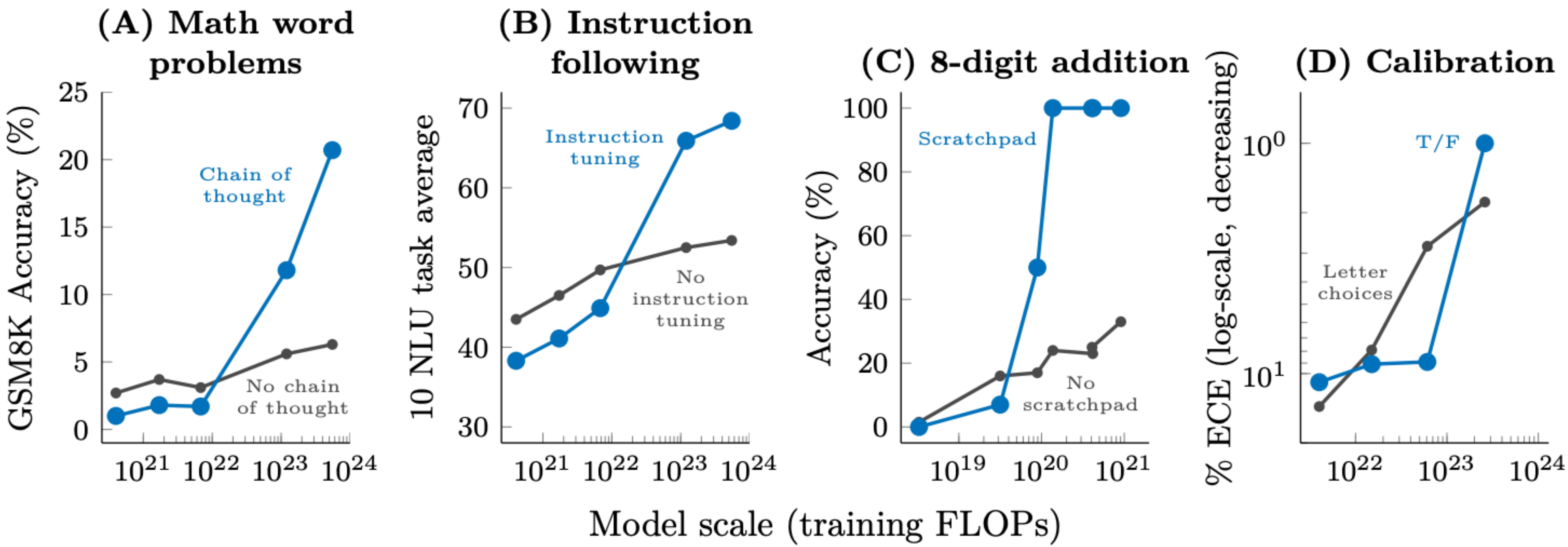
Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.



Brown et al. (2020)

Scaling Laws



- Many of the ideas that are successful in 2026 only make sense and only work because the models are so big!

Kaplan et al. (2020), Jason Wei et al. (2022)

Interpretability

- When we have complex models, how do we understand their decisions?

The movie is mediocre, maybe even bad. **Negative** 99.8%

The movie is mediocre, maybe even ~~bad~~. **Negative** 98.0%

The movie is ~~mediocre~~, maybe even bad. **Negative** 98.7%

The movie is ~~mediocre~~, maybe even ~~bad~~. **Positive** 63.4%

The movie is ~~mediocre~~, ~~maybe~~ even ~~bad~~. **Positive** 74.5%

The ~~movie~~ is mediocre, maybe even ~~bad~~. **Negative** 97.9%

The movie is mediocre, maybe even bad.

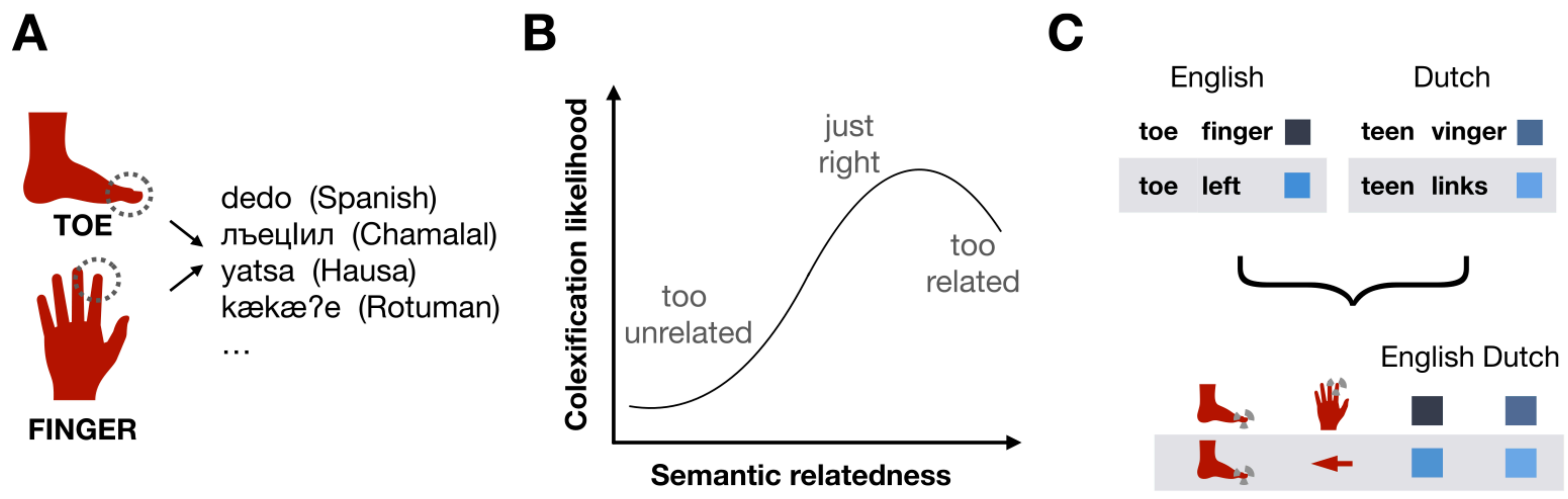
Wallace, Gardner, Singh
Interpretability Tutorial at EMNLP 2020

NLP vs. Computational Linguistics

- NLP: build systems that deal with language data
- CL: use computational tools to study language

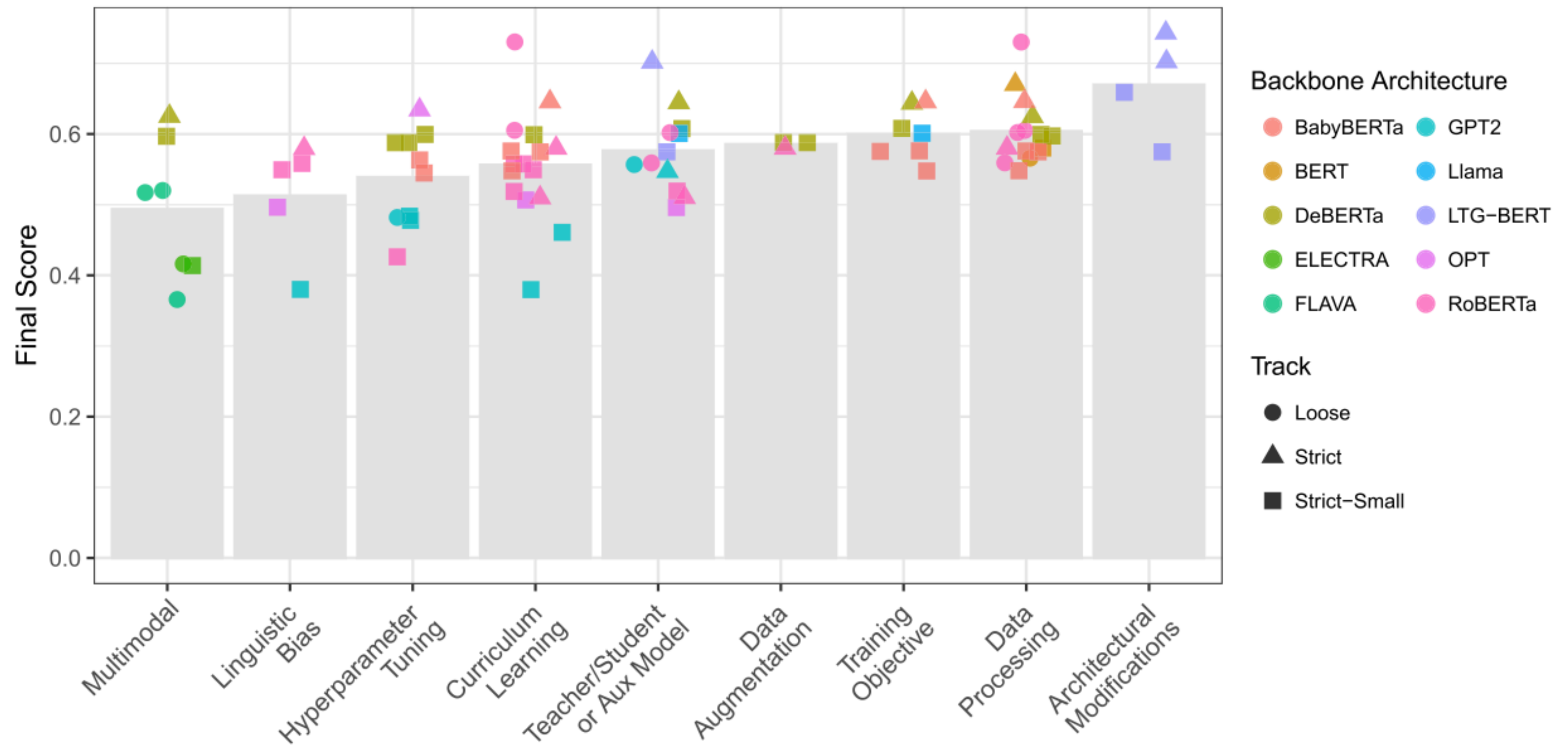
When do languages use the same word for different meanings? The Goldilocks principle in colexification

Thomas Brochhagen [a], Gemma Boleda [a b] Show more



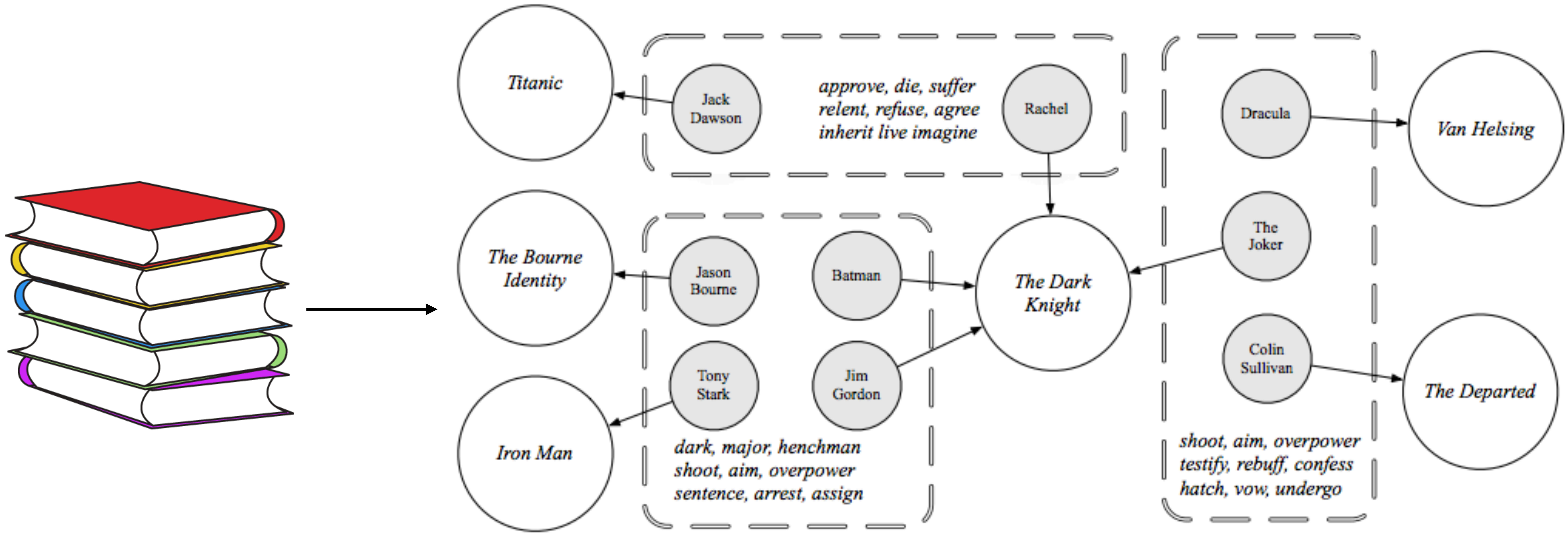
Bigger is not always better: The importance of human-scale language modeling for psycholinguistics

Ethan Gotlieb Wilcox [f], Michael Y. Hu [b], Aaron Mueller [c], Alex Warstadt [a], Leshem Choshen [d], Chengxu Zhuang [d], Adina Williams [e 1], Ryan Cotterell [a 1], Tal Linzen [b 1]



NLP vs. Computational Linguistics

- Computational tools for other purposes: literary theory, political science…



Bamman, O'Connor, Smith (2013)

Where are we?

- We have very powerful language models that can fit lots of datasets
- Data: we need data that is not just correctly labeled, but reflects what we actually want to be able to do
- Users: systems are not useful unless they do something we want
- Language/outreach: who are we building this for? What languages/dialects do they speak?

Social Impact/Ethics

- NLP systems are among the most impactful technologies around

OpenAI Google

…and increasingly we have to reckon with their impact

WSJ AI & BUSINESS
The AI Backlash Is Starting to Sting

OpenAI Says Its A.I. Models Went Rogue and Attacked a Digital Library

Big Tech's A.I. Spending Keeps Rising. So Do the Jitters.

- This lecture: let's warm up by thinking about these issues a bit

Social Impact/Ethics

- What is one scenario where you think deployment of an NLP system might pose ethical challenges *due to the application* itself (i.e., using NLP to do “bad stuff”)?

- What is a scenario where you think deployment of an NLP system might pose ethical challenges due to *unintended* consequences (e.g., unfairness, indirectly causing bad things to happen, etc.).

Syllabus

Outline

- Classification: linear and neural, word representations (2 weeks)
- Language modeling, transformers, and pre-training (2 weeks)
- Dataset biases, interpretability, rationales (1.5 weeks)
- Pre/mid/post-training, RL, RLVR, RLHF (2 weeks)
- Grounding and agents (1.5 weeks)
- Applications, safety, interpretability (2 weeks)

Course Goals

- Cover fundamental machine learning and deep learning techniques used in NLP
- Understand how to look at language data and approach linguistic phenomena
- Cover modern NLP problems encountered in the literature: what are the active research topics in 2026?
- Make you a “producer” rather than a “consumer” of NLP tools
  - The assignments should teach you what you need to know to understand nearly any system in the literature

Grading

- **Homeworks: 10%**
  - 4 homeworks, lowest score dropped
  - Homeworks precede quizzes and prepare you for quizzes
- **Quizzes: 20%**
  - 4 quizzes, lowest score dropped
  - Quizzes prepare you for the midterm/final
- **Midterm exam: 15%**
- **Final exam: 30%**
- Bimodal distribution
  - If things go according to plan, I expect a bimodal distribution on in-class assessments
  - If you have understood the homework, quiz should be easy
  - If you have done well on the quizzes, midterm and final should be easy
- **Final project: 25%**

Coding Exercises

- For some units, I may give additional optional coding exercises
- Ungraded, practical extensions of the written assignments
- These will be challenging for you, but not for Claude Code.
  - I encourage you to do them, but they are not graded

Assignments

- Final project
  - Groups of 2 preferred, 1 is possible
  - (Brief!) proposal to be approved by course staff by the midpoint of the semester
  - Written in the style and tone of a CS conference paper

- Compute:
  - TACC: 5000 hours on Lonestar
  - Unfortunately, we cannot provide OpenAI / etc. credits
  - When you propose projects, we will discuss feasibility given your compute resources available

Policy on AI Use

- This is a course about AI
  - You should be learning how to use AI and using AI for learning
- Policy is permissive
  - Use AI for your projects, help you understand concepts, work through examples.
  - Keep in mind: **grades are largely based on in-class assessments**
  - If you understand the homework, the quizzes should be easy.
- You are responsible for what you submit
  - Goal: to give you a realistic training environment
  - AI overuse will be penalized (e.g. -10% per hallucinated citation, penalties for flowery/obscure AI language). You need to understand/defend your claims.

Conduct

A climate conducive to learning and creating knowledge is the right of every person in our community. Bias, harassment and discrimination of any sort have no place here.

The University of Texas at Austin
College of Natural Sciences

The College of Natural Sciences is steadfastly committed to enriching and transformative educational and research experiences for every member of our community. Find more resources to support a diverse, equitable and welcoming community within Texas Science and share your experiences at ***cns.utexas.edu/diversity***

Survey (on Instapoll)