Linguistics and Symbolic Computation in a World of Large Language Models

Thomas Graf

Stony Brook University
mail@thomasgraf.net
https://thomasgraf.net

OFAI
December 20, 2023

The talk in a nutshell

- Have large language models (**LLMs**) eaten linguists' lunch?

General answer

No, and there's lots of opportunities for knowledge transfer in both directions.

Specific answer

No, and **subregular syntax** is a concrete example of how linguistics can inform neural approaches.

The talk in a nutshell

- Have large language models (**LLMs**) eaten linguists' lunch?

General answer

No, and there's lots of opportunities for knowledge transfer in both directions.

Specific answer

No, and **subregular syntax** is a concrete example of how linguistics can inform neural approaches.

The talk in a nutshell

- Have large language models (**LLMs**) eaten linguists' lunch?

General answer

No, and there's lots of opportunities for knowledge transfer in both directions.

Specific answer

No, and **subregular syntax** is a concrete example of how linguistics can inform neural approaches.

Outline

1. The lay of the land: Symbolic VS neural Computation
   - An old fight: Symbolic VS neural computation
   - A positive vision of knowledge transfer
2. Locality and n-grams over strings
   - Strict locality (SL)
   - Tier-based strict locality (TSL)
3. Subregular Syntax: SL & TSL over Trees
   - Sentence structure
   - Merge is SL
   - Move is TSL

The past-tense debate is older than me...

456 Opinion TRENDS in Cognitive Sciences Vol.6 No.11 November 2002

The Past-Tense Debate

The past and future of the past tense

Steven Pinker and Michael T. Ullman

What is the interaction between storage and computation in language processing? What is the psychological status of grammatical rules? What are the relative strengths of connectionist and symbolic models of cognition? How are the components of language implemented in the brain? The English past tense has served as an arena for debates on these issues. We defend the theory that irregular past-tense forms are stored in the lexicon, a division of declarative memory, whereas regular forms can be computed by a concatenation rule, which requires the procedural system. Irregulars have the psychological, linguistic and neuropsychological signatures of lexical memory, whereas regulars often have the signatures of grammatical processing. Furthermore, because regular inflection is rule-driven, speakers can apply it whenever memory fails.

For fifteen years, the English past tense has been the subject of a debate on the nature of language processing. The debate began with the report of a connectionist model by Rumelhart and McClelland [1] and a critique by Pinker and Prince [2], and has since been the subject of many papers, conferences and simulation models [3–7] (see also McClelland and Patterson in this issue [8]).

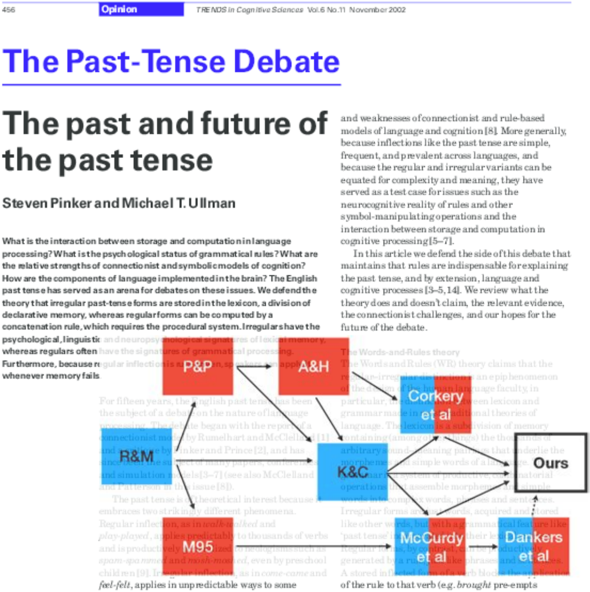
The past tense is of theoretical interest because it embraces two strikingly different phenomena. Regular inflection, as in *walk-walked* and *play-played*, applies predictably to thousands of verbs and is productively generalized to neologisms such as *spam-spammed* and *mosh-moshed*, even by preschool children [9]. Irregular inflection, as in *come-came* and *feel-felt*, applies in unpredictable ways to some and weaknesses of connectionist and rule-based models of language and cognition [8]. More generally, because inflections like the past tense are simple, frequent, and prevalent across languages, and because the regular and irregular variants can be equated for complexity and meaning, they have served as a test case for issues such as the neurocognitive reality of rules and other symbol-manipulating operations and the interaction between storage and computation in cognitive processing [5–7].

In this article we defend the side of this debate that maintains that rules are indispensable for explaining the past tense, and by extension, language and cognitive processes [3–5,14]. We review what the theory does and doesn't claim, the relevant evidence, the connectionist challenges, and our hopes for the future of the debate.

The Words-and-Rules theory

The Words and Rules (WR) theory claims that the regular–irregular distinction is an epiphenomenon of the design of the human language faculty, in particular, the distinction between lexicon and grammar made in most traditional theories of language. The lexicon is a subdivision of memory containing (among other things) the thousands of arbitrary sound–meaning pairings that underlie the morphemes and simple words of a language. The grammar is a system of productive, combinatorial operations that assemble morphemes and simple words into complex words, phrases and sentences. Irregular forms are just words, acquired and stored like other words, but with a grammatical feature like 'past tense' incorporated into their lexical entries. Regular forms, by contrast, can be productively generated by a rule, just like phrases and sentences. A stored inflected form of a verb blocks the application of the rule to that verb (e.g. *brought* pre-empts

The past-tense debate is older than me...



456 Opinion TRENDS in Cognitive Sciences Vol.6 No.11 November 2002

The Past-Tense Debate

The past and future of the past tense

Steven Pinker and Michael T. Ullman

What is the interaction between storage and computation in language processing? What is the psychological status of grammatical rules? What are the relative strengths of connectionist and symbolic models of cognition? How are the components of language implemented in the brain? The English past tense has served as an arena for debates on these issues. We defend the theory that irregular past-tense forms are stored in the lexicon, a division of declarative memory, whereas regular forms can be computed by a concatenation rule, which requires the procedural system. Irregulars have the psychological, linguistic and neuropsychological signatures of lexical memory, whereas regulars often have the signatures of grammatical processing. Furthermore, because regular inflection is a rule, speakers can apply it whenever memory fails.

For fifteen years, the English past tense has been the subject of a debate on the nature of language processing. The debate began with the report of a connectionist model by Rumelhart and McClelland [1] and a critique by Pinker and Prince [2], and has since been the subject of many papers, conferences and simulation models [3–7] (see also McClelland and Patterson in this issue [8]).

The past tense is of theoretical interest because it embraces two strikingly different phenomena. Regular inflection, as in *walk-walked* and *play-played*, applies predictably to thousands of verbs and is productively generalized to neologisms such as *spam-spammed* and *mosh-moshed*, even by preschool children [9]. Irregular inflection, as in *come-came* and *feel-felt*, applies in unpredictable ways to some

and weaknesses of connectionist and rule-based models of language and cognition [8]. More generally, because inflections like the past tense are simple, frequent, and prevalent across languages, and because the regular and irregular variants can be equated for complexity and meaning, they have served as a test case for issues such as the neurocognitive reality of rules and other symbol-manipulating operations and the interaction between storage and computation in cognitive processing [5–7].

In this article we defend the side of this debate that maintains that rules are indispensable for explaining the past tense, and by extension, language and cognitive processes [3–5, 14]. We review what the theory does and doesn't claim, the relevant evidence, the connectionist challenges, and our hopes for the future of the debate.

The Words-and-Rules theory

The Words and Rules (WR) theory claims that the regular-irregular distinction is an epiphenomenon of the design of the human language faculty, in particular, the distinction between lexicon and grammar made in traditional theories of language. The lexicon is a subdivision of memory containing (among other things) the thousands of arbitrary sound-meaning pairings that underlie the morphemes and simple words of a language. The grammar is a system of productive, combinatorial operations that assemble morphemes and simple words into complex words, phrases and sentences. Irregular forms are just words, acquired and stored like other words, but with a grammatical feature like 'past tense' in their lexical entries. Regular forms, by contrast, can be productively generated by a rule, just like phrases and sentences. A stored inflected form of a verb blocks the application of the rule to that verb (e.g. *brought* pre-empts

. . . and still haunts us. . .


The New York Times

OPINION

GUEST ESSAY

Noam Chomsky: The False Promise of ChatGPT

March 8, 2023

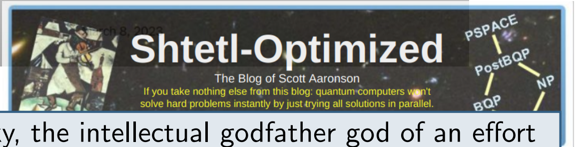
. . . and still haunts us. . .

. . . and still haunts us. . .


The New York Times

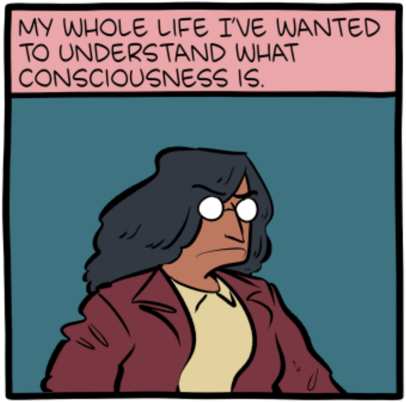
OPINION
GUEST ESSAY

Noam Chomsky: The False Promise of ChatGPT



"In this piece Chomsky, the intellectual godfather god of an effort that failed for 60 years to build machines that can converse in ordinary language, condemns the effort that succeeded."

. . . for a good reason



. . . for a good reason

. . . for a good reason

. . . for a good reason

Understanding matters: An alternate history of flight

- It is the 19th century, and airplanes aren't a thing yet.
- Three competitors at the first national flight competition:

Icarus Inc

Understanding matters: An alternate history of flight

- It is the 19th century, and airplanes aren't a thing yet.
- Three competitors at the first national flight competition:

Icarus Inc

Understanding matters: An alternate history of flight

- It is the 19th century, and airplanes aren't a thing yet.
- Three competitors at the first national flight competition:

Icarus Inc

Leonardo LLC

Understanding matters: An alternate history of flight

- It is the 19th century, and airplanes aren't a thing yet.
- Three competitors at the first national flight competition:

Icarus Inc

Leonardo LLC

Understanding matters: An alternate history of flight

- It is the 19th century, and airplanes aren't a thing yet.
- Three competitors at the first national flight competition:

Icarus Inc

Leonardo LLC

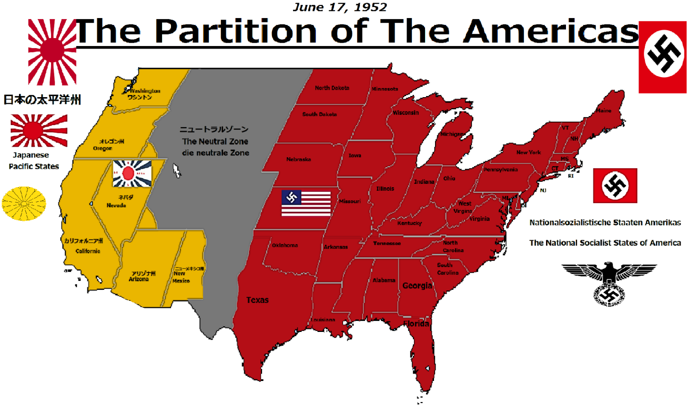
Timmy

An alternate history of flight [cont.]

- After Timmy's victory, **pogo sticks are all the rage**.
- Better and better pogo sticks hit the market.
- By 1930, the US is the world's leading pogo stick nation.

An alternate history of flight [cont.]

- By 1952, the US is occupied by Japan and Nazi Germany.
- Its pogo sticks were **no match for airplanes**.
- Nobody knows what happened to Timmy.



An emerging consensus

- LLMs work astonishingly well, but they are not perfect.
  hallucinations, bias, low resource languages
- LLMs are not just an engineering task.
- We need scientific understanding of these models.
- Linguistics and AI researchers are talking past each other.
- **What would positive collaboration look like?**

What do neural approaches do well?

- **LLMs model the actual** rather than the possible:

Phenomenon	**Linguistics**	**LLM**
lexical semantics	possible relations	actual relations
pronouns	possible referents	intended referent
word order	possible permutations	natural permutations

- (Generative) linguists focus on the possible range of variation
  $\Rightarrow$ lack of predictive, quantifiable theories of actual behavior

Example: Heavy NP shift

(1) a. Mary showed [$_{NP}$ the super-fancy car that she bought with her COVID relief check] to Sue.
b. Mary showed to Sue [$_{NP}$ the super-fancy car that she bought with her COVID relief check].

Linguists' focus precise theories of the syntactic configurations where a heavy NP may be shifted

Recent work non-syntactic factors that influence heavy NP shift

Still missing quantifiable theory of how those factors are weighted

- LLMs do the quantification, but we don't know their theory

Does linguistics still have anything to contribute?

- Linguists are already contributing by **assessing LLM performance**:
  - design linguistically challenging tasks
  - test on formal languages inspired by linguistic phenomena
  - probe LLM's internal representations
- **But**: this is all reactive instead of proactive
- **This talk**: use linguistic insights to break down complex phenomena into n-gram models for consumption by LLMs

The plan of action for the rest of the talk

1. n-grams are easy to work with for LLMs
2. n-grams cover a tremendous amount of ground in phonology
3. n-grams can be lifted from strings to trees for syntax
4. tree n-grams can be flattened back to string n-grams for LLMs

SL & TSL: (T)ier-Based Strictly Local

- Locality is perhaps the most central property of language.
- But locality can be measured in many ways.
- Formal language classes **SL** and **TSL** provide very strict notions of locality that
  - are grounded in n-grams, and
  - work well empirically.

Example: Word-final devoicing is SL-2

- Captured by forbidding voiced segments at the end of a word
- **German**: Don't have **z$** or **v$** or **d$** (where **$** = word edge).

Example: German

*$ r a **d** $

***z**$

***v**$

***d**$

Example: Word-final devoicing is SL-2

- Captured by forbidding voiced segments at the end of a word
- **German**: Don't have **z$** or **v$** or **d$** (where **$** = word edge).

Example: German

	*z$
* $ r a d $	*v$
	*d$

Example: Word-final devoicing is SL-2

- Captured by forbidding voiced segments at the end of a word
- **German**: Don't have **z$** or **v$** or **d$** (where **$** = word edge).

Example: German

*$\boxed{\$\ \text{r}}$ a **d** $

*z$

*v$

*d$

Example: Word-final devoicing is SL-2

- Captured by forbidding voiced segments at the end of a word
- **German**: Don't have **z$** or **v$** or **d$** (where **$** = word edge).

Example: German

* $ $\boxed{\text{r a}}$ **d** $

***z$**

***v$**

***d$**

Example: Word-final devoicing is SL-2

- Captured by forbidding voiced segments at the end of a word
- **German**: Don't have **z$** or **v$** or **d$** (where **$** = word edge).

Example: German

* $ r $\boxed{\text{a d}}$ $

***z$**

***v$**

***d$**

Example: Word-final devoicing is SL-2

- Captured by forbidding voiced segments at the end of a word
- **German**: Don't have **z$** or **v$** or **d$** (where **$** = word edge).

Example: German

* $ r a $\boxed{\textbf{d}\ \$}$

***z$**

***v$**

***d$**

Example: Word-final devoicing is SL-2

- Captured by forbidding voiced segments at the end of a word
- **German**: Don't have **z$** or **v$** or **d$** (where **$** = word edge).

Example: German

* **$** r a **d** **$**

***z$**

***v$**

***d$**

Example: Word-final devoicing is SL-2

- Captured by forbidding voiced segments at the end of a word
- **German**: Don't have **z$** or **v$** or **d$** (where **$** = word edge).

Example: German

	***z$**	
* **$** r a **d** **$**	***v$**	**$** r a t **$**
	***d$**	

Example: Word-final devoicing is SL-2

- Captured by forbidding voiced segments at the end of a word
- **German**: Don't have **z$** or **v$** or **d$** (where **$** = word edge).

Example: German

* $ r a **d** $

***z$**
***v$**
***d$**

$\boxed{\$\ \mathrm{r}}$ a t $

Example: Word-final devoicing is SL-2

- Captured by forbidding voiced segments at the end of a word
- **German**: Don't have **z$** or **v$** or **d$** (where $ = word edge).

Example: German

* $ r a **d** $

***z$**

***v$**

***d$**

$ r a t $

Example: Word-final devoicing is SL-2

- Captured by forbidding voiced segments at the end of a word
- **German**: Don't have **z$** or **v$** or **d$** (where **$** = word edge).

Example: German

* $ r a **d** $

***z$**
***v$**
***d$**

$ r $\boxed{\text{a t}}$ $

Example: Word-final devoicing is SL-2

- Captured by forbidding voiced segments at the end of a word
- **German**: Don't have **z$** or **v$** or **d$** (where **$** = word edge).

Example: German

	*z$	
* $ r a d $	*v$	$ r a $\boxed{\text{t \$}}$
	*d$	

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, . . .

Example: Northern Italian

* $ a s o l a $

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, . . .

Example: Northern Italian

* $ **a** **s** **o** l **a** $

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, . . .

Example: Northern Italian

* $ a s o l a $

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{\text{s,ʃ}\}$
  - $\mathrm{V} = \{\text{a,i,o,u}\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, . . .

Example: Northern Italian

* \$ $\boxed{\textbf{a s o}}$ l **a** \$

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, ...

Example: Northern Italian

* $ **a** **s** **o** l **a** $ $ **a** z **o** l **a** $

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{\text{s,ʃ}\}$
  - $\mathrm{V} = \{\text{a,i,o,u}\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, . . .

Example: Northern Italian

* $ **a** **s** **o** l **a** $ $\qquad$ $\boxed{\text{\$ a z}}$ **o** l **a** $

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, . . .

Example: Northern Italian

* $ **a s o** l a $ $ **a z o** l a $

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, ...

Example: Northern Italian

* $ a s o l a $ $ a z o l a $

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V[-voice]V}$
- **Suppose**:
  - $\mathrm{[-voice]} = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, ...

Example: Northern Italian

* $ **a** **s** **o** l **a** $ $ **a** z $\boxed{\text{o l a}}$ $

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, . . .

Example: Northern Italian

*$ **a** **s** **o** l **a** $ $ **a** z **o** $\boxed{\text{l a \$}}$

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, ...

Example: Northern Italian

*\$ **a s o** l **a** \$ \$ **a** z **o** l **a** \$ \$ **a** + **s o** c **i a** l **e** \$

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, . . .

Example: Northern Italian

*$ **a s o** l **a** $ $ **a** z **o** l **a** $ $\boxed{\$ \ \mathbf{a}\ +}$ **s o** c **i a** l **e** $

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, . . .

Example: Northern Italian

*$ **a** **s** **o** l **a** $ $ **a** z **o** l **a** $ $ $\boxed{\textbf{a} + \textbf{s}}$ **o** c **i** **a** l **e** $

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, . . .

Example: Northern Italian

*$ a s o l a $ $ a z o l a $ $ a $\boxed{+ \text{ s o}}$ c i a l e $

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, . . .

Example: Northern Italian

*$ a s o l a $ $ a z o l a $ $ a + $\boxed{\text{s o c}}$ i a l e $

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{\mathsf{s},\int\}$
  - $\mathrm{V} = \{\mathsf{a,i,o,u}\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, ...

Example: Northern Italian

*$ a s o l a $ $ a z o l a $ $ a + s $\boxed{\text{o c i}}$ a l e $

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, . . .

Example: Northern Italian

*$ a s o l a $ $ a z o l a $ $ a + s o c i a l e $

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{\mathsf{s},\int\}$
  - $\mathrm{V} = \{\mathsf{a,i,o,u}\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, . . .

Example: Northern Italian

*\$ a s o l a \$ \$ a z o l a \$ \$ a + s o c i a l e \$

Example: Intervocalic voicing is SL-3

- Captured by forbidding voiceless segments between vowels
- **Description:** don't have $\mathrm{V}[-\mathrm{voice}]\mathrm{V}$
- **Suppose**:
  - $[-\mathrm{voice}] = \{$s,ʃ$\}$
  - $\mathrm{V} = \{$a,i,o,u$\}$
- **Compiled out**: don't have **asa**, **aʃa**, **asi**, **aʃi**, . . .

Example: Northern Italian

*\$ a s o l a \$ \$ a z o l a \$ \$ a + s o c i a l e \$

A problem: Samala sibilant harmony

- If multiple sibilants occur in the same word, they must all be [+anterior] (s,z) or [−anterior] (ʃ,ʒ).
- In other words: Don't mix **purple** and **teal**.

  *sʃ *sʒ *zʃ *zʒ
  *ʃs *ʒs *ʃz *ʒz

- **But:** Sibilants can be arbitrarily far away from each other!

Example: Samala (Applegate 1972)

* $ h a s x i n t i l a w a ʃ $

$ h a ʃ x i n t i l a w a ʃ $

A problem: Samala sibilant harmony

- If multiple sibilants occur in the same word, they must all be [+anterior] (s,z) or [−anterior] (ʃ,ʒ).
- In other words: Don't mix **purple** and **teal**.

  *sʃ *sʒ *zʃ *zʒ
  *ʃs *ʒs *ʃz *ʒz

- **But:** Sibilants can be arbitrarily far away from each other!

Example: Samala (Applegate 1972)

*$ h a s x i n t i l a w a ʃ $

$ h a ʃ x i n t i l a w a ʃ $

A problem: Samala sibilant harmony

- If multiple sibilants occur in the same word, they must all be [+anterior] (s,z) or [−anterior] (ʃ,ʒ).
- In other words: Don't mix **purple** and **teal**.

  *sʃ *sʒ *zʃ *zʒ
  *ʃs *ʒs *ʃz *ʒz

- **But:** Sibilants can be arbitrarily far away from each other!

Example: Samala (Applegate 1972)

*$ h a s x i n t i l a w a ʃ $

$ h a ʃ x i n t i l a w a ʃ $

A problem: Samala sibilant harmony

- If multiple sibilants occur in the same word, they must all be [+anterior] (s,z) or [−anterior] (ʃ,ʒ).
- In other words: Don't mix **purple** and **teal**.

  *sʃ *sʒ *zʃ *zʒ
  *ʃs *ʒs *ʃz *ʒz

- **But:** Sibilants can be arbitrarily far away from each other!

Example: Samala (Applegate 1972)

*$ h a s x i n t i l a w a ʃ $

$ h a ʃ x i n t i l a w a ʃ $

A problem: Samala sibilant harmony

- If multiple sibilants occur in the same word, they must all be [+anterior] (s,z) or [−anterior] (ʃ,ʒ).
- In other words: Don't mix **purple** and **teal**.

  *sʃ *sʒ *zʃ *zʒ
  *ʃs *ʒs *ʃz *ʒz

- **But:** Sibilants can be arbitrarily far away from each other!

Example: Samala (Applegate 1972)

*$ h a s x i n t i l a w a ʃ $

$ h a ʃ x i n t i l a w a ʃ $

*$ s t a j a n o w o n o w a ʃ $

A problem: Samala sibilant harmony

- If multiple sibilants occur in the same word, they must all be [+anterior] (s,z) or [−anterior] (ʃ,ʒ).
- In other words: Don't mix **purple** and **teal**.

  *sʃ *sʒ *zʃ *zʒ
  *ʃs *ʒs *ʃz *ʒz

- **But:** Sibilants can be arbitrarily far away from each other!

Example: Samala (Applegate 1972)

* $ h a $\boxed{\text{s x i n t i l a w a ʃ}}$ $

$ h a $\boxed{\text{ʃ x i n t i l a w a ʃ}}$ $

* $ $\boxed{\text{s t a j a n o w o n o w a ʃ}}$ $

Making long-distance dependencies local

- Let's take a clue from phonology: create locality with **tiers**. (Heinz et al. 2011)
- comparable to **skip n-grams**

Jeff Heinz

Example: Samala Revisited

1. Project sibilant tier
2. *sʃ, *sʒ, *zʃ, *zʒ, *ʃs, *ʒs, *ʃz, *ʒz

* $ h a s x i n t i l a w a ʃ $ $ h a ʃ x i n t i l a w a ʃ $

Making long-distance dependencies local

- Let's take a clue from phonology: create locality with **tiers**. (Heinz et al. 2011)
- comparable to **skip n-grams**

Jeff Heinz

Example: Samala Revisited

1. Project sibilant tier
2. *sʃ, *sʒ, *zʃ, *zʒ, *ʃs, *ʒs, *ʃz, *ʒz

*$ h a s x i n t i l a w a ʃ $ $ h a ʃ x i n t i l a w a ʃ $

Making long-distance dependencies local

- Let's take a clue from phonology: create locality with **tiers**. (Heinz et al. 2011)
- comparable to **skip n-grams**

Jeff Heinz

Example: Samala Revisited

1. Project sibilant tier
2. *sʃ, *sʒ, *zʃ, *zʒ, *ʃs, *ʒs, *ʃz, *ʒz

$ s ʃ $

*$ h a s x i n t i l a w a ʃ $ $ h a ʃ x i n t i l a w a ʃ $

Making long-distance dependencies local

- Let's take a clue from phonology: create locality with **tiers**. (Heinz et al. 2011)
- comparable to **skip n-grams**

Jeff Heinz

Example: Samala Revisited

1. Project sibilant tier
2. *sʃ, *sʒ, *zʃ, *zʒ, *ʃs, *ʒs, *ʃz, *ʒz

$ s ʃ $

*$ h a s x i n t i l a w a ʃ $

$ ʃ ʃ $

$ h a ʃ x i n t i l a w a ʃ $

Making long-distance dependencies local

- Let's take a clue from phonology: create locality with **tiers**. (Heinz et al. 2011)
- comparable to **skip n-grams**

Jeff Heinz

Example: Samala Revisited

1. Project sibilant tier
2. *sʃ, *sʒ, *zʃ, *zʒ, *ʃs, *ʒs, *ʃz, *ʒz

$ s ʃ $ $ ʃ ʃ $

*$ h a s x i n t i l a w a ʃ $ $ h a ʃ x i n t i l a w a ʃ $

Making long-distance dependencies local

- Let's take a clue from phonology: create locality with **tiers**. (Heinz et al. 2011)
- comparable to **skip n-grams**

Jeff Heinz

Example: Samala Revisited

1. Project sibilant tier
2. *sʃ, *sʒ, *zʃ, *zʒ, *ʃs, *ʒs, *ʃz, *ʒz

```
 $   s   ʃ $          $  ʃ          ʃ $
*$ h a s x i n t i l a w a ʃ $    $ h a ʃ x i n t i l a w a ʃ $
```

Making long-distance dependencies local

- Let's take a clue from phonology: create locality with **tiers**. (Heinz et al. 2011)
- comparable to **skip n-grams**

Jeff Heinz

Example: Samala Revisited

1. Project sibilant tier
2. *sʃ, *sʒ, *zʃ, *zʒ, *ʃs, *ʒs, *ʃz, *ʒz

$ s ʃ $ $\qquad$ $\boxed{\$\ ʃ}$ ʃ $

* $ h a s x i n t i l a w a ʃ $ $\qquad$ $ h a ʃ x i n t i l a w a ʃ $

Making long-distance dependencies local

- Let's take a clue from phonology: create locality with **tiers**. (Heinz et al. 2011)
- comparable to **skip n-grams**

Jeff Heinz

Example: Samala Revisited

1. Project sibilant tier
2. *sʃ, *sʒ, *zʃ, *zʒ, *ʃs, *ʒs, *ʃz, *ʒz

$ s ʃ $

* $ h a s x i n t i l a w a ʃ $

$ ʃ ʃ $

$ h a ʃ x i n t i l a w a ʃ $

Making long-distance dependencies local

- Let's take a clue from phonology: create locality with **tiers**. (Heinz et al. 2011)
- comparable to **skip n-grams**

Jeff Heinz

Example: Samala Revisited

1. Project sibilant tier
2. *sʃ, *sʒ, *zʃ, *zʒ, *ʃs, *ʒs, *ʃz, *ʒz

$ s ʃ $ $\qquad$ $ ʃ ʃ $

*$ h a s x i n t i l a w a ʃ $ $\qquad$ $ h a ʃ x i n t i l a w a ʃ $

Tiers have many interpretations

1. **Separate structure**

$$\$ \xrightarrow{\triangleleft} \mathbf{z} \xrightarrow{\triangleleft} \mathbf{d} \xrightarrow{\triangleleft} \int \xrightarrow{\triangleleft} \$$$

2. **Additional relation**

$$\begin{tikzcd}
\$ \arrow[r, "\triangleleft"] & \mathbf{z} \arrow[r, "\triangleleft"] \arrow[rr, bend left, "\triangleleft_T"] & \mathrm{i} \arrow[r, "\triangleleft"] & \mathbf{d} \arrow[r, "\triangleleft"] \arrow[rr, bend left, "\triangleleft_T"] & \mathrm{a} \arrow[r, "\triangleleft"] & \int \arrow[r, "\triangleleft"] & \$
\end{tikzcd}$$

3. **Substructures**

- without tier: \$z, zi, id, da, aʃ, ʃ\$
  n-grams
- with tier: \$z, zd, dʃ, ʃ\$
  ≈ **skip n-grams**

Metaphor

SL + TLS =
LLM with encoder that extracts n-grams and specific skip-ngrams

Outline

Sentences have structure

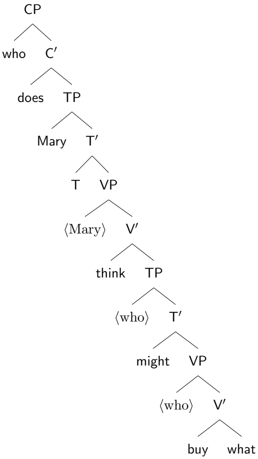
(2) Who does Mary think might buy what?

- **Selection**: head-argument relations
  what is the argument of *buy*
- **Displacement**: elements are not always pronounced where they are interpreted
  who is the subject of *buy* but pronounced further left

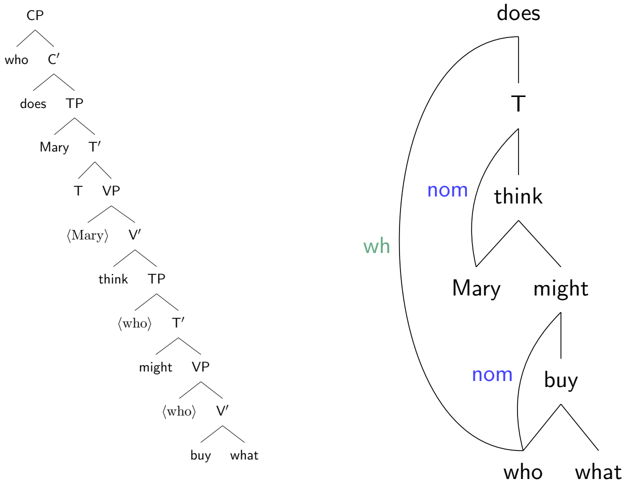
Chomskyan terminology

- Selection $\equiv$ Merge
- Displacement $\equiv$ Move

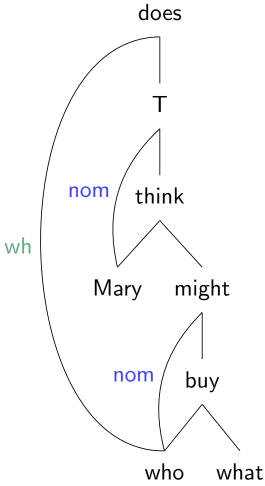
Representing sentence structure



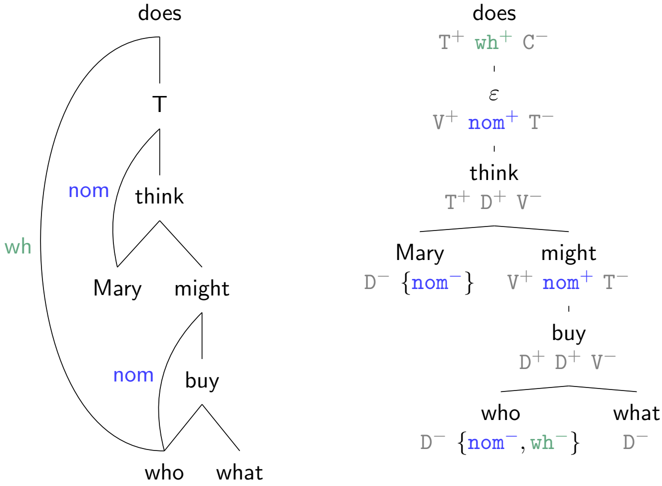
Representing sentence structure



Using diacritics to arborize dependency graphs



Using diacritics to arborize dependency graphs



Merge is SL-2

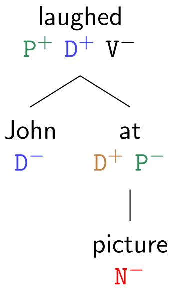
Merge is SL-2 over trees because we only need to look at

1. the mother and
2. its string of daughters

Merge is SL-2

Merge is SL-2 over trees because we only need to look at

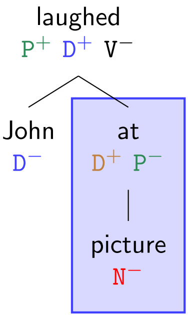
1. the mother and
2. its string of daughters



Merge is SL-2

Merge is SL-2 over trees because we only need to look at

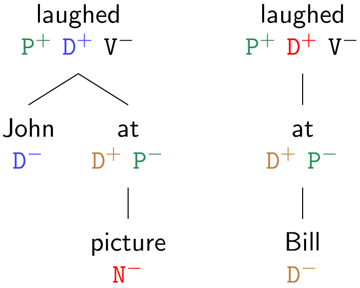
1. the mother and
2. its string of daughters



Merge is SL-2

Merge is SL-2 over trees because we only need to look at

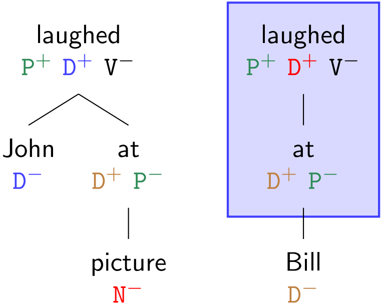
1. the mother and
2. its string of daughters



Merge is SL-2

Merge is SL-2 over trees because we only need to look at

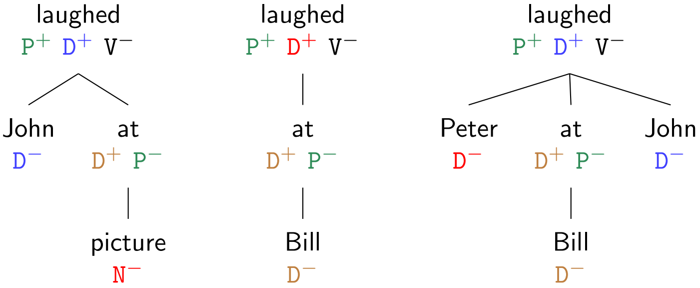
1. the mother and
2. its string of daughters



Merge is SL-2

Merge is SL-2 over trees because we only need to look at

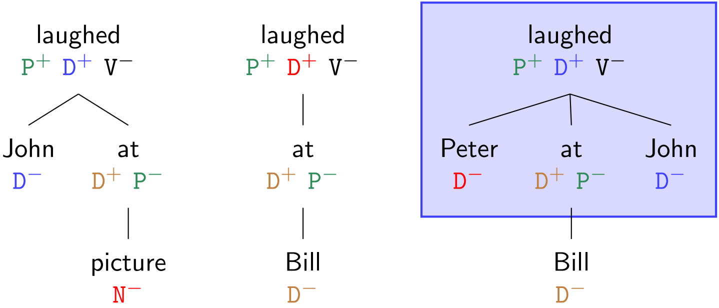
1. the mother and
2. its string of daughters



Merge is SL-2

Merge is SL-2 over trees because we only need to look at

1. the mother and
2. its string of daughters



Stringifying Merge

1. Add a relation connecting each head to its string of daughters.
2. This gives you a list of *Merge n-grams*.
3. Check that there are no illicit Merge n-grams.

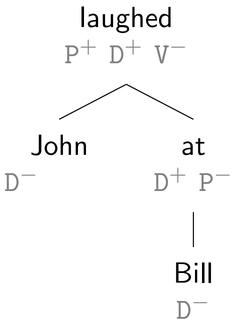
Merge n-grams for
Mary, John doubts that Bill doubts that Sue doubts

- **0 arguments**
  Mary \$, John \$, Bill \$, Sue \$
- **1 argument**
  that doubts (occurs twice), \$ doubts
- **2 arguments**
  doubts John that, doubts Bill that, doubts Sue Mary

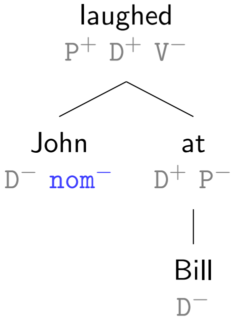
(empty heads and features are omitted for the ease of exposition)

⇒ **selection only needs Merge n-grams, not full tree**

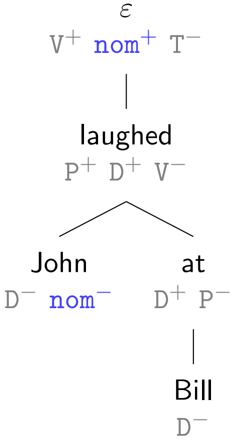
Move is local over tree tiers



Move is local over tree tiers

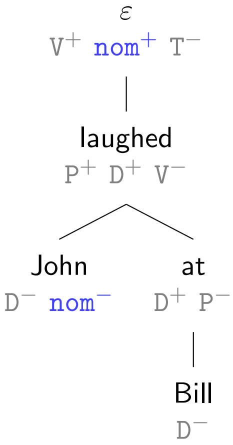


Move is local over tree tiers

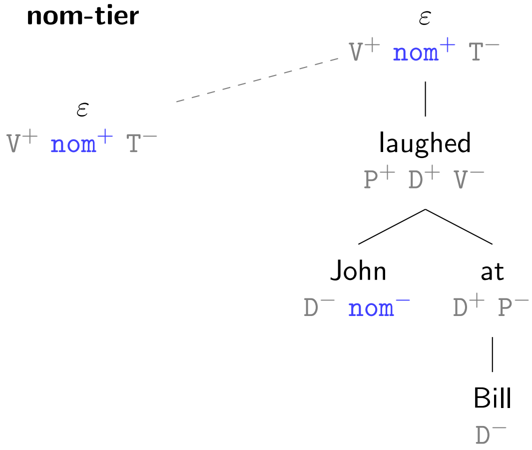


Move is local over tree tiers

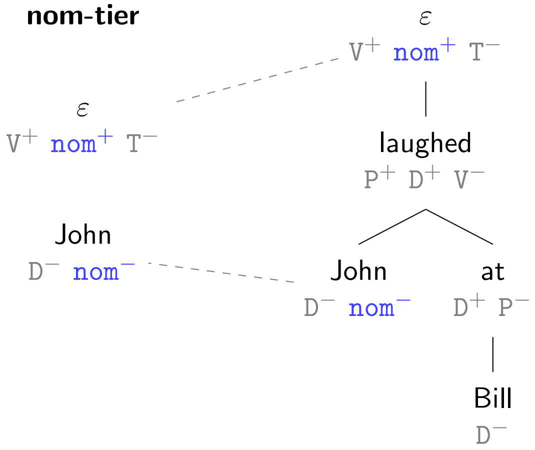
nom-tier



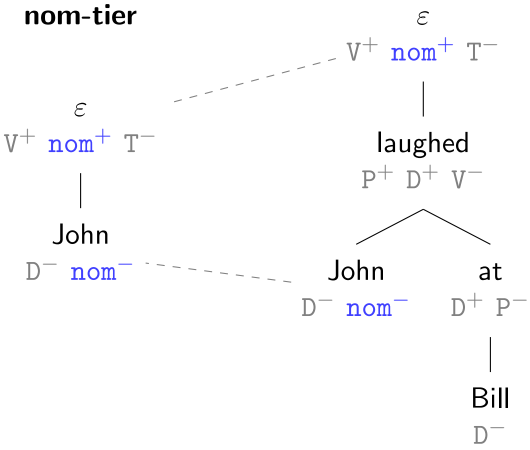
Move is local over tree tiers



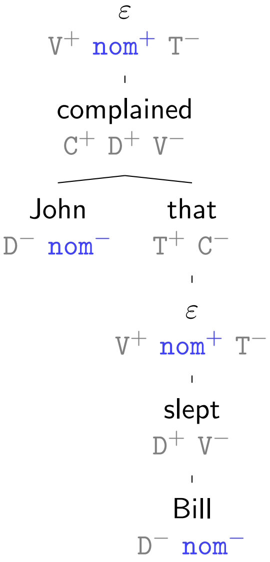
Move is local over tree tiers



Move is local over tree tiers

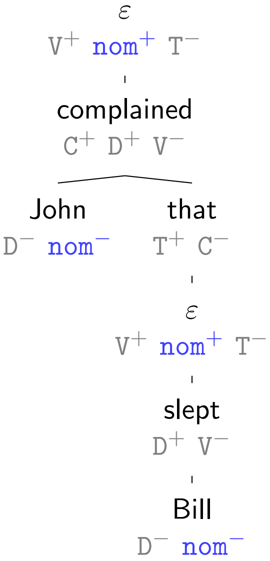


Tier with multiple movers



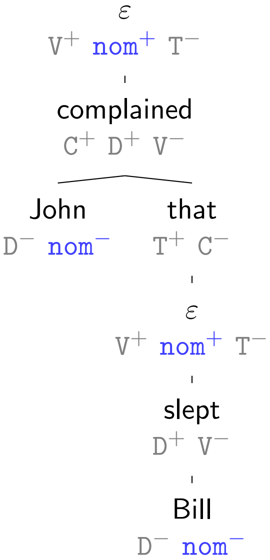
Tier with multiple movers

nom-tier



Tier with multiple movers

nom-tier



Tier with multiple movers

nom-tier

ε

$\mathtt{V}^{+}\ \mathtt{nom}^{+}\ \mathtt{T}^{-}$

John

$\mathtt{D}^{-}\ \mathtt{nom}^{-}$

ε

$\mathtt{V}^{+}\ \mathtt{nom}^{+}\ \mathtt{T}^{-}$

complained

$\mathtt{C}^{+}\ \mathtt{D}^{+}\ \mathtt{V}^{-}$

John

$\mathtt{D}^{-}\ \mathtt{nom}^{-}$

that

$\mathtt{T}^{+}\ \mathtt{C}^{-}$

ε

$\mathtt{V}^{+}\ \mathtt{nom}^{+}\ \mathtt{T}^{-}$

slept

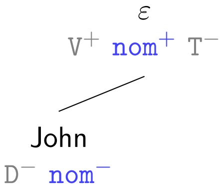
$\mathtt{D}^{+}\ \mathtt{V}^{-}$

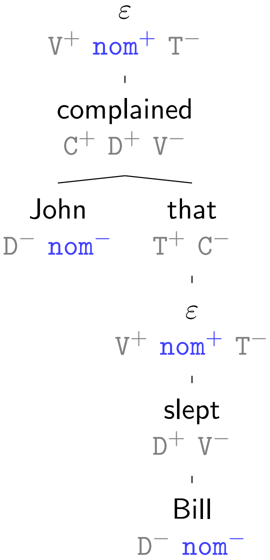
Bill

$\mathtt{D}^{-}\ \mathtt{nom}^{-}$

Tier with multiple movers

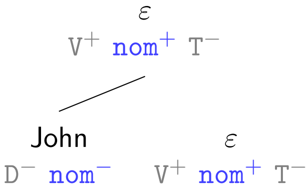
nom-tier

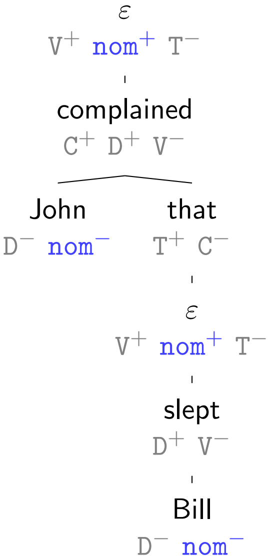




Tier with multiple movers

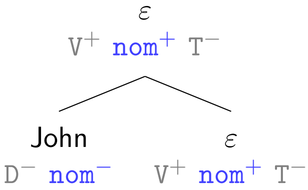
nom-tier

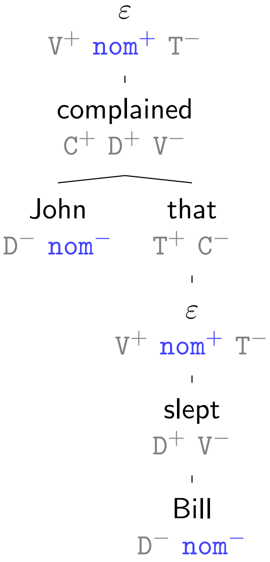




Tier with multiple movers

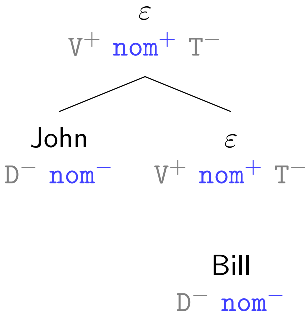
nom-tier

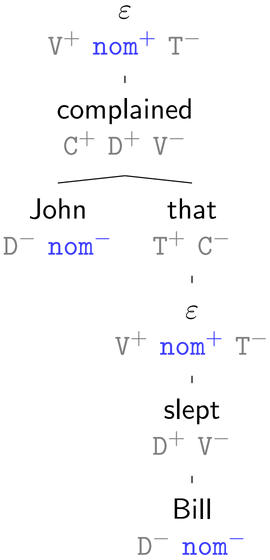




Tier with multiple movers

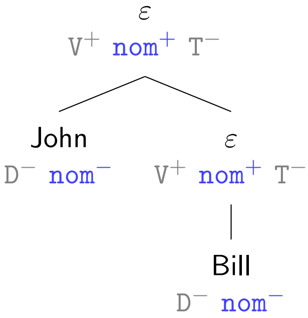
nom-tier

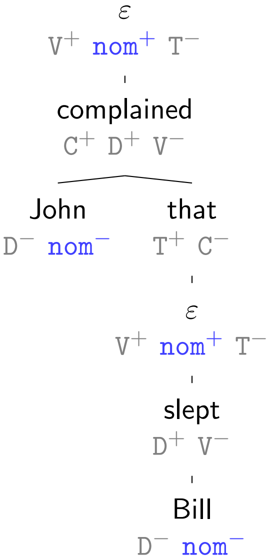




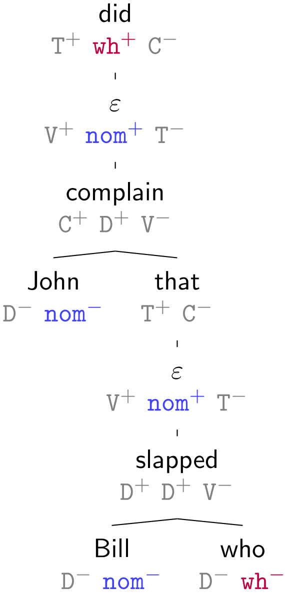
Tier with multiple movers

nom-tier



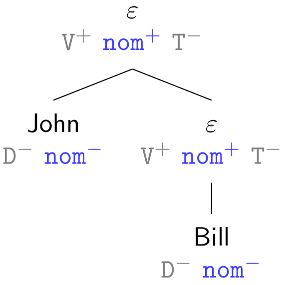


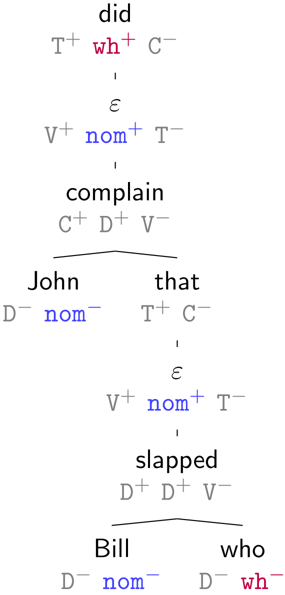
Separate Tier for Each Movement Type



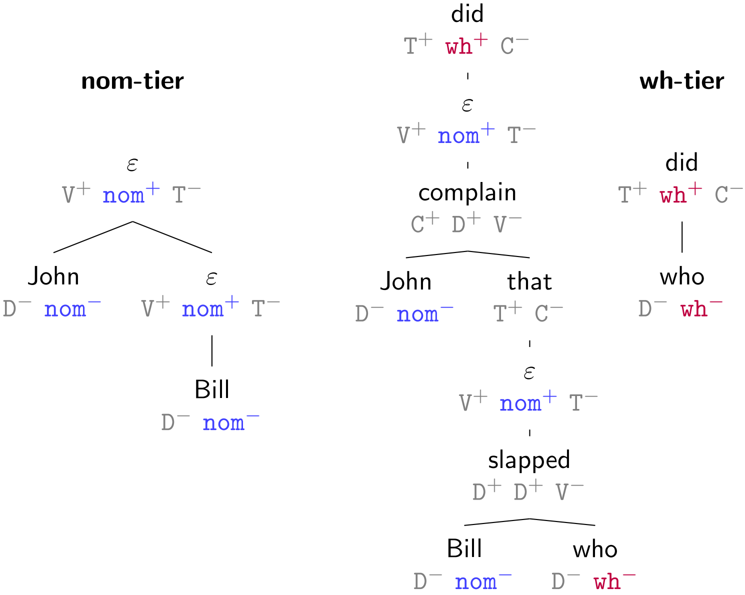
Separate Tier for Each Movement Type

nom-tier





Separate Tier for Each Movement Type



Move is TSL-2

- We now know how to construct movement tiers.
- Licit movement only creates tiers of a specific shape.
- **Move is TSL-2 over trees**:
  1. Only $\mathtt{f}^+$ has a tier daughter string containing $\mathtt{f}^-$.
  2. Every $\mathtt{f}^+$ has exactly one $\mathtt{f}^-$ in its tier daughter string.

Stringifying Move on the `f`-tier

- Add a relation that
  1. connects each node on `f`-tier to its string **s** of `f`-tier daughters,
  2. **except** that **s** is filtered down to only those daughters that carry $\mathtt{f}^-$ (i.e. the $\mathtt{f}^-$ string tier)

Example

`wh` n-grams for **Who what did John say that* $\langle who \rangle$ *bought* $\langle what \rangle$

- \$ did[$\mathtt{wh}^+$]
- did[$\mathtt{wh}^+$] who[$\mathtt{wh}^-$] what[$\mathtt{wh}^-$]
- who[$\mathtt{wh}^-$] \$
- what[$\mathtt{wh}^-$] \$

$\Rightarrow$ **displacement only needs Move n-grams, not full tree**

How far does stringification go?

- Stringification of syntax is not limited to Merge and Move.
- All of the following also can be stringified:
  1. island effects
  2. agreement (upward, downward, defective intervention)
  3. case assignment
  4. wh-agreement in Irish
  5. extraction morphology in Wolof, Chumash, . . .
  6. German wh-copying
  7. multiple wh-movement in Bulgarian, Russian, . . .
  8. NPI licensing
  9. binding

The general upshot

- Most of syntax does not need full tree structure.
- All we need is n-grams with more sophisticated relations than string successor.

But don't we still have tree structure in there?

No we only have n-grams,
distinct trees can have the same n-grams

Yes we don't know how to compute those n-grams
without creating the trees first

Next step

- Maybe LLMs are using Merge/Move n-grams instead of trees
- What if we took a treebank, converted it to Merge/Move n-grams, and trained an LLM on that instead?

Conclusion

- LLMs and linguistics can and do inform each other.
- LLMs provide models of the actual behavior,
  rather than the possible behavior.
- Linguistics is not limited to assessing LLMs' performance.
- Linguistic analysis provides new ways of looking at language
  that are suitable for LLMs.

Acknowledgments

This work is supported by the National Science Foundation under Grant No. BCS-1845344.

Another TSL boon: Blocking

- TSL can also handle blocking effects.
- **Slovenian sibilant harmony with blocking**
  1. **[−ant]** ... **[+ant]** is illicit,
  2. unless **t** or **d** intervenes.
- **TSL-2 account**
  1. project all **[−ant]**, **[+ant]**, **t**, and **d**
  2. don't have **[−ant] [+ant]**

Example: Slovenian (Jurgec 2011; McMullin 2016)

*$ **s** p i **ʃ** $ $ **z** i **d** a **ʃ** $

Another TSL boon: Blocking

- TSL can also handle blocking effects.
- **Slovenian sibilant harmony with blocking**
  1. [−ant] . . . [+ant] is illicit,
  2. unless t or d intervenes.
- **TSL-2 account**
  1. project all [−ant], [+ant], t, and d
  2. don't have [−ant] [+ant]

Example: Slovenian (Jurgec 2011; McMullin 2016)

$ s ʃ $

*$ s p i ʃ $ $ z i d a ʃ $

Another TSL boon: Blocking

- TSL can also handle blocking effects.
- **Slovenian sibilant harmony with blocking**
  1. **[−ant]** . . . **[+ant]** is illicit,
  2. unless **t** or **d** intervenes.
- **TSL-2 account**
  1. project all **[−ant]**, **[+ant]**, **t**, and **d**
  2. don't have **[−ant]** **[+ant]**

Example: Slovenian (Jurgec 2011; McMullin 2016)

$ **s** ʃ $	$ **z** **d** ʃ $
*$ **s** p i ʃ $	$ **z** i **d** a ʃ $

Islands come for free

Two Fundamental Questions of Syntax

- Why do islands exist?
- Why do island exceptions exist?

A computational argument

1. Movement requires the power of TSL-2.
2. TSL-2 can model islands as blocking effects.
3. The cognitive ability for movement
   entails the cognitive ability for islands.

Island examples

(3) * Which car did John complain [**because** Bill damaged t].

(4) * Which car did John deny [the **fact** that Bill damaged t].

(5) Which car did John drive Mary crazy [**while** trying to fix t].

Island examples

(3) * Which car did John complain [**because** Bill damaged t].

(4) * Which car did John deny [the **fact** that Bill damaged t].

(5) Which car did John drive Mary crazy [**while** trying to fix t].

did
$\mathtt{T}^{+}\ \mathtt{wh}^{+}\ \mathtt{C}^{-}$
|
because
$\mathtt{T}^{+}\ \mathtt{V}^{\sim}$
|
which
$\mathtt{D}^{-}\ \mathtt{wh}^{-}$

Island examples

(3) * Which car did John complain [**because** Bill damaged t].

(4) * Which car did John deny [the **fact** that Bill damaged t].

(5) Which car did John drive Mary crazy [**while** trying to fix t].

did
$\mathtt{T}^{+}\ \mathtt{wh}^{+}\ \mathtt{C}^{-}$
|
because
$\mathtt{T}^{+}\ \mathtt{V}^{\sim}$
|
which
$\mathtt{D}^{-}\ \mathtt{wh}^{-}$

did
$\mathtt{T}^{+}\ \mathtt{wh}^{+}\ \mathtt{C}^{-}$
|
fact
$\mathtt{C}^{+}\ \mathtt{N}^{-}$
|
which
$\mathtt{D}^{-}\ \mathtt{wh}^{-}$

Island examples

(3) * Which car did John complain [**because** Bill damaged t].

(4) * Which car did John deny [the **fact** that Bill damaged t].

(5) Which car did John drive Mary crazy [**while** trying to fix t].

did
$\mathtt{T}^{+}\ \mathtt{wh}^{+}\ \mathtt{C}^{-}$
|
because
$\mathtt{T}^{+}\ \mathtt{V}^{\sim}$
|
which
$\mathtt{D}^{-}\ \mathtt{wh}^{-}$

did
$\mathtt{T}^{+}\ \mathtt{wh}^{+}\ \mathtt{C}^{-}$
|
fact
$\mathtt{C}^{+}\ \mathtt{N}^{-}$
|
which
$\mathtt{D}^{-}\ \mathtt{wh}^{-}$

did
$\mathtt{T}^{+}\ \mathtt{wh}^{+}\ \mathtt{C}^{-}$
|
which
$\mathtt{D}^{-}\ \mathtt{wh}^{-}$

Impossible Islands

- Islands arise when a blocker is projected onto a tier.
- Tier projection only considers lexical item itself, not its structural context
- TSL-2 theory of islands hence rules out:
    - **Gang-up islands**
      "A mover can escape n islands, but not more than that."
    - **Configurational islands**
      "An adjunct is an island iff it is inside an embedded clause."
    - **Cowardly islands**
      "An adjunct is an island iff there are at least two adjuncts in the clause."
    - **Rationed islands**
      "Only one adjunct per clause can be an island."
    - **Discerning islands**
      "Adjuncts only block movers that contain an adjective."

References I

Applegate, Richard B. 1972. *Ineseño Chumash grammar*. Doctoral Dissertation, University of California, Berkeley.

Heinz, Jeffrey, Chetan Rawal, and Herbert G. Tanner. 2011. Tier-based strictly local constraints in phonology. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 58–64. URL `http://www.aclweb.org/anthology/P11-2011`.

Jurgec, Peter. 2011. *Feature spreading 2.0: A unified theory of assimilation*. Doctoral Dissertation, University of Tromsø.

McMullin, Kevin. 2016. *Tier-based locality in long-distance phonotactics: Learnability and typology*. Doctoral Dissertation, University of British Columbia.