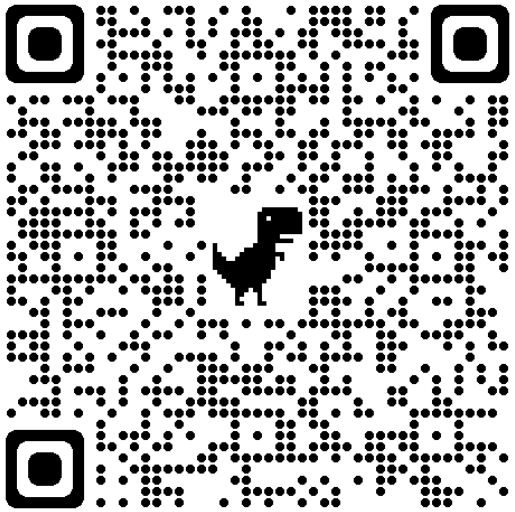
CS6120: Lecture 7
Prompting

Kenneth Church

https://kwchurch.github.io/

Homework



Assignment 4: ChatGPT

Please use ChatGPT, Google and whatever other tools might be useful to do this assignment. The point is not so much to solve the problems, but to learn how to use these tools effectively, and to discover their strengths and weaknesses.

Question 1

Suppose you were a high school student and you were asked to write an essay on *witch hunts*. Please use whatever tools you can find to write an essay that covers the following questions:

1. What does Trump mean when he refers to *witch hunts*?
2. What did *witch hunt* mean in Salem Massachusetts in the late 1600s?
3. What did it mean in Europe starting a few hundred years earlier?
4. What did it mean in Arthur Miller's *The Crucible*?
5. What does *Red Scare* mean, and what is the connection between Arthur Miller and *Red Scare*?

In addition to the essay, please explain how you did what you did.

1. Were you already familiar with this material, or did you have to use some tools to answer these questions?
2. Did you use ChatGPT?
3. What prompts did you use?
4. Did you use the output as is, or did you modify it?
5. Was it useful?
6. What did it do well, and are there any opportunities for improvement?
7. If you used Google, what queries did you use?
8. If you used any other tools, which tools did you use, and how?
9. A well written essay should have a high level structure that flows naturally from the beginning to the end, with smooth transitions from one topic to the next. Do the tools mentioned above help with the high level structure and the flow, or did you have to fix that by hand?

Question 2

Sports metaphors often do not translate well from one English speaking country to another. Use the tools mentioned above to explain what the following terms mean. Which sports are these terms from? What do they mean in that sport? How are they used

Prompt Engineering
(and disintermediating web-search)

- Super-Popular (100+ million users)
  - Most successful (rapid) adoption of any web app ever
- Super-Easy
  - Easier than Fine-Tuning (and Inference)
- Use Cases
  - “Helping” with homework:
    - Cheating (?)
  - Documentation:
    - Alternative to stack overflow

“Helping” with homework: Cheating (?)

- ***Collaborate*** with students on essays
  - *You have no idea how much we’re using ChatGPT*
  - Cheating?
- ChatGPT is better for some tasks
  - Good: thesis statements, outlines
  - Bad: capture student’s voice
  - Worse: quotes
- Learning opportunity:
  - How to decompose writing to subtasks
  - Collaboration is great,
    - but student is responsible for end-product

Disintermediating Google

- Google → Stack Overflow → Instant Answers → ChatBot

stackoverflow About Products For Teams [faiss]

Home

PUBLIC

Questions

Tags

Users

Companies

COLLECTIVES

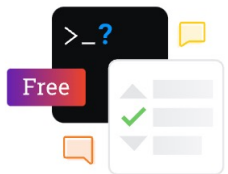
Explore Collectives

LABS

Discussions

TEAMS

Stack Overflow for Teams – Start collaborating and sharing organizational knowledge.



Questions tagged [faiss]

Ask Question

The faiss tag has no usage guidance, but it has a tag wiki.

Learn more... Top users Synonyms

88 questions

Newest | Active | Bountied | Unanswered | More | Filter

8 votes
✓ 3 answers
29k views

faiss ERROR: Could not find a version that satisfies the requirement faiss (from versions: none)

When Running Installation: pip install faiss I am getting this error: ERROR: Could not find a version that satisfies the requirement faiss (from versions: none) ERROR: No matching distribution ...

python python-3.x pip anaconda faiss

predactor **782** asked Nov 20, 2019 at 14:51

7 votes
2 answers
4k views

Installing Faiss on Lambda Stack

I want to install Faiss-GPU on Lambda Stack by conda install -c pytorch faiss-gpu but there is no conda installed. I tried the solution mentioned here: Installing faiss on Google Colaboratory with ...

python faiss

MAMProgr **410** asked Nov 3, 2019 at 11:10

6 votes
✓ 4 answers
10k views

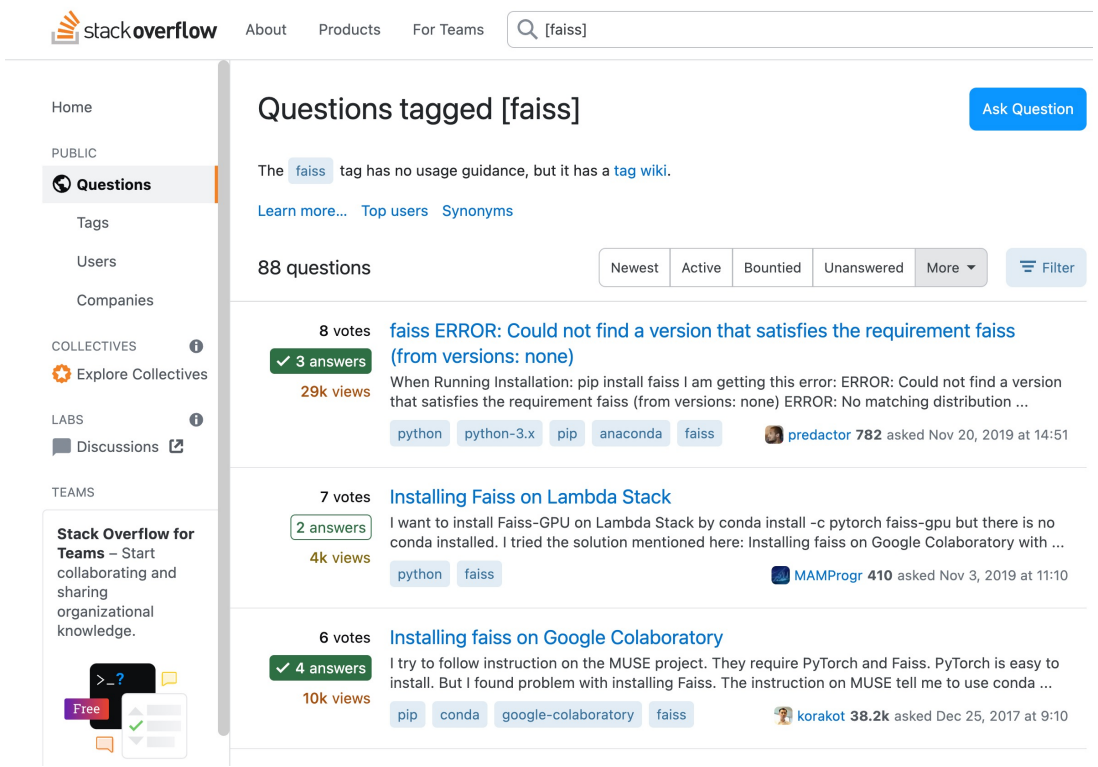
Installing faiss on Google Colaboratory

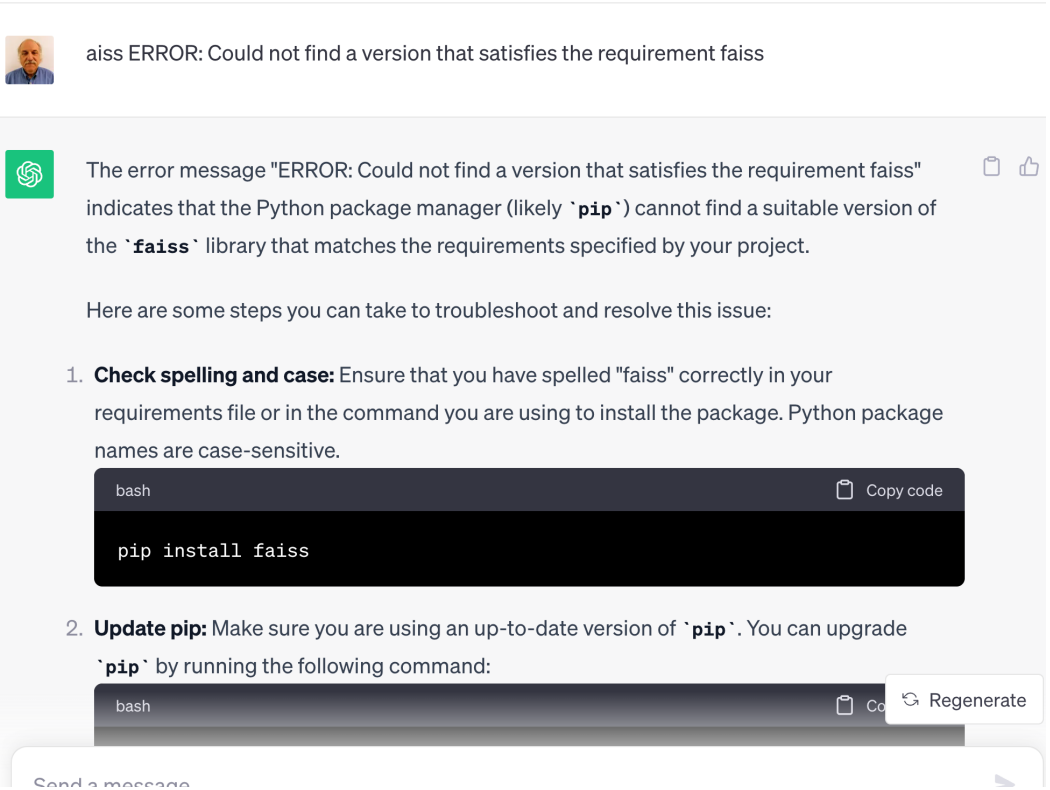
I try to follow instruction on the MUSE project. They require PyTorch and Faiss. PyTorch is easy to install. But I found problem with installing Faiss. The instruction on MUSE tell me to use conda ...

pip conda google-colaboratory faiss

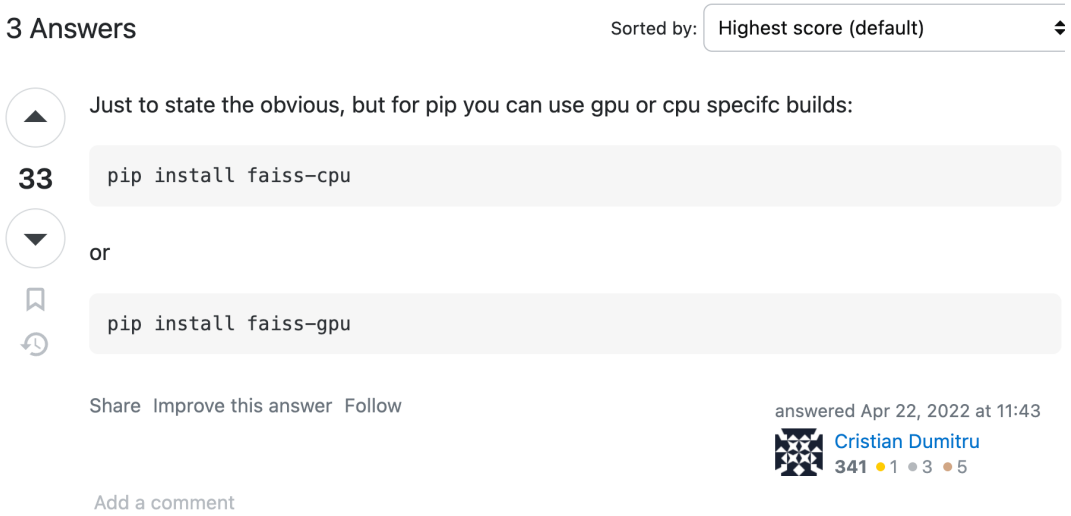
korakot **38.2k** asked Dec 25, 2017 at 9:10

Replacing Stack Overflow with ChatGPT





I may be old-fashioned, but I still use Stack Overflow…



- Advantages of Stackoverflow
  - Behavioral signals: logs, votes
  - Wisdom of the crowd
  - Feedback to developers
    - Bug Reports
      - with stats (prioritization)
      - and workarounds (with votes)
- Web search
  - Solitaire
  - Multi-player game (auction)
- If we lose stackoverflow
  - Developers will suffer
- Where does ChatGPT get its content?
  - Stackoverflow?

Decomposing a big problem into subtasks

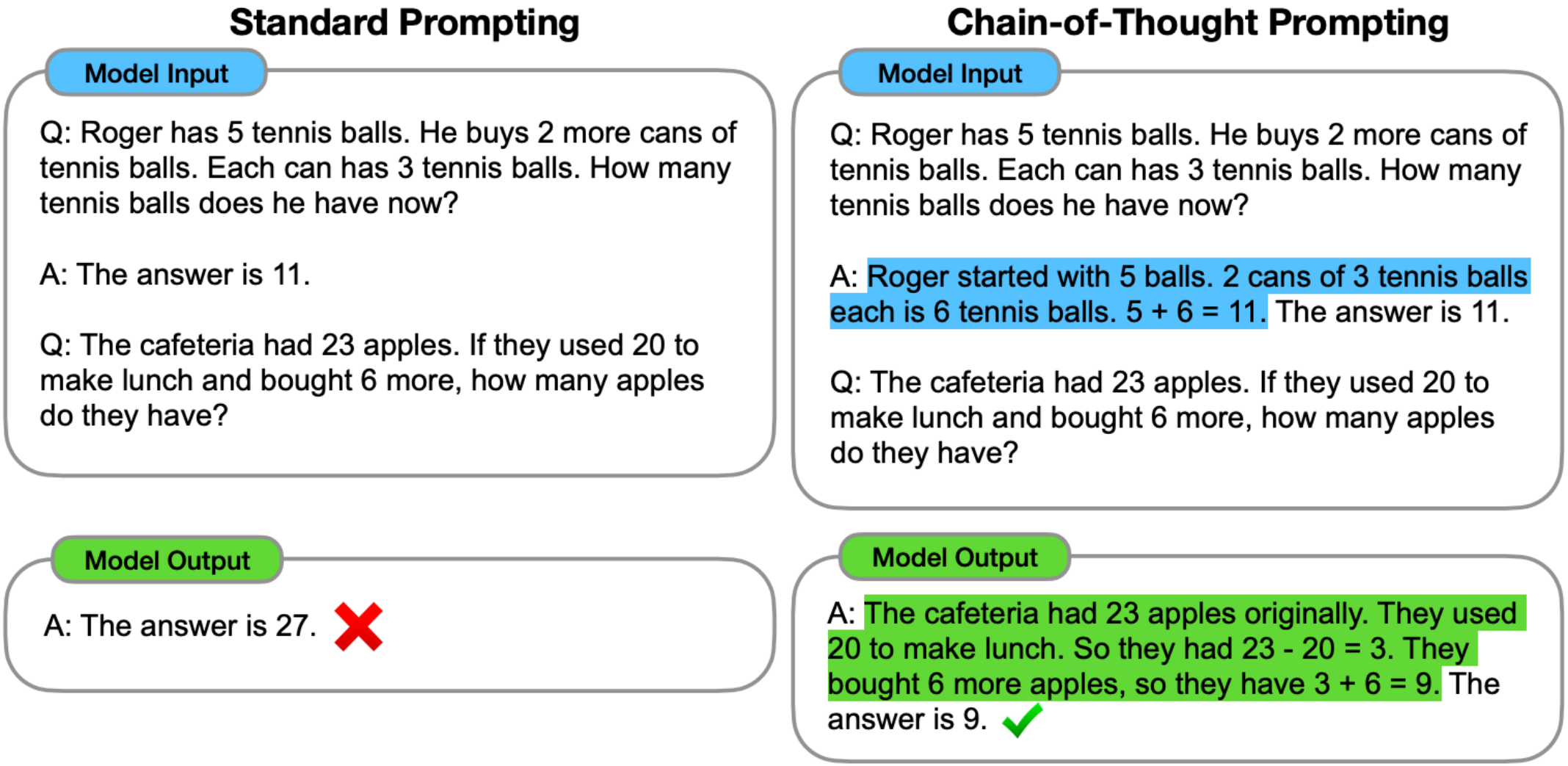


Figure 1: Chain-of-thought prompting enables large language models to tackle complex arithmetic, commonsense, and symbolic reasoning tasks. Chain-of-thought reasoning processes are highlighted.

Limitations

Decomposing Problems

- Chain-of-Thought Prompting
  - works because...
  - ChatGPT needs help decomposing problems into subtasks

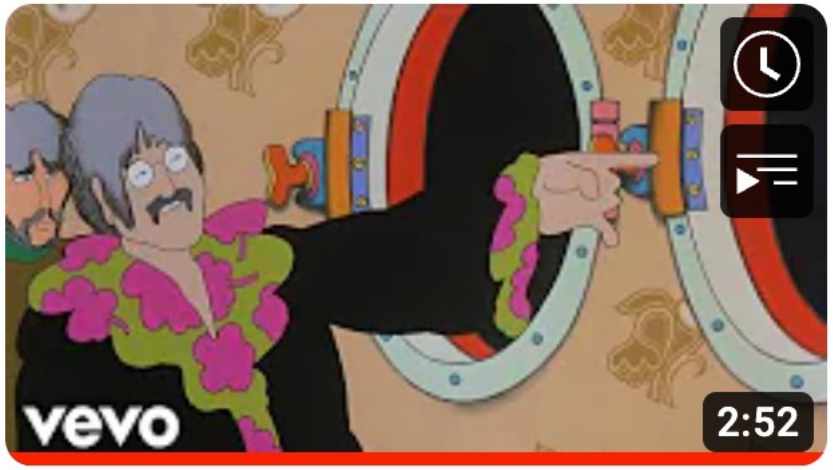
Hypothetical

- Suppose ChatGPT
  - can add two small numbers
- But for large numbers,
  - it makes up answers
- After each release,
  - “small” gets “bigger”
- (We will return to this later)

Test of time: Inference is cool (for now)

https://youtu.be/rLl9XBg7wSs

Test of time: Inference is cool (for now)
Will you still love me when I'm Sixty-Four?



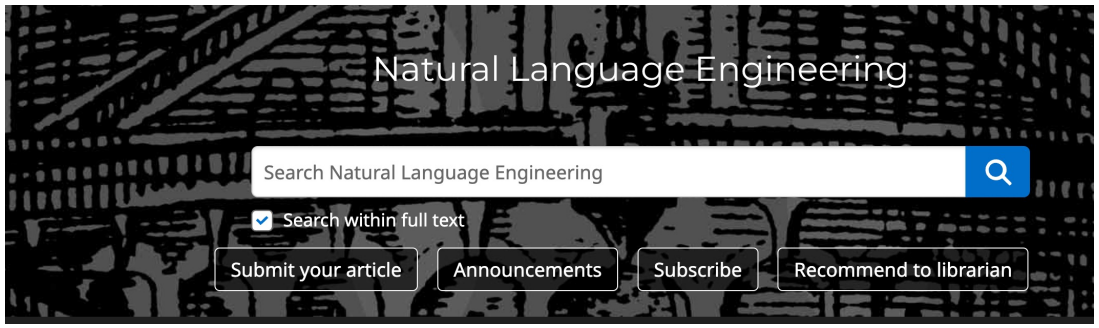
The Beatles - When I'm Sixty Four (Official Video)

TheFreddyShow • 585K views

"When I'm Sixty-Four" is a song by the English rock band The Beatles, written by Paul McCartney (credited to Lennon–McCartney) and...

https://www.youtube.com/watch?v=wUDRIC5RSX4

Shameless Plug: Smooth-Talking Machines



Most read

This page lists the top ten most read articles for this journal based on the number of full text views and downloads recorded on Cambridge Core over the last 30 days. This list is updated on a daily basis.

Emerging trends: Smooth-talking machines
Kenneth Ward Church, Richard Yue
Published online by Cambridge University Press: 11 September 2023, pp. 1402-1410
Article · Access · Open access · PDF · HTML · Export citation
View abstract

GPT-3: What's it good for?
Robert Dale
Published online by Cambridge University Press: 15 December 2020, pp. 113-118
Article · Access · Open access · PDF · HTML · Export citation
View abstract

Word2Vec
KENNETH WARD CHURCH
Published online by Cambridge University Press: 16 December 2016, pp. 155-162
Article · Access · Open access · PDF · HTML · Export citation
View abstract

Challenge: Fluency ≠ Trustworthiness

- A number of articles on ChatGPT
  - lead with amazing successes
    - that seem too good to be true,
  - and end with back-peddling
- ChatGPT has
    - amazing strengths (fluency)
  - as well as
    - amazing weaknesses (trustworthiness)

- Many people assume
  - Fluency ≈ Intelligence
  - IQ testing:
    - Measure vocabulary size
    - Large vocabulary → Fluent
- Fluency is particularly important
  - on first impression
- But eventually → disappointment
  - Weaknesses will become clear
- What is the difference between
  - a hallucination and a con?

What should we do next?

- Three paths forward:
  - Low road:
    - Give up (hallucinations)
  - Middle road:
    - Fact-checking with search
  - High road:
    - Revive rationalism (“AI Complete”)
    - Minsky & Chomsky

- Recommendations
  - Short-term:
    - Middle Road: Search
    - “Good apps for Crummy MT”
      - Find apps for what we have
      - given strengths and weaknesses
  - Long-term:
    - High road may be necessary
      - But it is very ambitious
    - Inclusiveness:
      - Interdisciplinary Collaboration
      - Growth opportunities
        - (Low Resource Languages)

Hypothesis:
NLP History → Strengths & Weaknesses

- Deep Nets are
  - more fluent
  - than trustworthy
- Pendulum Swung Too Far (2011)
  - Empiricism (1950s)
  - Rationalism (1970s)
  - Empiricism (1990s)
  - Deep Nets (2010s)

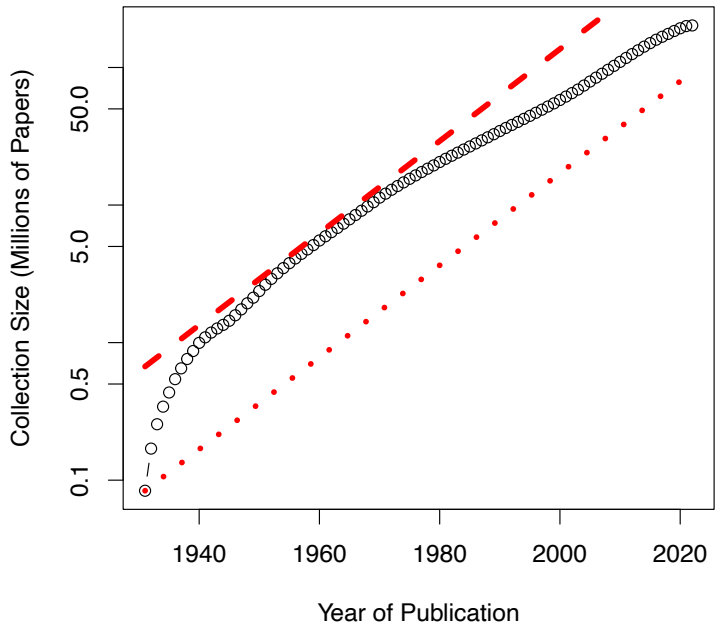
Truth

Fluency

Massive Growth → Mistaken Impression that everything is new (and there is no history)

- What's new
  - The world is taking notice (in AI)
  - Fluency is much improved
- What's not new
  - Chatbots (and much of the tech)
  - SOTA-Chasing
  - New/better shiny objects
    - (with same old quality?)
    - Does SOTA-Chasing → Progress?
  - Trustworthiness is still wide open

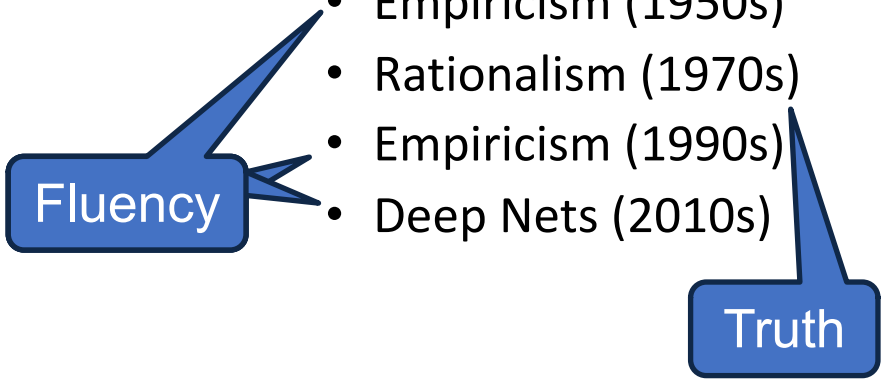
Scientific Literature doubles every 9 years (90% written since I started PhD)



Personal History

- Strengths (fluency) and weaknesses (trustworthiness)
  - may be a consequence of choices we made in 1990s
- We started EMNLP in 1990s for pragmatic reasons
  - Field had been attempting to do too much
    - and was accomplishing too little
    - (during a funding winter)
- We chose to stop working on hard problems
  - (trustworthiness)
  - in order to make relatively quick progress on fluency
  - by reviving empirical methods from the 1950s
    - (Shannon, Skinner, Firth)

- Deep Nets are
  - more fluent
  - than trustworthy
- Pendulum Swung Too Far (2011)
  - Empiricism (1950s)
  - Rationalism (1970s)
  - Empiricism (1990s)
  - Deep Nets (2010s)



Pendulum Swung Too Far

ChatGPT's strengths (fluency) and weaknesses (trustworthiness) may be a consequence of choices we made in 1990s

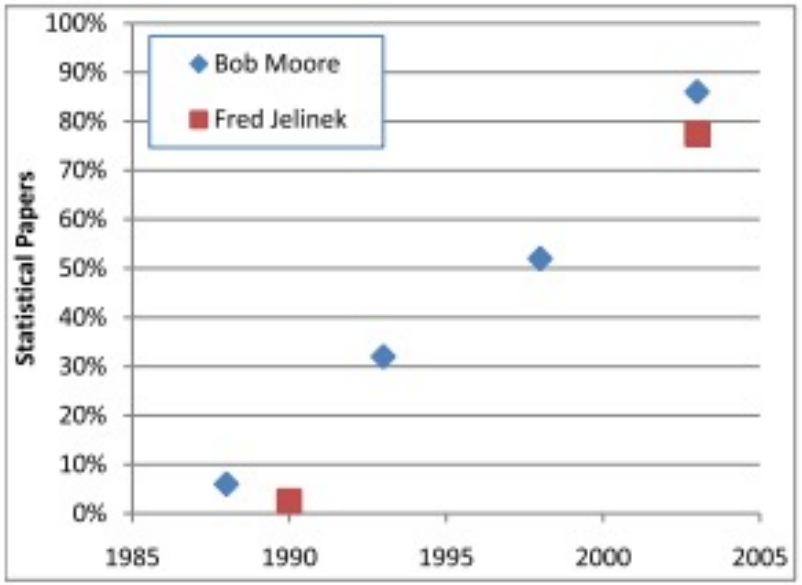


FIGURE 1 The shift from Rationalism to Empiricism is striking (and no longer controversial). This plot is based on two independent surveys of ACL meetings by Bob Moore and Fred Jelinek (personal communication).

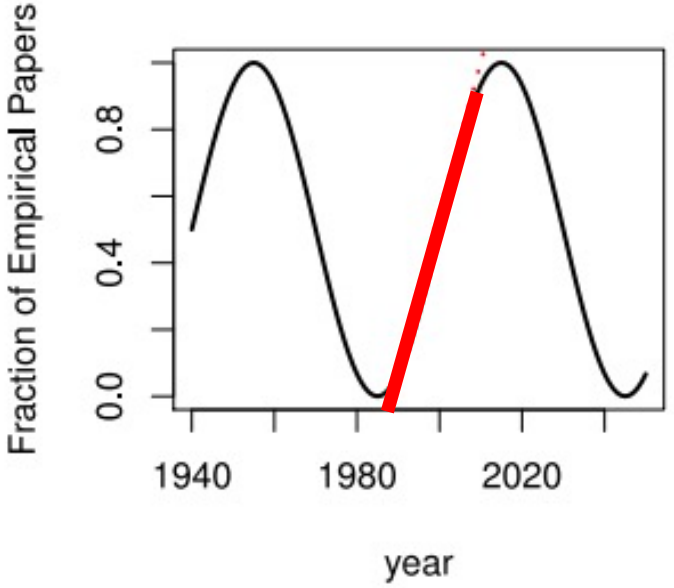


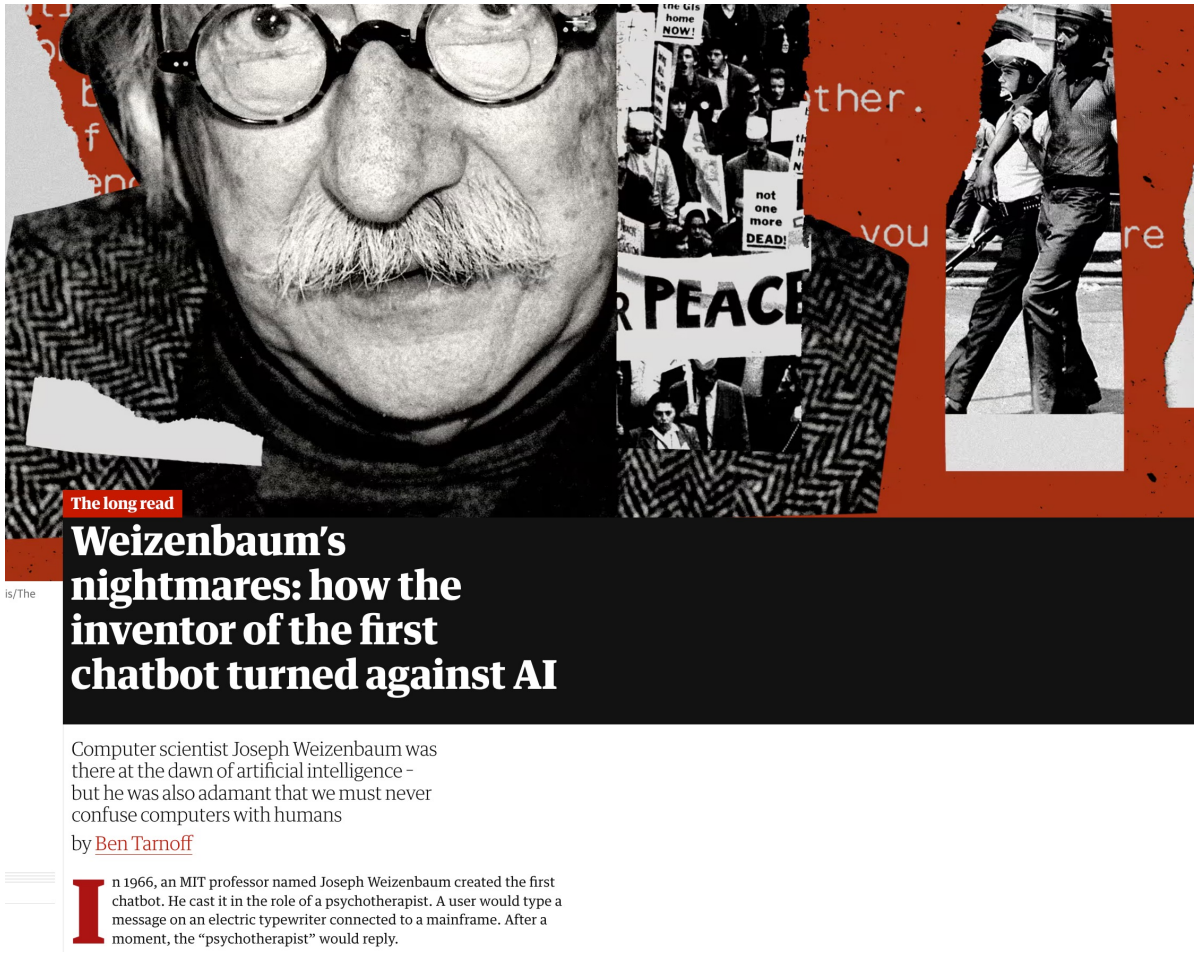
FIGURE 2 An extreme view of the literature, where the trend in Figure 1 (denoted by a dashed red line) is dominated by the larger oscillation every couple of decades. Note that that line is fit to empirical data, unlike the oscillation which is drawn to make a point.

- 1950s: Empiricism (Shannon, Skinner, Firth, Harris)
- 1970s: Rationalism (Chomsky, Minsky)
- 1990s: Empiricism (IBM Speech Group, AT&T Bell Labs)

Mechanism

We started EMNLP

History does not repeat itself (but it rhymes)



- Recent article in Guardian
  - Compared ChatGPT and Weizenbaum's ELIZA
  - https://www.theguardian.com/technology/2023/jul/25/joseph-weizenbaum-inventor-eliza-chatbot-turned-against-artificial-intelligence-ai
- As a TA for Weizenbaum (1978),
  - I know just how horrified he was
  - by how seriously people took ELIZA
- Responsible AI
- His views were not popular at MIT
  - at the time

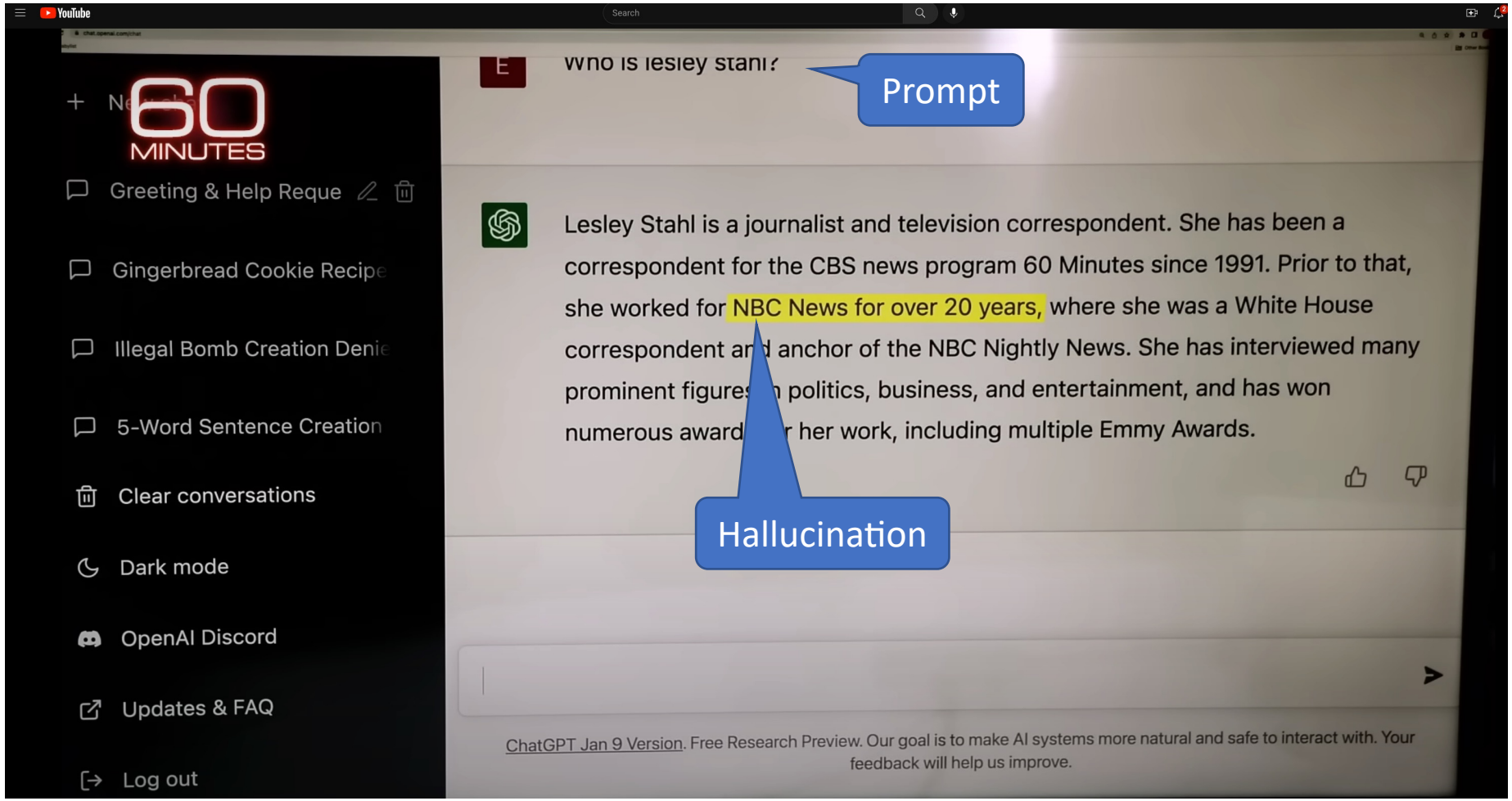
How Hard are Hallucinations? Yogi Berra

- Possible answers:
  - Soon: “See next release”
  - Eventually, but not soon:
    - “Next Year in Jerusalem”
    - History of Machine Translation (MT)
      - When will MT be practical?
      - Prediction from 1950s: 5 years
  - Never

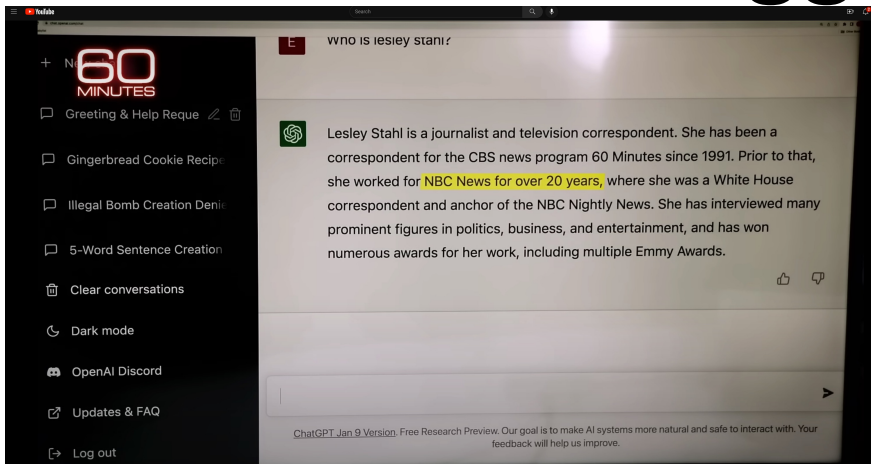
- Assuming ChatGPT’s strengths & weaknesses
  - are a consequence of our choices from 1990s
  - and it took three decades to do well on fluency
  - and fluency ≪ trustworthiness
- then not soon

ChatGPT Hallucinates on CBS ``60 Minutes''

https://www.youtube.com/watch?v=1wzPr4cUoMQ&t=463s



Constructive Suggestions for Hallucinations



1. Low Road:
   - Give up
2. Middle Road
   - Fact-checking with search
3. High Road
   - Revive Rationalism

Query: *Lesley Stahl works for which company*

Challenges for Fact-Checking

- Which claims need to be checked?
- How do we create queries?
- When we get search results, then what?

Acronyms: A Simple Case for Fact-Checking

- Acronyms are easier
  - Google Translate is better on long forms (LFs) than short forms (SFs)
    - Use Google to translate LFs to target language (English)
    - Generate candidate SFs in target
    - Use search to verify candidates
- Co-author (Richard Yue)
  - used to be a professional translator
- Metrics matter:
  - Terminology:
    - important to translators
    - (But less so for BLEU)

- *De nombreux facteurs de risque participent au développement de cette pathologie, parmi lesquels les acides gras trans (AGT).*
- *... une diminution d'expression de 12 gènes mutés dans l'anémie de Fanconi (AF)*

Table 1. Opportunity for Fact-Checking: Translation of Acronyms

Input French		Output English		
LF	**SF**	**LF**	**SF (gold)**	**SF (Google)**
Acides gras trans	AGT	Trans fatty acids	TFA	TGA
Anémie de Fanconi	AF	Fanconi Anemia	FA	AF

Table 2. Search will find more documents matching the good combinations than the bad combinations

Good Combinations	Bad Combinations
Trans fatty acids (TFA)	Trans fatty acids (TGA)
Fanconi Anemia (FA)	Fanconi Anemia (AF)

Fact-Checking Take-Aways

- Short-term patch for hallucinations
- Acronyms:
  - Easy special case of hallucinations
- Don’t re-invent the wheel
  - Re-use existing tools: Search
- Standard Loss/Metrics
  - Insufficient Penalty for Fatal Errors
    - Terminology
    - Responsible AI

- Agreement:
  - hyp = gold
- Verification:
  - Search finds 2+ examples in publications

Method	Agreement	Verified
Identity Baseline	21.5%	0.06%
Reverse Baseline	28.5%	14.6%
Google Baseline	54.3%	29.2%
Gold Labels	100%	42%
Proposed (Section 2.3)	**62.6%**	**42.8%**

Table 7: Proposed method outperforms baselines.

Simple Example: Arithmetic
(Chomsky: Capturing Relevant Generalizations)

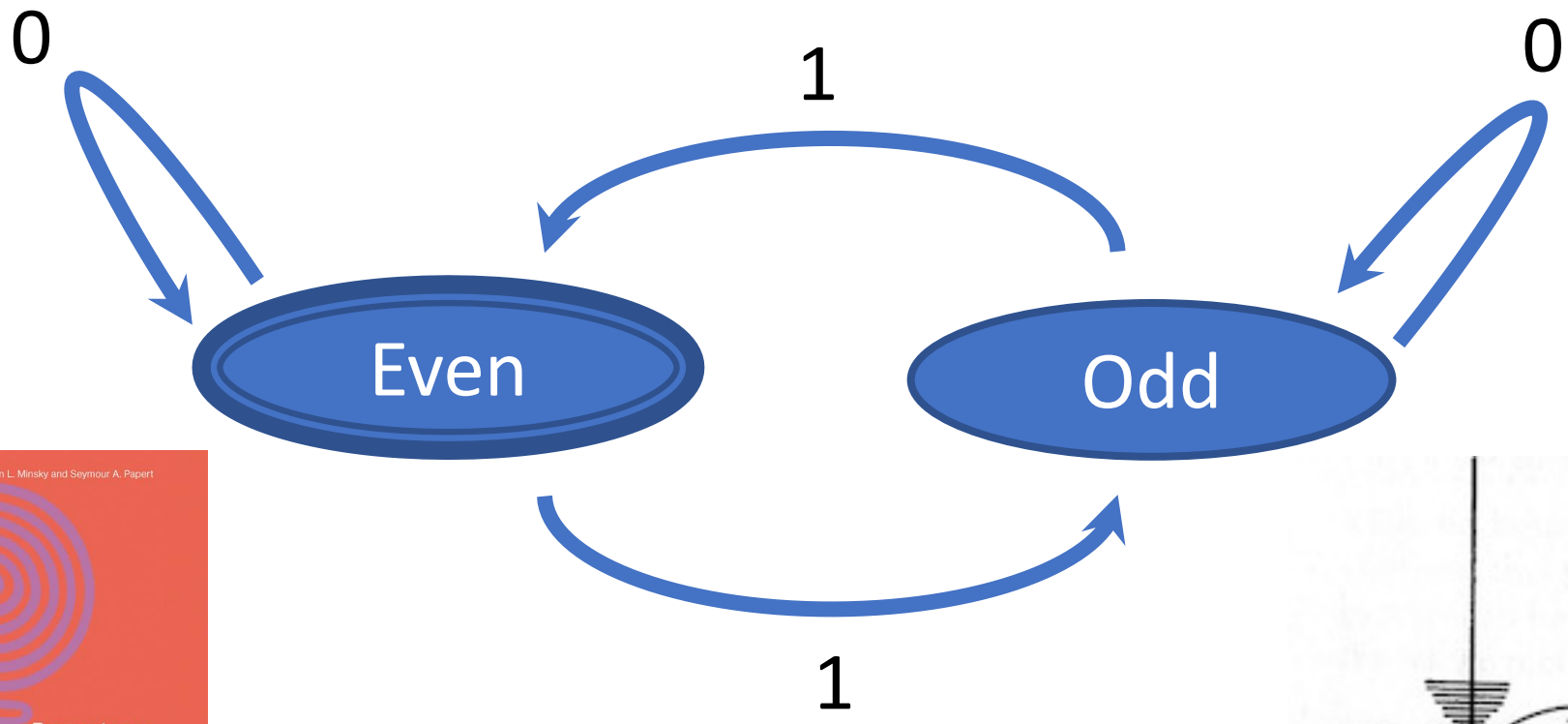
Hypothetical

- Suppose ChatGPT
  - can add two small numbers
- But for large numbers,
  - it makes up answers
- After each release,
  - “small” gets “bigger”

Is this progress?

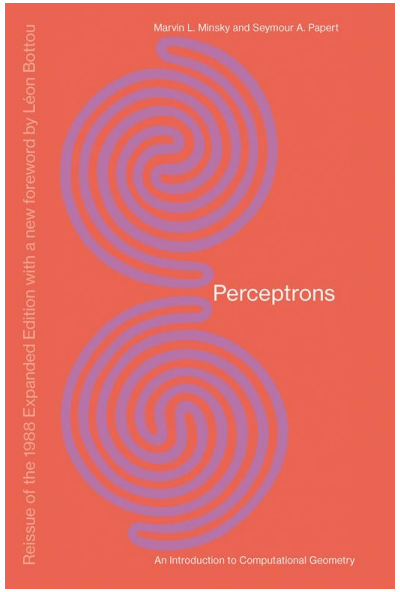
- Empiricism
  - According to SOTA-Chasing,
    - Yes
- Rationalism
  - According to Minsky & Chomsky,
    - No
    - ChatGPT is not mastering concepts
    - “Stochastic Parrots”

Perceptron: "Can't Compute Parity" (Capturing Relevant Generalizations)

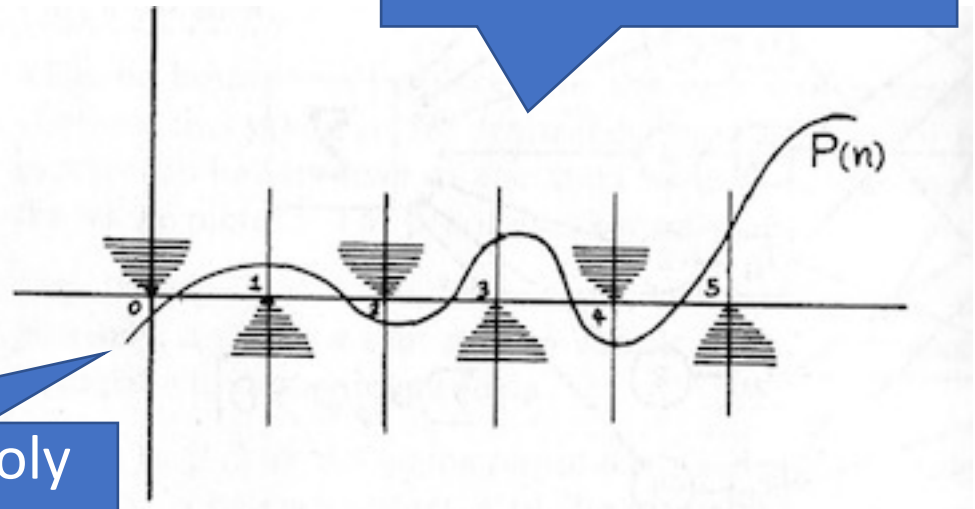


All Cases in $O(1)$ Space

Some Cases in $\Omega(n)$ Space

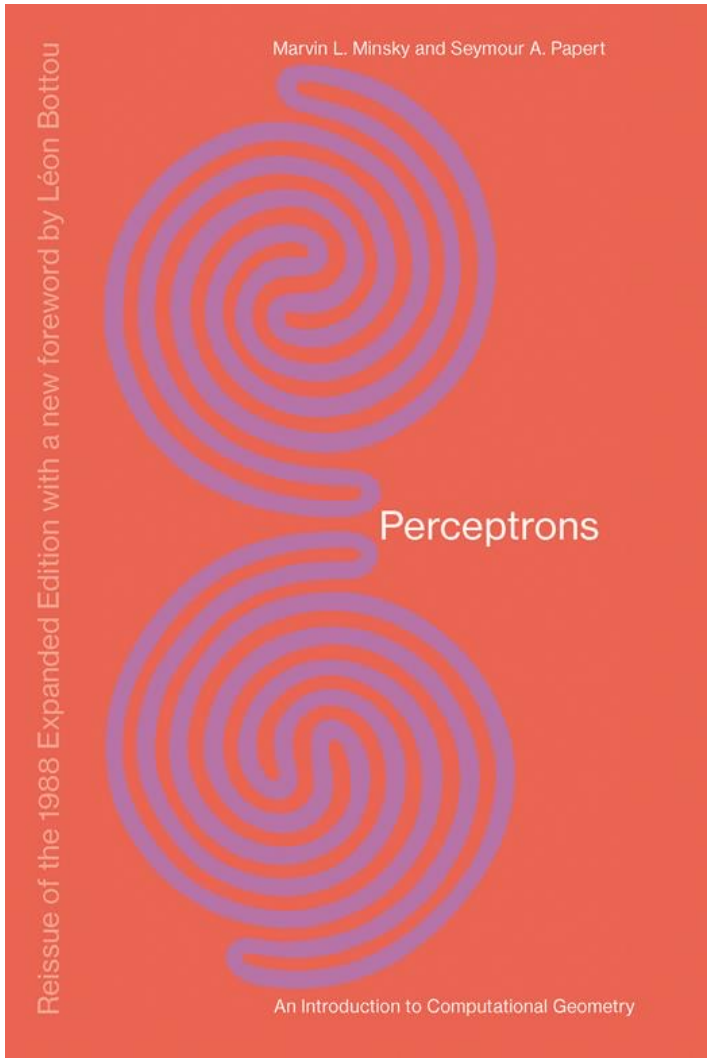


Minsky rejects deep nets (1969)



Regression: n^{th} order poly requires n parameters

Minsky rejects deep nets (1969)



N. Chomsky, "Three models for the description of language," in *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 113-124, September 1956, doi: 10.1109/TIT.1956.1056813

Whatever the other interest of statistical approximation in this sense may be, it is clear that it can shed no light on the problems of grammar. There is no general relation between the frequency of a string (or its component parts) and its grammaticalness. We can see this most clearly by considering such strings as

(14) colorless green ideas sleep furiously

Introduced Chomsky Hierarchy:
Finite State → Turing Machines

Capturing Generalizations Argument

Pros

- Long-Term Focus
  - Can’t get to moon by
  - Incremental short-term local optimization

Cons

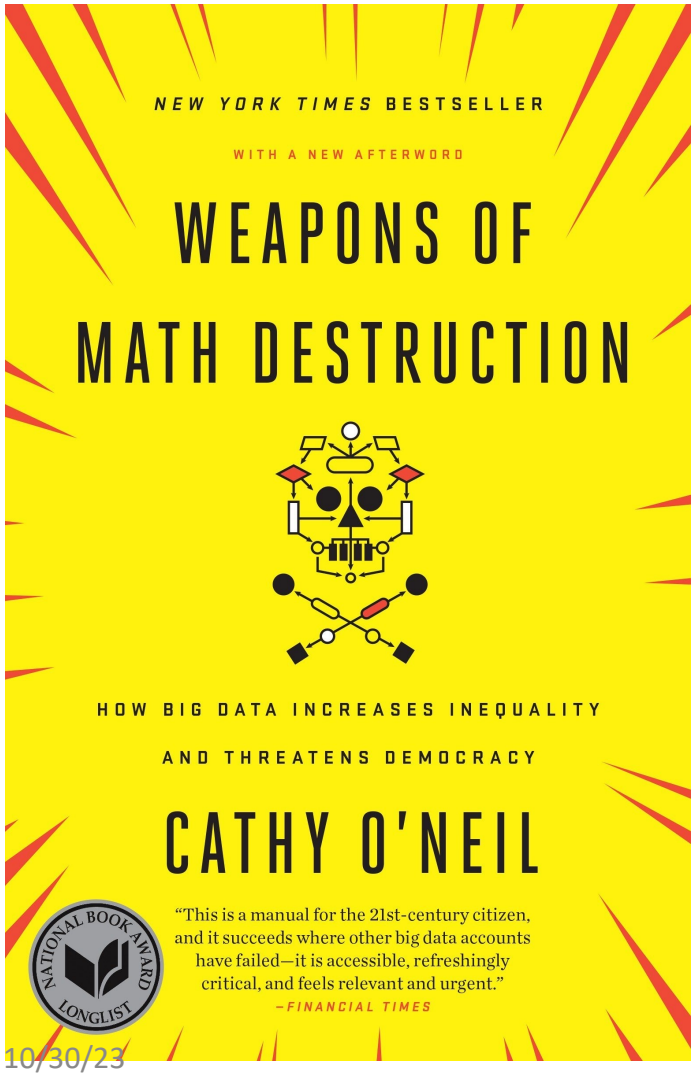
- Dismisses Short-term Progress
  - Sometimes it is useful
  - to solve some simple special cases
    - (addition of small numbers)
- Not constructive:
  - “MIT School of Negativity”

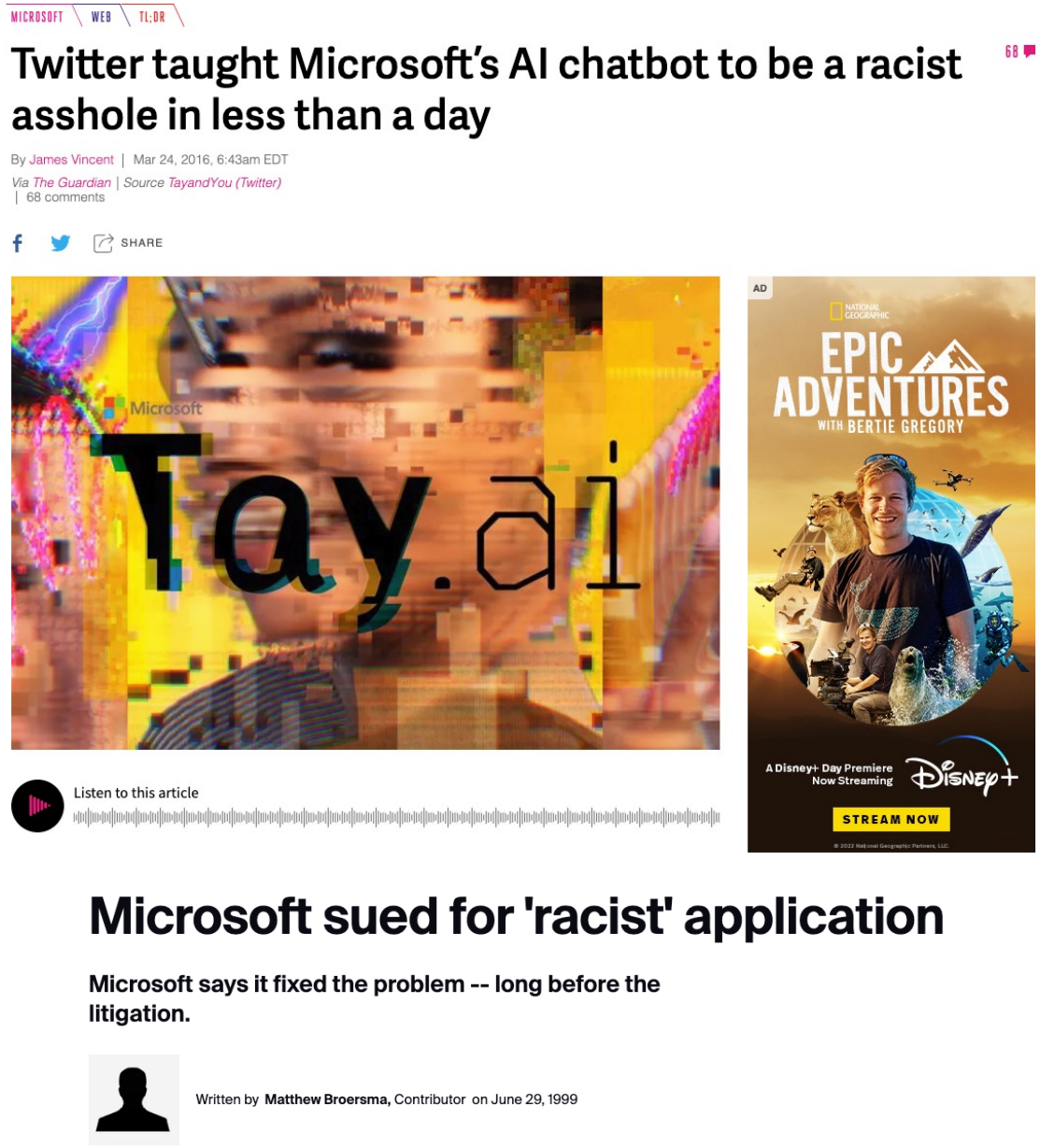
The Easy, the Hard and the Ugly

- ✓Easy: Exciting Eco-system
  - ✓Prompting
  - ✓Inference (*fit*)
  - ✓Fine-Tuning (*predict*)
- ✓Hard: Large Companies
  - ✓Pre-training
- ➢**Ugly (Responsible AI)**
  - Bias
  - Toxicity
  - Misinformation
  - Hallucinations
  - Plagiarism

History of Irresponsible AI

Risk (5 years ago) Product gets canceled

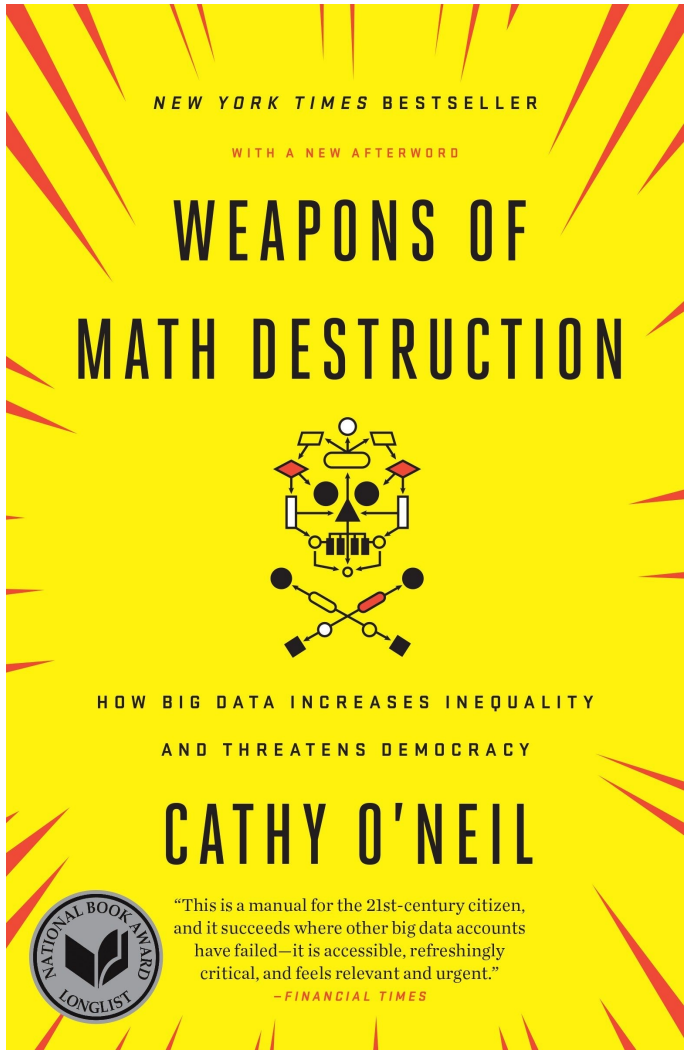




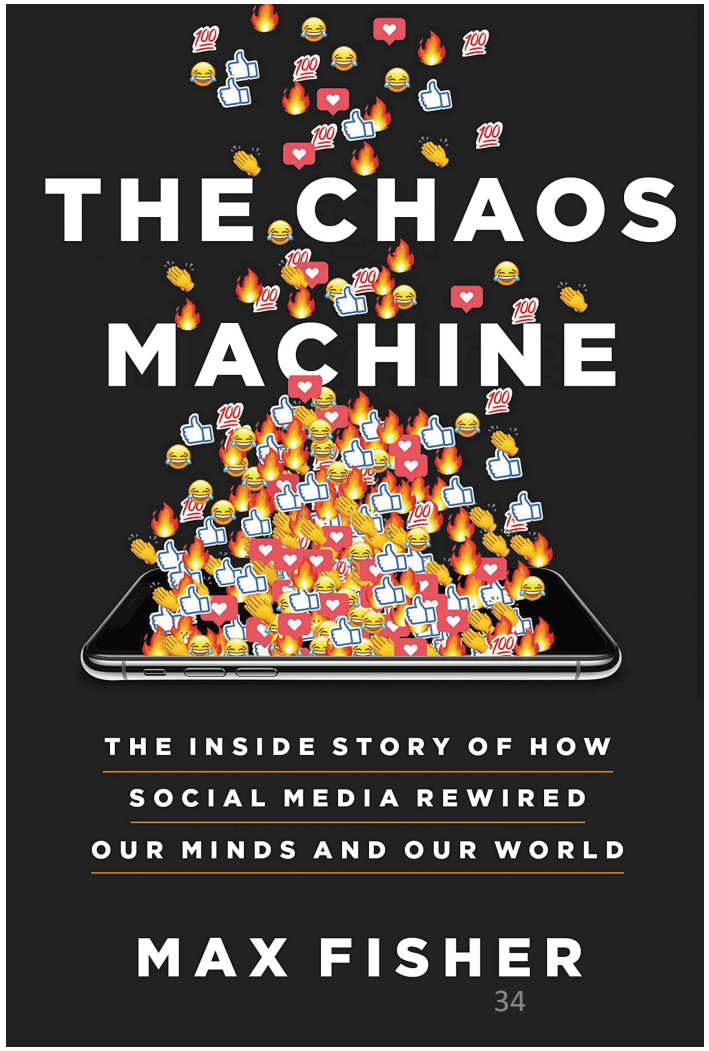
Are we losing ground??? Are we at fault???

- Risks 1.0: (work in progress)
  - Bias, Fairness
- Risks 2.0: (bigger than us)
  - Genocide, Insurrection
  - Root causes:
    - ML + Social Media → Addiction
    - Max Engagement → Dangerous
    - Insanely profitable:
      - Companies & Countries
  - Long book, but no mention of our efforts to address old risks

2016

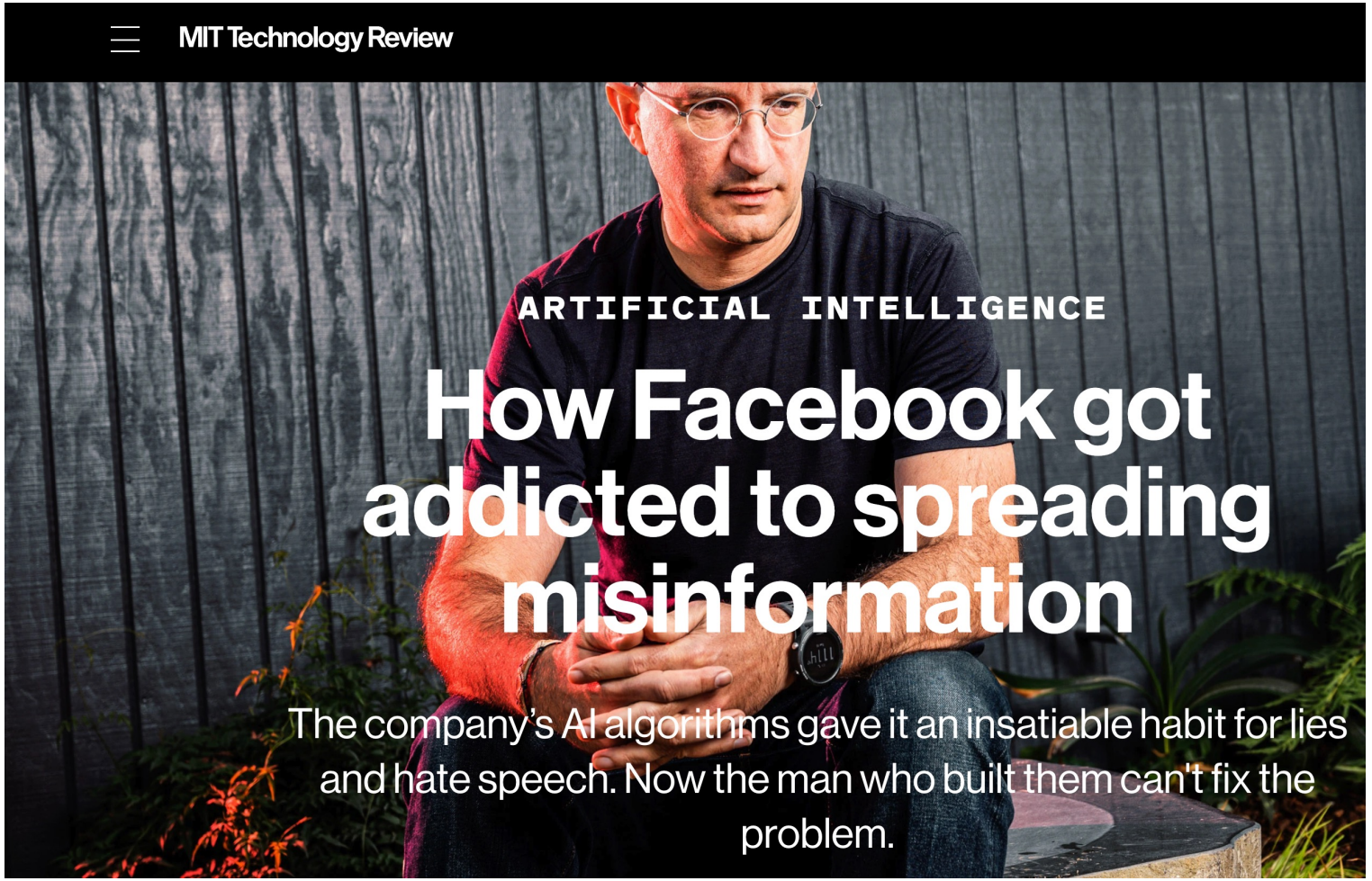


2022



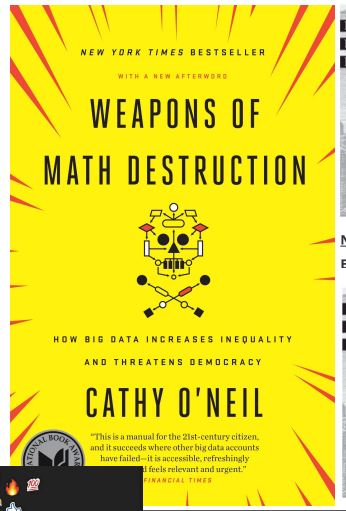
Reporter wanted to talk about Risks 2.0; Accused Facebook of pivoting to Risks 1.0

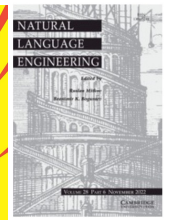
https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/



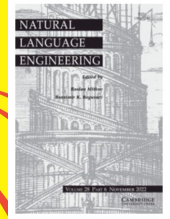
Ugly: Responsible AI

- Incentives matter
  - Risks 1.0 (2016)
    - Unfair, Biased
  - Risks 2.0 (2022)
    - Addictive, dangerous, deadly
    - and insanely profitable
  - Risks 3.0 (2023)
    - Malware
    - Spyware
- Challenge for Regulation
    - Business case ≠ Public Interest (Health, National Security)
  - Tobacco companies maximize sales; ditto for fast food & junk food
  - Risks 2.0 (Toxicity): Good for social media companies; we ❤️ click bait
  - Risks 3.0 (Conflict): Good for defense industry





Natural Language Engineering



Emerging trends: Unfair, biased, addictive, dangerous, deadly, and insanely profitable

Published online by Cambridge University Press: **19 December 2022**

Kenneth Church, Annika Schoene, John E. Ortega, Raman Chandrasekar and Valia Kordoni

Show author details

Article | Figures | Metrics

Save PDF | Share | Cite | Rights & Permissions

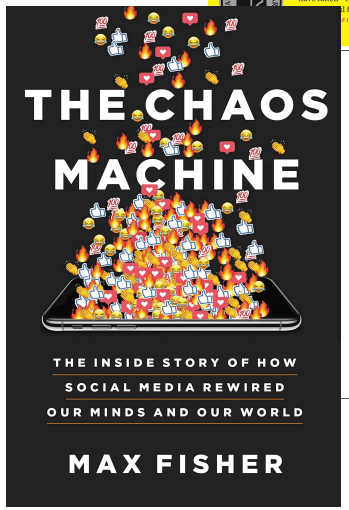
Emerging trends: Risks 3.0 and proliferation of spyware to 50,000 cell phones

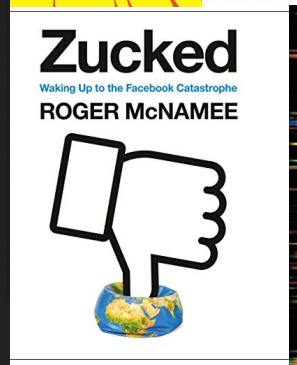
Published online by Cambridge University Press: **19 May 2023**

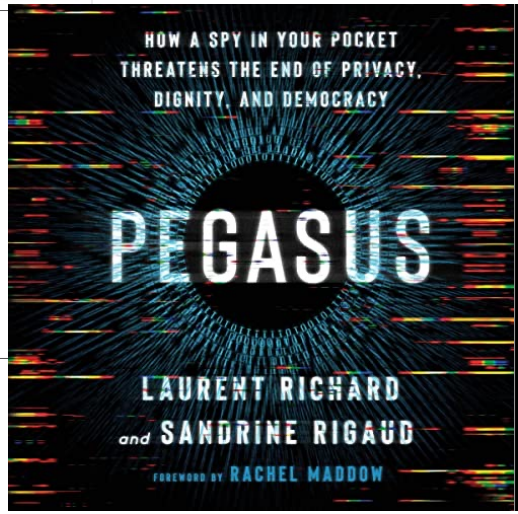
Kenneth Ward Church and Raman Chandrasekar

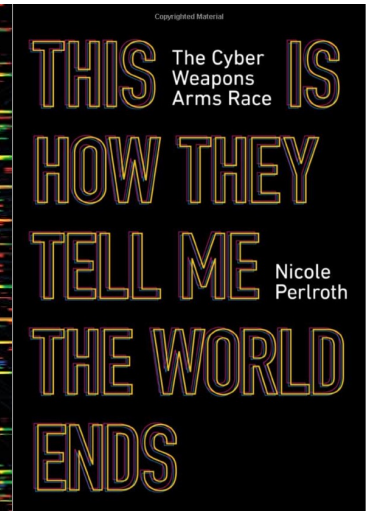
Show author details

Article | Metrics









Winter is Coming

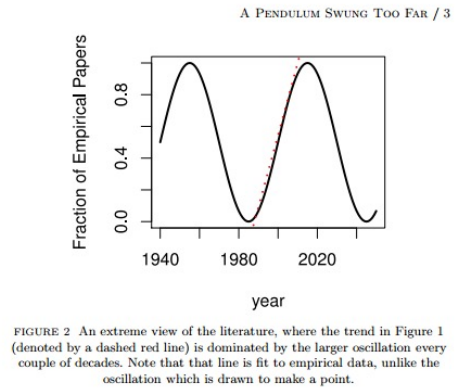
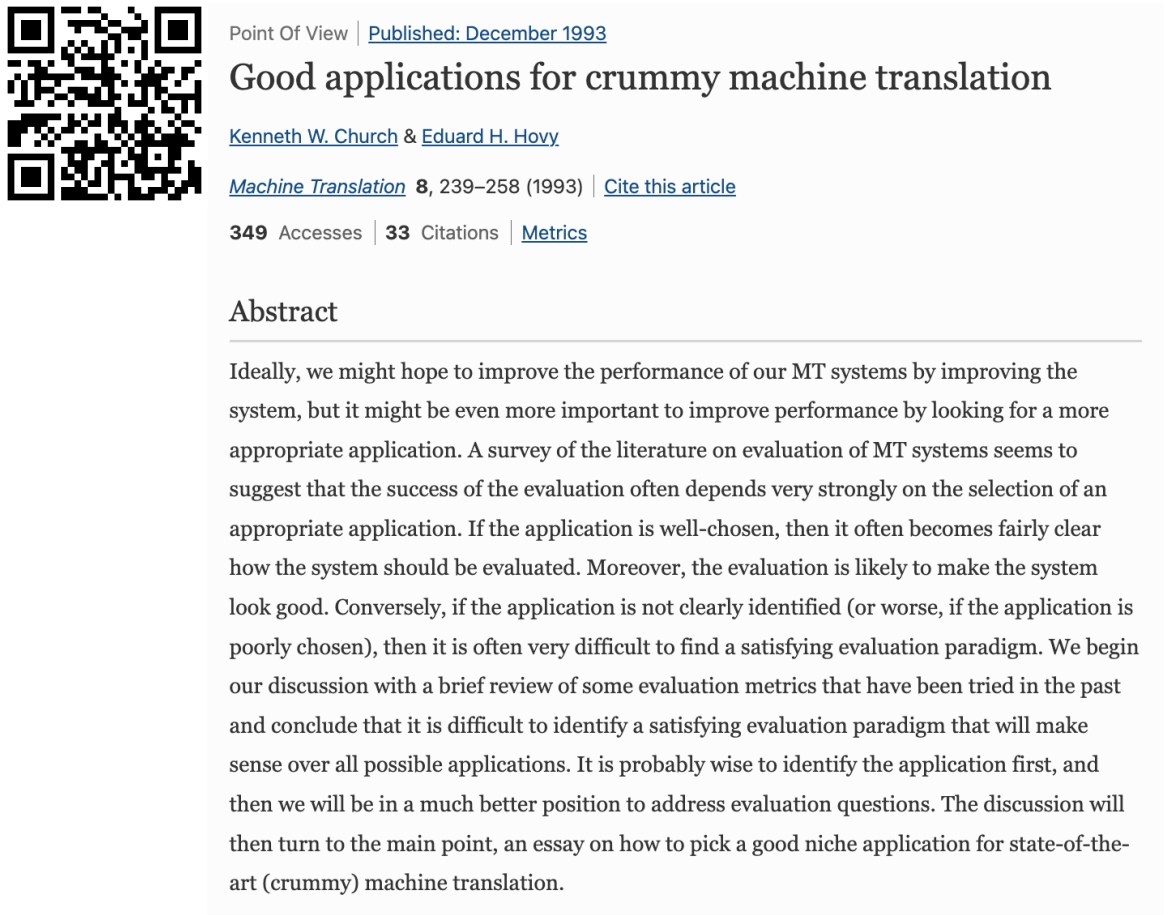


FIGURE 2 An extreme view of the literature, where the trend in Figure 1 (denoted by a dashed red line) is dominated by the larger oscillation every couple of decades. Note that that line is fit to empirical data, unlike the oscillation which is drawn to make a point.

- Pendulum Swung Too Far
  - There have been many AI Winters
  - Often, after ``irrational exuberance''
  - (like current excitement with nets)

- We tend to be impressed by people that speak/write well
  - Fluency → well-read → success → smart
- Machines are better than people on many tasks (spelling),
  - Now that machines are more fluent than people, ***are they smarter***?
- Fear: AI Winter
  - there will be disappointment
  - when public realizes
    - ***fluency ≠intelligence***

Recommendation:
Go for Singles (Not Home Runs)



Point Of View | Published: December 1993

Good applications for crummy machine translation

Kenneth W. Church & Eduard H. Hovy

Machine Translation **8**, 239–258 (1993) | Cite this article

Abstract

Ideally, we might hope to improve the performance of our MT systems by improving the system, but it might be even more important to improve performance by looking for a more appropriate application. A survey of the literature on evaluation of MT systems seems to suggest that the success of the evaluation often depends very strongly on the selection of an appropriate application. If the application is well-chosen, then it often becomes fairly clear how the system should be evaluated. Moreover, the evaluation is likely to make the system look good. Conversely, if the application is not clearly identified (or worse, if the application is poorly chosen), then it is often very difficult to find a satisfying evaluation paradigm. We begin our discussion with a brief review of some evaluation metrics that have been tried in the past and conclude that it is difficult to identify a satisfying evaluation paradigm that will make sense over all possible applications. It is probably wise to identify the application first, and then we will be in a much better position to address evaluation questions. The discussion will then turn to the main point, an essay on how to pick a good niche application for state-of-the-art (crummy) machine translation.

- Need some short-term successes for ChatBots
  - that take advantage of strengths
  - and avoid weaknesses
- Suggestion
  - ***Collaborate*** with students on essays
    - *You have no idea how much we're using ChatGPT*
    - Cheating?
  - ChatGPT is better for some tasks
    - Good: thesis statements, outlines
    - Bad: capture student's voice
    - Worse: quotes
  - Learning opportunity:
    - How to decompose writing to subtasks
    - Collaboration is great,
      - but student is responsible for end-product
    - Factoring example

What should we do next?

- Three paths forward:
  - Low road:
    - Give up (hallucinations)
  - Middle road:
    - Fact-checking with search
  - High road:
    - Revive rationalism (“AI Complete”)
    - Minsky & Chomsky

- Recommendations
  - ➢Short-term:
    - Middle Road: Search
    - “Good apps for Crummy MT”
      - Find apps for what we have
      - given strengths and weaknesses
  - Long-term:
    - High road may be necessary
      - But it is very ambitious
    - Inclusiveness:
      - Interdisciplinary Collaboration
      - Growth opportunities
        - (Low Resource Languages)

Big Picture

(Linguistics Perspective)

- Speech (JM28)
  - Acoustics
  - Digital Signal Processing
  - Phonetics: phonemes, distinctive features
  - Prosody: Pitch, duration, energy
  - Phonology: Stress Assignment
- Morphology
  - Regular Inflection, level 1, level 2
  - Compounding
- The Lexicon (JM23)
- Syntax
  - Parsing
    - Chomsky Hierarchy:
      - Finite-State, Context-Free (JM17)
      - Context-Sensitive, Turing Equivalent
  - Variable Binding
    - Pronouns, Quantifier Scope
    - WH-movement
  - Predicate-argument structure (JM24)
- Semantics
  - Logical Form (JM19)
- Pragmatics, Discourse & Dialogue (JM27)
  - Gricean Maxims
  - Indirect speech acts: *do you have the time?*
  - Diarization (who spoke when)
  - Filled pauses, restarts, corrections

Big Picture

(Computational Perspective)

- Speech (JM16)
  - Digital Signal Processing
  - Speech to Text (STT; ASR)
  - Text to Speech (TTS; synthesis)
  - Diarization
- Morphology
  - Tokenization (Subwords) (JM2)
- The Lexicon (JM23)
  - Spelling Correction
    - (JM appendix B)
- Syntax
  - Parsing (JM17 & JM18)
  - Token classification (JM8)
    - Part of speech tagging
    - NER (named entity recognition)
  - Coreference (JM26)
- Semantics
  - Logical Form (JM19)
  - Temporal Reasoning (JM22)
- Pragmatics, Discourse & Dialogue
  - Chatbots (JM15)
  - Question Answering & Information Retrieval (JM14)

Topics not well covered

- Conjunction
  - Scope: narrow vs. wide
  - Ellipsis, gapping, etc.
- Long-distance dependencies
  - WH-movement
- Belief contexts, quotations, etc.
- Historical Linguistics
- Sociolinguistics
- Psycholinguistics:
  - reaction time
  - memory limitation
- Topology: SVO, SOV, etc.

Phonetics

- Distinctive Features
  - Voicing:
    - Unvoiced, Voiced
  - Place:
    - Labial, Dental, Velar
  - Manner:
    - Stop, Fricative, Vowel

Place	Unvoiced	Voiced
Labial	p	b
Dental	t	d
Velar	k	g

ARPAbet Symbol	IPA Symbol	Word	ARPAbet Transcription	ARPAbet Symbol	IPA Symbol	Word	ARPAbet Transcription
[p]	[p]	parsley	[p aa r s l iy]	[iy]	[i]	lily	[l ih l iy]
[t]	[t]	tea	[t iy]	[ih]	[ɪ]	lily	[l ih l iy]
[k]	[k]	cook	[k uh k]	[ey]	[eɪ]	daisy	[d ey z iy]
[b]	[b]	bay	[b ey]	[eh]	[ɛ]	pen	[p eh n]
[d]	[d]	dill	[d ih l]	[ae]	[æ]	aster	[ae s t axr]
[g]	[g]	garlic	[g aa r l ix k]	[aa]	[ɑ]	poppy	[p aa p iy]
[m]	[m]	mint	[m ih n t]	[ao]	[ɔ]	orchid	[ao r k ix d]
[n]	[n]	nutmeg	[n ah t m eh g]	[uh]	[ʊ]	wood	[w uh d]
[ng]	[ŋ]	baking	[b ey k ix ng]	[ow]	[oʊ]	lotus	[l ow dx ax s]
[f]	[f]	flour	[f l aw axr]	[uw]	[u]	tulip	[t uw l ix p]
[v]	[v]	clove	[k l ow v]	[ah]	[ʌ]	butter	[b ah dx axr]
[th]	[θ]	thick	[th ih k]	[er]	[ɝ]	bird	[b er d]
[dh]	[ð]	those	[dh ow z]	[ay]	[aɪ]	iris	[ay r ix s]
[s]	[s]	soup	[s uw p]	[aw]	[aʊ]	flower	[f l aw axr]
[z]	[z]	eggs	[eh g z]	[oy]	[oɪ]	soil	[s oy l]
[sh]	[ʃ]	squash	[s k w aa sh]	[ax]	[ə]	pita	[p iy t ax]
[zh]	[ʒ]	ambrosia	[ae m b r ow zh ax]				
[ch]	[tʃ]	cherry	[ch eh r iy]				
[jh]	[dʒ]	jar	[jh aa r]				
[l]	[l]	licorice	[l ih k axr ix sh]				
[w]	[w]	kiwi	[k iy w iy]				
[r]	[r]	rice	[r ay s]				
[y]	[j]	yellow	[y eh l ow]				
[h]	[h]	honey	[h ah n iy]				

Figure 28.1 ARPAbet and IPA symbols for English consonants (left) and vowels (right).

Phonetics

- Distinctive Features
  - Voicing:
    - Unvoiced, Voiced
  - Place:
    - Labial, Dental, Velar
  - Manner:
    - Stop, Fricative, Vowel

Place	Unvoiced	Voiced
Labial	p	b
Dental	t	d
Velar	k	g

Consonants: Place of Articulation

Because consonants are made by restricting airflow, we can group them into classes by their point of maximum restriction, their **place of articulation** (Fig. 28.3).

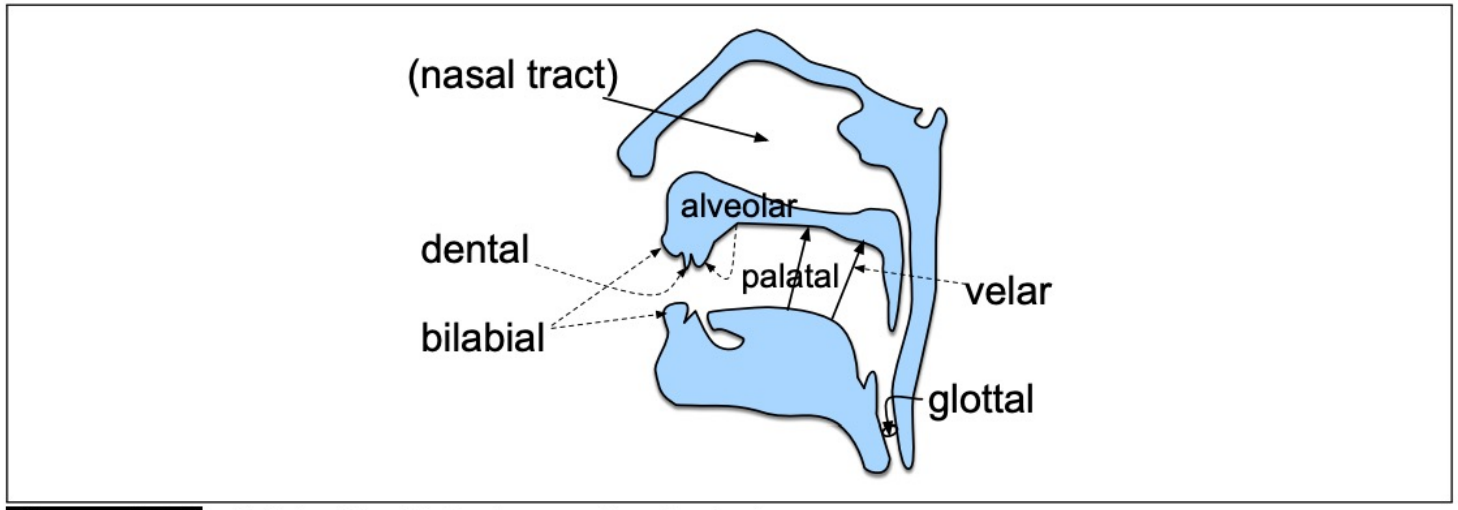


Figure 28.3 Major English places of articulation.

Distinctive Features: Vowels

	High	Low
Front	beet	bat
Back	boot	bought

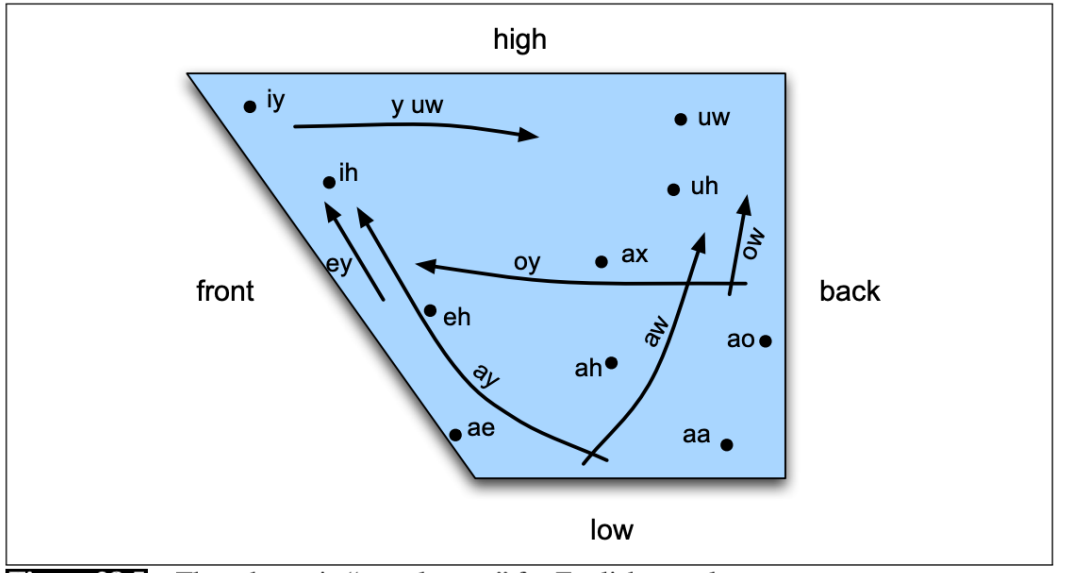


Figure 28.5 The schematic "vowel space" for English vowels.

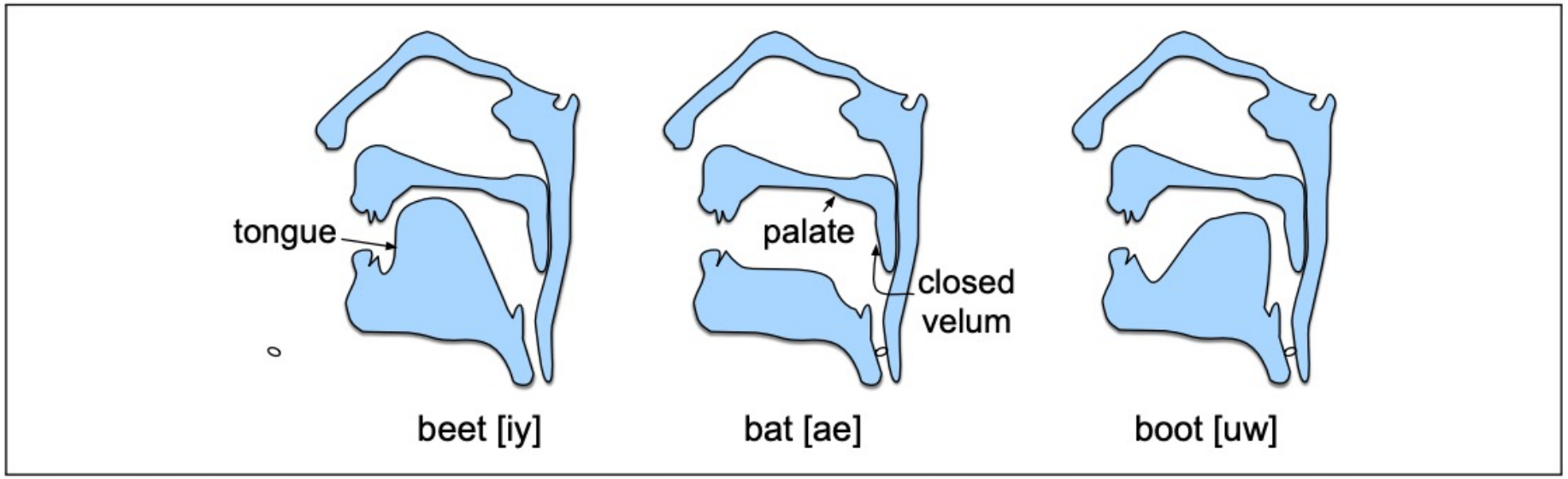


Figure 28.4 Tongue positions for English high front [iy], low front [ae] and high back [uw].

Distinctive Features: Vowels

	High	Low
Front	beet	bat
Back	boot	bought

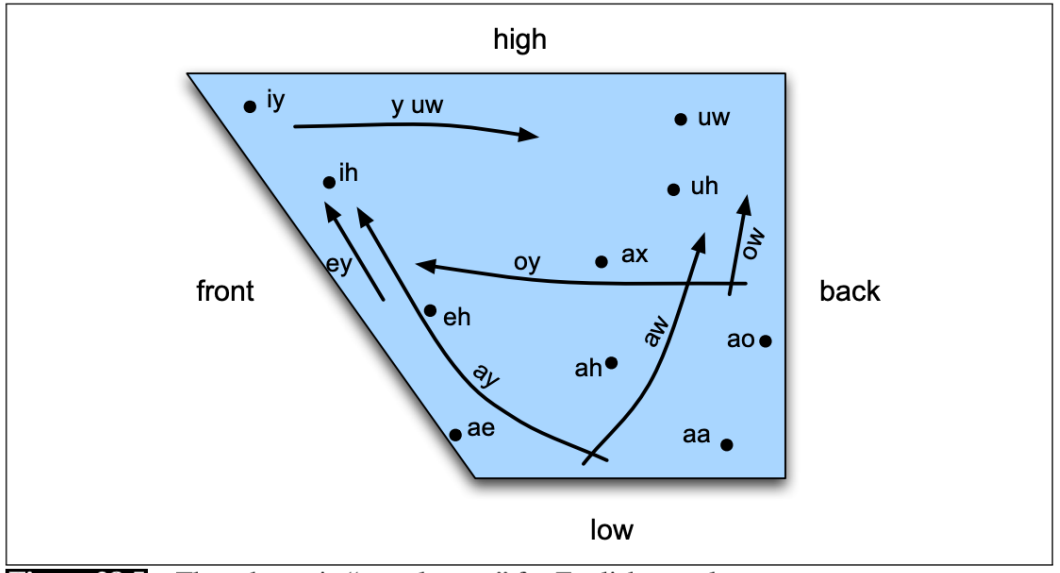


Figure 28.5 The schematic "vowel space" for English vowels.

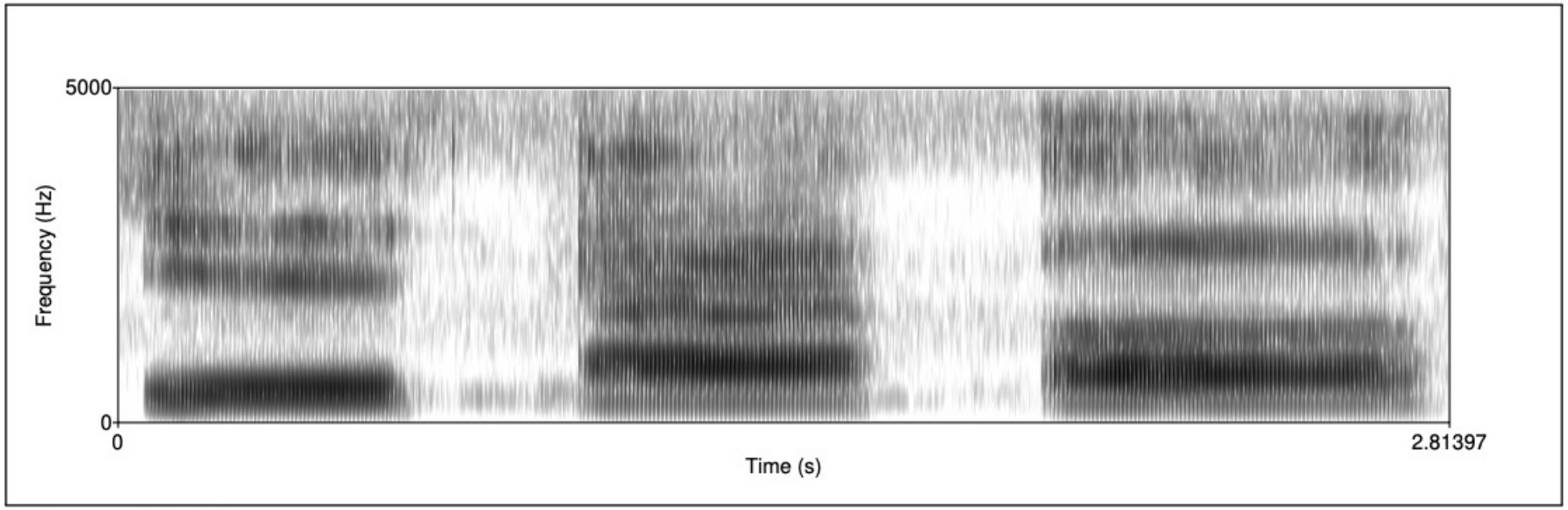


Figure 28.19 Spectrograms for three American English vowels, [ih], [ae], and [uh]

Distinctive Features: Vowels

	High	Low
Front	beet	bat
Back	boot	bought

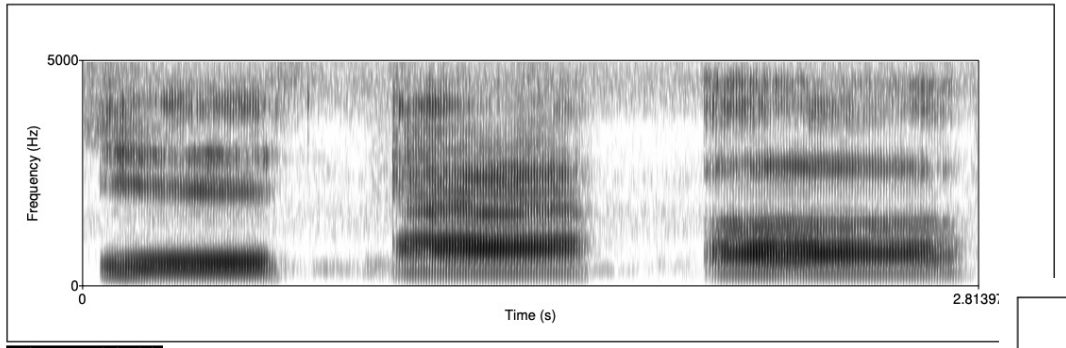


Figure 28.19 Spectrograms for three American English vowels, [ih], [ae], and [uh]

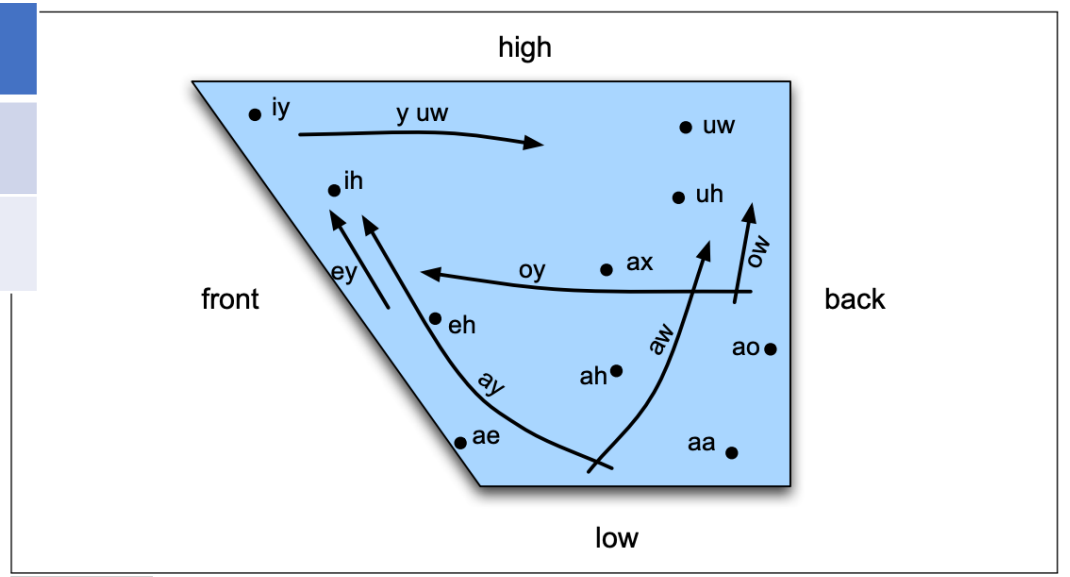


Figure 28.5 The schematic "vowel space" for English vowels.

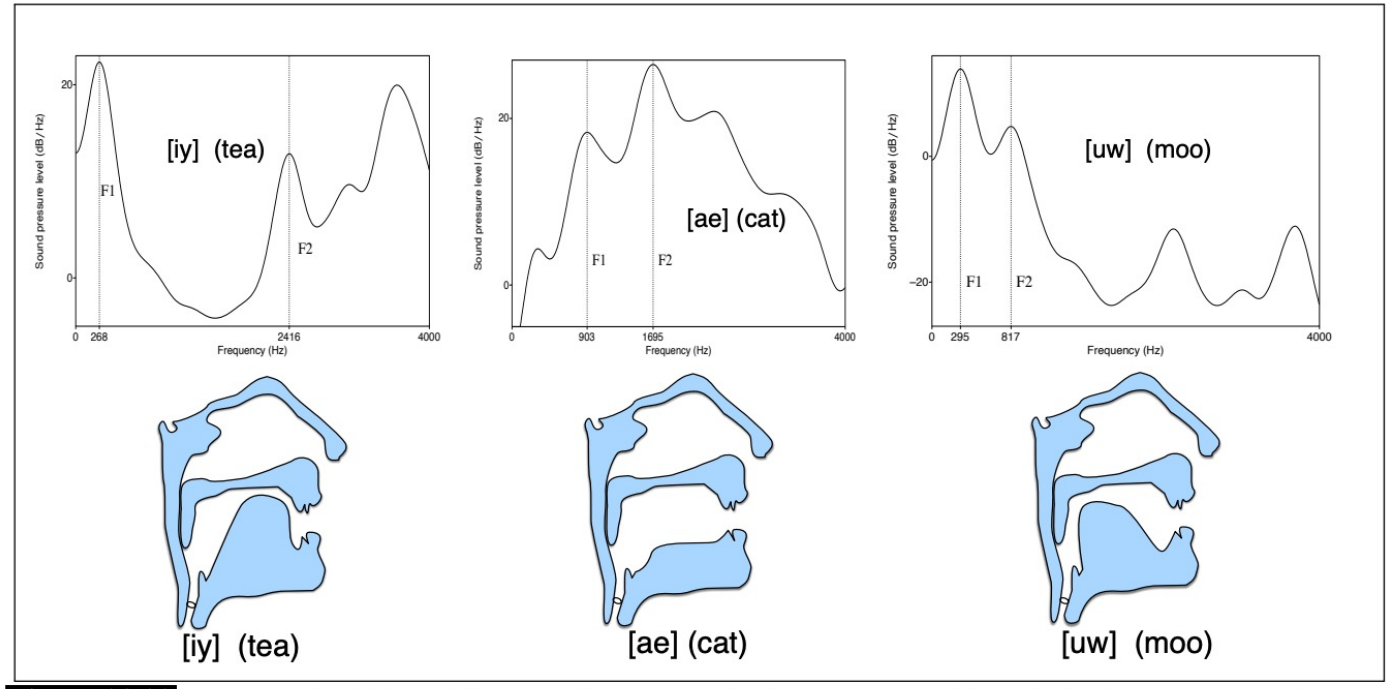


Figure 28.22 Visualizing the vocal tract position as a filter: the tongue positions for three English vowels and the resulting smoothed spectra showing F1 and F2.

Multiple Perspectives:

F1 & F2 are poles of 2nd order differential equation

Perception

Production

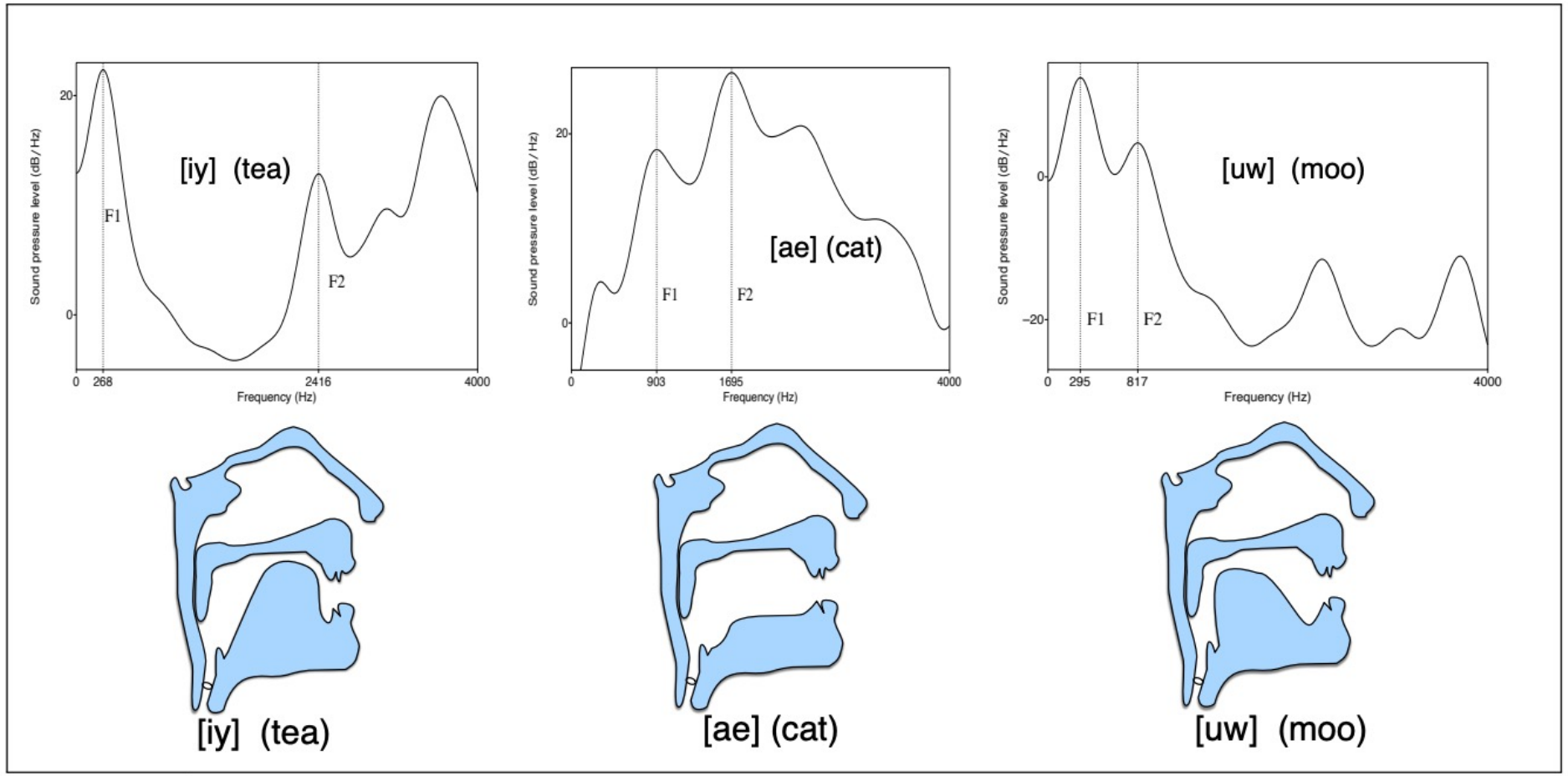
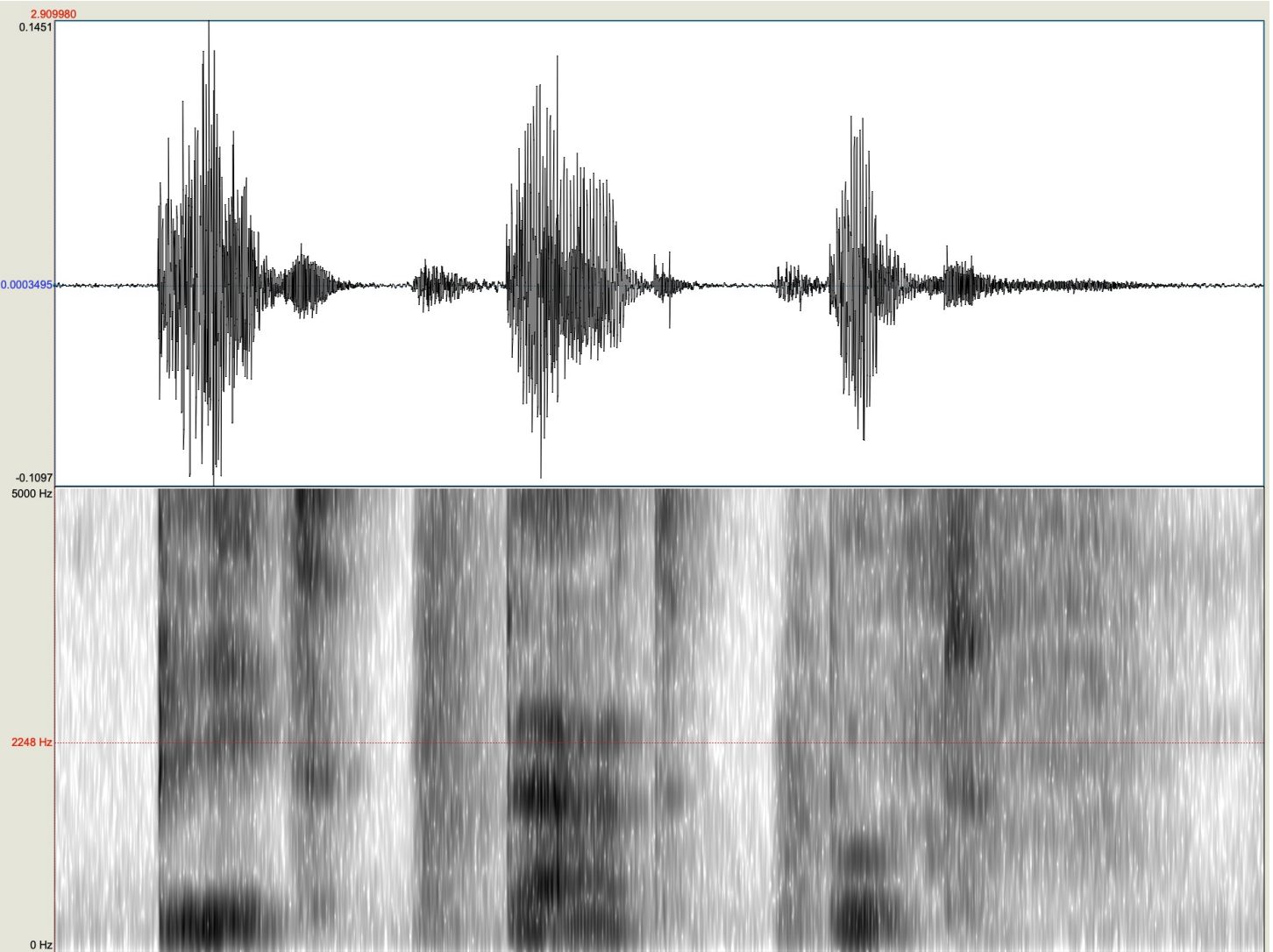


Figure 28.22 Visualizing the vocal tract position as a filter: the tongue positions for three English vowels and the resulting smoothed spectra showing F1 and F2.

Multiple Perspectives:

F1 & F2 are poles of 2nd order differential equation



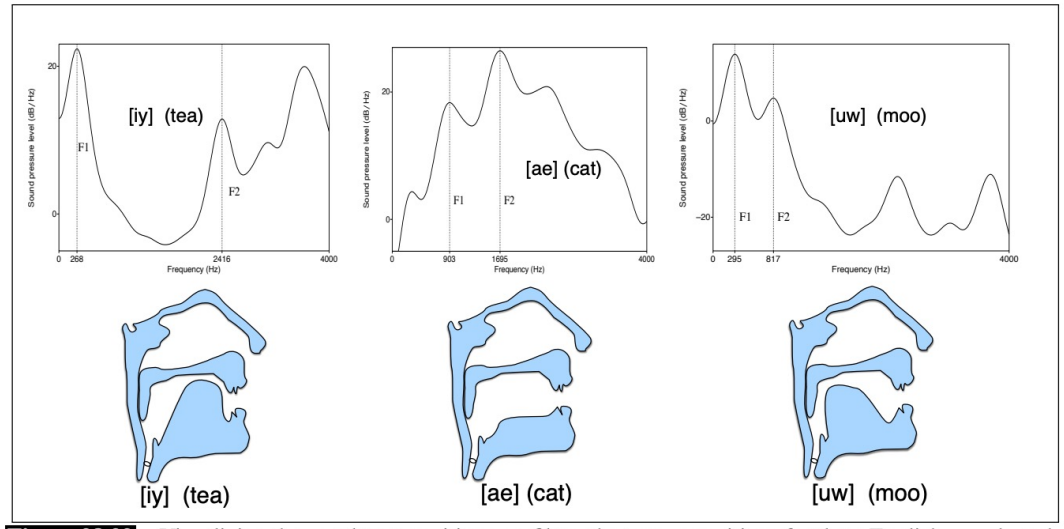


Figure 28.22 Visualizing the vocal tract position as a filter: the tongue positions for three English vowels and the resulting smoothed spectra showing F1 and F2.

Big Picture

(Computational Perspective)

- Speech (JM16)
  - Digital Signal Processing
  - Speech to Text (STT; ASR)
  - Text to Speech (TTS; synthesis)
  - Diarization
- Morphology
  - Tokenization (Subwords) (JM2)

➢The Lexicon (JM23)

- Spelling Correction
  - (JM appendix B)

- Syntax
  - Parsing (JM17 & JM18)
  - Token classification (JM8)
    - Part of speech tagging
    - NER (named entity recognition)
  - Coreference (JM26)
- Semantics
  - Logical Form (JM19)
  - Temporal Reasoning (JM22)
- Pragmatics, Discourse & Dialogue
  - Chatbots (JM15)
  - Question Answering & Information Retrieval (JM14)

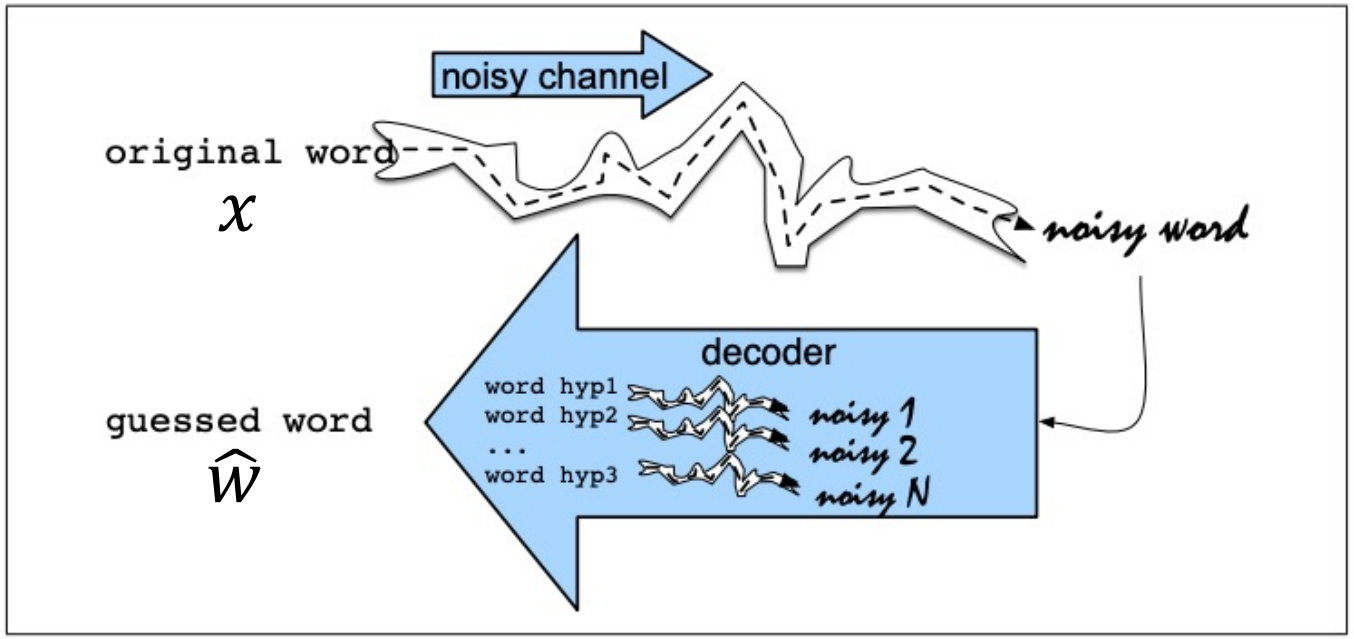
Word Sense Disambiguation

- bank → money bank | river bank
- translations
  - money bank → banque
  - river bank → banc
- classification task
  - predict when *bank* will be translated as *banque* or *banc*

Spelling Correction

B.1 The Noisy Channel Model

In this section we introduce the noisy channel model and show how to apply it to the task of detecting and correcting spelling errors. The noisy channel model was applied to the spelling correction task at about the same time by researchers at AT&T Bell Laboratories (Kernighan et al. 1990, Church and Gale 1991) and IBM Watson Research (Mays et al., 1991).



$$\hat{w} = \underset{w \in C}{\operatorname{argmax}} \overbrace{P(x|w)}^{\text{channel model}} \overbrace{P(w)}^{\text{prior}} \tag{B.5}$$

Candidate Corrections

		Transformation			
Error	Correction	Correct Letter	Error Letter	Position (Letter #)	Type
acress	actress	t	—	2	deletion
acress	cress	—	a	0	insertion
acress	caress	ca	ac	0	transposition
acress	access	c	r	2	substitution
acress	across	o	e	3	substitution
acress	acres	—	s	5	insertion
acress	acres	—	s	4	insertion

Figure B.3 Candidate corrections for the misspelling *acress* and the transformations that would have produced the error (after Kernighan et al. (1990)). "—" represents a null letter.

Prior

w	count(w)	p(w)
actress	9,321	.0000231
cress	220	.000000544
caress	686	.00000170
access	37,038	.0000916
across	120,844	.000299
acres	12,874	.0000318

Channel Model

$$P(x|w) = \begin{cases} \frac{\text{del}[x_{i-1}, w_i]}{\text{count}[x_{i-1} w_i]}, & \text{if deletion} \\ \frac{\text{ins}[x_{i-1}, w_i]}{\text{count}[w_{i-1}]}, & \text{if insertion} \\ \frac{\text{sub}[x_i, w_i]}{\text{count}[w_i]}, & \text{if substitution} \\ \frac{\text{trans}[w_i, w_{i+1}]}{\text{count}[w_i w_{i+1}]}, & \text{if transposition} \end{cases}$$

Candidate Correction	Correct Letter	Error Letter	x|w	P(x|w)	P(w)	10^9*P(x|w)P(w)
actress	t	-	c|ct	.000117	.0000231	2.7
cress	-	a	a|#	.00000144	.000000544	0.00078
caress	ca	ac	ac|ca	.00000164	.00000170	0.0028
access	c	r	r|c	.000000209	.0000916	0.019
across	o	e	e|o	.0000093	.000299	2.8
acres	-	s	es|e	.0000321	.0000318	1.0
acres	-	s	ss|s	.0000342	.0000318	1.0

Figure B.5 Computation of the ranking for each candidate correction, using the language model shown earlier and the error model from Fig. B.4. The final score is multiplied by 10^9 for readability.

The computations in Fig. B.5 show that our implementation of the noisy channel model chooses `across` as the best correction, and `actress` as the second most likely word.

Unfortunately, the algorithm was wrong here; the writer's intention becomes clear from the context: *...was called a "stellar and versatile* **acress** *whose combination of sass and glamour has defined her..."*. The surrounding words make it clear that `actress` and not `across` was the intended word.

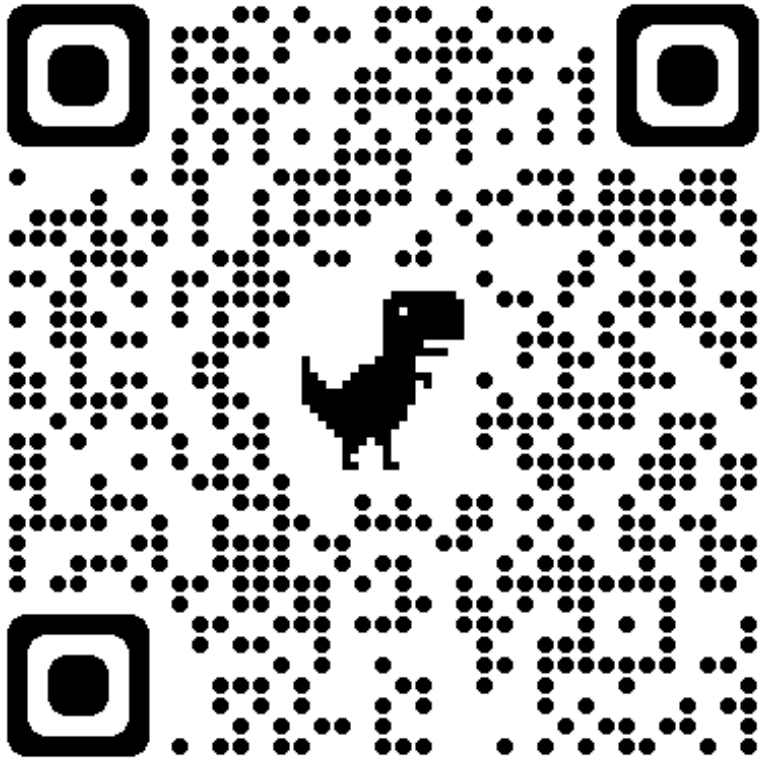
Big Picture

(Computational Perspective)

- Speech (JM16)
  - Digital Signal Processing
  - Speech to Text (STT; ASR)
  - Text to Speech (TTS; synthesis)
  - Diarization
- Morphology
  - Tokenization (Subwords) (JM2)
- The Lexicon (JM23)
  - Spelling Correction
    - (JM appendix B)

- Syntax
  - Parsing (JM17 & JM18)
  - **Token classification (JM8)**
    - **Part of speech tagging**
    - **NER (named entity recognition)**
  - Coreference (JM26)
- Semantics
  - Logical Form (JM19)
  - Temporal Reasoning (JM22)
- Pragmatics, Discourse & Dialogue
  - Chatbots (JM15)
  - Question Answering & Information Retrieval (JM14)

Part of Speech Tagging



Parsing & Context-Free Grammars

Grammar

Grammar Rules			Examples
S	→	*NP VP*	I + want a morning flight
NP	→	*Pronoun*	I
	|	*Proper-Noun*	Los Angeles
	|	*Det Nominal*	a + flight
Nominal	→	*Nominal Noun*	morning + flight
	|	*Noun*	flights
VP	→	*Verb*	do
	|	*Verb NP*	want + a flight
	|	*Verb NP PP*	leave + Boston + in the morning
	|	*Verb PP*	leaving + on Thursday
PP	→	*Preposition NP*	from + Los Angeles

Figure 17.3 The grammar for $\mathscr{L}_0$, with example phrases for each rule.

Parse Tree

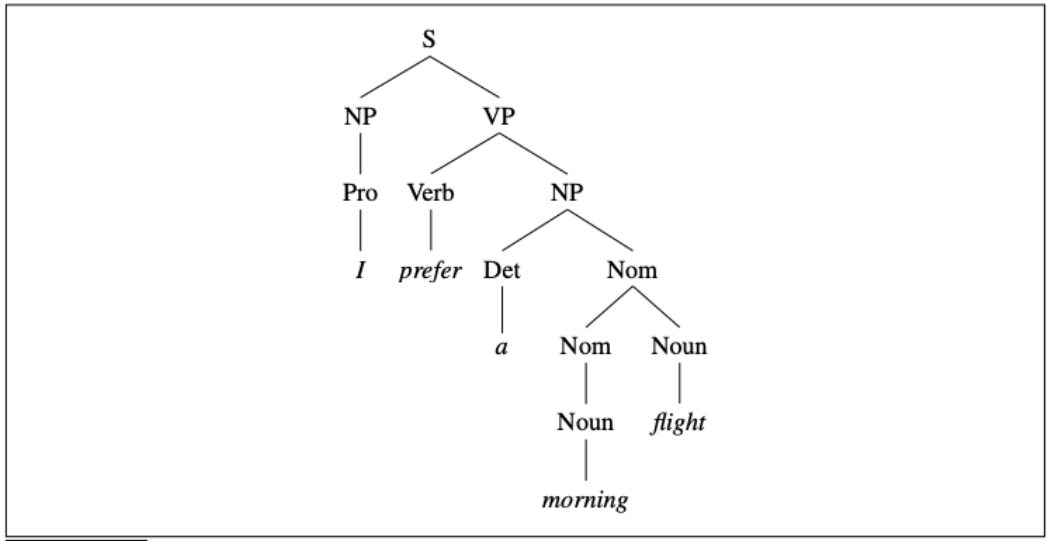


Figure 17.4 The parse tree for "I prefer a morning flight" according to grammar $\mathscr{L}_0$.

Parse Tree from Penn TreeBank

```
((S
   (NP-SBJ (DT That)
     (JJ cold) (, ,)
     (JJ empty) (NN sky) )
   (VP (VBD was)
     (ADJP-PRD (JJ full)
       (PP (IN of)
         (NP (NN fire)
           (CC and)
           (NN light) ))))
   (. .) ))
```

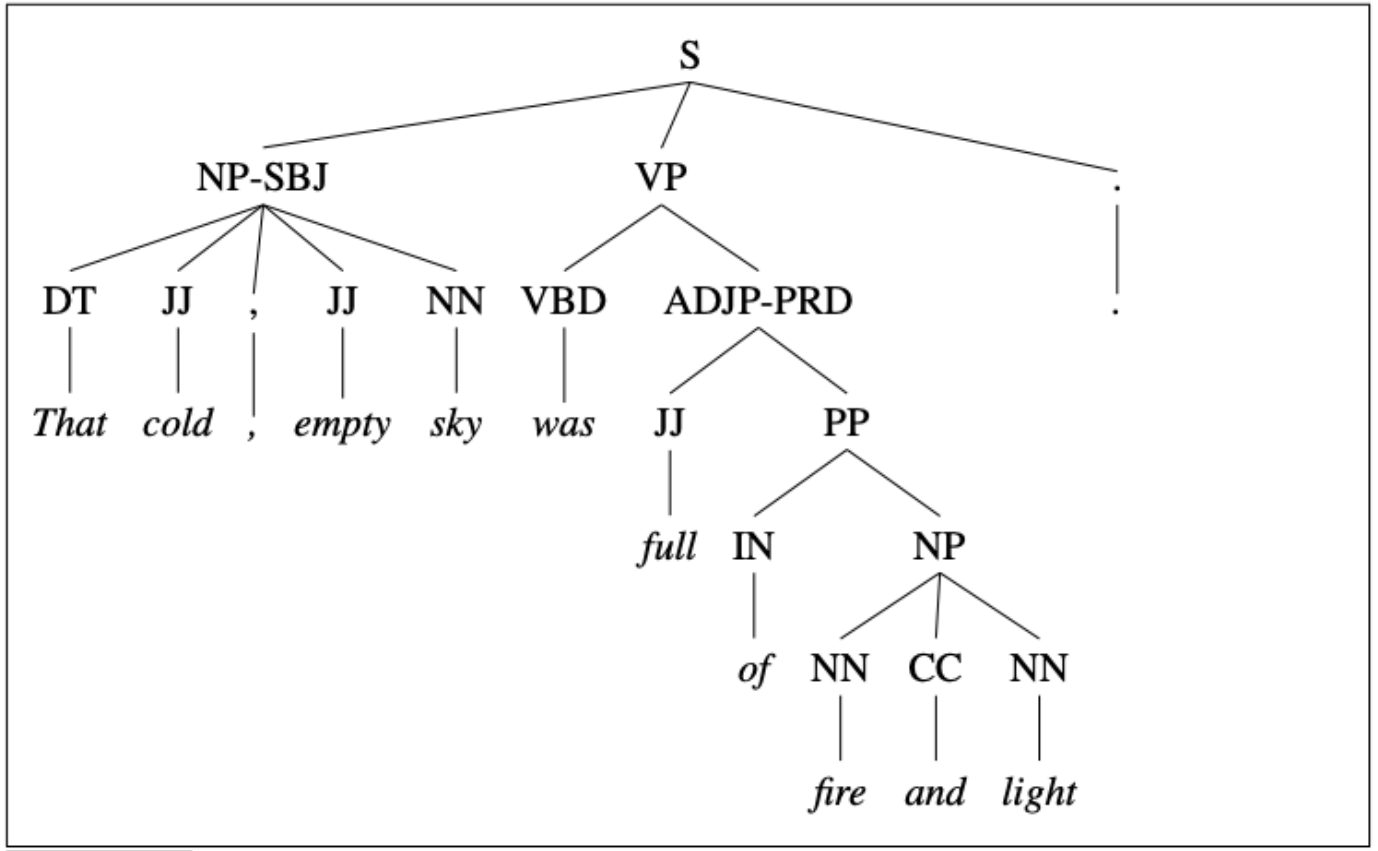


Figure 17.6 The tree corresponding to the Brown corpus sentence in the previous figure.

Ambiguity

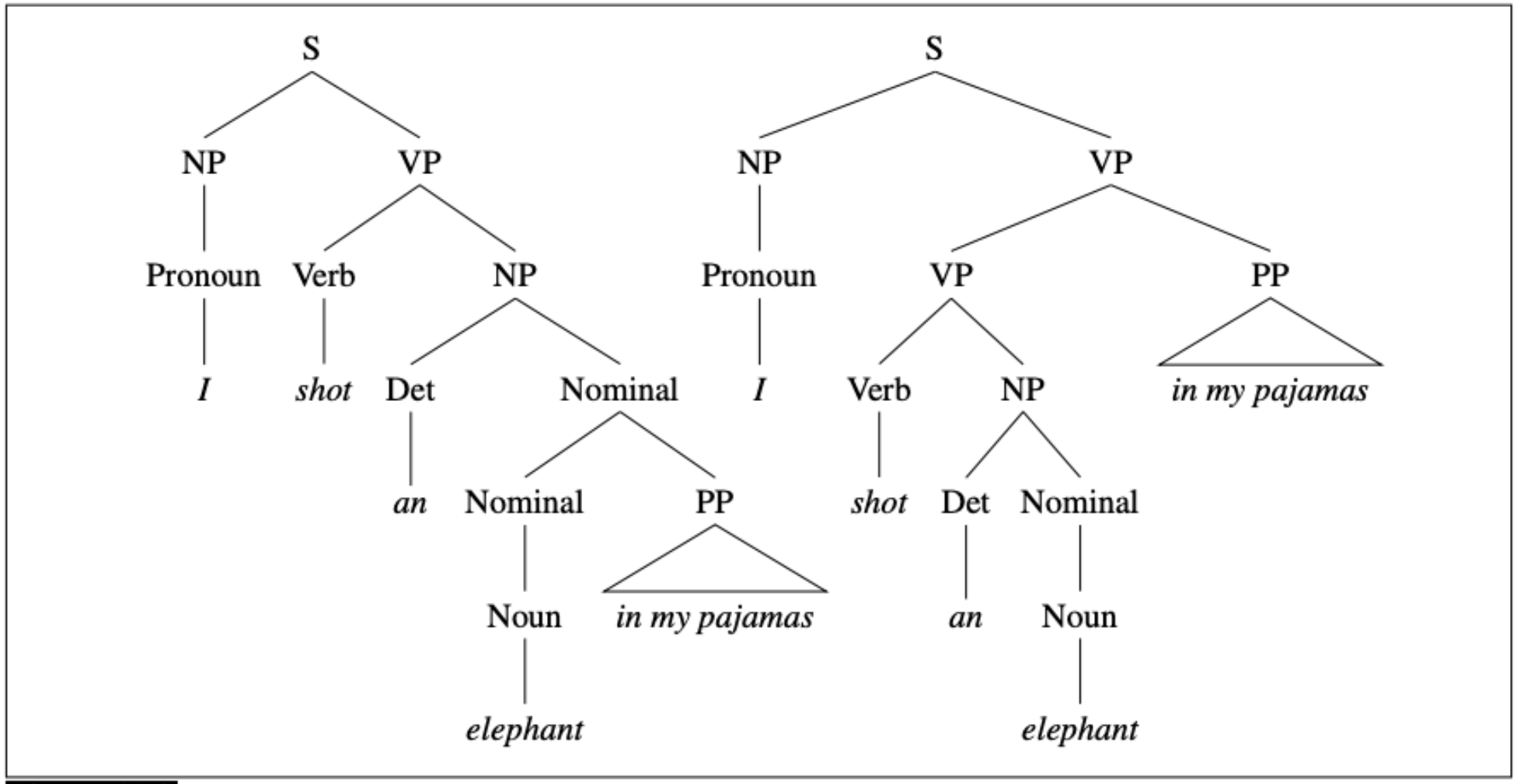
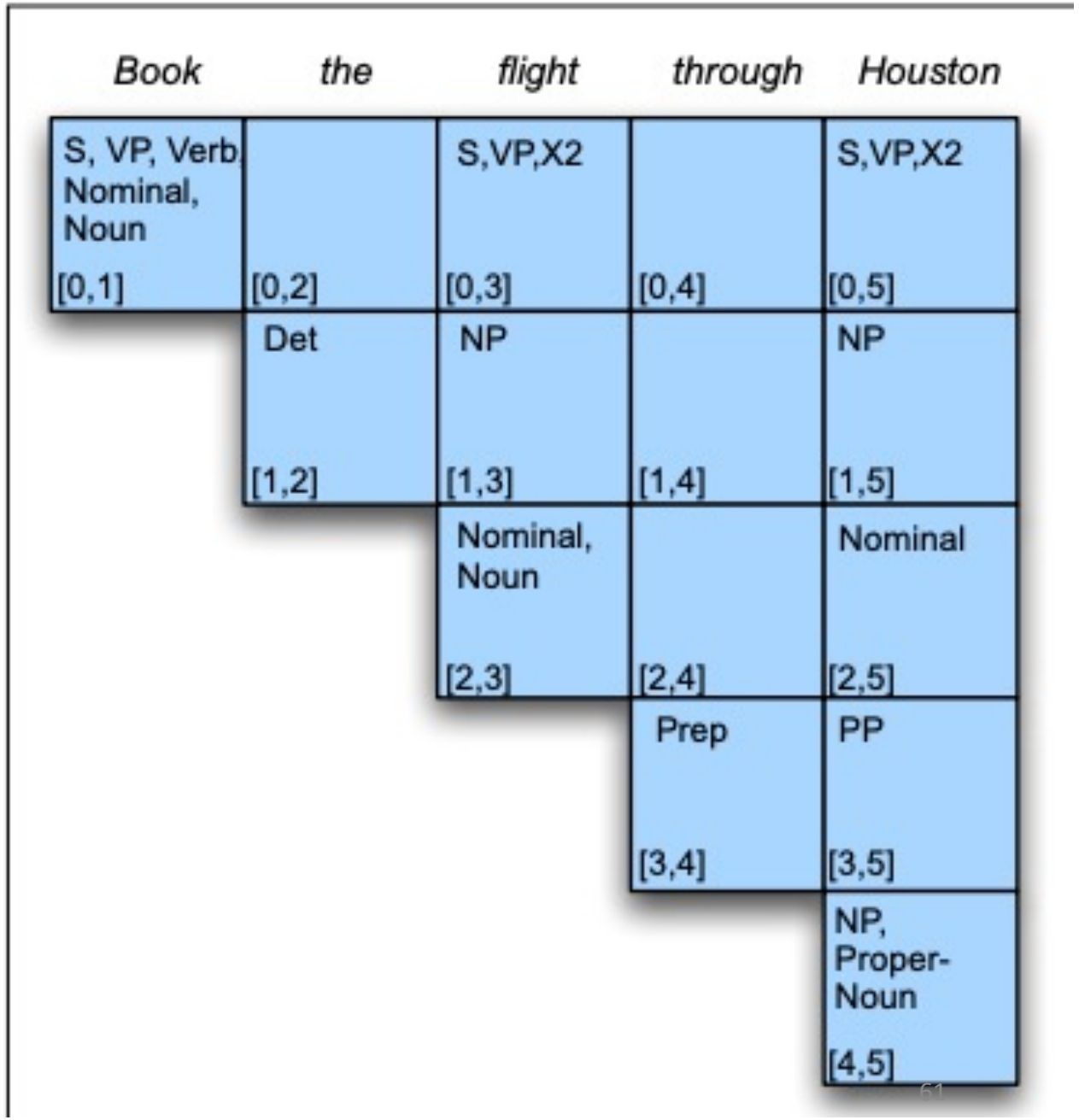


Figure 17.9 Two parse trees for an ambiguous sentence. The parse on the left corresponds to the humorous reading in which the elephant is in the pajamas, the parse on the right corresponds to the reading in which Captain Spaulding did the shooting in his pajamas.

Parse table for: *Book the flight through Houston*



Big Picture

(Linguistics Perspective)

- Speech (JM28)
  - Acoustics
  - Digital Signal Processing
  - Phonetics: phonemes, distinctive features
  - Prosody: Pitch, duration, energy
  - Phonology: Stress Assignment
- Morphology
  - Regular Inflection, level 1, level 2
  - Compounding
- The Lexicon (JM23)
- Syntax
  - Parsing
    - Chomsky Hierarchy:
      - Finite-State, Context-Free (JM17)
      - Context-Sensitive, Turing Equivalent
  - Variable Binding
    - Pronouns, Quantifier Scope
    - WH-movement
  - Predicate-argument structure (JM24)
- Semantics
  - Logical Form (JM19)
- Pragmatics, Discourse & Dialogue (JM27)
  - ➢ **Gricean Maxims**
  - Indirect speech acts: *do you have the time?*
  - Diarization (who spoke when)
  - Filled pauses, restarts, corrections

Grice's Maxims

- Maxim of quantity (length)
  - Be as informative as possible
  - Do not say more than is required
- Maxim of quality (truth)
  - Do not say what you believe is false
  - Do not say that for which you lack adequate evidence
- Maxim of relation (relevance)
  - Be relevant
- Maxim of manner (clarity)
  - Avoid obscurity and ambiguity
  - Be brief and orderly

High-level Rhetorical Structures

- Description
- Exposition
- Narration
- Persuasive

- Beginning → End
  - Plot
  - Character Arc
  - Punch line (after setup)

- Academic style:
  - Say everything three times
    - Promise
    - Connect the dots between promise and delivery
    - Delivery
- Newspaper style:
  - Lead with the lead
- Who-done-it:
  - End with the solution
  - First suspect never did it