

# (Dis)harmonic word orders and the role of case for parsing efficiency

---

Alexandra Simonenko & Yiming Liang

June 4, 2025

Universiteit Gent

DELPASIMO3, Université de Lille



# Table of contents

1. (Dis)harmony
2. Minimizing domain
3. The role of case

(Dis)harmony

---

# (Dis)harmony and typological universalia

- Among Greenberg's 31 syntactic universalia, about half are implicational statements about **harmony** in constituent headedness (U2, U3, U4, U5, U9, U10?, U13, U15, U16, U17, U18, U21, U22, U23, U24?):
  - If constituent XP is head-final/head-initial, so is constituent YP.
- Of these, 8 are statistical: the implication holds with “an overwhelmingly more than chance frequency”.
  - If constituent XP is head-final/head-initial, so is constituent YP with an overwhelmingly more than chance frequency.

# The TP-VP harmony

- U15: “In expressions of volition and purpose, a subordinate verbal form always follows the main verb as the normal order except in those languages in which the nominal object always precedes the verb”

$$[_{VP} V_{head} O] \rightarrow [_{TP} Modal_{head} V]$$
$$[_{VP} O V_{head}] \rightarrow [_{TP} V Modal_{head}]$$

- U16: “In languages with dominant order VSO, an inflected auxiliary always precedes the main verb. In languages with dominant order SOV, an inflected auxiliary always follows the main verb.”

$$[_{VP} V_{head} O] \rightarrow [_{TP} Aux_{head} V]$$
$$[_{VP} O V_{head}] \rightarrow [_{TP} V Aux_{head}]$$

Because of the relatively small size of Greenberg's sample, some absolute universalia were later found to be statistical. In Matthew Dryer's (p.c.) sample of 122 genera (236 languages), about 10% are disharmonic with respect to TP/VP headedness in main clauses.

	Africa	Eurasia	SEAsia&Oc	NAmer	SAmer	#gen.	#lgs.
OV&VAux	9	21	14	8	13	65 (0.53)	131
VO&VAux	1	1	1	1	1	5 ( <b>0.04</b> )	6
OV&AuxV	1	3	0	1	1	6 ( <b>0.05</b> )	7
VO&AuxV	23	9	7	6	1	46 (0.38)	92

**Table 1:** Aux,V and V,O orderings in main clauses, M. Dryer (p.c)

# The disharmonic languages in Dryer's sample

## VO&VAux

- Africa. Atlantic-Congo: KWA: Akan [aka]
- Eurasia. Indo-European: SLAVIC: Czech [ces], Serbian-Croatian [hbs].
- Oceania. Tiwi: TIWIAN: Tiwi [tiw].
- North America. Chinookan: CHINOOKAN: Chinook (Lower) [chh].
- South America. Arawakan: CARIBBEAN ARAWAKAN: Garífuna [cab].

# The disharmonic languages in Dryer's sample

## OV&AuxV

- Africa. Siamou: SIAMOU: Seme [sif].
- Eurasia. Indo-European: SLAVIC: Sorbian [hsb]. Indo-European: IRANIAN: Zazaki [diq]. Sino-Tibetan: BODIC: Balti [bft].
- North America. Oto-Manguean: CHICHIMEC: Chichimeca-Jonaz [pei].
- South America. Chibchan: GUAYMIIC: Ngäbere [gym], Buglere [sab].



The disharmony is not categorically prohibited, not even the Final-over-Initial type (contra the FOFC, Biberauer et al. 2008 a.m.o.). They are, however, very rare, especially the Final-over-Initial type.

A solution at the interfaces with performance systems?

## Minimizing domain

---

# Hawkins' processing account

Disharmony leads to an increased parsing effort, equated with how many words need to be processed for an XP to be constructed (Hawkins 1990, 1994, 2014):

- **I[mmEDIATE]C[onstituent]-to-word** ratio within the Constituent Recognition Domain (CRD): “[f]or each word and its dominating IC, the IC total is divided by the word total at that point, and the result is expressed as percentage (e.g. 2/3, or the second IC of the domain divided by the third word, i.e. 67%)”
- “[a]n **aggregate IC-to-word ratio** for the whole domain is ... computed by averaging the percentages for all the words in the domain”.

Minimizing word distance between heads – with a full headedness harmony being the extreme minimal case – means to a higher IC-to-word ratio.

# Why does disharmony survive?

Hawkins (1990): disharmonic systems, which can result from independent changes, are predicted to be more short-lived than harmonic ones. H. shows the prediction to be borne out in a sample of ancient Indo-European Languages.

However, assuming that children converge on a particular probability distribution over competing grammatical analyses (in production) in function of grammars' respective success in analyzing the input data (Yang 2002): why do disharmonic grammars “survive” across generations at all?

# A case study: Old English

Old English: a mixture of harmonic (VO&AuxV) and disharmonic (OV&AuxV) alignments.

## VO&AuxV

- (1) Nu sceole we gehyran þæt halige godspell mid onbryrdnysse  
now shall we hear the holy Gospel with zeal  
“We shall now eagerly hear the holy Gospel”  
(coaelhom,+AHom\_8:12.1168)

## OV&AuxV

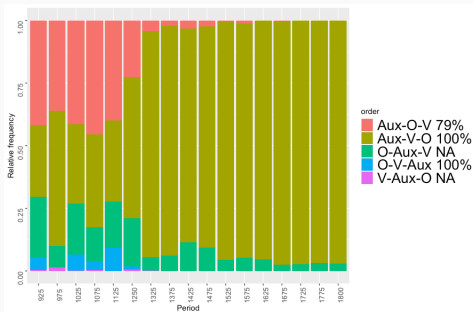
- (2) Æfter þon sceal man wryt drenc sellan  
after that shall man herbal drink administer  
“After that, man should administer a herbal drink”  
(colaece,Lch\_II\_[1]:1.13.3.83)

## A case study: Old English

- (3) þæt ure Drihten **sceolde** þa wundra **wyrca**n ...  
that our lord should those miracles work ...  
“that our lord should perform those miracles...”  
(coaelhom,+AHom\_4:12.522, YCOE)
- (4) a. þæt ure Drihten [<sub>TP</sub> **sceolde** [<sub>VP</sub> [<sub>NP</sub> þa wundra] **wyrca**n]]  
Aux-O-V IC-to-word %79
- b. þæt ure Drihten [<sub>TP</sub> [<sub>VP</sub> [<sub>NP</sub> þa wundra] **wyrca**n] **sceolde**]  
O-V-Aux IC-to-word %100
- c. þæt ure Drihten [<sub>TP</sub> **sceolde** [<sub>VP</sub> **wyrca**n [<sub>NP</sub> þa wundra]]]  
Aux-V-O IC-to-word %100

# A case study: Old English

The disharmony remained possible until Middle English.



**Figure 1:** TP-VP alignments in main clauses (N = 14575) in historical English

TP's aggregate IC-to-word is not counted for O-aux-V and V-aux-O: either O or V vacates the TP.

## The role of case

---



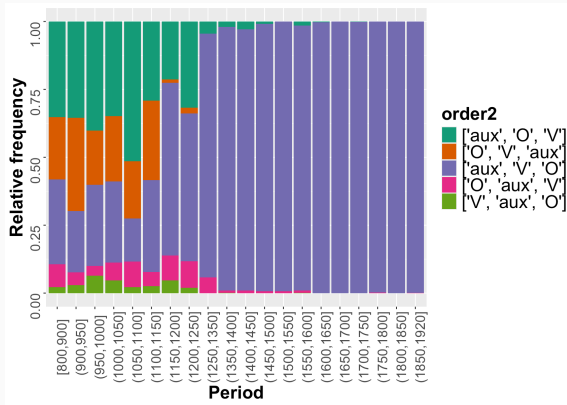
Hawkins (2014, p. 91) considers heads as a sub-case of a mother-node:

- Mother Node Construction (Hawkins 1994, p. 62) In the left-to-right parsing of a sentence, **if any word of syntactic category C uniquely determines a phrasal mother node M**, in accordance with the PS rules of the grammar, then M is immediately constructed over C.

The accusative case morpheme fits the definition as it uniquely determines VP.

# Case and OV&AuxV systems

## Disharmonic OV&AuxV systems in OE



**Figure 2:** TP-VP alignments in subordinate clauses (N=10 536) in historical English.

→ “slow processing” **Aux-O-V** accounts for roughly 30% in Old English

# Disharmonic OV&AuxV systems in OE

**Hypothesis:** the accusative case immediately constructs VP → Aux-O<sub>acc</sub>-V has an IC-to-word ratio of 100%

- (5) þæt ure Drihten [<sub>TP</sub> **sceolde** [<sub>VP</sub> [<sub>NP</sub> þ**one** wundor] wyrca]]  
“that our lord should perform those miracles...”

## Corpus study:

- **Data:** historical English prose (Kroch, 2020; Taylor et al., 2003)
- **Extraction:** subordinate clauses with an **object NP** (N=10 536)
- **Annotation:** accusative case marking based on determiners, adjectives, and nouns
  - **unambobj:** the object marking itself being unambiguously accusative
  - **unambobj\_bysbj:** unambiguous object marking as implied by a previously seen subject with unambiguous nominative case
  - **ambobj:** ambiguous marking (case marking does not allow for identifying an object)

# Disharmonic OV&AuxV systems in OE

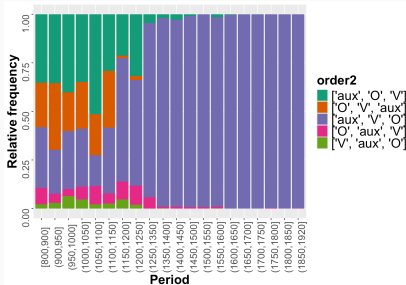


Figure 3: Word orders

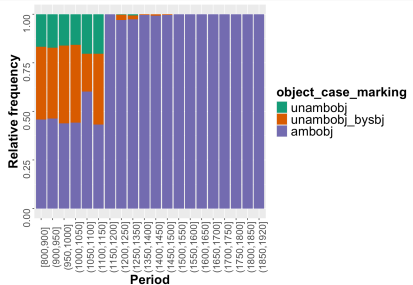


Figure 4: Object case marking

- object case marking was lost after 1150, while the Aux-O-V order disappeared after 1250
- all cases of object case marking (**unambobj** + **unambobj\_bysbj**) (right) roughly correspond to the share of **Aux-O-V** (left)

# Case?

In OE: case made the “survival” of disharmony possible.

What about other OV&AuxV systems?

- 7 languages in Dryer’s sample
  - Seme (Siamou):** no case, no object agreement, fixed S(Aux)OV
  - Sorbian: case
  - Zazaki: case
  - Balti: case
  - Chichimeca-Jonaz: polysynthetic, verb object agreement
  - [Ngäbere: misclassified? Tense/aspect particles *follow* V, it’s a harmonic OV&VAux, Payne 1981, p. 45]
  - [Buglere: misclassified? A closely related Chibchan language]
- Swiss German (embedded only): case on articles
- Yiddish (embedded only, moves to VO&AuxV with nominal arguments (reviewer)): case on articles

What about Dutch, Afrikaans, Seme?

- (6) Een vriend heeft een kennis            aan mij voorgesteld.  
a    friend has an acquaintance to me introduced  
“A friend introduced to me an acquaintance.”

- In Middle Dutch we find harmonic/disharmonic variation:  
VO&AuxV vs. OV&AuxV, disharmony wins in main clauses, and  
the case is lost.

Does *fixed* O-V order also make disharmony possible?

# Case or full predictability

Hypothesis: OV&AuxV disharmony is possible if

- there is case or object agreement adjacent to Aux OR
- there is no VO/OV variation

Conjecture: the disappearance of case leads to the disappearance of VO/OV variation.

Future task: supplement IC-to-word ratio with a measure of *uncertainty* for the following constituent.

# Thank you!

The work is funded by the European Union (ERC, CAUSALITY, grant number 101042427). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Alexandra Simonenko  
alexandra.simonenko@ugent.be

Yiming Liang  
yiming.liang@ugent.be





## References

---



Biberauer, T., Holmberg, A., & Roberts, I. **Structure and linearization in disharmonic word orders.** In: In *Proceedings of the 26th West Coast Conference on Formal Linguistics*, Charles. B. Chang and Hannah J. Haynie (eds.) 2008, 96–104.







Hawkins, J. (1990). **Seeking motives for change in typological variation.** In S. K. William Croft Keith Denning (Ed.), *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday* (pp. 95–128). John Benjamins Amsterdam.



Hawkins, J. (1994). **A performance theory of order and constituency.** Cambridge University Press.



Hawkins, J. (2014). **Cross-linguistic variation and efficiency.** Oxford University Press.

-  Kroch, A. (2020). *Penn Parsed Corpora of Historical English LDC2020T16*. (Web download.).
-  Payne, D. L. (1981). *The pragmatics of word order: Information and structure in ngäbere (guaymí)* [Ph.D. dissertation]. University of California, Berkeley.
-  Taylor, A., Warner, A., Pintzuk, S., & Beths, F. (2003). *The York-TorontoHelsinki Parsed Corpus of Old English Prose (YCOE)*..
-  Yang, C. (2002). **Grammar competition and language change**. In D. Lightfoot (Ed.), *Syntactic effects of morphological change* (pp. 367–409). Cambridge University Press.

# Corpus study: disharmonic system in OE (main clauses)

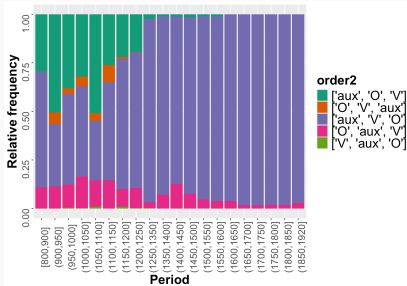


Figure 5: Word orders

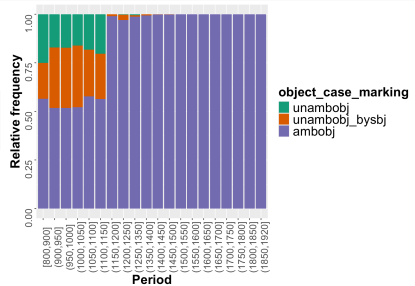


Figure 6: Object case marking