How much overt agreement is needed for polysynthesis? Quantitative evidence from Cherokee.

Erin Humphreys & Brian Hsu (UNC-Chapel Hill) Penn Linguistics Conference 48 March 2024



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

1. Overview

- 2. Polysynthesis & agreement
- 3. Cherokee word order & agreement
- 4. Quantitative analysis
- 5. Implications & conclusion



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Two properties that cluster in polysynthetic languages (Jelinek 1984, Baker 1996, Mithun 2017 a.o.):

Flexible clausal word order

- Often with free omission of argument NPs, nonconfigurational properties (Hale 1983).
- Polypersonal agreement: Coindexation of multiple thematic arguments in verbal agreement morphology



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

This paper re-examines a standard generative proposal on the causal relationship between these properties (Jelinek 1984; Baker 1996; 2006).

We focus on key grammatical properties of **Cherokee** (Southern Iroquoian: Oklahoma and North Carolina).



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

1. Overview

Key grammatical properties of Cherokee:

(i) Highly flexible clausal word order, with known predictive factors (Scancarelli 1987; Hsu & Frey 2023)

Example: Agent arguments of transitive predicates can precede or follow verbs.

(1) *Subject < Verb*

[gitli] ogi-sdawadvs-v.

dog 1.PL.EXCL-follow-EXPP

'The dog followed us.' (Feeling et al. 2017; 101)



1. Overview

Key grammatical properties of Cherokee:

(i) Highly flexible clausal word order, with known predictive factors (Scancarelli 1987; Hsu & Frey 2023)

Example: Agent arguments of transitive predicates can precede or follow verbs.

(2) Verb < Subject
a-n-adasdelis-g-o [yvwi j-u-n-asdi].
3-PL-help-PROG-HAB people DST-3-PL-little
'The little people help (others).' (Feeling et al. 2017; 43)



Key grammatical properties of Cherokee:

(ii) Fewer NP arguments coindexed by verbal agreement than in other polysynthetic languages

- (3) [sagwu=no] [j-un-atana ahwi] d-a-hih-e.
 one=CN DST-3.PL-big deer DST-3.sG-kill-REPP
 'One (of the hunters) killed big deer.' (Feeling et al. 2017; 53)
 - Features of only one 3rd-person argument expressed on the pronominal prefix *a*-.



We present a **quantitative investigation** of the relation between overt coindexation of arguments and NP placement in Cherokee clauses.

Main result: Argument NP placement does *not* vary, based on whether its features are overtly coindexed.

Having fewer coindexed arguments does not lead to reduced word order flexibility relative to other polysynthetic languages (ex. Mohawk)



Proposal: The threshold for maintaining this type of word order flexibility is *abstract polypersonal agreement,* rather than *surface polypersonal agreement* (our terms)

Surface polypersonal agreement

Agreement morphology directly expresses features of multiple arguments of a verb.

Abstract polypersonal agreement

Agreement morphology systematically depends on relative properties of multiple arguments of a verb (not all argument features directly expressed).



1. Overview

2. Polysynthesis & agreement

- 3. Cherokee word order & agreement
- 4. Quantitative analysis
- 5. Implications & conclusion



Next slides: Summary of Baker's (1996) analysis of Mohawk (Northern Iroquoian), and the relation between polypersonal agreement and flexible word order.



Mohawk: Flexible clausal word order.

S V (4)a. Sak ra-núhwe'-s Sak MsS-like-нав 'Sak likes her dress'

O ako-[a]tyá'tawi FsP-dress

(Baker 1996; 10)



Mohawk: Flexible clausal word order.

V S O (4)b. ra-núhwe'-s Sak ako-[a]tyá'tawi MsS-like-нав Sak FsP-dress 'Sak likes her dress'

(Baker 1996; 10)



Mohawk: Flexible clausal word order.

SOV(4)c.Sakako-[a]tyá'tawira-núSakFsP-dressMsS-li'Sak likes her dress'

<mark>ra-núhwe'-s</mark> MsS-like-нав

(Baker 1996; 10)



Mohawk: Flexible clausal word order.

V O S (4)d. ra-núhwe'-s ako-[a]tyá'tawi Sak MsS-like-HAB FsP-dress Sak 'Sak likes her dress'

(Baker 1996; 10)



Mohawk: Flexible clausal word order.

O V S (4)e. ako-[a]tyá'tawi ra-núhwe'-s Sak FsP-dress MsS-like-HAB Sak 'Sak likes her dress'

(Baker 1996; 10)



Mohawk: Flexible clausal word order.

O S V (4)f. ako-[a]tyá'tawi Sak ra-núhwe'-s FsP-dress Sak MsS-like-нав 'Sak likes her dress'

(Baker 1996; 10)



Mohawk: Flexible clausal word order.

Other non-configurational properties (Baker 1996):

- Absence of weak cross-over effects
- Free omission of non-pronominal NP arguments



Mohawk: Verbal agreement prefixes coindex features of up to two thematic arguments (including two 3rd-person arguments)

- Shako- expresses gender, number properties of two arguments (masc. sg. subject, fem. object)
- (5) Shako-nuhwe'-s (ne owira'a) Mohawk (Baker 1996; 21)
 MsS/FO-like-нав NE baby
 'He likes them (babies)'



Mohawk: Productive noun incorporation (cf. Mithun 2017b), in complementary distribution with pronominal object agreement

- (6) Ra-wir-a-nuhwe'-s Mohawk (Baker 1996; 21)
 MsS-baby-ø-like-нав 'He likes babies'
- Baker (1996): Incorporated objects fill the "slot" taken by object agreement.



Baker's proposed condition on agreement, argument licensing in polysynthetic languages:

The Morphological Visibility Condition (MVC) A phrase X is visible for θ-role assignment from a head Y only if it is coindexed with a morpheme in the word containing Y via:

- An agreement relationship, or
- A movement relationship

(Baker 1996; 17)



Baker (1996): If this condition is met, all nonpronominal NPs pattern as adjuncts: they can be freely omitted, freely ordered.

[In a certain class of languages], a verb [Y] agrees with an overt NP [X] if and only if [X] is in a dislocated, adjunct position.

(Baker 2006: 290)



Verbs do not overtly coindex all argument features in all polysynthetic languages with flexible word order.

How does the MVC accommodate this?

 The MVC tolerates some degree of phonologically null agreement (Baker 1996):

 Some argument features, or combinations thereof, can be expressed by a null (Ø) agreement morpheme.

 $_{\odot}\,$ i.e. Particular cells in the paradigm are expressed as Ø



Verbs do not overtly coindex all argument features in all polysynthetic languages with flexible word order.

How does the MVC accommodate this?

 The MVC tolerates some degree of phonologically null agreement (Baker 1996):

 The structural position of arguments indexed by null (Ø) agreement morphemes does not differ from arguments indexed overtly (Baker 2006).



Our question: How much overt agreement is necessary to maintain general word order freedom, general polysynthetic characteristics?

Next sections: Quantitative evidence from Cherokee shows that *abstract polypersonal agreement* is sufficient to satisfy the MVC.



- 1. Overview
- 2. Polysynthesis & agreement

- 4. Quantitative analysis
- 5. Implications & conclusion



Language background: Southern Iroquoian language, once spoken throughout lower Appalachian regions.

Most speakers (~2000 native/fluent) now reside in Oklahoma and Western North Carolina.

Next slides: Key aspects of clausal word order, agreement morphology.



Flexible clausal word order:

- All orders of subject, object, verb are grammatical (Pulte & Feeling 1975, Scancarelli 1987).
- (7) Subject < Verb
 [gitli] ogi-sdawadvs-v.
 dog 1.PL.EXCL-follow-EXPP
 'The dog followed us.' (Feeling et al. 2017; 101)



Flexible clausal word order:

- All orders of subject, object, verb are grammatical (Pulte & Feeling 1975, Scancarelli 1987).
- (8) Verb < Subject

 a-n-adasdelis-g-o [yvwi j-u-n-asdi].
 3-PL-help-PROG-HAB people DST-3-PL-little
 'The little people help (others).' (Feeling et al. 2017; 43)



Flexible clausal word order:

- All orders of subject, object, verb are grammatical (Pulte & Feeling 1975, Scancarelli 1987).
- (9) *Object < Verb*

Nokil[am]ji-todis-g-oThenuntilwater1-heat.water-PROG-HAB'Then I heat some water.' (Feeling et al. 2017; 129)



Flexible clausal word order:

 All orders of subject, object, verb are grammatical (Pulte & Feeling 1975, Scancarelli 1987).

(10) Verb < Object
u-sdu-hnv [galohisdi?i].
3-close-EXPP door
'(he) closed the door.' (Feeling et al. 2017; 35)



Flexible clausal word order:

- Non-configurational properties shared with Mohawk (Beghelli 1996)
 - Absence of weak cross-over effects
 - Free omission of non-pronominal NP arguments



Recent findings (Hsu & Frey 2023):

- No subject/object asymmetry in word order:
 - NP placement insensitive to number of NPs in clause:
 - Ex: Themes of transitive verbs show same placement as themes of unaccusative verbs.
- Placement of each NP determined probabilistically by its properties related to *contrast*, *referential* accessibility and thematic role.



Recent findings (Hsu & Frey 2023):

- Word order principles qualitatively similar to clausal word order in related Nothern Iroquoian languages:
 - Mohawk (Baker 1996; Mithun 2017b)
 - Cayuga (Mithun 1992)
 - Tuscarora (Mithun 1995)



Pronominal agreement system:

- All verbs have one pronominal prefix
 - Person and number properties of at least one argument are indexed
 - No paradigmatic gaps, all combinations of person and number are available.
- If both arguments are local (1st person, 2nd person), features of both, and their relative thematic prominence, are expressed by the prefix



Pronominal agreement system:

 If both arguments are local (1st or 2nd person): features of both, and their relative thematic prominence, are expressed by the prefix

(11) hi? da-jv-nej-el-i.
this FUT-1sg/2PL-tell-DAT-FUT
'I will tell you about this'

(Feeling et al. 2017; 75)


Pronominal agreement system:

- If one argument is local, and the other is not: the prefix expresses features of the local argument
 - Regardless of relative thematic prominence of the arguments



Pronominal agreement system:

- If one argument is local, and the other is not: the prefix expresses features of the local argument
 - Regardless of relative thematic prominence of the arguments
- (12) [gitli] ogi-sdawadvs-v.
 dog 1.PL.EXCL-follow-EXPP
 'The dog followed us.'

(Feeling et al. 2017; 101)



Pronominal agreement system:

- If both arguments are 3rd person: The prefix expresses number features of the more animate argument
 - If they are equally animate, the more thematically prominent argument is overtly agreed with
- (13) [sagwu=no] [j-un-atana ahwi] d-a-hih-e.
 one=CN DST-3.PL-big deer DST-3.sG-kill-REPP
 'One (of the hunters) killed big deer.'



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL (Feeling et al. 2017; 53)

Pronominal agreement system:

- If both arguments are 3rd person: The prefix expresses number features of the more animate argument.
 - If they are equally animate, the more thematically prominent argument is overtly agreed with
- (14) [ijul] y-uni-kevs-e
 both NONFIN-**3.PL-**chase-REPP
 '(The bull) would chase them both.'



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL (Feeling et al. 2017; 114)

- Mohawk: Features of two 3rd person arguments can be expressed on a pronominal prefix.
 (Baker 1996; 191; Mithun 2017b; 33)
- Cherokee: Features of two 3rd person arguments are *never* simultaneously expressed on a prefix.
 - Pronominal prefixes on transitive verbs are only partially informative about the features of the noncoindexed argument.
 - Ex. A 3rd person prefix on a transitive verb only indicates that the other argument is not local



- Cherokee only consistently expresses features of one argument of a transitive verb.
 - Unlike Mohawk
- There is no productive noun incorporation
 Also unlike Mohawk



A proposed distinction among polypersonal agreement systems:

Surface polypersonal agreement

Agreement morphology directly expresses features of multiple arguments of a verb

Abstract polypersonal agreement

Agreement morphology systematically depends on relative properties of multiple arguments of a verb (not all argument features directly expressed)



Pronominal prefixes in Cherokee typically express features of only one argument of transitive verb.

Few patterns of surface polypersonal agreement, compared to Mohawk.

However, the choice of which argument to coindex depends on relative properties of *more than one* thematic argument.

Evidence for abstract polypersonal agreeement.



Next section: Does this reduced amount of visible argument features on verbs result in a distinct structural representation of NPs in the clause?

Or, is abstract polypersonal agreement sufficient to maintain word order flexibility?



Interim summary

- Cherokee has less overt coindexation than the other Iroquoian languages
 - Taken as the "prototypical examples of polysynthesis" (Mithun 2017b; 747)
- We know what predicts word order in Cherokee

We use Cherokee to test hypotheses about the relationship between word order flexibility and morphology



If *Surface Polypersonal Agreement* is needed to satisfy the MVC...

Whether or not an argument is *overtly* agreed with will significantly affect its position in the clause.

If *Abstract Polypersonal Agreement* meets the agreement threshold for the MVC...

Whether an argument is *overtly* agreed with will have *no* significant effect on its position in the clause.



- 1. Overview
- 2. Polysynthesis & agreement
- 3. Cherokee word order & agreement
- 4. Quantitative analysis
- 5. Implications & conclusion



Hsu and Frey (2023) identify factors that best predict clausal word order in Cherokee:

 All non-pronominal NPs in a corpus are tagged for a range of grammatical properties.

Two word order factors found to be significant in predicting the placement of NPs:

- Thematic role
- Referential accessibility (a.k.a. newness / givenness)



Probabilistic tendencies of Cherokee word order in relation to thematic role and referential accessibility:

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

(Hsu & Frey 2023)



This paper: We expand the Hsu & Frey corpus by tagging all NPs for whether their features are coindexed by a pronominal prefix on the verb.

We can thus identify whether overt coindexation on a pronominal prefix influences placement of NP arguments of transitive verbs.

 On intransitive verbs, features of the sole argument NP are always coindexed.



We control for factors that influence NP placement, and/or argument coindexation:

- Thematic role affects word order and coindexation
- Argument animacy affects only coindexation
- *Referential accessibility* affects only **word order**

(Montgomery-Anderson 2015; Hsu & Frey 2023)



Agent NPs are coindexed significantly more than themes are. It also has a word order effect: agent NPs are likelier to precede verbs than themes (H&F 2023)

X-squared = 67.803, df = 1, p-value < $2.2e^{-16}$





When controlling for **thematic role:** No significant effect of coindexation on NP position.

Figure: Placement of *theme NPs* only, by coindexed vs. non-coindexed status.

X-squared = 0.040695, df = 1, p-value = 0.8401





Animacy significantly affects which arguments are coindexed but not word order.

X-squared = 170.01, df = 2, p-value < 2.2e-16





When controlling for **animacy contrasts**: No significant effect of coindexation on NP position.

Figure: Placement of *animate NPs only*, by coindexed vs. non-coindexed status.

X-squared = 0.021634, df = 1, p-value = 0.8831





Referential accessibility does not significantly affect which argument is indexed. It does have word order effects; discourse-new NPs likelier to precede verbs than given NPs (H&F 2023)

```
X-squared = 0.5467,
df = 1,
p-value = 0.4597
```





When controlling for **referential accessibility:** No significant effect of coindexation on NP position.

Figure: Placement of *discoursegiven NPs only*, by coindexed vs. non-coindexed status.

X-squared = 0.0038223, df = 1, p-value = 0.9507





The placement of an argument NP in relation to a verb does **not** depend on whether its features are **visibly coindexed** by a pronominal prefix

In Cherokee, non-coindexed arguments appear to have **the same structural status** as coindexed arguments



- 1. Overview
- 2. Polysynthesis & agreement
- 3. Cherokee word order & agreement
- 4. Quantitative analysis
- 5. Implications & conclusion



Cherokee does not have less word order freedom than other polysynthetic languages, while showing overt agreement with systematically fewer arguments.

Abstract polypersonal agreement is sufficient to meet the agreement threshold for maintaining general polysynthetic properties.



What does this mean for the MVC?

Baker (2006): Nonvisibly-coindexed NPs show flexible placement if their features happen to be expressed by a null affix (Ø represents a slot in the agreement paradigm)

Cherokee agreement is hard to characterize as such: Whether features of an NP are coindexed depends on relative properties of two arguments.



What does this mean for the MVC?

Crucially, verbal agreement in Cherokee requires all NP arguments to be visible to the verb, in order to determine which features are coindexed.

 Can be modeled by Multiple Agree between verbs and argument NPs (Oxford 2019).

A one argument to one affix correspondence is not strictly required.



The Morphological Visibility Condition (MVC)

A phrase X is **visible for θ-role assignment** from a head Y only if it is **coindexed** with **a morpheme** in the word containing Y via:

An agreement relationship*, or

A movement relationship

(Baker 1996; 17)

**abstract polypersonal agreement* is sufficient to meet this threshold



Summary of main results:

- NP placement in the Cherokee clause does not depend on whether its features are visibly coindexed on a pronominal prefix.
- Abstract polypersonal agreement is sufficient to maintain free word order in polysynthetic languages



Future directions:

- More quantitative analysis of word order predictors in polysynthetic languages and/or languages with flexible word order.
- How much can languages vary in the movementtriggering propensities of particular features? Is it predicted by any (morphological) properties?



Thank you!



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL Slides online: <u>hsub.web.unc.edu</u>

Adger, David., Daniel Harbour, and Laurel J. Watkins. 2009. Mirrors and macroparameters. Cambridge University Press.

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.

Baker, Mark. 1996. The polysynthesis parameter. Oxford: Oxford University Press.



- Baker, Mark. 2006. On zero agreement and polysynthesis. In *Arguments and Agreement*, ed. Peter Ackema, Patrick Brandt, Maaike Schoorlemmer, and Fred Weerman, 289-342. Oxford University Press.
- 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.
- 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.
- 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.

- Hale, Kenneth. 1983. Warlpiri and the grammar of non-configurational languages. *Natural Language & Linguistic Theory* 1, 5–47.
- 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.
- Hsu, Brian, and Benjamin Frey. 2023. Word order in Cherokee: information structure, thematic structure, and variability. Ms., University of North Carolina at Chapel Hill.



Jelinek, Eloise. 1984. Empty categories, case, and configurationality. NLLT 2: 39–76.

Jelinek, Eloise. 2006. The Pronominal Argument Parameter. In *Arguments and Agreement*, ed. Peter Ackema, Patrick Brandt, Maaike Schoorlemmer, and Fred Weerman, 261–288. Oxford University Press.

- 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.
- Mithun, Marianne. 2017a. Argument marking in the polysynthetic verb and its implications. In The Oxford handbook of polysynthesis, ed. Michael Fortescue, Marianne Mithun, and Nicholas Evans, 30–58. Oxford University Press



- Mithun, Marianne. 2017b. 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.
- Oxford, Will. 2019. Inverse marking and multiple agree in algonquin: Complementarity and variability. *Natural Language & Linguistic Theory*, *37*(3), 955–996.
- Posit team. 2023. RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.


References

R Core Team. 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

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.

Wickham, H. et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, doi:10.21105/joss.01686



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL