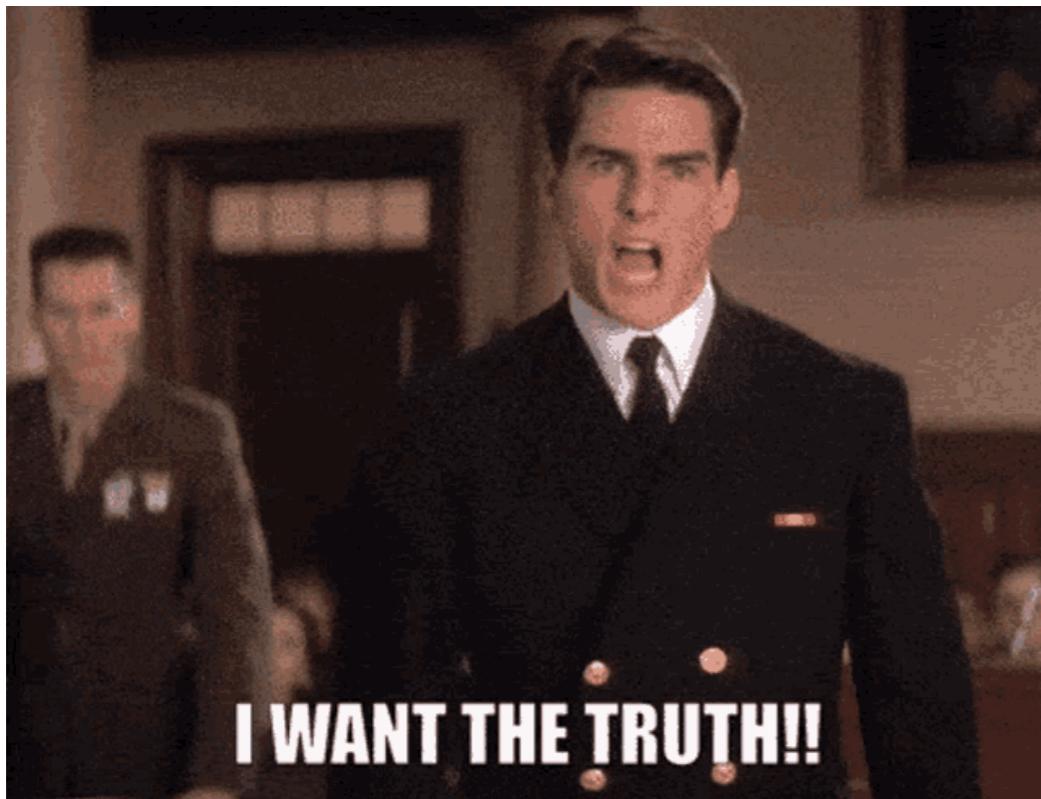


AI for Legal Interpretation!?

Legal Interpretation @ NASSLI 2025: Day 5

Perception:

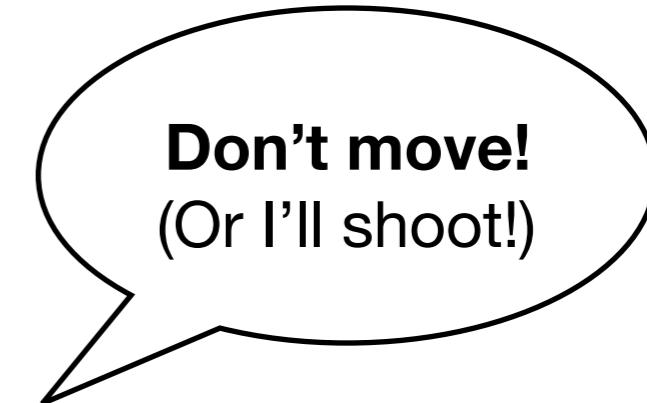


Reality:



US Sentencing Guidelines:

If a victim was **physically restrained**... [increase prison sentence].





“... was
physically restrained
by police...”

“... tried to *physically restrain* him...”

“... *physically restrained* in his
movements...”



Because they cast their nets so widely, LLMs can provide useful statistical predictions about how, in the main, ordinary people ordinarily use words and phrases in ordinary life.

Judge Kevin Newsom
Snell v. United Speciality Ins. Co.
(11th Cir. 2024)



LLMs for legal interpretation?

Skepticism

Optimism

Lee & Egbert (2025)

Choi (ms.)

Arbel & Hoffman (2024)

Engel & McAdams (2024)

Martínez (ms.)

Judge Kevin Newsom
*(Snell v. United Speciality Insurance,
Deleon v. US)*

Snell v. United Specialty Ins. Co.:



Is installing an in-ground trampoline “landscaping”?

Yes, installing an in-ground trampoline can be considered a part of landscaping. Landscaping involves altering the visible features of an outdoor area for aesthetic or practical purposes, and adding an in-ground trampoline would modify the appearance and function of the space...

ChatGPT (Queried by Judge Newsom)

Deleon v. US:



What is the ordinary meaning of “physically restrained”?

... The term [“physically restrained”] implies an active and physical intervention to control or confine a person’s movements...

ChatGPT (Queried by Judge Newsom)

“I do think—and increasingly so—that LLMs may well serve a valuable auxiliary role as we aim to triangulate ordinary meaning.”

- Judge Newsom, *Deleon v. US*



Nathan Schneider



Ethan Wilcox



Amir Zeldes



Kevin Tobia

B

What is the ordinary meaning of "physically restrained"?

"Direct Querying"

The ordinary meaning of "physically restrained" refers to when someone or something is held back or restricted from moving or acting freely.

Our headline claim: Judges should not do this.

Claude 3.7 Sonnet ▾ Choose style ▾



Waldon, Schneider, Wilcox, Zeldes, & Tobia (Georgetown Law Journal, forthcoming)

Part 1: Four myths about LLMs

- LLM training data reflects *naturalistic* language use.
- LLMs are optimized to produce correct statements about ‘ordinary’ language.
- LLM chatbots state accurate generalizations of their training data.
- LLMs are less biased than alternative interpretive tools.

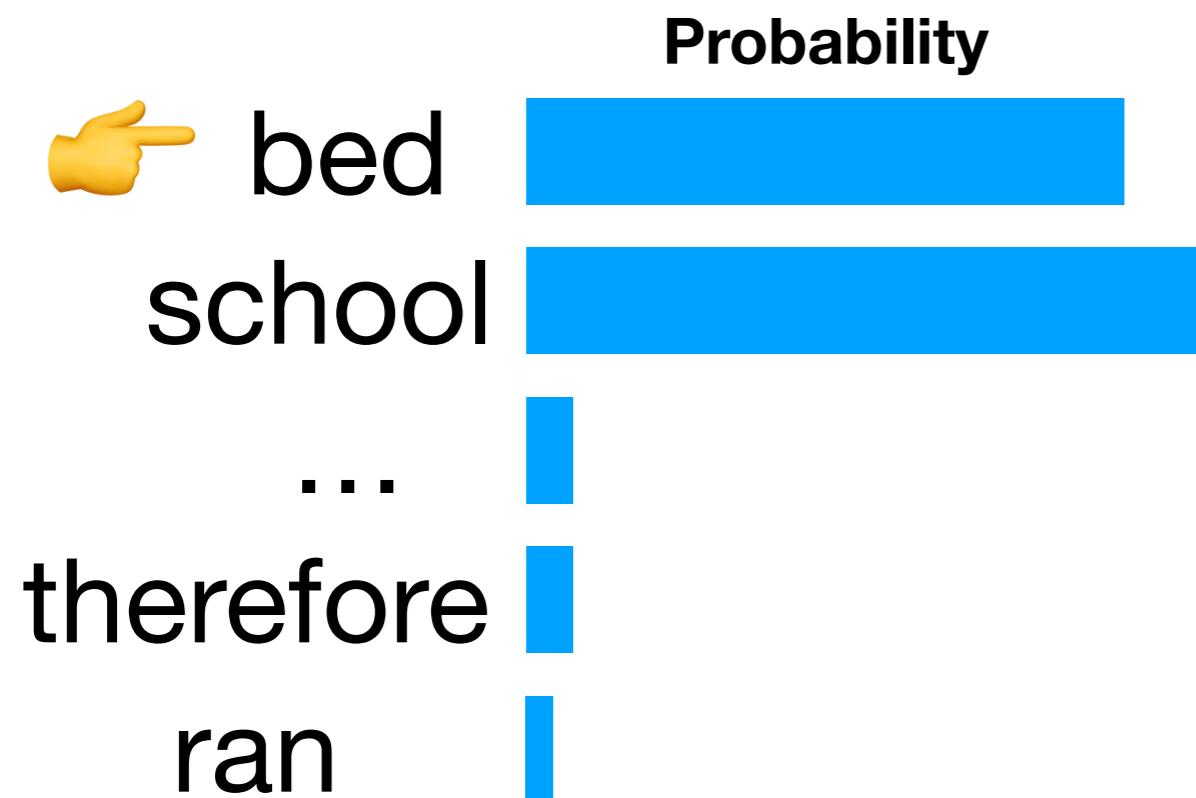
Part 2: Towards better practices

- Engage LLM-based applications as **dialectical partners, not linguistic authorities.**
- Case study:
The interpretation of *physically restrained*.

Language Model Systems

Input Context

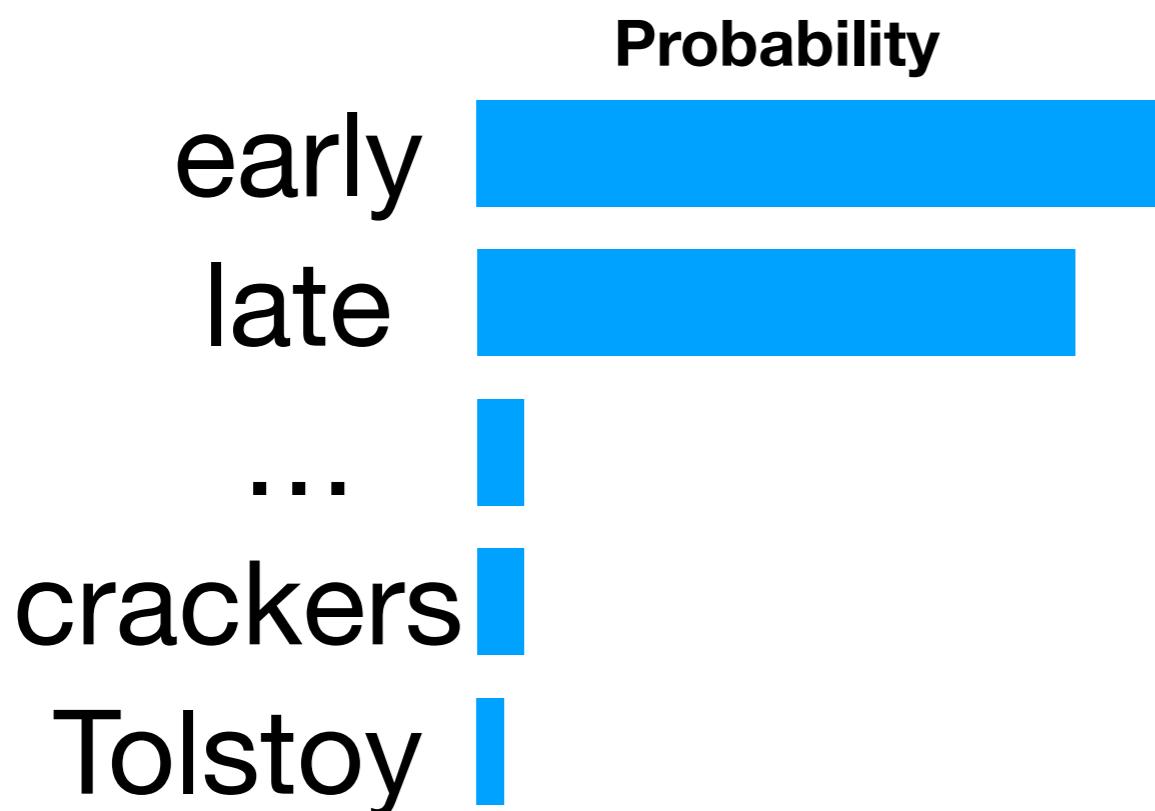
For a long time I went to _____



Language Model Systems

Input Context

For a long time I went to bed _____



LLM system:

- Text-probability mapping
- Sampling strategy

Language Model Platforms

ChatGPT ▾

Temporary

B

What can I help with?

[User Prompt]



Search

Reason

...



Input Context

[System Prompt:

????????????

] +

[User Prompt]

Text-probability mapping?

Sampling strategy?

Four myths about LLMs

- LLM training data reflects *naturalistic* language use.
- LLMs are optimized to produce correct statements about ‘ordinary’ language.
- LLM chatbots state accurate generalizations of their training data.
- LLMs are less biased than alternative interpretive tools.

Myth: LLM training data reflects *naturalistic* language use.

[T]he best reason to think that LLMs might provide useful information to those engaged in the interpretive enterprise... [is that] LLMs are quite literally “taught” using data that aim to reflect and capture how individuals use language in their everyday lives.

Specifically, [LLMs] train on a mind-bogglingly enormous amount of raw data taken from the internet... those data run the gamut from the highest-minded to the lowest, from Hemmingway novels and Ph.D. dissertations to gossip rags and comment threads.



Judge Kevin Newsom
Snell v. United Speciality Ins. Co.

Pre vs. post-training

Prompt: Write a haiku about the USC Gamecocks.

Pre-trained
LLM

Plausible system responses:

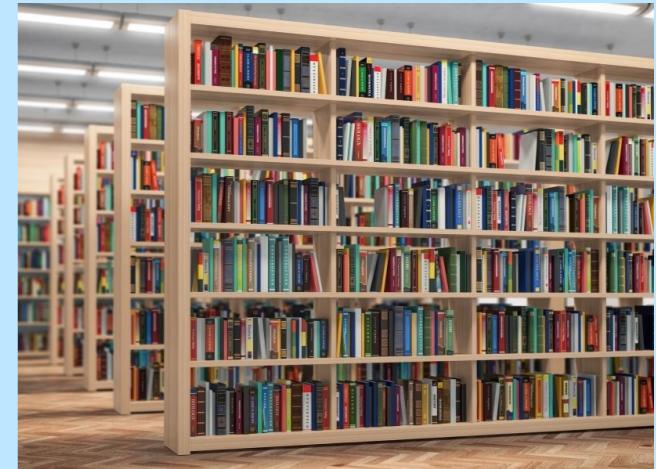
- *I need a first draft by 11pm tonight.*
 - *You could win up to \$500...*
- *Or write a limerick about the Clemson Tigers.*

Post-trained
LLM

Plausible system response:

*Spurs strike, feathers fly,
Williams-Brice echoes with pride—
Gamecocks never yield.*

Pre-training



Post-training

Instruction
tuning +

Preference
alignment



Huggingface H4 Instruction Dataset



Give me the answer to this trivia question: what team was LeBron James playing for when he broke the NBA all time scoring record?

Los Angeles
Lakers



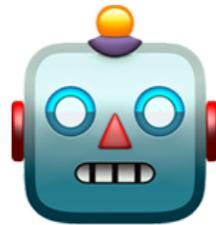
= written by a developer

Helpsteer Preference Dataset (NVIDIA)



What is the most useful resource in the world today?

The most useful resource in the world today is still water. It is essential for sustaining life, powering energy sources, and sustaining the environment.



annotator rating:

(Max:)

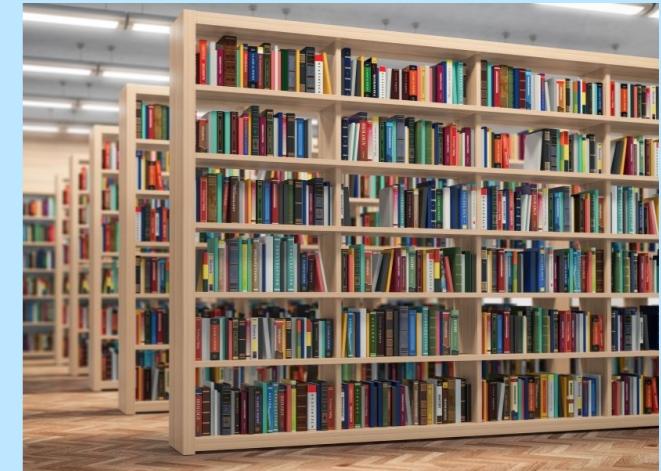


= developer



= LLM generated

Pre-training



Post-training

Instruction
tuning +

Preference
alignment

Synthetic data:

- Limited range of contexts.
- Produced by developers (and/or machines).
- Not intended to reflect ‘ordinary’ use.

For critical perspectives: Zhang et al. (2025); Juzek et al. (2025)

Four myths about LLMs

- LLM training data reflects *naturalistic* language use.
- LLMs are optimized to produce correct statements about ‘ordinary’ language.
- LLM chatbots state accurate generalizations of their training data.
- LLMs are less biased than alternative interpretive tools.

Huggingface H4 Instruction Dataset

Think of alternatives and paraphrases
for the underlined word.

Input: what we have expected

what we have projected
what we have anticipated
what we have waited
what we have hoped for
what we have accustomed

Helpsteer Preference Dataset (NVIDIA)

What are some words
that rhyme with “flung”?

Ring, Wing, King,
and Stung

Annotator rating: 
(Max: 

Four myths about LLMs

- LLM training data reflects *naturalistic* language use.
- LLMs are optimized to produce correct statements about ‘ordinary’ language.
- LLM chatbots state accurate generalizations of their training data.
- LLMs are less biased than alternative interpretive tools.

Because they cast their nets so widely, LLMs can provide useful statistical predictions about how, in the main, ordinary people ordinarily use words and phrases in ordinary life.

Judge Kevin Newsom
Snell v. United Speciality Ins. Co.

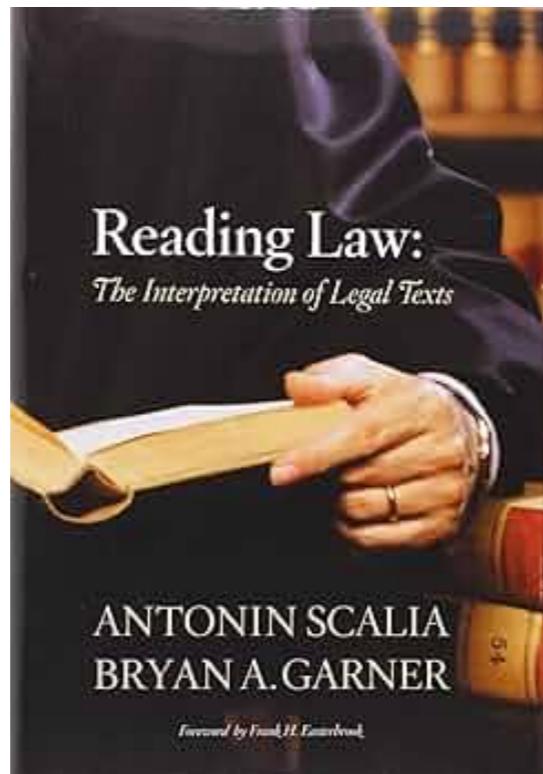
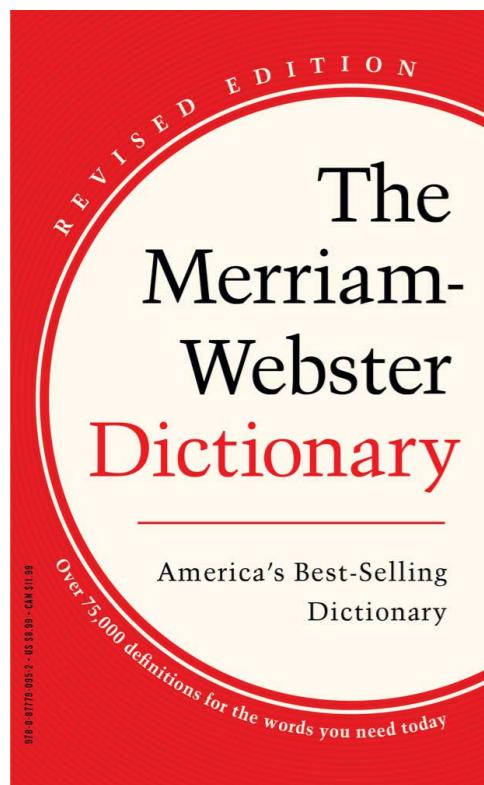


(ELQA Corpus, Behzad et al. 2023)

Metalinguistic statements produced by LLMs don't reflect underlying model probabilities (Hu & Levy, 2023)

Four myths about LLMs

- LLM training data reflects *naturalistic* language use.
- LLM chatbots are optimized to produce correct statements about ordinary language.
- LLM chatbots produce accurate generalizations of their training data.
- LLMs are less biased than alternative interpretive tools.



Dictionaries/canons can be cherry-picked.

(Llewellyn, 1950; Brudney & Baum, 2013;
inter alia)

• •

Newsom claims...



More transparent
(if one discloses prompt/model).

More consistent
(different models, ~similar responses).

(Cf. Lee & Egbert, 2025)

What can I help with?

[User Prompt]



Search



Reason

...



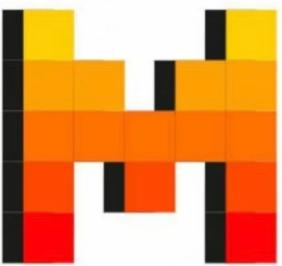
System prompt?

Text-probability mapping?

Sampling strategy?

Not reproducible!

Solution for transparency: use open-source models?



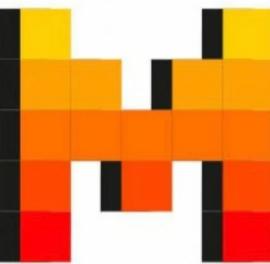
MISTRAL
AI_



**User has full knowledge/control of
text-probability mapping and sampling strategy.**

Samples are fully reproducible.

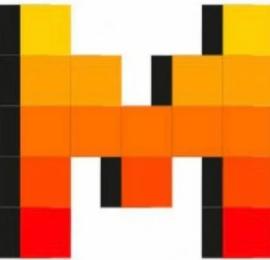
But problems remain.



Is installing an in-ground trampoline *landscaping*?

Installing an in-ground trampoline can be considered a form of landscaping...



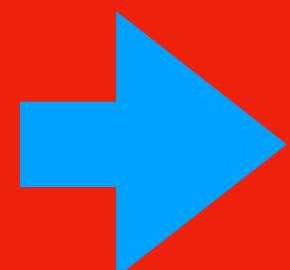


Is installing an in-ground trampoline *landscaping*?
If not, what would you call it?

Installing an in-ground trampoline is
not typically considered “landscaping”
in the traditional sense...



Subtle prompt variations can yield
qualitatively different responses.



LLMs are susceptible to
cherry-picking.

Part 1: Four myths about LLMs

- LLM training data reflects *naturalistic* language use.
- LLMs are optimized to produce correct statements about ‘ordinary’ language.
- LLM chatbots state accurate generalizations of their training data.
- LLMs are less biased than alternative interpretive tools.

Part 2: Towards better practices

- Engage LLM-based applications as **dialectical partners, not linguistic authorities.**
- Case study:
The interpretation of *physically restrained*.

A dialectical legal AI approach

- Direct querying treats LLMs as linguistic authorities to which we should show deference. **Dangerous!**
- ... But we can use LLMs as intellectual sparring partners (with ourselves, with each other) to encourage more sophisticated thinking about language. **Safer!**



Judge

Law clerk

Part 2: Towards better practices

- Engage LLM-based applications as **dialectical partners, not linguistic authorities.**
- Case study:
The interpretation of *physically restrained*.

US Sentencing Guidelines:

If a victim was **physically restrained**... [increase prison sentence].



'physically restrained' → use of physical force

Thesis:

What's the strongest argument **for** the claim?

Antithesis:

What's the strongest argument **against** the claim?

Synthesis: analysis (possibly) informed
by LLM interaction.

‘physically restrained’ → use of physical force

What's the strongest argument **for** the claim?

...The phrase “The police physically restrained the protester” implies that the police used physical force to prevent the protester from moving.

...[T]he phrase “police physically restrained the crowd” would not make sense if it meant that the police merely talked the crowd into dispersing.



‘physically restrained’ → use of physical force

What’s the strongest argument **against** the claim?

[A] child being told to stay in their room is “physically restrained” even if no physical force is used, and a person feeling unable to leave a crowded room due to social norms is also “physically restrained” in this broader sense.



'physically restrained' → use of physical force

Thesis:

What's the strongest argument **for** the claim?

Antithesis:

What's the strongest argument **against** the claim?

Synthesis: *physically restrained* is ambiguous!

‘Adv- restrained’ in the wild

Sen. Markwayne Mullin had to be **verbally restrained** by... Committee Chairman Bernie Sanders from getting into a fight with the president of the Teamsters during a heated committee hearing.

McConnell: Not My ‘Responsibility’ to Police Aggressive Behavior of Other Republicans, The Hill (Nov. 14, 2023)

verbally = means of restraint

restrained = passive participle verb

‘Adv- restrained’ in the wild

Because of an ill-considered comment by Frank Francisco, the Mets’ shaky closer, there was a heightened sense of drama in the ninth inning at Citi Field on Friday...

But the game was really decided eight innings earlier when a more pensive and **verbally restrained** pitcher [Andy Pettitte] had the worst inning of his season.

Mets Make Opening Statement Hold Up, N.Y. Times (June 22, 2012).

verbally = domain of restraint

restrained = (gradable) adjective

scenario:



physically:

Means of restraint

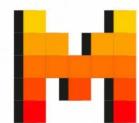
Domain of restraint

restrained:

Verb

Adjective

examples:

 MISTRAL
AI_

Police physically restrained the crowd.

Bill was (very) physically restrained at the party.

(4) (A) If any person was abducted to facilitate commission of the offense or to facilitate escape, increase by 4 levels; or (B) if any person was physically restrained to facilitate commission of the offense or to facilitate escape, increase by 2 levels.

US Sentencing Guidelines §2B3.1(b)(4)

Context strongly supports the means-of-restraint reading.



United States Sentencing Commission

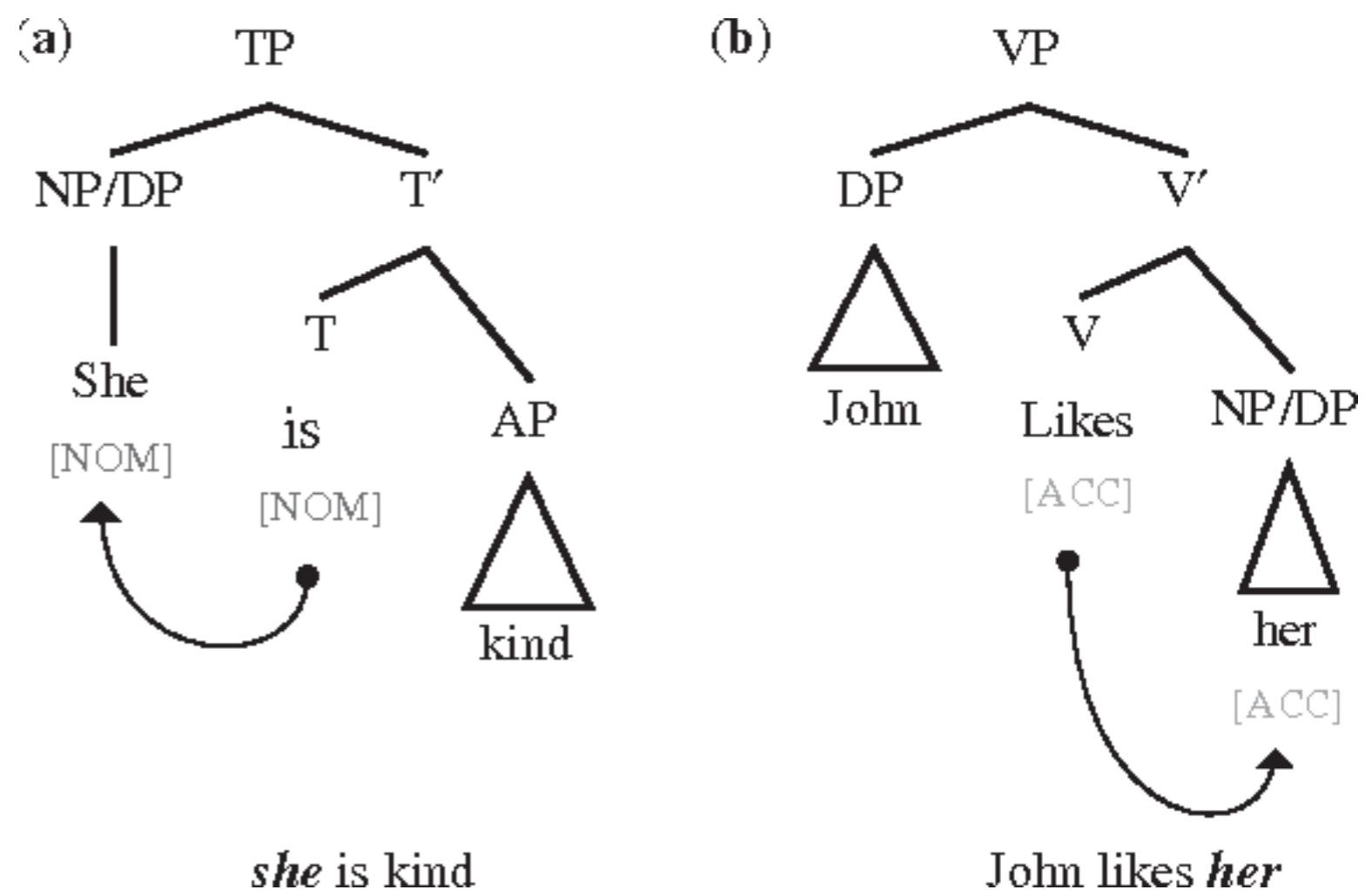
2024-2025 Amendment Cycle

Reply Public Comment on Certain
Proposed Amendments

90 FR 128

***Physically restrained
is ambiguous!***

**... And context resolves
the ambiguity in the
Sentencing Guidelines.**



Linguistics = the science that asks:
What does it mean to know a language?