CS6120: Lecture 11
Topics that may become great (again)

Kenneth Church

https://kwchurch.github.io/

Your Projects

- Oral:
  - 20-40 minutes
- Written
  - 5-20 pages
  - lots of references
- Say everything three times
  - Winston: How to Speak
  - https://www.youtube.com/watch?v=Unzc731iCUY
- Examples:
  - https://aclanthology.org/2022.acl-long.60
  - http://34.204.188.58/cgi-bin/similar?embedding=s2_recommendations&limit=20&search=An+Information-theoretic+Approach+to+Prompt+Engineering+Without+Ground+Truth+Labels
  - https://aclanthology.org/events/acl-2022/

Say everything three times
Winston: How to Speak

- Introduction: Promise
  - Say what you will say
- Body:
  - Say it
- Conclusion: Delivery
  - Say what you said

Paper Outline

https://aclanthology.org/2022.acl-long.60.pdf

- Abstract
- Introduction
- Related Work
- Methods
    - Task Definition
    - Mutual Information
- Experimental Setup
    - Datasets
    - Models
- Results
    - Template Selection Performance
    - Correlation between Template Mutual Information and Accuracy
    - Compared to Few Labeled Examples
    - Method Robustness and Ensembling
    - Transferability across Models
- Conclusion
- Ethics
- Limitations (missing)
- Acknowledgements
- References
- Appendices

References

- Google Scholar
- Semantic Scholar
- JSALT-2023
  - http://34.204.188.58/similar.html
  - http://34.204.188.58/cgi-bin/similar?embedding=s2_recommendations&limit=20&search=An+Information-theoretic+Approach+to+Prompt+Engineering+Without+Ground+Truth+Labels
  - http://34.204.188.58/cgi-bin/similar?CorpusId=CorpusId:236493269&embedding=s2_recommendations&limit=20
  - http://34.204.188.58/cgi-bin/similar?CorpusId=CorpusId:236493269&embedding=proposed&limit=20

Foundation Models

https://hai.stanford.edu/news/what-foundation-model-explainer-non-experts
https://arxiv.org/pdf/2108.07258.pdf

On the Opportunities and Risks of Foundation Models

Rishi Bommasani* Drew A. Hudson Ehsan Adeli Russ Altman Simran Arora
Sydney von Arx Michael S. Bernstein Jeannette Bohg Antoine Bosselut Emma Brunskill
Erik Brynjolfsson Shyamal Buch Dallas Card Rodrigo Castellon Niladri Chatterji
Annie Chen Kathleen Creel Jared Quincy Davis Dorottya Demszky Chris Donahue
Moussa Doumbouya Esin Durmus Stefano Ermon John Etchemendy Kawin Ethayarajh
Li Fei-Fei Chelsea Finn Trevor Gale Lauren Gillespie Karan Goel Noah Goodman
Shelby Grossman Neel Guha Tatsunori Hashimoto Peter Henderson John Hewitt
Daniel E. Ho Jenny Hong Kyle Hsu Jing Huang Thomas Icard Saahil Jain
Dan Jurafsky Pratyusha Kalluri Siddharth Karamcheti Geoff Keeling Fereshte Khani
Omar Khattab Pang Wei Koh Mark Krass Ranjay Krishna Rohith Kuditipudi
Ananya Kumar Faisal Ladhak Mina Lee Tony Lee Jure Leskovec Isabelle Levent
Xiang Lisa Li Xuechen Li Tengyu Ma Ali Malik Christopher D. Manning
Suvir Mirchandani Eric Mitchell Zanele Munyikwa Suraj Nair Avanika Narayan
Deepak Narayanan Ben Newman Allen Nie Juan Carlos Niebles Hamed Nilforoshan
Julian Nyarko Giray Ogut Laurel Orr Isabel Papadimitriou Joon Sung Park Chris Piech
Eva Portelance Christopher Potts Aditi Raghunathan Rob Reich Hongyu Ren
Frieda Rong Yusuf Roohani Camilo Ruiz Jack Ryan Christopher Ré Dorsa Sadigh
Shiori Sagawa Keshav Santhanam Andy Shih Krishnan Srinivasan Alex Tamkin
Rohan Taori Armin W. Thomas Florian Tramèr Rose E. Wang William Wang Bohan Wu
Jiajun Wu Yuhuai Wu Sang Michael Xie Michihiro Yasunaga Jiaxuan You Matei Zaharia
Michael Zhang Tianyi Zhang Xikun Zhang Yuhui Zhang Lucia Zheng Kaitlyn Zhou
Percy Liang*[1]

Center for Research on Foundation Models (CRFM)
Stanford Institute for Human-Centered Artificial Intelligence (HAI)
Stanford University

AI is undergoing a paradigm shift with the rise of models (e.g., BERT, DALL-E, GPT-3) trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks. We call these models foundation models *to underscore their critically central yet incomplete character. This report provides a thorough account of the opportunities and risks of foundation models, ranging from their capabilities (e.g., language, vision, robotic manipulation, reasoning, human interaction) and technical principles (e.g., model architectures, training procedures, data, systems, security, evaluation, theory) to their applications (e.g., law, healthcare, education) and societal impact (e.g., inequity, misuse,*

The Easy, the Hard and the Ugly

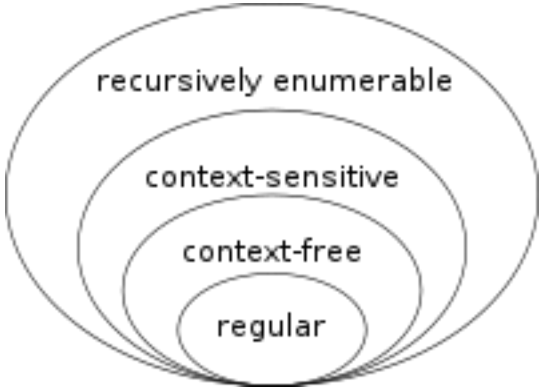
- Easy
  - Prompt Engineering
  - Inference (*fit*)
  - Fine-Tuning (*predict*)
- Hard
  - Pre-training
- Ugly (Responsible AI)
  - Bias
  - Toxicity
  - Misinformation
  - Hallucinations
  - Plagiarism

Agenda

- Chomsky Hierarchy:
  - Generative Capacity
- Parsing
  - (every way ambiguous structures)
  - PP-attachment & Conjunction
- Speech
  - Prosody (Pitch, Duration, Energy)
  - Chinese Tone
- Discourse
  - Rhetoric
  - Grice's Maxims

- Interdisciplinary Collaboration
  - Linguistics
  - Lexicography
  - Psychology
  - Statistics
  - Medicine
  - Law

Chomsky Hierarchy



Grammar	Languages	Recognizing Automaton	Production rules (constraints)*	Examples[5][6]
Type-3	Regular	Finite state automaton	$A \to \mathrm{a}$ and $A \to \mathrm{a}B$	$L = \{a^n \mid n \geq 0\}$
Type-2	Context-free	Non-deterministic pushdown automaton	$A \to \alpha$	$L = \{a^n b^n \mid n > 0\}$
Type-1	Context-sensitive	Linear-bounded non-deterministic Turing machine	$\alpha A \beta \to \alpha \gamma \beta$	$L = \{a^n b^n c^n \mid n > 0\}$
Type-0	Recursively enumerable	Turing machine	$\gamma \to \alpha$ (γ non-empty)	$L = \{w \mid w$ describes a terminating Turing machine$\}$

* Meaning of symbols:

- a = terminal
- A, B = non-terminal
- α, β, γ = string of terminals and/or non-terminals

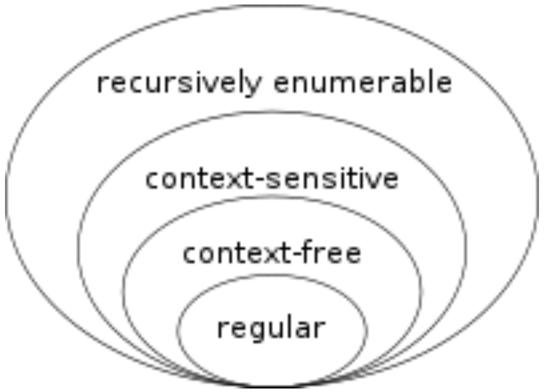
Chomsky Hierarchy

- Language ≡ Set of Strings, S
- Grammar ≡ Set of Parse Trees, G
- Automaton ≡ Recognizer
  - Machine: s → $\{1, 0\}$
  - Is $s \in S$?

- Equivalence: Strong vs. Weak
  - Weak: $S_1 = S_2$
  - Strong: $G_1 = G_2$

Grammar	Languages	Recognizing Automaton	Production rules (constraints)*	Examples[5][6]
Type-3	Regular	Finite state automaton	$A \to \mathrm{a}$ and $A \to \mathrm{a}B$	$L = \{a^n \mid n \geq 0\}$
Type-2	Context-free	Non-deterministic pushdown automaton	$A \to \alpha$	$L = \{a^n b^n \mid n > 0\}$
Type-1	Context-sensitive	Linear-bounded non-deterministic Turing machine	$\alpha A \beta \to \alpha\gamma\beta$	$L = \{a^n b^n c^n \mid n > 0\}$
Type-0	Recursively enumerable	Turing machine	$\gamma \to \alpha$ (γ non-empty)	$L = \{w \mid w$ describes a terminating Turing machine$\}$

* Meaning of symbols:
- a = terminal
- A, B = non-terminal
- α, β, γ = string of terminals and/or non-terminals



Examples:

- Finite-State:
  - ngrams,
  - Markov Processes
  - grep
- Context-Free:
  - parentheses
  - programming languages (syntax)
  - ww^R
- Context-Sensitive: ww
- Turing Machines:
  - Lambda Calculus
  - Programming languages
    - (including semantics)

Compilers

https://www.lonelyscholar.com/knuth

Home Turing Jackson P=NP Dawn Knuth Naur Future

Lonely Scholar scientific books

Home

The Essential Knuth

Donald E. Knuth lived two separate lives in the late 1950s. During daylight he ran down the visible and respectable lane of mathematics. During nighttime, he trod the unpaved road of computer programming and compiler writing.

Both roads intersected! — as Knuth discovered while reading Noam Chomsky's book Syntactic Structures on his honeymoon in 1961.

"Chomsky's theories fascinated me, because they were mathematical yet they could also be understood with my programmer's intuition. It was very curious because otherwise, as a mathematician, I was doing integrals or maybe was learning about Fermat's number theory, but I wasn't manipulating symbols the way I did when I was writing a compiler. With Chomsky, wow, I was actually doing mathematics and computer science simultaneously."

How, when, and why did mathematics and computing converge for Knuth? To what extent did logic and Turing machines appear on his radar screen?

The early years of convergence ended with the advent of Structured Programming in the late 1960s. How did that affect his later work on TeX? And what did "structure" come to mean to Knuth?

Shedding light on where computer science stands today by investigating Knuth's past — that's what this booklet is about.

Compilers & Interpreters

- Compiler vs. Interpreter
  - C vs. python
  - Which is more valuable (according to a patent lawyer)?
- Compiler: source code (foo.c) → object code (a.out)
  - Parse (step 1): source code → parse tree
  - Code generation (step 2): parse tree → object code
- Intermediate representation: Parse Tree

- https://www.cs.auckland.ac.nz/courses/compsci220s1t/archive/compsci220ft/lectures/GGlectures/220ch4_gramm.pdf

Agenda

- ✓ Chomsky Hierarchy:
  - ✓ Generative Capacity
- ➢ Parsing
  - (every way ambiguous structures)
  - PP-attachment & Conjunction
- Speech
  - Prosody (Pitch, Duration, Energy)
  - Chinese Tone
- Discourse
  - Rhetoric
  - Grice's Maxims

- Interdisciplinary Collaboration
  - Linguistics
  - Lexicography
  - Psychology
  - Statistics
  - Medicine
  - Law

Example of Parsing: https://web.stanford.edu/~jurafsky/slp3/17.pdf

Grammar

17.2 • CONTEXT-FREE GRAMMARS 5

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

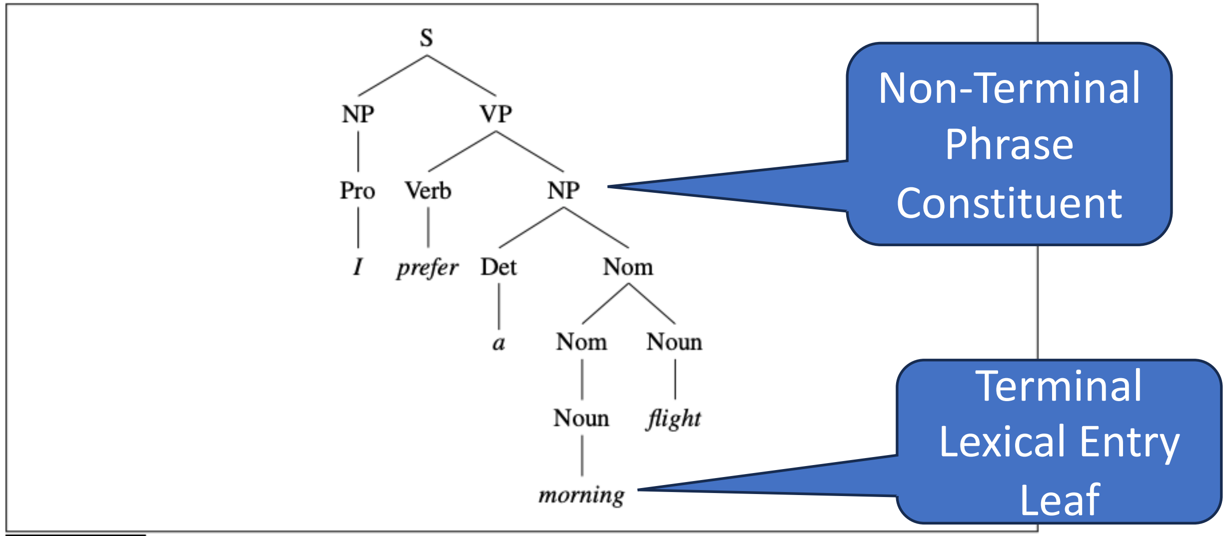


Figure 17.4 The parse tree for "I prefer a morning flight" according to grammar $\mathscr{L}_0$.

Should a Corpus be Balanced?

Penn TreeBank: https://aclanthology.org/J93-2004.pdf

Table 4
Penn Treebank (as of 11/92).

Description	Tagged for Part-of-Speech (Tokens)	Skeletal Parsing (Tokens)
Dept. of Energy abstracts	231,404	231,404
Dow Jones Newswire stories	3,065,776	1,061,166
Dept. of Agriculture bulletins	78,555	78,555
Library of America texts	105,652	105,652
MUC-3 messages	111,828	111,828
IBM Manual sentences	89,121	89,121
WBUR radio transcripts	11,589	11,589
ATIS sentences	19,832	19,832
Brown Corpus, retagged	1,172,041	1,172,041
Total:	4,885,798	2,881,188

Syntactic Ambiguity

Programming Languages

(Considered Bad)

- *if … then … else …*
  - https://en.wikipedia.org/wiki/Dangling_else
  - **if** a **then** (**if** b **then** s) **else** s2
  - **if** a **then** (**if** b **then** s **else** s2)

Natural Languages

(Reality or Academic Non-Issuue?)

- *I shot an elephant in my pajamas*
  - *I shot* [*an elephant in my pajamas*]
  - *I shot* [*an elephant*][*in my pajamas*]
- *Time flies like an arrow*
- *Fruit flies like a banana*
- *https://en.wikipedia.org/wiki/List_of_linguistic_example_sentences*

Ivan Sag, 1976, PhD Thesis

Constituency

- Conjunction:
  - $X \rightarrow X \,\&\, X$
  - NP → NP & NP
  - S → S & S
- Wh-movement
  - I saw an elephant
  - What did I see?
  - Wh(*t*) [I saw *t*]

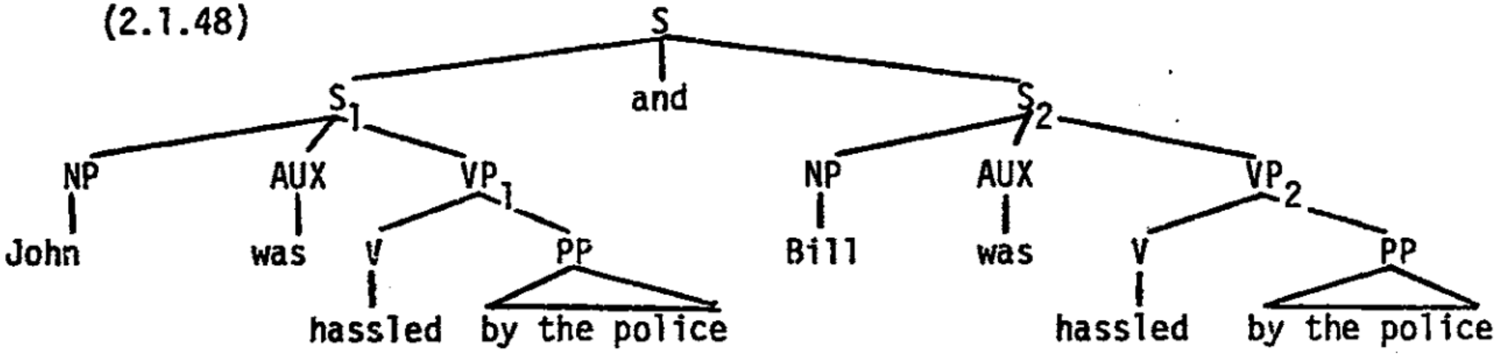
(2.1.47)(a) John was hassled by the police, and Sam was ø, too.

(b) Betsy seems to me to be unhappy, and Sandy does ø, too.

(c) Betsy is easy to talk to, and Peter is ø, too.

would not provide the appropriate identity to predict the possibility of deletion. The problem is actually a form of the "correspondence problem" we noted earlier. To take one example, (2.1.47)(a) would have a standard theory deep structure roughly as in (2.1.48) (the position of AUX is not of concern here).

(2.1.48)



https://web.archive.org/web/20220120200245/https://dspace.mit.edu/bitstream/handle/1721.1/16401/03491413.pdf?sequence=1

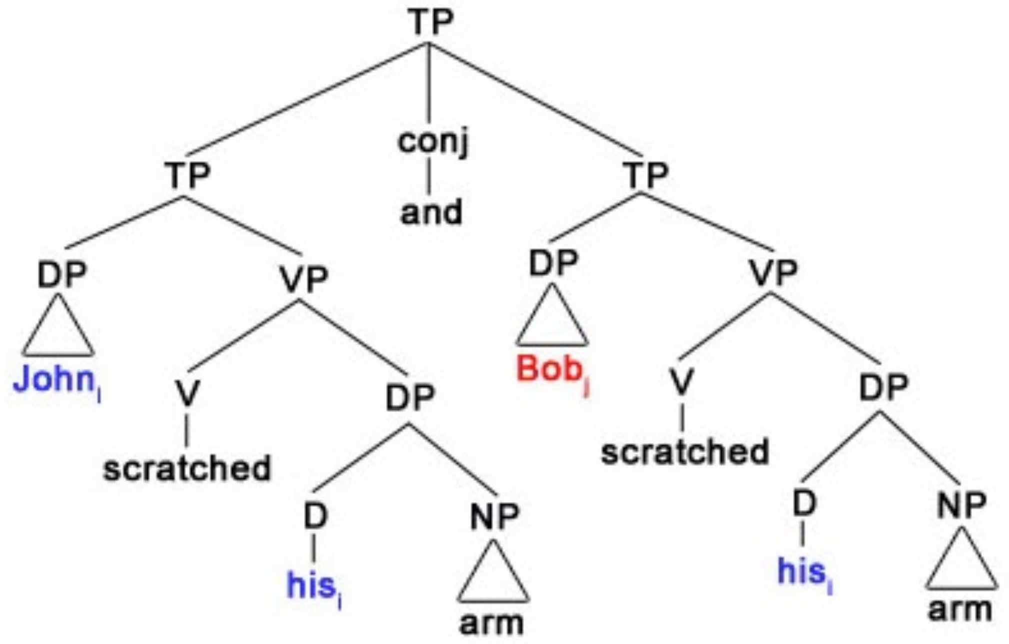
(2.1.60)(a) The children are ready to eat [i.e. to partake of food] and the chickens are ready to eat [i.e. to be eaten] too.

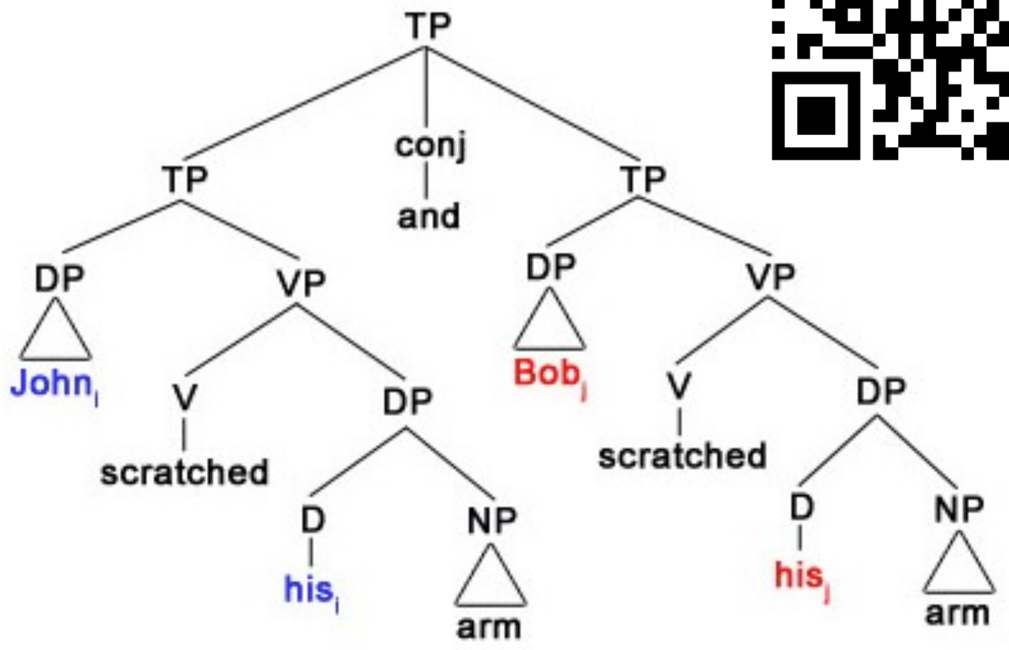
(b) The children, $\lambda x(\text{ready}(x,[\Delta, \lambda y(y \text{ eat})]))$ & the chickens, $\lambda x(\text{ready}(x,[\Delta, \lambda y(y \text{ eat } x)]))$

1) *John scratched his arm and Bob did too.*

a. **Strict reading:** *John$_i$ scratched his$_i$ arm and Bob$_j$ [scratched his$_i$ arm] too.*

b. **Sloppy reading:** *John$_i$ scratched his$_i$ arm and Bob$_j$ [scratched his$_j$ arm] too.*





https://en.wikipedia.org/wiki/Sloppy_identity

Ivan Sag, 1976, PhD Thesis

CHAPTER THREE

Gapping

3.1 Syntactic Overview

3.1.0 Introduction

This chapter will deal with the well-known elliptical phenomenon illustrated by the sentences in (3.1.1).

(3.1.1)(a) Sandy played shortstop, and Betsy ø first base.

[ø = played]

(b) Alan ran to second base, and Betsy ø to first base.

[ø = ran]

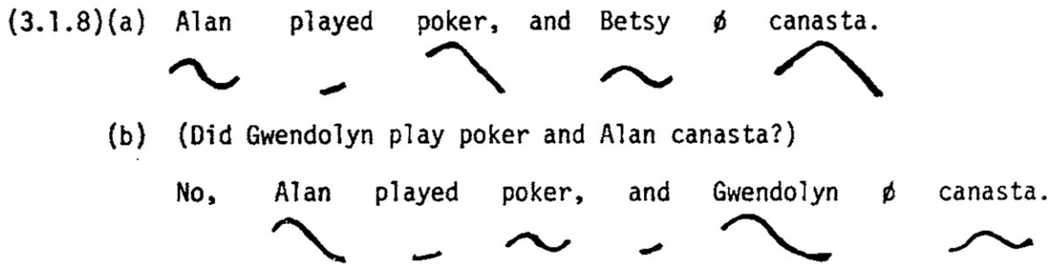
(c) Betsy plays first base for the first four innings, and Peter ø for the rest of the game.

[ø = plays first base]

(d) Sandy wanted to begin to write a novel, and Betsy ø a short story. [ø = wanted to begin to write]

https://web.archive.org/web/20220120200245/https://dspace.mit.edu/bitstream/handle/1721.1/16401/03491413.pdf?sequence=1

Conjunction & Prosody



- Contrastive Stress
- Phrase Final Lengthening
- Accents

- Practical Opportunity
  - Books on Tape
    - https://www.openslr.org/12
  - Construct parallel corpus of speech and text
  - Use constraints from both to annotate conjunctions

Agenda

- Chomsky Hierarchy:
  - Generative Capacity
- Parsing
  - (every way ambiguous structures)
  - PP-attachment & Conjunction
- ➢ Speech
  - **Prosody (Pitch, Duration, Energy)**
  - Chinese Tone
- Discourse
  - Rhetoric
  - Grice's Maxims
- Interdisciplinary Collaboration
  - Linguistics
  - Lexicography
  - Psychology
  - Statistics
  - Medicine
  - Law

LANGUAGE AND SPEECH, 1994, 37(3), 237 — 250

DURATIONAL EFFECTS IN FINAL LENGTHENING, GAPPING, AND CONTRASTIVE STRESS*

ROCHELE BERKOVITS
Tel Aviv University

Lengthening in utterance-final position and in contrastive stress was examined in Hebrew, focusing on the distribution of the durational effect across syllables and within the final syllable. Initially-stressed and finally-stressed bisyllabic key words were read in sentence-final *versus* nonfinal position, and in contrastive stress *versus* nonstressed constructions. The results were compared with data from an earlier study of verb gapping. Contrastive stress showed a smaller effect than final lengthening and verb gapping, consistent with the claim that other acoustic parameters are more prominently involved in this process. Utterance-final lengthening and verb gapping principally affected the final syllable regardless of stress, whereas contrastive stress primarily lengthened the stressed syllable. The pattern of progressively greater lengthening within the utterance-final syllable, previously found for stressed syllables, applied to unstressed syllables as well. The finding that target syllables in sentence-final position are characterized by progressive lengthening, unlike those in contrastive stress and gapping, supports the suggestion that utterance-final lengthening is a reflection of deceleration at the end of motor activity. Durational measures of individual syllables within the key word, and of segments in addition to the vocalic portion of the final syllable, reveal differences in the acoustic implementation of different lengthening processes.

Key words: final lengthening, syllable duration, stress, Hebrew

A **prosody tutorial** for **investigators** of **auditory sentence processing**

S Shattuck-Hufnagel, AE Turk - Journal of psycholinguistic research, 1996 - Springer

… , has proven useful in capturing the **prosodic** structures behind spoken utterances in a …
prosody into account in **auditory sentence processing** research. A great many aspects of **prosody** …

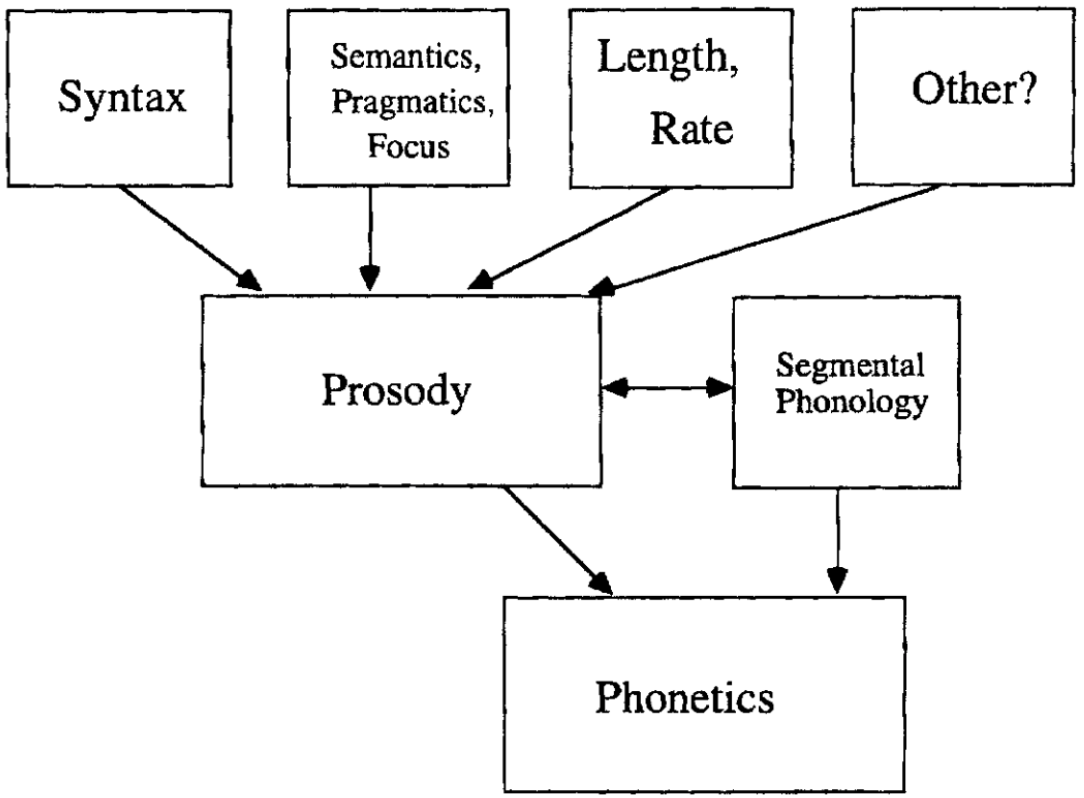


Fig. 5. One view of the role of the prosodic component of the grammar.

Prosody in Speech

- Prosody → Pitch, Duration, Energy
- Prosodic Constraints involving Constituents
  - English: Stress, Phrase Final Lengthening
  - Chinese: Tone (https://en.wikipedia.org/wiki/Tone_sandhi)

A **prosody tutorial** for **investigators** of **auditory sentence processing**

S Shattuck-Hufnagel, AE Turk - Journal of psycholinguistic research, 1996 - Springer

… , has proven useful in capturing the **prosodic** structures behind spoken utterances in a … **prosody** into account in **auditory sentence processing** research. A great many aspects of **prosody** …

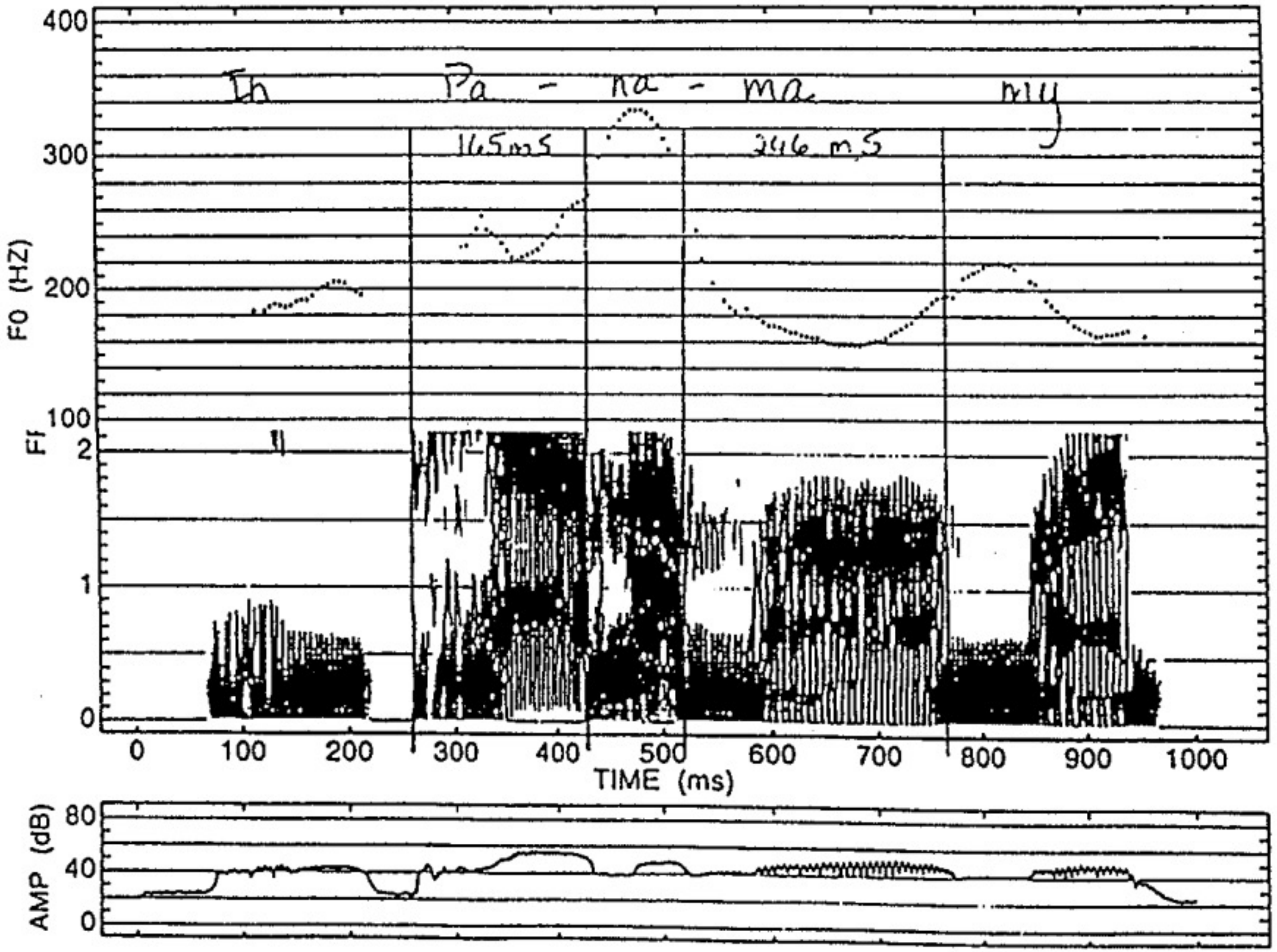


Fig. 6. F0 track and spectrogram of the first part of an utterance of *In Panama my plane was late*, produced as two separate Full Intonational Phrases (digitized utterance available via anonymous ftp from lexic.mit.edu). Note the delayed F0 peak on the reduced syllable -na-, following the prominent syllable Pa-.

https://aclanthology.org/W16-6008.pdf

$C^2D^2E^2$: Using Call Centers to Motivate the Use of Dialog and Diarization in Entity Extraction

Kenneth Church, Weizhong Zhu and **Jason Pelecanos**
IBM, Yorktown Heights, NY, USA
{kwchurch, zhuwe, jwpeleca}@us.ibm.com

Diarization

Abstract

This paper introduces a deceptively simple entity extraction task intended to encourage more interdisciplinary collaboration between fields that don't normally work together: diarization, dialog and entity extraction. Given a corpus of 1.4M call center calls, extract mentions of trouble ticket numbers. The task is challenging because first mentions need to be distinguished from confirmations to avoid undesirable repetitions. It is common for agents to say part of the ticket number, and customers confirm with a repetition. There are opportunities for dialog (given/new) and diarization (who said what) to help remove repetitions. New information is spoken slowly by one side of a conversation; confirmations are spoken more quickly by the other side of the conversation.

1 Extracting Ticket Numbers

Much has been written on extracting entities from text (Etzioni et al., 2005), and even speech (Kubala et al., 1998), but less has been written in the context of dialog (Clark and Haviland, 1977) and diarization

t0	t1	S1	S2
278.16	281.07	I do have the new hardware case number for you when you're ready	
282.60	282.85		*okay*
284.19	284.80	**nine**	
285.03	285.86	**zero**	
286.22	286.74	**two**	
290.82	291.30		*nine*
292.87	293.95	*zero two*	
297.87	298.24		*okay*
299.30	300.49	**M. as in Mike**	
301.97	303.56	**D. as in delta**	
304.89	306.31	**Y. as in Yankee**	
307.50	308.81	**K. as in kilo**	
310.14	310.57		*okay*
310.77	311.70		*nine zero two*
311.73	312.49		*M. D.*
312.53	313.18		*Y. T.*
313.75	314.21	*correct*	
314.21	317.28	and thank you for calling IBM is there anything else I can assist you with	

Table 1: A ticket dialog: 7 bytes (902MDYK) at 1.4 bps. First mentions (**bold**) are slower than confirmations (*italics*).

Speech: JM16

- Speech-to-Text (STT): Recognition (ASR)
- Text-to-Speech (TTS): Synthesis

Agenda

- Chomsky Hierarchy:
  - Generative Capacity
- Parsing
  - **(every way ambiguous structures)**
  - PP-attachment & Conjunction
- Speech
  - Prosody (Pitch, Duration, Energy)
  - Chinese Tone
- Discourse
  - Rhetoric
  - Grice's Maxims
- Interdisciplinary Collaboration
  - Linguistics
  - Lexicography
  - Psychology
  - Statistics
  - Medicine
  - Law

Catalan Numbers: Everyway ambiguous https://aclanthology.org/J82-3004.pdf

(3a) Put the block [[in the box on the table] in the kitchen].
(3b) Put the block [in the box [on the table in the kitchen]].
(3c) Put [[the block in the box] on the table] in the kitchen.
(3d) Put [the block [in the box on the table]] in the kitchen.
(3e) Put [the block in the box] [on the table in the kitchen].

The first few Catalan numbers are 1, 1, 2, 5, 14, 42, 132, 469, 1430, 4862. They are generated by the closed form expression:[7]

$$(10)\quad \mathrm{Cat}_n = \binom{2n}{n} - \binom{2n}{n-1}$$

index	*trees*
0	{[John]}
1	{[John and John]}
2	{[[John and John] and John], [John and [John and John]]}
⋮	

The table would be more general if it did not specify the lexical items at the leaves. Let us replace the table above with

index	*trees*
0	{[x]}
1	{[x x]}
2	{[[x x] x], [x [x x]]}
⋮	

Conjunction

ACC = argument cluster coordination; RNR = right node raising

(1) a. [We gave Jan a cake] and [we gave Yo a book]. (Constituent Coordination)

b. We [gave Jan a cake] and [gave Yo a book]. (Constituent Coordination)

c. We gave [Jan a cake] and [Yo a book]. (ACC)

d. We visited [Jan on Monday] and [Yo on Tuesday]. (ACC)

e. [Jan visited and Yo refused to visit] my stepmother's father. (RNR)

f. [Kim told Pat that Jan visited] and [Sandy that Yo refused to visit] [my stepmother's father] (ACC+RNR)

Beavers, John, & Sag, Ivan A. (2004). *Coordinate Ellipsis and Apparent Non-Constituent Coordination*

Argument for Non-Standard Constituents

$$X \rightarrow X\ and\ X$$

(2)

$$\begin{array}{ccccccc}
\text{We} & \text{gave} & \text{Jan} & \text{a record} & \text{and} & \text{Yo} & \text{a book} \\
\hline
NP & DTV & NP & NP & CONJ & NP & NP \\
 & & \overline{\quad}^{<\mathbf{T}} & \overline{\quad}^{<\mathbf{T}} & & \overline{\quad}^{<\mathbf{T}} & \overline{\quad}^{<\mathbf{T}} \\
 & & TV\backslash DTV & VP\backslash TV & & TV\backslash DTV & VP\backslash TV \\
 & & \multicolumn{2}{c}{\overline{\qquad\qquad}^{<\mathbf{B}}} & & \multicolumn{2}{c}{\overline{\qquad\qquad}^{<\mathbf{B}}} \\
 & & \multicolumn{2}{c}{VP\backslash DTV} & & \multicolumn{2}{c}{VP\backslash DTV} \\
 & & \multicolumn{5}{c}{\overline{\qquad\qquad\qquad\qquad}^{<\Phi>}} \\
 & & \multicolumn{5}{c}{VP\backslash DTV} \\
 & \multicolumn{6}{c}{\overline{\qquad\qquad\qquad\qquad}^{<}} \\
 & \multicolumn{6}{c}{VP} \\
\multicolumn{7}{c}{\overline{\qquad\qquad\qquad\qquad}^{<}} \\
\multicolumn{7}{c}{S}
\end{array}$$

(where VP = (S\NP), TV = (VP/NP), and DTV = (TV/NP))

Agenda

- ✓ Chomsky Hierarchy:
  - ✓ Generative Capacity
- ✓ Parsing
  - ✓ **(every way ambiguous structures)**
  - ✓ PP-attachment & Conjunction
- ✓ Speech
  - ✓ Prosody (Pitch, Duration, Energy)
  - ✓ Chinese Tone
- ➢ **Discourse**
  - Rhetoric
  - [Grice's Maxims]

- Interdisciplinary Collaboration
  - Linguistics
  - Lexicography
  - Psychology
  - Statistics
  - Medicine
  - Law

Grice's Maxims

1. **The maxim of quantity**, where one tries to be as informative as one possibly can, and gives as much information as is needed, and no more.
2. **The maxim of quality**, where one tries to be truthful, and does not give information that is false or that is not supported by evidence.
3. **The maxim of relation**, where one tries to be relevant, and says things that are pertinent to the discussion.
4. **The maxim of manner**, when one tries to be as clear, as brief, and as orderly as one can in what one says, and where one avoids obscurity and ambiguity.

Genre: Texts have beginning, Middle & End

- Academic Writing:
  - Say everything three times
    - Promise (intro): Say what you will say
    - Delivery (conclusion): Say what you said
    - Body: Say it (connect dots between promise and delivery)
- News:
  - Lead with lead
  - No conclusion (your article will get cut to fit space constraints)
- *Who-done-it*
  - No lead (don't spill the punch line)
  - First suspect never did it
  - Conclusion (punch line)
- Greek tragedy
  - Tragic hero starts out with a tragic flaw that ultimately leads to his tragic downfall
- Character arc

Writing 101: The 12 Literary Archetypes

https://www.masterclass.com/articles/writing-101-the-12-literary-archetypes

- **Lover:** The romantic lead who's guided by the heart.
  - **Strengths**: humanism, passion, conviction
  - **Weaknesses**: naivete, irrationality
  - Example: Romeo and Juliet
- **The Hero**: The protagonist who rises to meet a challenge and saves the day.
  - **Strengths**: courage, perseverance, honor
  - **Weaknesses**: overconfidence, hubris
  - Example: Achilles (*The Iliad*)
- **The Magician**: A powerful figure who has harnessed the ways of the universe to achieve key goals.
  - **Strengths**: omniscience, omnipotence, discipline
  - **Weaknesses**: corruptibility, arrogance
  - Example: Darth Vader (*Star Wars*)
- **The Outlaw**: The rebel who won't abide by society's demands.
  - **Strengths**: independent thinking, virtue, owes no favors
  - **Weaknesses**: self-involved, potentially criminal
  - Example: Batman (*The Dark Knight*)
- **The Explorer**: A character naturally driven to push the boundaries of the status quo and explore the unknown.
  - **Strengths**: curious, driven, motivated by self-improvement
  - **Weaknesses**: restless, unreliable, never satisfier
  - Odysseus (*The Odyssey*)
- **The Sage:** A wise figure with knowledge for those who inquire. The mother figure or mentor is often based on this archetype.
  - **Strengths**: wisdom, experience, insight
  - **Weaknesses**: cautious, hesitant to actually join the action
  - Athena (*The Odyssey*)
- **The Innocent**: A morally pure character, often a child, whose only intentions are good.
  - **Strengths**: morality, kindness, sincerity
  - **Weaknesses**: vulnerable, naive, rarely skilled
  - Example: Tiny Tim (*A Christmas Carol*)
- **The Creator:** A motivated visionary who creates art or structures during the narrative.
  - **Strengths**: creativity, willpower, conviction
  - **Weaknesses**: self-involvement, single-mindedness, lack of practical skills
  - Example: Zeus (*The Iliad*)
- **The Ruler**: A character with legal or emotional power over others.
  - **Strengths**: omnipotence, status, resources
  - **Weaknesses**: aloofness, disliked by others, out of touch
  - Example: Tony Soprano (*The Sopranos*)
- **The Caregiver**: A character who continually supports others and makes sacrifices on their behalf.
  - **Strengths**: honorable, selfless, loyal
  - **Weaknesses**: lacking personal ambition or leadership
  - Example: Mary Poppins (*Mary Poppins*)
- **The Everyman**: A relatable character who feels recognizable from daily life.
  - **Strengths**: grounded, salt-of-the-earth, relatable
  - **Weaknesses**: lacking special powers, often unprepared for what's to come
  - Example: Bilbo Baggins (*The Hobbit*)
- **The Jester**: A funny character or trickster who provides comic relief, but may also speak important truths.
  - **Strengths**: funny, disarming, insightful
  - **Weaknesses**: can be obnoxious and superficial
  - Example: King Lear's Fool

Formal Linguistics

- Constituency
- Predicate-Argument Structure
  - Subject-Verb-Object Relations
  - Raising / Control (Equi)
- Logical Form (Binding Theory)
  - https://en.wikipedia.org/wiki/Government_and_binding_theory
  - Wh-movement
  - Quantifier Scope
  - Pronoun Binding (Co-reference)

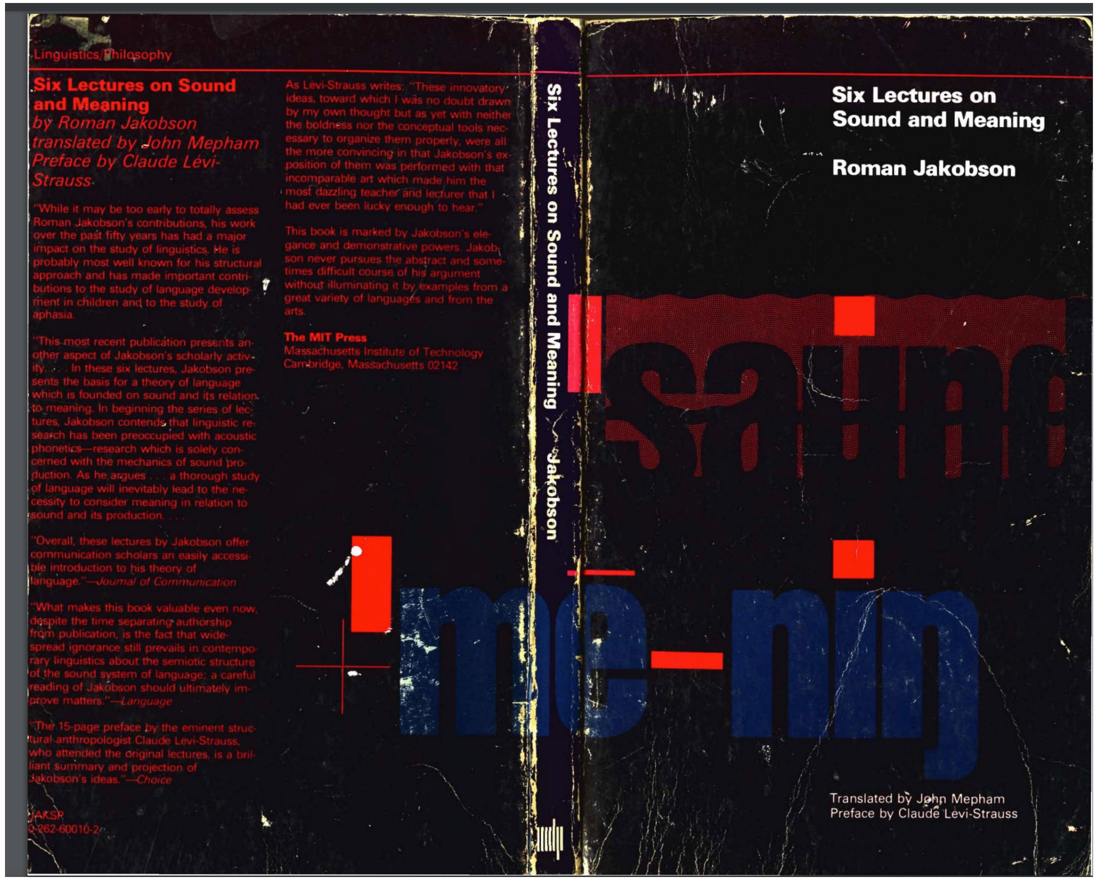
- Sound & Meaning
  - But not:
    - Spelling
    - Lexicography
    - Historical Linguistics
    - Sociolinguistics
    - Distribution
      - (PMI, Word2vec, BERT, perplexity)
    - Psycholinguistics
      - (Reaction Times, Memory Limitations, Errors)
    - Education
    - Statistics

Raising vs. Control

https://www.ling.upenn.edu/courses/ling150/ch9.html#idiom-chunks

- Idiom chunks:
  - *The cat is out of the bag*
  - *The cat* ***seems*** *to be out of the bag*
  - *The cat* ***wants*** *to be out of the bag*
- Implications
  - *Want* $\Rightarrow$ There exists a cat
  - *Seems* $\nRightarrow$ There exists a cat
- Lexical entries
  - Want: $f(subj, pred)$
  - Seems: $f(S)$

Semantics & Phonology



- Sound & Meaning
  - But not:
    - Spelling
    - Lexicography
    - Historical Linguistics
    - Sociolinguistics
    - Distribution
      - (PMI, Word2vec, BERT, perplexity)
    - Psycholinguistics
      - (Reaction Times, Memory Limitations, Errors)
    - Education
    - Statistics

Competence vs. Performance

https://en.wikipedia.org/wiki/Linguistic_competence

- In linguistics, **linguistic competence** is the system of unconscious knowledge
  - that one knows when they know a language.
- It is distinguished from linguistic performance,
  - which includes all other factors that allow one to use one's language in practice.

https://en.wikipedia.org/wiki/Linguistic_performance

- The term **linguistic performance** was used by Noam Chomsky in 1960 to describe
  - "the actual use of language in concrete situations".

How Many Multiword Expressions Do People Know?

KENNETH CHURCH, IBM

What is a multiword expression (MWE) and how many are there? Mark Liberman gave a great invited talk at ACL-89, titled "How Many Words Do People Know?" where he spent the entire hour questioning the question. Many of the same questions apply to multiword expressions. What is a word? An expression? What is many? What is a person? What does it mean to know? Rather than answer these questions, this article will use them as Liberman did, as an excuse for surveying how such issues are addressed in a variety of fields: computer science, Web search, linguistics, lexicography, educational testing, psychology, statistics, and so on.

https://dl.acm.org/doi/pdf/10.1145/2483691.2483693

(1) What do we mean by *many*? Is there a limit like 20,000 or 1M or 13.6M or does vocabulary size keep growing with experience (larger corpora (N) → larger vocabulary size (V))? Should we be talking about coverage (entropy) instead of size?
(2) What is a *word*? Is a word defined in terms of meaning? Sound? Syntax? Spelling? White space? Distribution? Etymology? Learnability? Word formation rules? What is the difference between a word and a phrase (multiword expression)?
(3) What is a *person*? Child? Adult? Native speaker? Language learner?
(4) What does it mean to *know* something? Active knowledge is different from passive knowledge. What is (artificial) intelligence? Is vocabulary size a measure of intelligence (or experience) [Terman et al. 1918]? Is Google smarter than we are, or just more experienced?

Rather than answer these questions, we will use them as Liberman did, as an excuse for surveying how such issues are addressed in a variety of fields: computer science, Web search, linguistics, lexicography, educational testing, psychology, statistics, and so on.

Lexicography Perspective

- General Vocabulary
- Technical Terminology
- Regionalisms
- Gazetteers

2.6. Regionalisms

Dictionaries tend to avoid regionalisms such as sports metaphors. Collins has better coverage of British sports metaphors than of Canadian ones.

—British: sticky wicket, red card, punt (a kind of boat);
—American: punt (on 4th down), kick the can down the road, off base, drop the ball, out in left field, hit the ball out of the park, strike out;
—Canadian: ragging the puck, put the puck in the net, lost the puck.

Although sports metaphors don't work everywhere, they seem to be limited more by geography than language. "Ragging the puck" is common in Canadian parliamentary debates in both official languages.[17]

niaiser avec le puck	2
faites traîner la rondelle	2
tergiverser	2
tricoter avec la rondelle	2
taquiner la rondelle	1
renvoyer la balle	1
lofer	1

faites traîner la rondelle 2

An Hon. Member: You are ragging the puck.	Une voix: Vous faites traîner la rondelle.
I am not ragging the puck on anybody.	Je ne faits traîner la rondelle au sujet de personne.

1 / 1 Pages

Statistics

Biometrika (1976), **63**, 3, *pp.* 435–47
With 3 text-figures
Printed in Great Britain

Estimating the number of unseen species: How many words did Shakespeare know?

BY BRADLEY EFRON AND RONALD THISTED
Department of Statistics, Stanford University, California

SUMMARY

Shakespeare wrote 31 534 different words, of which 14 376 appear only once, 4343 twice, etc. The question considered is how many words he knew but did not use. A parametric empirical Bayes model due to Fisher and a nonparametric model due to Good & Toulmin are examined. The latter theory is augmented using linear programming methods. We conclude that the models are equivalent to supposing that Shakespeare knew at least 35 000 more words.

Some key words: Empirical Bayes; Euler transformation; Linear programming; Negative binomial; Vocabulary.

8. CONCLUSIONS

Figure 3 displays the different estimates of $\Delta(t)$. Our experience with the Shakespeare data can be summarized as follows.

(i) Estimate $\hat{\Delta}(\infty) = 35\,000$ is a reasonably conservative lower bound for the amount of vocabulary Shakespeare knew but did not use.

(ii) An estimate of $\Delta(t)$ can be made very accurately for $t \leqslant 1$, but the uncertainties magnify quickly as t grows larger. Without a parametric model the data give very little additional information for t larger than 10.

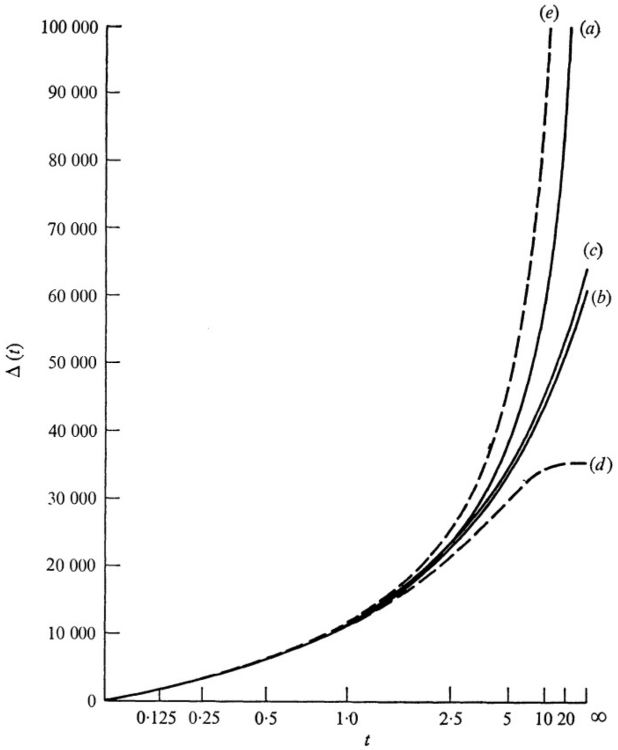


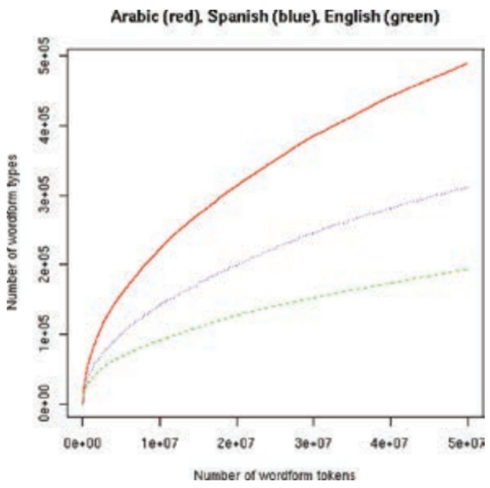
Fig. 3. Different estimates of $\Delta(t)$ for the Shakespeare data: (a) Fisher's negative binomial model with parameters (3·4); (b) Euler transformation (4·4), $x_0 = 9$, $\hat{\xi}_y$ from $\hat{\eta}_x = n_x$; (c) as (b), but with $\hat{\xi}_y$ from maximum likelihood values (3·2) and (3·4); (d) lower bound estimates from linear program (7·4) and (7·6), $c = 1$; (e) upper bound, as (d).

- *to be or not to be*
  - 4 types
  - 6 tokens

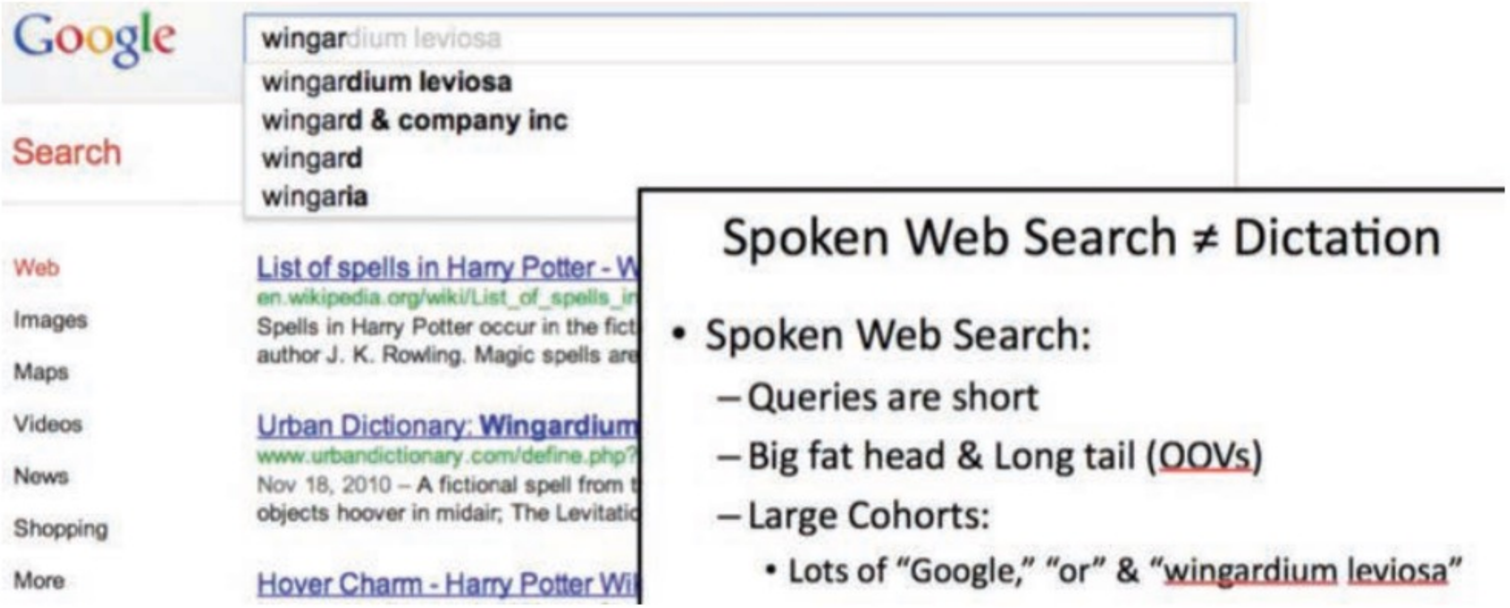
3.4. Does V Have an Upper Bound or Does It Go Up and Up and Up with N?

While I had no reason to question Google's estimate, I was reluctant to make a strong statement, given Efron and Thisted [1976]. Efron and Thisted studied a similar question: How many words did Shakespeare know (but didn't use)? Consider the phrase "to be or not to be." There are N = 6 tokens (instances of words) and V = 4 types (distinct words). How does V grow as a function of N? That is, as we consider larger and larger corpora (samples of Shakespeare's work), does V reach an asymptote or does V continue to grow with N (and if so, how quickly)? Efron and Thisted concluded that we can extrapolate N a little bit (a factor of two), but not too much (an order of magnitude). With a little bit of extrapolation, different estimates of V are reasonably close to one another, but if you extrapolate N by an order of magnitude, then the estimates of V depend more on the assumptions of the statistical model than on the empirical observations. Since Google is working with corpora that are much larger than what I have (by several orders of magnitude), it would require way too much extrapolation to answer Norvig's question based on my relatively limited experience.

Many people share the (mistaken) intuition that there is an upper bound on the size of the vocabulary. At ACL-1989, Liberman presented plots like the one below, demonstrating that V goes up and up and up with N, with no end in sight.[24] There is no upper bound on V. 20k isn't enough. Nor is 400k, or even 13.6 million... V is probably unbounded (infinite).



http://itre.cis.upenn.edu/~myl/languagelog/archives/005514.html



Dictionaries are used for correcting documents, but not for Web queries. General vocabulary is essential for documents but useless for Web queries. See Kukich [1992] for a comprehensive survey on spelling correction (before the Web) and Cucerzan and Brill [2004] for a discussion of Web queries such as these:

—*albert einstein* (4834)
—*albert einstien* (525)
—*albert einstine* (149)
—*albert einsten* (27)
—*albert einsteins* (25)
—and many many more. . .

Similar questions come up when sizing all sorts of things. How big is language? How big is the Web? How many Web queries are there? Two influential (and controversial)[28] answers are that

(1) Language is infinite [Chomsky 1956;1957].
(2) 1.3 bits per character [Shannon 1951].

Obviously, these two answers are talking about different things: size $\neq$ coverage. Chomsky [1956, p. 115] used pumping constructions like the following to argue that "there is an infinite set of sentences."

(1) If S1, then S2.
(2) Either S3, or S4.
(3) The man who said that S5 is arriving today.

Should we be counting size or coverage? If there is a word in the dictionary and no one sees it, did it make a sound? Similar questions come up when sizing all sorts of things. How big is language? How big is the Web? Chomsky argues that language is an infinite set; Shannon showed that entropy (coverage) is small and manageable. In the retail business, sales (coverage) are a good thing and costs (inventory) are not. For many practical applications, it is important to do more with less; we want to maximize sales and minimize costs.

Rather than answer these questions, this article used these questions as Liberman did, as an excuse for surveying how such issues are addressed in a variety of fields: computer science, Web search, linguistics, lexicography, educational testing, psychology, statistics, and so on. Going forward, research on MWEs should focus on methods that will scale up with the rising tide of data. Dictionaries compiled by hand such as Collins and Wordnet are extremely valuable resources, but Google has found ways to make use of much more experience (Web crawls and click logs and more). Future research ought to find ways to harness the wisdom of the crowd in order to keep up with language change, as well as diversity.

Agenda

- Chomsky Hierarchy:
  - Generative Capacity
- Parsing
  - (every way ambiguous structures)
  - PP-attachment & Conjunction
- Speech
  - Prosody (Pitch, Duration, Energy)
  - Chinese Tone
- Discourse
  - Rhetoric
  - Grice's Maxims
- Interdisciplinary Collaboration
  - Linguistics
  - Lexicography
  - Psychology
  - Education
  - Statistics
  - Medicine
  - Law