

---

# Clausal word order in Cherokee: a corpus approach and its formal implications

Brian Hsu and Benjamin Frey (UNC-Chapel Hill)  
UC Davis Language Group  
November 19, 2021



# Organization of the talk

---

1. Overview
2. Key questions in Cherokee syntax
3. The corpus of narratives
4. Probabilistic patterns in word order
5. Implications for generative theories
6. Conclusion



# 1. Overview

---

**Big question:** What are the principles that determine word order in Cherokee clauses?

- It is well known that the ordering of major constituents in the clause is **highly flexible** (Feeling and Pulte 1975; Scancarelli 1986; Montgomery-Anderson 2015; Akkus 2018)
- However, the principles that determine speakers' preferences among possible orders are not as well studied.



# 1. Overview

---

**Big question:** What are the principles that determine word order in Cherokee clauses?

- Conflicting claims in prior descriptions:
  - Order is determined by information structure alone (Scancarelli 1986).
  - Order is influenced by thematic structure (Feeling and Pulte 1975).
- Both views are challenged by substantial variability in the orders produced by speakers.



# 1. Overview

---

**Proposal:** We obtain a superior description and explanation of Cherokee grammar by examining *quantitative variability* in a corpus of narratives.

- Word order is determined probabilistically by features related to **both information structure and thematic role**.
- Multiple factors can interact **cumulatively** to influence word order preferences.



# 1. Overview

---

Implications for language documentation and conservation efforts:

- A more accurate description of Cherokee word order principles beyond “flexibility” alone.
- Regular trends in corpus patterns can inform pedagogical materials (Frey 2020).
- The corpus approach can be replicated to describe and compare other indigenous American languages with understudied word order variability.



# 1. Overview

---

Implications for **generative syntactic theory**:

- The Cherokee pattern of *probabilistic and cumulative* variability supports a theory that integrates:
  - (i) Structures and principles of generative syntax.
  - (ii) A probabilistic weighted constraint grammar (e.g. Maximum Entropy Harmonic Grammar)



# Organization of the talk

---

1. Overview
2. Key questions in Cherokee syntax
3. The corpus of narratives
4. Probabilistic patterns in word order
5. Implications for generative theories
6. Conclusion



## 2. Key questions in Cherokee syntax

---

Cherokee (Southern Iroquoian) is indigenous to western North Carolina and adjacent areas.

- Currently severely endangered:
  - ~2000 speakers in Oklahoma
  - ~180 speakers in North Carolina (as of 2021)

Source: <https://mountainx.com/news/ Cherokee-language/>



## 2. Key questions in Cherokee syntax

---

Often classified as **polysynthetic**:

- Verbs are *highly inflected* for multiple properties, including: (combined) subject and object agreement, tense, aspect, and evidentiality.
- Single words can express full propositions.

(1) ᵼᵼᵼ4ᵼᵼᵼᵼᵼ

y-oj-agasesdan-el-e

NEG-1.EXCL-pay.attention-DAT-REPP

‘We didn’t pay much attention to it.’ (Feeling et al. 2017: 12)



## 2. Key questions in Cherokee syntax

---

Most existing work on Cherokee focuses on its morphology and phonology; there are few descriptions of syntax above the word level (Feeling & Pulte 1975, Scancarelli 1986, Beghelli 1996, Montgomery-Anderson 2015, Akkus 2018).

All descriptions concur that the **ordering of major constituents** (main verbs, nominal and prepositional phrases, adverbials) in a clause is highly flexible.





## 2. Key questions in Cherokee syntax

---

Many grammatical descriptions focus on ordering restrictions on subjects (S), objects (O), verbs (V)

- A quick glance shows that all orders of S, O, V are possible (Scancarelli 1986)

(3) YG      OʔPE  
kilo    utvsohnv    u-dlv-g-v  
some old.man    3B-sick-PROG-EXP  
'an old man was sick'                      (Feeling 2017: 22)



## 2. Key questions in Cherokee syntax

---

Many grammatical descriptions focus on ordering restrictions on subjects (S), objects (O), verbs (V)

- A quick glance shows that all orders of S, O, V are possible (Scancarelli 1986)

(4) <b>ShLŃ</b>	<b>ŃAʔ</b>	<b>hoǃŃŃ</b>
ka-níitaʔtívʔi	wi-uu-kooh-éʔi	jíistvna
3A-tail	TRN-3B-saw.CMP-NXP	crawdad
‘the crawdad saw the wolf’s tail.’ (M-A 2008:552)		



## 2. Key questions in Cherokee syntax

---

Scancarelli (1986): Cherokee word order is determined by **information structure** properties:

- Ex: Items that denote *new information* or *contrast* occur earlier in the clause than those that do not.



## 2. Key questions in Cherokee syntax

---

Representative excerpt: In this clause, *gitli* ‘dogs’ is **new information** introduced for the first time in this story, and precedes the verb and subject.

- (1a) **YCO**      hAꞑ      Oꞑꞑꞑꞑꞑꞑ      Oꞑꞑꞑꞑꞑꞑꞑꞑ      Joꞑꞑ  
**gitli-hnv**    nigolv    julsihnvd    u-n-adeytohdih-e    gusd.  
dog-top    always    nightly      3-pl-bother-repP    something  
‘Every night something bothered their dogs.’



## 2. Key questions in Cherokee syntax

---

Representative excerpt: In a later clause of the narrative, *gitli* 'dogs' is now **discourse-given**, and it follows the verb.

(1b)	LS0 <sup>o</sup>	D4	DθhP <sup>o</sup> θI <sup>o</sup>	<b>YC</b>
	Hleg-hnv	ase	a-n-anhdlvs-g-e	<b>gitli.</b>
	While-topic	maybe	3-pl-lie.down-prog-repP	dog
	'The dogs would lie down for a while.'			



## 2. Key questions in Cherokee syntax

---

Feeling and Pulte (1975) describe “ordinary” word order in **thematic terms**.

"In simple declarative sentences in Cherokee, the **subject of the sentence ordinarily precedes the verb** with its modifiers and objects. In addition, objects of verbs ordinarily precede the verb, resulting in **subject-object-verb** word order."

(p. 353)

- This order can be “rearranged” for information-structure purposes.



## 2. Key questions in Cherokee syntax

---

Feeling and Pulte (1975) describe “ordinary” word order in **thematic terms**:

- “Subjects” refer to **agents** (of transitive verbs): entities that actively cause or initiate an event.
- “Objects” refer to **themes** (of transitive verbs): entities that are affected by an event, or undergo a change of state.



## 2. Key questions in Cherokee syntax

---

Example of *agent* > *theme* > *verb* order: In this sentence, the agent and theme are equally accessible in the discourse, and both bear contrast.

ᎠᏍᎠᏲᏍᎠ	ᎠᏍᎠᏲᏍᎠ	ᎠᏍᎠᏲᏍᎠ	ᎠᏍᎠᏲᏍᎠ
[sagwu-no]	[j-u-n-atana ahwi]		d-a-hih-e.
one-TOP	PL-3-PL-big	deer	3-PL-kill-REPP

“One (of the men) killed big deer.”





## 2. Key questions in Cherokee syntax

---

**Unresolved question:** To what extent do these factors determine word order, and how do they interact?

- Scancarrelli (1986): Cherokee follows the **newsworthiness principle** (Mithun 1992):  
  
“In a number of languages, **the order of constituents does not reflect their syntactic functions at all**, but rather their pragmatic functions. ” (Mithun 1992; 58)
- This denies any role of thematic role in ordering.



## 2. Key questions in Cherokee syntax

---

**Methodological issues in prior work:** It is hard to identify which factor(s) determine word order in a given sentence, even if discourse context is considered.

- All nominal phrases have **multiple grammatical properties** related to animacy, thematic role, newness, contrast, etc.
- It is difficult or impossible to isolate which property among many is responsible for a particular observed word order.



## 2. Key questions in Cherokee syntax

---

**Methodological issues in prior work:** It is hard to identify which factor(s) determine word order in a given sentence, even if discourse context is considered.

- Grammatical rules can show **probabilistic variability**; they may not always apply, even when all input properties are constant.
- Word order restrictions can result from **cumulative interactions** of  $>1$  property, in both *probabilistic* (Ellsiepen and Bader 2018) and *categorical* (Murphy 2018) patterns.



## 2. Key questions in Cherokee syntax

---

We investigate word order in Cherokee by building and analyzing an **annotated corpus** of narratives:

- A large data set makes it possible to *observe* and *quantify* the propensities of various grammatical properties to determine word order, using statistical measures (including a logistic regression)
- Similar studies on other languages: Tonhauser and Colijn (2010) on Paraguayan Guaraní; Harbour, Watkins, and Adger (2012) on Kiowa.



# Organization of the talk

---

1. Overview
2. Key questions in Cherokee syntax
- 3. The corpus of narratives**
4. Probabilistic patterns in word order
5. Implications for generative theories
6. Conclusion



# 3. The corpus of narratives

---

The corpus contains 15 narratives: 12 from Feeling et al. (2017), 3 from Montgomery-Anderson (2008):

- All transcribed, spoken narratives of a variety of topics, by 9 speakers (8 from OK, 1 from NC)
- Both works provide a *morphologically segmented gloss*, and English translation for all texts.
- These consist primarily of personal narratives and folk tales, each told by a single speaker.



# 3. The corpus of narratives

---

In these narratives, we tagged every major constituent that is:

- **A nominal expression**, which can refer to an identifiable entity in the world.
- **A thematic element** that has a thematic relation with the verbal predicate (these items were not necessarily referential or nominal).



# 3. The corpus of narratives

---

The corpus has 580 total sentences. A large majority of them have only one tagged major constituent other than the verb (we ignored sentences with only a verb).

- Sentences with one major constituent: 410 (~71%)
- Sentences with two major constituents: 140
- Sentences with three major constituents: 23
- Sentences with four major constituents 7



# 3. The corpus of narratives

---

519 constituents were tagged for identifiable values of the following properties (the independent variables of the quantitative analysis):

- INFORMATION STRUCTURE: Does the NP denote new information, given information, an accessible item, or a nonspecific referent.
- THEMATIC ROLE: Agent, theme, goal, time, location, predicate subject, predicate object.



# 3. The corpus of narratives

---

519 constituents were tagged for identifiable values of the following properties (the independent variables of the quantitative analysis):

- Most annotation procedures and tags are based on guidelines in Dipper, Götze, and Skopeteas (2007)



# 3. The corpus of narratives

---

Each item was tagged for 2 word order values, the dependent variables of the analysis:

- PREVERBAL, preceding the verb.
- POSTVERBAL, following the verb.



# Organization of the talk

---

1. Overview
2. Key questions in Cherokee syntax
3. The corpus of narratives
- 4. Probabilistic patterns in word order**
5. Implications for generative theories
6. Conclusion



# 4. Probabilistic patterns in word order

---

There is an overall preference for most phrases to precede verbs, across all types of information status and thematic role.

Out of 519 total tagged major constituents:

- 365 (70%) preverbal.
- 154 (30%) postverbal.



# 4. Probabilistic patterns in word order

---

Tags for INFORMATION STRUCTURE: **New** entities mentioned for the first time.

- Context: The narrative has discussed people staying at a house. Dogs are being mentioned for the first time.

**gitli-hnv** nigolv julsihnvd u-n-adeytohdih-e gusd.  
dog-TOP always nightly 3-PL-bother-REPP something  
'Every night something bothered their dogs.'

(Feeling et al. 2017; Transformation)



# 4. Probabilistic patterns in word order

---

Tags for INFORMATION STRUCTURE: **Given** entities that were previously mentioned explicitly

- Context: The dogs in this story have been previously mentioned.

galjode      w-u-ni-deysdih-e      **gitli.**  
house      tr-3-pl-run.around-repP      dog  
'The dogs would run around the house'

(Feeling et al. 2017; Transformation)



# 4. Probabilistic patterns in word order

---

Tags for INFORMATION STRUCTURE: **Given** entities that were previously mentioned explicitly

Further sub-classifications:

- GIVEN-ACTIVE: Mentioned in the same, or previous sentence.
- GIVEN-INACTIVE: Mentioned before the previous sentence.

While the distinction is relevant for patterns in some languages (Bader 2020), it does not have a noticeable or significant effect in our Cherokee corpus.



# 4. Probabilistic patterns in word order

---

Tags for INFORMATION STRUCTURE: **Accessible** entities have not been mentioned, but reference can be inferred from a relation with a given entity, or general knowledge.

- Context: A group of four women has been previously mentioned, but none of them individually.

No        **sowu**        j-ajiy-anvh-e  
Then    one        PL-3.PAS-call-EXP

'One (of the women) was called.'

(Feeling et al. 2017; Spearfinger)



# 4. Probabilistic patterns in word order

---

Tags for INFORMATION STRUCTURE: **Accessible** entities have not been mentioned, but reference can be inferred from a relation with a given entity, or general knowledge.

- Context: The previous sentence first mentions guns. Bullets are inferable as a typical associate/subpart of them.

**Na**      **gani**      de-g-vwanih-v

The      bullet      PL-3-hit-EXPP

'The bullets struck (her).'

(Feeling et al. 2017; Spearfinger)



## 4. Probabilistic patterns in word order

---

Tags for INFORMATION STRUCTURE: **Accessible** entities have not been mentioned, but reference can be inferred from a relation with a given entity, or general knowledge.

- Context: The narrator is describing the story's setting. The moon has not previously been mentioned, but is presumably part of the listener's general knowledge.

[**Svnoyi**    **eh-i**    **nvda**] vsgwu igahi            u-tisd-v?i  
Evening    be-AG    sun            also            brightly            3-shine-EXP  
'The moon was shining brightly.'

(Feeling et al. 2017; Ball of Fire)



# 4. Probabilistic patterns in word order

---

Tags for INFORMATION STRUCTURE: **Nonreferential** items do not refer an identifiable entity. These are typically generics, or references to events.

- Context: The pronoun and relative clause are thematic arguments of the verb, but they refer to events rather than entities.

[**Vsgi-hno**] n-u-lstani-dol-v

[**hi?a** **ji-ji-noheh-a**]

This-topic SPEC-3-happen-around-EXP this REL-1-LIVE-PRES.

'This is what happened in this story.'

(Feeling et al. 2017; Cat Meowing)



# 4. Probabilistic patterns in word order

---

**Effect of information structure:** NEW and NON-REFERENTIAL phrases are more likely to occur pre-verbally than GIVEN and ACCESSIBLE items.

	Accessible	Given-active	Given-inactive	New	Non-referential	Total
Postverbal	29	64	34	24	14	165
Preverbal	70	95	55	100	84	404
<b>Percent preverbal</b>	<b>70%</b>	<b>59%</b>	<b>62%</b>	<b>80%</b>	<b>86%</b>	



# 4. Probabilistic patterns in word order

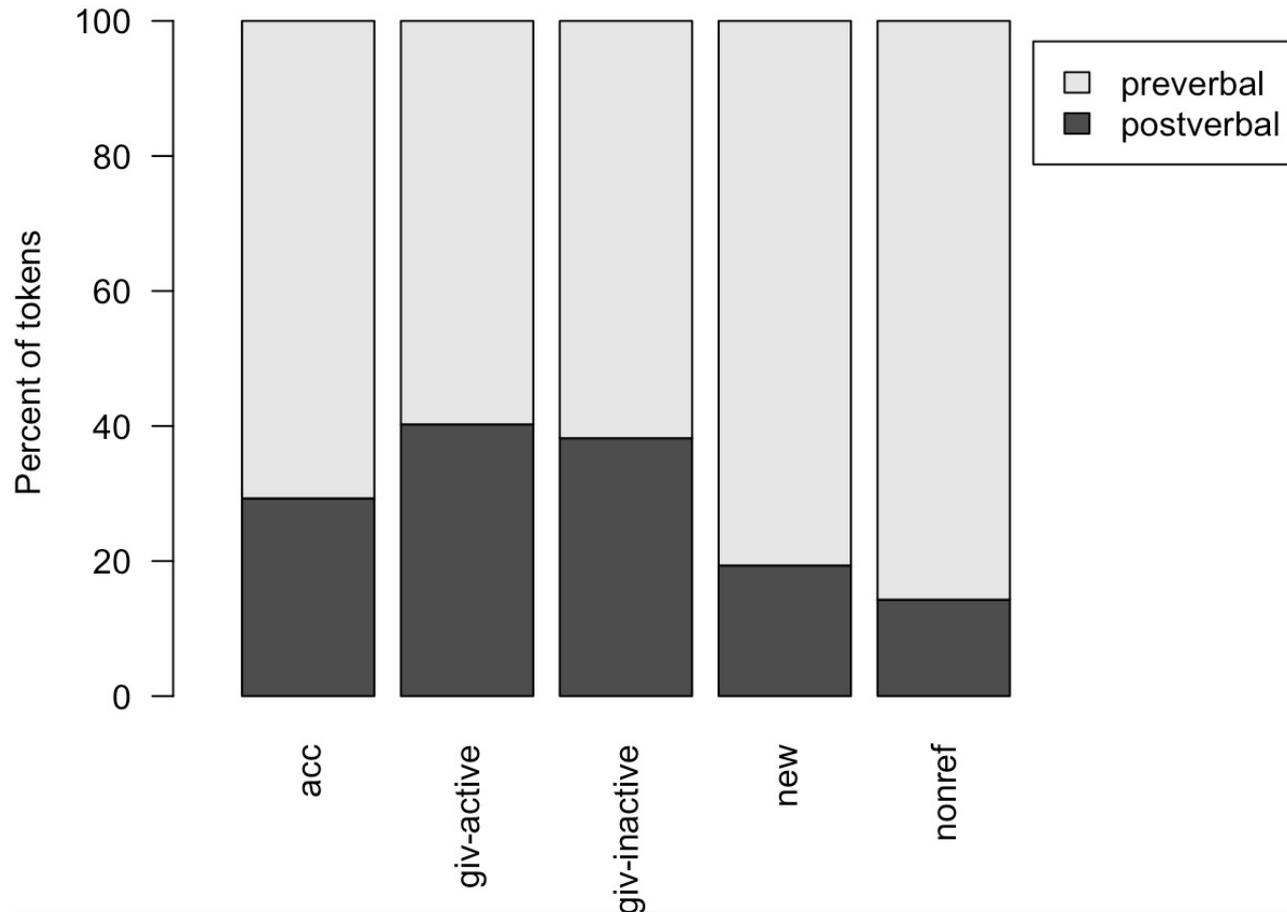
---

The regression model finds a significant difference between ACCESSIBLE vs. NEW and NONREFERENTIAL, no significant difference between accessible and given.

	Accessible	Given-active	Given-inactive	New	Non-referential	Total
Postverbal	29	64	34	24	14	165
Preverbal	70	95	55	100	84	404
<b>Percent preverbal</b>	<b>70%</b>	<b>59%</b>	<b>62%</b>	<b>80%</b>	<b>86%</b>	



# 4. Probabilistic patterns in word order



# 4. Probabilistic patterns in word order

---

Key tags for THEMATIC ROLE: **Agents** refer to entities that actively cause or initiate an event.

[**hi?a**   **a-ni-ta?li**   **a-ni-sgaya**]      a-ni-gawehih-e  
this      3-PL-two      3-PL-man      3-PL-paddle-REPP  
'Two men were paddling.'

(Feeling et al. 2017; Water Beast)



# 4. Probabilistic patterns in word order

---

Tags for THEMATIC ROLE: **Themes** are entities that are affected by an event.

- Includes *objects of transitive verbs*, and sole arguments of change-of-location (go, arrive etc.), change-of-state (fall, die, break etc.) predicates.

No kil **am** ji-todis-g-o

Then until water 1-heat.water-PROG-HAB

'then I heat some water.'

(Feeling et al. 2017; How to Make Chestnut Bread)



# 4. Probabilistic patterns in word order

---

Tags for THEMATIC ROLE: **Themes** are entities that are affected by an event.

- Includes objects of transitive verbs, and sole *arguments of change-of-location* (go, arrive etc.), change-of-state (fall, die, break etc.) *predicates*.

w-u-n-vsgoj-v-gwu

TR-3-PL-go.out-EXP-just

'My brother just went out.'

**j-osd-adanvli**

PL-1.DUAL.EXCL-brother

(Feeling et al. 2017; Transformation)



# 4. Probabilistic patterns in word order

---

Tags for THEMATIC ROLE: **Themes** are entities that are affected by an event.

- Includes objects of transitive verbs, and sole *arguments of change-of-location* (go, arrive etc.), *change-of-state* (fall, die, break etc.) *predicates*.

d-u-hlihgwadinel-e **jiyu.**                      d-u-ni-gvje-hno                      **a-ni-sgay.**

pl-3-turn.over-repP canoe                      pl-3-pl-fall.in-topic                      3-pl-man

'The canoe turned over, and the men fell (into the water).'

(Feeling et al. 2017; Transformation)



# 4. Probabilistic patterns in word order

---

Tags for THEMATIC ROLE: **Goals** are the recipient of an action, or endpoint in space of an event.

[**Jog**                      **akti**]                      a-ni-gawehih-e  
Upstream                      toward                      3-PL-paddle-REPP  
'They were paddling upstream.'      (Feeling et al. 2017: Water Beast)

[**Didanelv**]                      w-awadinvs-v  
Home                      TR-throw-EXP  
'I threw (it) towards home.'              (Feeling et al. 2017: Throw it Home)



## 4. Probabilistic patterns in word order

---

Tags for THEMATIC ROLE: **Predicate-subjects** are the “modified” entities of predicative or equative verbs like *to be* or *be named*.

An sgwu d-u-do?-v [agi-ji ji-ges-v]  
Ann also PL-3-be.named-EXP 1.POS-mother REL-be-EXP  
“My mother was also named Ann.”

(Feeling et al. 2017: Transformation)

[Sgi-hnv] asuhnidoh ge-hv  
He-TOP fisher be-EXPP  
“He was a fisher.”

(Feeling et al. 2017: Throw it Home)



## 4. Probabilistic patterns in word order

---

Tags for THEMATIC ROLE: **Predicate-objects** are the objects of predicative or equative verbs like *to be* or *be named*.

[An] sgwu d-u-do?-v                      agi-ji                      ji-ges-v  
Ann also PL-3-be.named-EXP      1.POS-mother      REL-be-EXP  
“My mother was also named Ann.”  
(Feeling et al. 2017: Transformation)

Sgi-hnv      [asuhnidoh]      ge-hv  
He-TOP      fisher                      be-EXPP  
“He was a fisher.”                      (Feeling et al. 2017: Throw it Home)



# 4. Probabilistic patterns in word order

---

Tags for THEMATIC ROLE: **times** refer to the delimiting time or duration of an event.

[**Hleg-hnv ase**]      a-n-anhdlvs-g-e      gitli.  
While-topic    maybe    3-PL-lie.down-PROG-REPP    dog  
'The dogs would lie down for a while'

(Feeling et al. 2017: Transformation)



# 4. Probabilistic patterns in word order

---

**Effect of thematic role:** The "core" thematic roles differ in propensity to precede the verb, in a cross-linguistically common hierarchy *agent* > *theme*, *goal*.

	<b>Agent</b>	<b>Goal</b>	Location	Pred-obj	Pred-sub	<b>Theme</b>	Time
Postverbal	<b>17</b>	<b>25</b>	15	0	7	<b>89</b>	12
Preverbal	<b>68</b>	<b>56</b>	62	23	18	<b>166</b>	124
<b>Percent preverbal</b>	<b>80%</b>	<b>69%</b>	<b>80%</b>	<b>100%</b>	<b>72%</b>	<b>65%</b>	<b>91%</b>



# 4. Probabilistic patterns in word order

---

**Effect of thematic role:** Consistent with the claim of Feeling and Pulte (1975). We find a similar postverbal tendency for themes of intransitive verbs.

	<b>Agent</b>	<b>Goal</b>	Location	Pred-obj	Pred-sub	<b>Theme</b>	Time
Postverbal	<b>17</b>	<b>25</b>	15	0	7	<b>89</b>	12
Preverbal	<b>68</b>	<b>56</b>	62	23	18	<b>166</b>	124
<b>Percent preverbal</b>	<b>80%</b>	<b>69%</b>	<b>80%</b>	<b>100%</b>	<b>72%</b>	<b>65%</b>	<b>91%</b>



# 4. Probabilistic patterns in word order

---

PREDICATE OBJECTS in the corpus uniformly preverbal, consistent with previous descriptions based on acceptability judgments (Akkus 2018).

	Agent	Goal	Location	<b>Pred-obj</b>	Pred-sub	Theme	Time
Postverbal	17	25	15	<b>0</b>	7	89	12
Preverbal	68	56	62	<b>23</b>	18	166	124
<b>Percent preverbal</b>	<b>80%</b>	<b>69%</b>	<b>80%</b>	<b>100%</b>	<b>72%</b>	<b>65%</b>	<b>91%</b>



# 4. Probabilistic patterns in word order

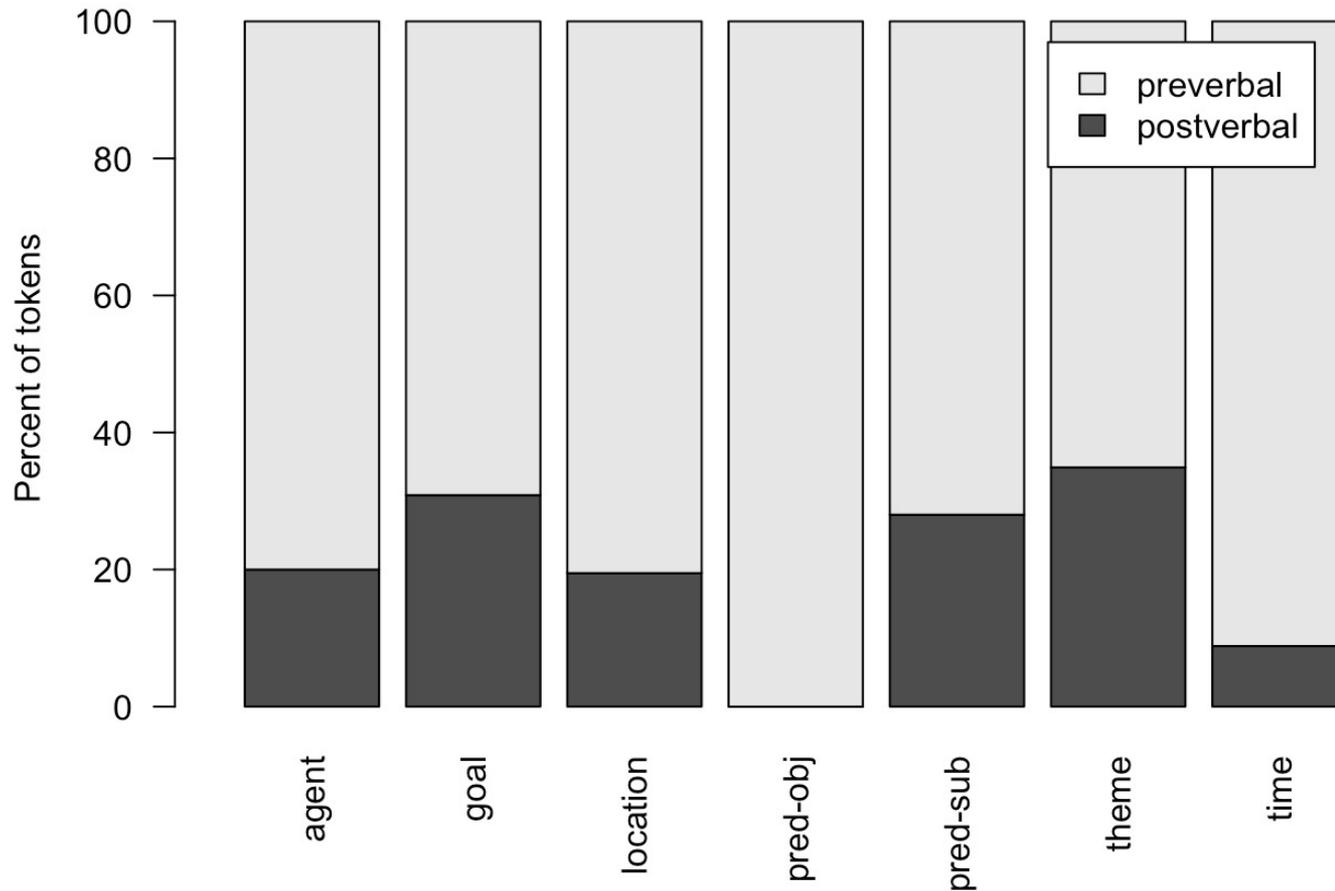
---

The regression model finds significant differences between AGENT vs. PRED-SUB, THEME. (TIME not included in the model)

	Agent	Goal	Location	Pred-obj	Pred-sub	Theme	Time
Postverbal	17	25	15	0	7	89	12
Preverbal	68	56	62	23	18	166	124
<b>Percent preverbal</b>	<b>80%</b>	<b>69%</b>	<b>80%</b>	<b>100%</b>	<b>72%</b>	<b>65%</b>	<b>91%</b>



# 4. Probabilistic patterns in word order



# 4. Probabilistic patterns in word order

---

We use a **regression model** to identify information-structure and thematic factors that are statistically significant predictors of word order.

- We use a subset of the corpus (70%) with *sentences that contain only one constituent other than the verb*, to avoid the chance that sentences with multiple items follow different principles.
- TIME items are excluded, as there is no clear way to annotate them for an information structure value.



# 4. Probabilistic patterns in word order

---

Modeling was done with the `glm()` function in R:

- Information structure and thematic role tags are the independent variables, with position (preverbal vs. postverbal) as the dependent variable.
- The model presented here does not include interaction terms. We found **no significant interaction effects** in an alternative model.



# 4. Probabilistic patterns in word order

---

Notes on reading the results on next slide:

- The reference category (intercept) consists of items that are **accessible** and **agents**.
- *Positive coefficient* = More likely preverbal placement than the reference category;
- *Negative coefficient* = More likely postverbal placement than the reference category
- Significance codes:

$p < 0.001$  '\*\*'

$p < 0.01$  '\*'

$p < 0.05$  '.'



# 4. Probabilistic patterns in word order

---

Coefficients:

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	2.0142	0.6159	3.270	0.00107 **
IS: given-active	-0.4002	0.4019	-0.996	0.31928
IS: given-inactive	-0.4529	0.4592	-0.986	0.32400
IS: new	0.9314	0.4581	2.033	0.04202 *
IS: nonreferential	1.6250	0.6206	2.618	0.00883 **
THETA: goal	-0.9435	0.6628	-1.423	0.15459
THETA: location	-1.4226	0.7245	-1.963	0.04960 *
THETA: pred-obj	13.5820	694.3753	0.020	0.98439
THETA: pred-sub	-1.7650	0.8600	-2.052	0.04015 *
THETA: theme	-1.5503	0.5730	-2.706	0.00682 **



# 4. Probabilistic patterns in word order

---

The model confirms **independent** effects of information structure and thematic properties on word order.

- The results strongly challenge the claim that thematic role plays no role in Cherokee or Iroquoian word order (cf. Mithun 1992; 2017).



# 4. Probabilistic patterns in word order

---

The **cumulative** nature of these factors is seen in *cross pair tables* with thematic and information structure tags.

	<i>NP is an agent</i>	<i>NP is a theme</i>
<i>NP is new</i>	<b>92%</b> preverbal (12/13)	<b>73%</b> preverbal (36/49)
<i>NP is given</i>	<b>76%</b> preverbal (37/49)	<b>50%</b> preverbal (57/114)

- Agents occur earlier than themes (left vs. right column)
- New information occurs earlier than old information (top vs. bottom row).



# Organization of the talk

---

1. Overview
2. Key questions in Cherokee syntax
3. The corpus of narratives
4. Probabilistic patterns in word order
- 5. Implications for generative theories**
6. Conclusion



# 5. Implications for generative theories

---

**Question:** How can we account for the Cherokee pattern in generative (Minimalist) approaches to syntax?

There is a simple way to analyze the *structure* of Cherokee clauses, based on cross-linguistically robust patterns.

- Postverbal XPs occur in a relatively low “base” position below the verb, within vP.
- Preverbal placement results from **feature-driven movement**, triggered by higher functional projections.
  - E.g. CP (information structure) and InflP (argument structure, thematic relations).



# 5. Implications for generative theories

---

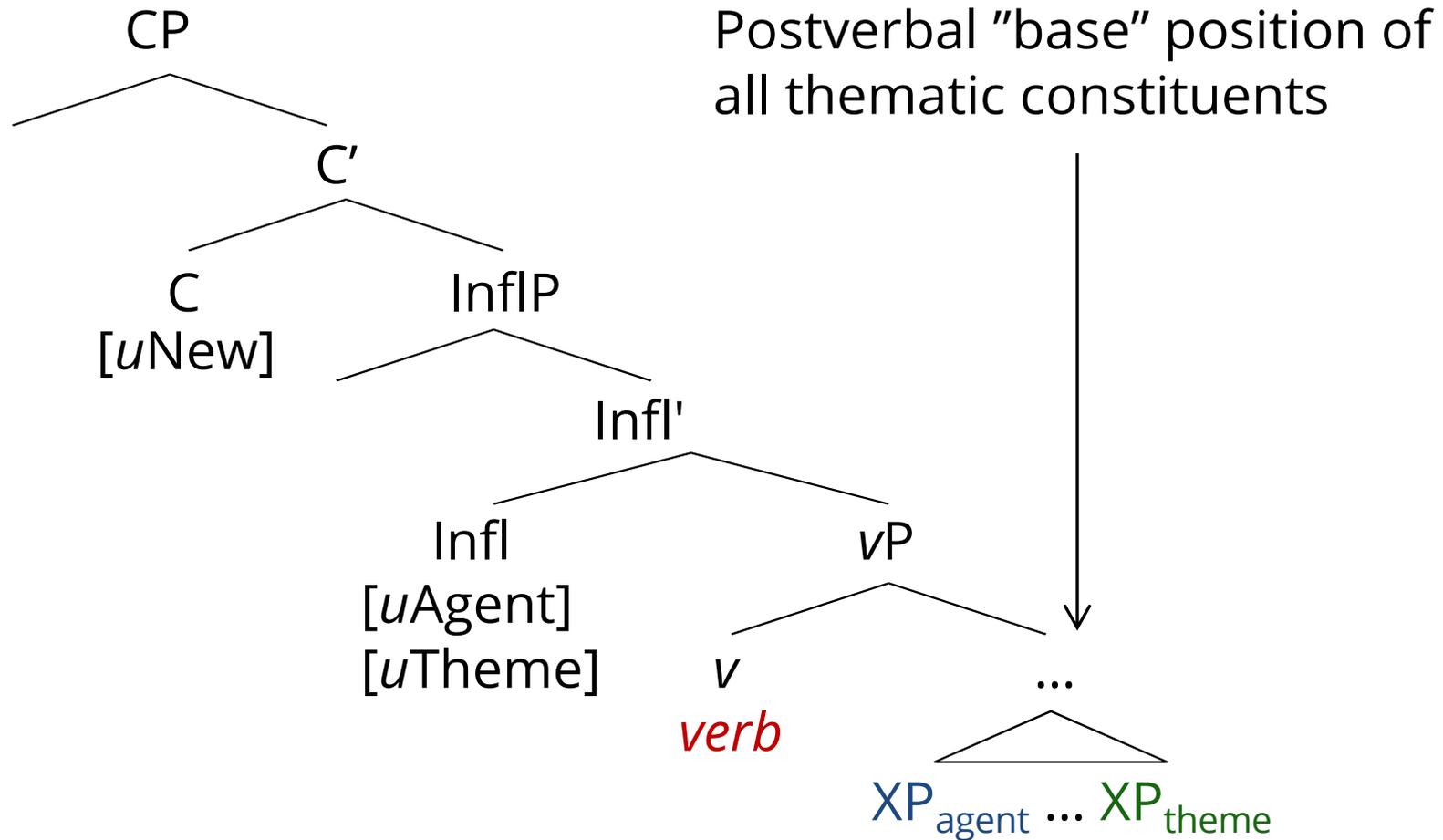
**Question:** How can we account for the Cherokee pattern in generative (Minimalist) approaches to syntax?

There is a straightforward way to understand the *structure* of Cherokee clauses:

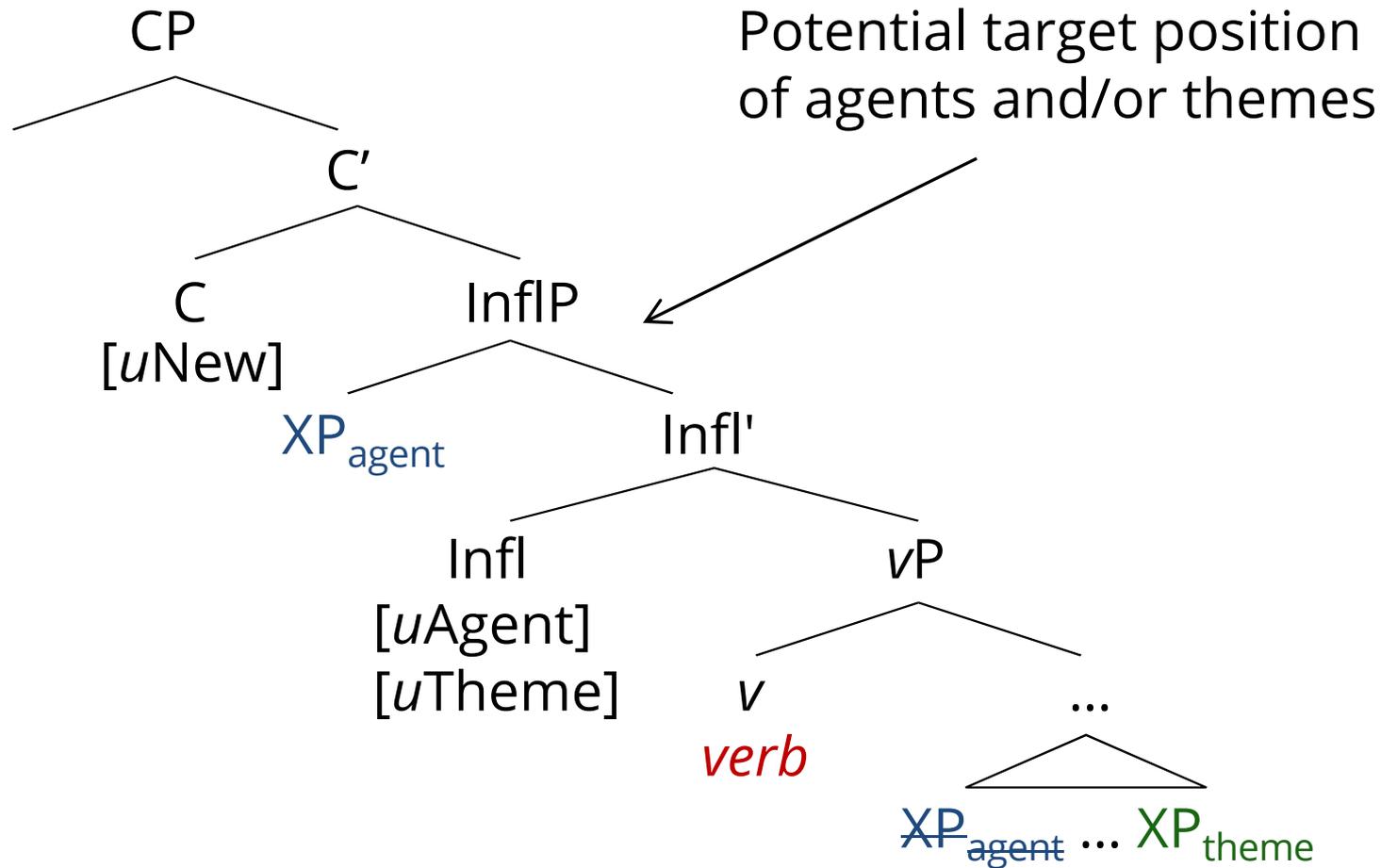
- Illustrated on next slides in simplified clause structure with three features:
  - [*uNew*] on C attracts new information XPs.
  - [*uAgent*] on Infl attracts agent XPs.
  - [*uTheme*] on Infl attracts theme XPs.



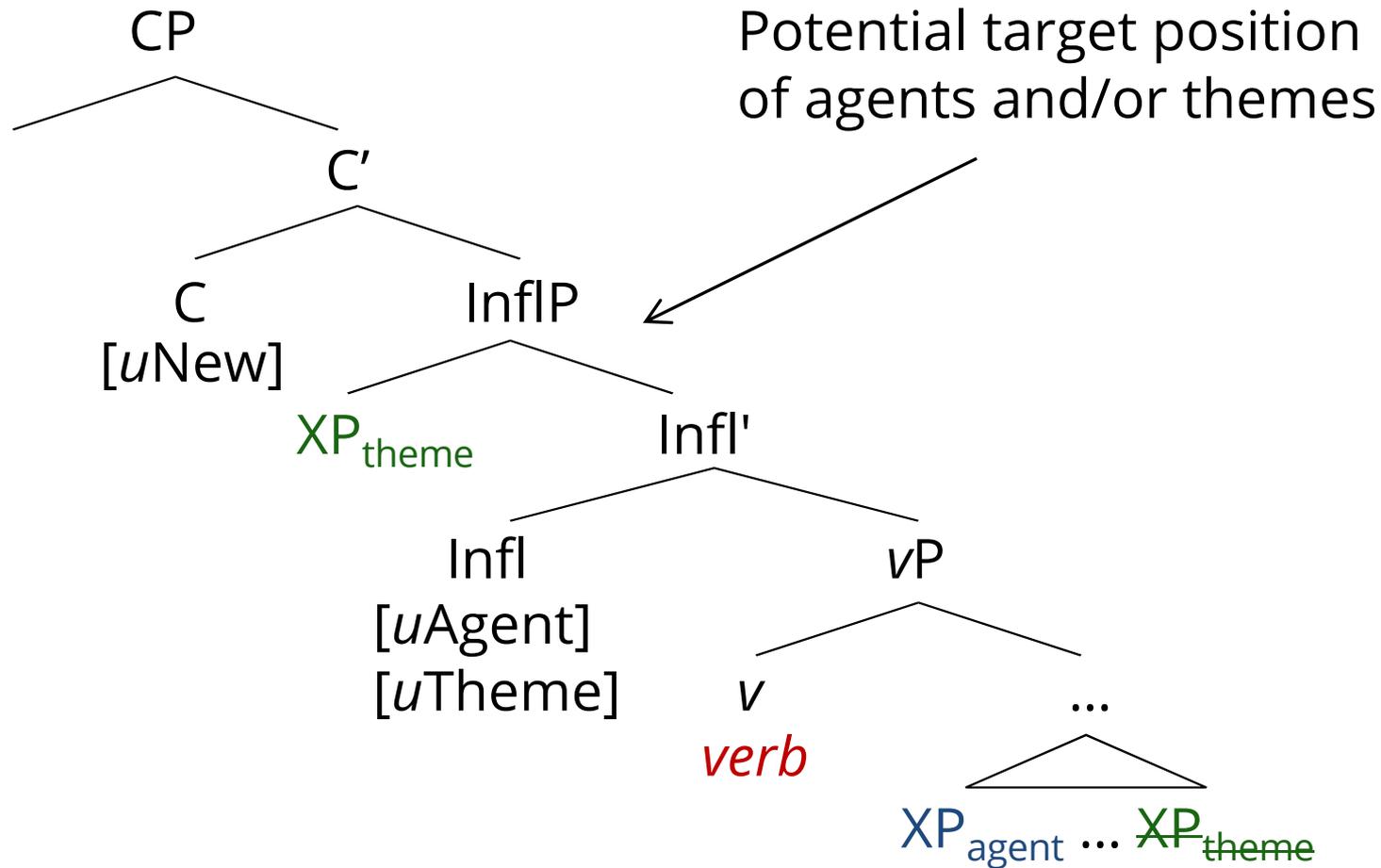
# 5. Implications for generative theories



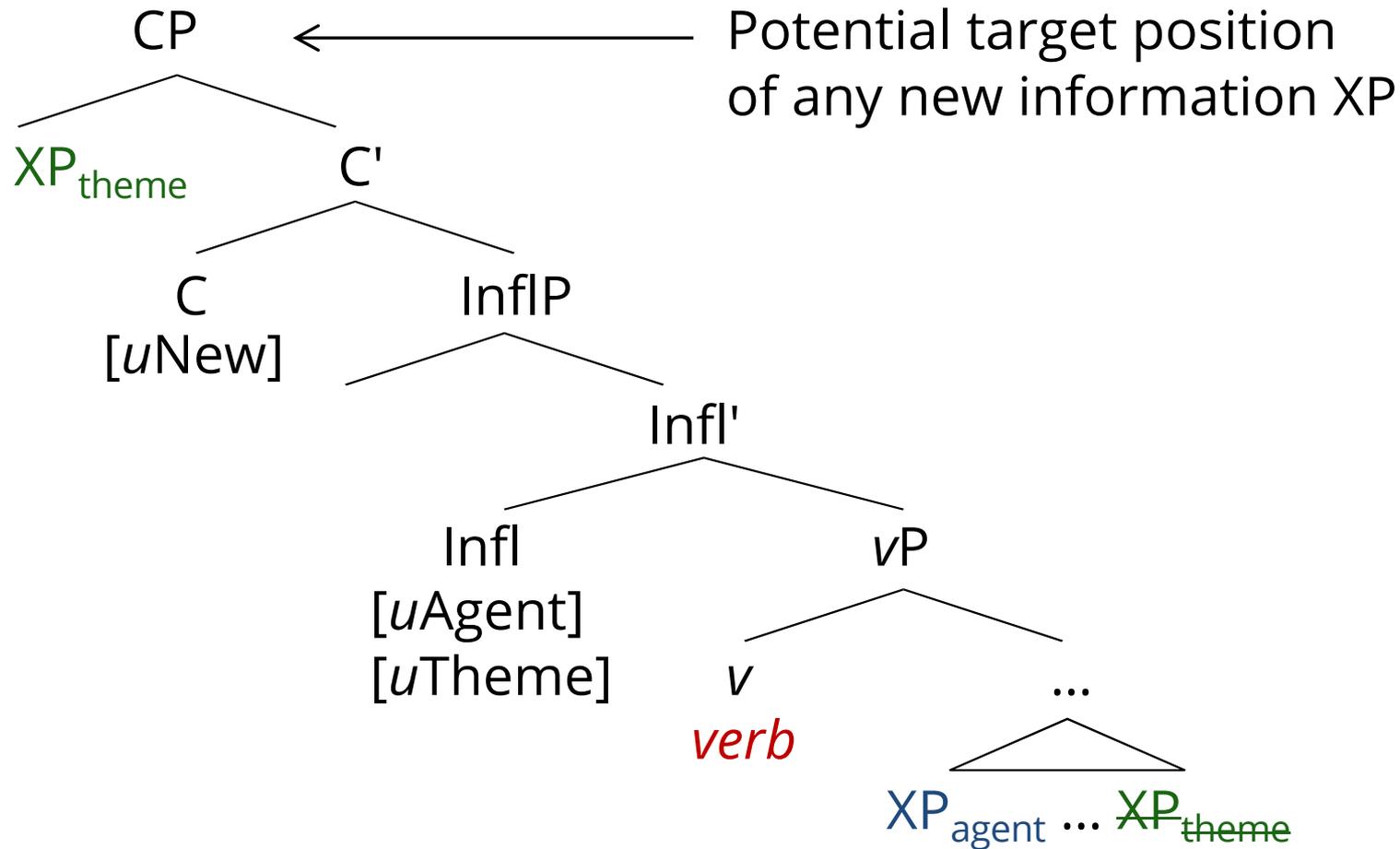
# 5. Implications for generative theories



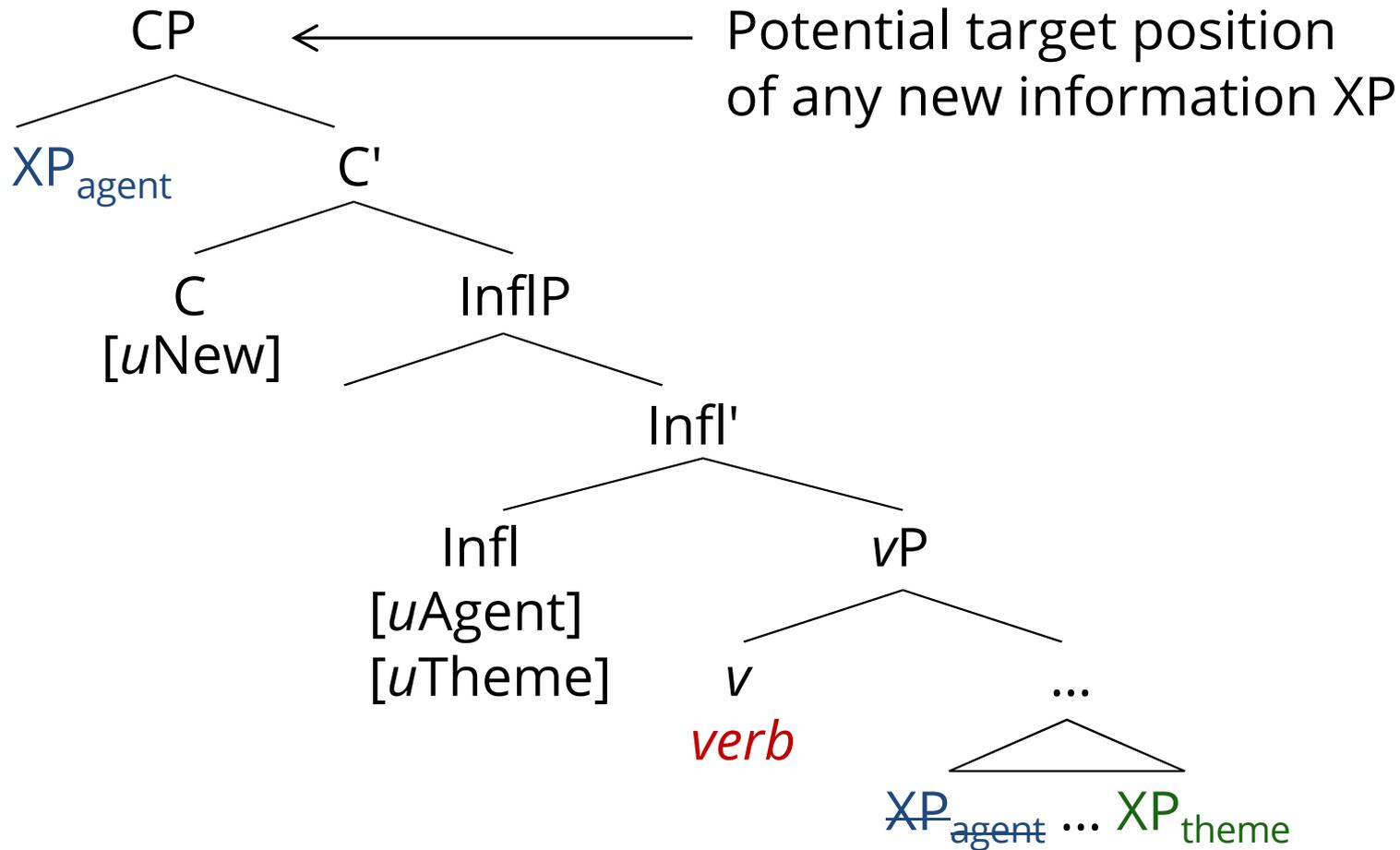
# 5. Implications for generative theories



# 5. Implications for generative theories



# 5. Implications for generative theories



# 5. Implications for generative theories

---

However, “standard” Minimalism has no clear way to explain **optional/probabilistic movement**.

For each combination of features and lexical items in a derivation, the grammar is expected to generate exactly one output structure.



# 5. Implications for generative theories

---

**Proposal:** We can account for these patterns by integrating Minimalist derivations with a **constraint-based model of grammatical computation** (Heck & Müller 2003, 2013; Murphy 2017; Hsu *to appear*).

- Syntactic structures are built derivationally from the bottom up (Chomsky 1993, et seq.).



# 5. Implications for generative theories

---

**Proposal:** We can account for these patterns by integrating Minimalist derivations with a **constraint-based model of grammatical computation** (Heck & Müller 2003, 2013; Murphy 2017; Hsu *to appear*).

At each step of the derivation:

- The grammar examines the existing (input) structure.
- It compares output candidates that each apply a syntactic operation (like movement)
- It selects an optimal candidate.



# 5. Implications for generative theories

---

Phrasal movement satisfies the **FEATURE CONDITION** constraint (Heck & Müller 2003):

MERGE CONDITION (general):

For each  $[uF]$  and XP with matching  $[F]$ , the XP occurs in the specifier of the head with  $[uF]$ .



# 5. Implications for generative theories

---

We propose multiple versions of this constraint, indexed to each movement trigger, and weighted separately.

FEATURE CONDITION (NEW):

For each [*uNew*] and XP with matching [New], the XP occurs in the specifier of the head with [*uNew*].

FEATURE CONDITION (AGENT):

For each [*uAgent*] and XP with matching [Agent], the XP occurs in the specifier of the head with [*uAgent*].

FEATURE CONDITION (THEME):

For each [*uTheme*] and XP with matching [Theme], the XP occurs in the specifier of the head with [*uTheme*].



# 5. Implications for generative theories

---

Constraint evaluation occurs in a **Maximum Entropy Harmonic Grammar** (MaxEnt: Goldwater and Johnson 2000; Hayes and Wilson 2008):

- Constraints have *numerical weights* (Legendre et al. 1990), rather than strict rankings (Prince and Smolensky 1993).
- Probabilities of output types are computed from their harmony scores.
- Less well-formed candidates are not categorically banned, but *less likely to surface*.



# 5. Implications for generative theories

---

On next slides:

- Tableaux showing output selection for sentences with (i) new information agent, (ii) given agent, (iii) new theme, (iv) given theme [cf. Table 1]
- Tableaux show violation profiles, harmony scores ( $H$ ), and predicted probabilities ( $P$ ) of candidates.
- Constraint weights  $w$  identified with the MaxEnt Grammar Tool learner (Wilson & George 2009), trained on cross pair data in Table 1.



# 5. Implications for generative theories

---

On next slides:

- Note: We abstract away from the separate derivational steps that build InflP and CP, and show them in one step.



# 5. Implications for generative theories

Tableau (i): Clause with a new agent

$[C/Infl [_{VP} \text{verb} \dots XP_{[New][Agent]} ]]$	FC (NEW) $w=1.04$	FC (AGEN) $w=1.16$	FC (THM) $w=0$	$H$	$p$
$[C/InflP XP [C/Infl [_{VP} \text{verb} \dots XP ]]$ <i>Preverbal agent</i>				0	<b>.90</b>
$[C/InflP [C/Infl [_{VP} \text{verb} \dots XP ]]$ <i>Postverbal agent (no movement)</i>	-1	-1		-2.2	<b>.10</b>



# 5. Implications for generative theories

Tableau (ii): Clause with a given agent

$[_{C/Infl} [_{VP} \text{verb} \dots XP_{[Agent]} ]$	FC (NEW) $w=1.04$	FC (AGEN) $w=1.16$	FC (THM) $w=0$	$H$	$p$
$[_{C/InflP} XP [_{C/Infl} [_{VP} \text{verb} \dots XP ]$ <i>Preverbal agent</i>				0	<b>.76</b>
$[_{C/InflP} [_{C/Infl} [_{VP} \text{verb} \dots XP ]$ <i>Postverbal agent (no movement)</i>		-1		-1.16	<b>.24</b>



# 5. Implications for generative theories

Tableau (iii): Clause with a new theme

$[C/Infl [_{VP} \text{verb} \dots XP_{[New][Theme]} ]]$	FC (NEW) $w=1.04$	FC (AGEN) $w=1.16$	FC (THM) $w=0$	$H$	$p$
$[C/InflP XP [C/Infl [_{VP} \text{verb} \dots XP ]]$ <i>Preverbal theme</i>				0	<b>.74</b>
$[C/InflP [C/Infl [_{VP} \text{verb} \dots XP ]]$ <i>Postverbal theme (no movement)</i>	-1		-1	-1.04	<b>.26</b>



# 5. Implications for generative theories

Tableau (iii): Clause with a given theme

$[_{C/Infl} [_{VP} \text{verb} \dots XP_{[Theme]} ]]$	FC (NEW) $w=1.04$	FC (AGEN) $w=1.16$	FC (THM) $w=0$	$H$	$p$
$[_{C/InflP} XP [_{C/Infl} [_{VP} \text{verb} \dots XP ]]$ <i>Preverbal theme</i>				0	<b>.50</b>
$[_{C/InflP} [_{C/Infl} [_{VP} \text{verb} \dots XP ]]$ <i>Postverbal theme (no movement)</i>			-1	0	<b>.50</b>



# Organization of the talk

---

1. Overview
2. Key questions in Cherokee syntax
3. The corpus of narratives
4. Probabilistic patterns in word order
5. Implications for generative theories
- 6. Conclusion**



# 6. Conclusion

---

## Summary of main results:

- Analysis of quantitative variability in an annotated corpus allows a more comprehensive description of Cherokee word order principles.
- A range of features of both information structure and thematic role influence Cherokee word order in principled, probabilistic ways.
- Cumulative effects support integration of formal theory with probabilistic, constraint-based computation.



# 6. Conclusion

---

## Future directions:

- Examine possible word order effects of other grammatical properties in Cherokee, such as person, animacy, clause type, agreement morphology.
- Examine Cherokee clauses with more than 1 major constituent – are there different principles involved?
- Compare with other languages with high word order flexibility. How much can languages vary in the movement propensities of particular features?



---

# Thank you!



THE UNIVERSITY  
*of* NORTH CAROLINA  
*at* CHAPEL HILL

# Rereferences

---

- Akkus, Faruk. 2018. Copular constructions and clausal syntax in Cherokee. In *Proceedings of the Workshop on the Structure and Constituency of Languages of the Americas 21, University of British Columbia Working Papers in Linguistics 46*, ed. Megan Keough, Natalie Weber, Andrei Anghelescu, Sihwei Chen, Erin Guntly, Khia Johnson, Daniel Reisinger, and Oksana Tkachman.
- Beghelli, Filippo. 1996. Cherokee clause structure. In *Cherokee papers from UCLA*, ed. Filippo Beghelli, Barbara Blankenship, Michael Dukes, Edward S. Flemming, Pamela Munro, Brian Potter, Robert S. Williams, and Richard Wright, 105–114. Los Angeles: Department of Linguistics, University of California Los Angeles.
- Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The View from Building 20*, ed. Kenneth Hale and Samuel Jay Keyser, 1–52. Cambridge, MA: MIT Press.



# Rereferences

---

- Dipper, Stefanie, Michael Götze, and Stavros Skopeteas, ed. 2007. *Information structure in cross-linguistic corpora: Annotation guidelines for phonology, morphology, syntax, semantics, and information structure (ISIS Working Papers of the SFB 632)*. Potsdam: Universitätsverlag.
- Ellsiepen, Emilia, and Markus Bader. 2018. Constraints on Argument Linearization in German. *Glossa: a journal of general linguistics* 3: 6. 1-36.
- Feeling, Durbin, and William Pulte. 1975. *Cherokee-English dictionary*. Talequah: Cherokee Nation of Oklahoma.
- Frey, Benjamin. 2020. "Data is nice:" Theoretical and pedagogical implications of an Eastern Cherokee corpus. In *Collaborative approaches to the challenge of language documentation and conservation: Selected papers from the 2018 Symposium on American Indian Languages (SAIL)*, ed. Wilson de Lima Silva and Katherine Riestenberg, 38–53. Honolulu: University of Hawai'i Press.



# Rereferences

---

- Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Workshop on Variation within Optimality Theory*, ed. Jennifer Spenader, Anders Eriksson, and Osten Dahl, 111–120. Stockholm: Stockholm University.
- Harbour, Daniel, Laurel J Watkins, and David Adger. 2012. Information structure, discourse structure, and noun phrase position in Kiowa. *International Journal of American Linguistics* 78: 97–126.
- Hayes, Bruce, and Colin Wilson. 2008. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry* 39: 379–440.
- Heck, Fabian, and Gereon Müller. 2003. Derivational optimization of wh-movement. *Linguistic Analysis* 33: 97–148.
- Heck, Fabian, and Gereon Müller. 2013. Extremely local optimization. In *Linguistic derivations and filtering*, ed. Hans Broekhuis and Ralf Vogel, 136–165. Sheffield: Equinox.



# Rereferences

---

- Hsu, Brian. To appear. Harmonic Grammar in phrasal movement: an account of probe competition and blocking. In *NELS 51: Proceedings of the 51st Meeting of the North East Linguistic Society*, ed. Alessa Farinella and Angelica Hill. Amherst, MA: GLSA.
- Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky. 1990. Can connectionism contribute to syntax?: Harmonic Grammar, with an application. In *Proceedings of the 26th regional meeting of the Chicago Linguistic Society*, ed. Karen Deaton, Manuela Noske, and Michael Ziolkowski, 237–252. Chicago: Chicago Linguistic Society.
- Mithun, Marianne. 1992. Is basic word order universal? In *Pragmatics of word order flexibility*, ed. Doris L. Payne, 15–62. Amsterdam: John Benjamins.
- Mithun, Marianne. 1995. Morphological and prosodic forces shaping word order. In *Word order in discourse*, ed. Pamela A. Downing and Michael Noonan, 387–423. Amsterdam/Philadelphia.



# Rereferences

---

- Mithun, Marianne. 2017. The Iroquoian Language Family. In *The Cambridge handbook of linguistic typology*, ed. Alexandra Y. Aikhenvald, 747–781. Cambridge University Press.
- Montgomery-Anderson, Brad. 2008. A reference grammar of Oklahoma Cherokee. Ph.D dissertation. University of Kansas.
- Montgomery-Anderson, Brad. 2015. *Cherokee Reference Grammar*. Norman: University of Oklahoma Press.
- Murphy, Andrew. 2017. Cumulativity in syntactic derivations. Ph.D dissertation, Universität Leipzig.
- Prince, Alan, and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell.
- Scancarelli, Janine. 1986. Pragmatic Roles in Cherokee Grammar. In *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*, ed. Vassiliki Nikiforidou, Mary VanClay, Mary Niepokuj, and Deborah Feder, 224–234. Berkeley: Berkeley Linguistics Society.



# Rereferences

---

- Tonhauser, Judith, and Erika Colijn. 2010. Word Order In Paraguayan Guaraní. *International Journal of American Linguistics* 76: 255–288.
- Wilson, Colin, and Ben George. 2009. Maxent grammar tool. <https://linguistics.ucla.edu/people/hayes/MaxentGrammarTool/>

