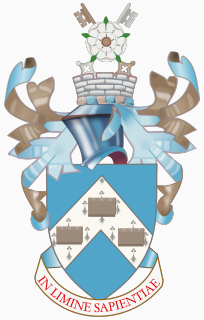


Topicalization is (Still) Disappearing: information uniformity as a dimension of specialization

DiGS 25, Workshop: *The Legacy of the Penn Corpora*

Joel C. Wallenberg, University of York

Today's Talk

Topicalization as competing grammars

Topicalization and Accent Clash

Topicalization and Informational Uniformity

- Methods:
  testing information spread as a dimension of specialization

Conclusions

Topicalization as competing grammars

Topicalization and Accent Clash

Topicalization and Informational Uniformity

Methods:
testing information spread as a dimension of specialization

Conclusions

Further Work: Non-V2 Object Topicalization

- English object topicalization is possible in two pragmatic/IS contexts (Prince, 1985, 1998, 1999):

Further Work: Non-V2 Object Topicalization

- English object topicalization is possible in two pragmatic/IS contexts (Prince, 1985, 1998, 1999):
  - **contrastive topic** with focus elsewhere in clause.

Further Work: Non-V2 Object Topicalization

- English object topicalization is possible in two pragmatic/IS contexts (Prince, 1985, 1998, 1999):
  - **contrastive topic** with focus elsewhere in clause.
  - **focus with scalar contrast** (cf. also "true A'-movement of Frey 2006 and subs.).

Further Work: Non-V2 Object Topicalization

- English object topicalization is possible in two pragmatic/IS contexts (Prince, 1985, 1998, 1999):
  - **contrastive topic** with focus elsewhere in clause.
  - **focus with scalar contrast** (cf. also "true A'-movement of Frey 2006 and subs.).
- However, the movement is always optional even in these contexts.

Further Work: Non-V2 Object Topicalization

- English object topicalization is possible in two pragmatic/IS contexts (Prince, 1985, 1998, 1999):
  - **contrastive topic** with focus elsewhere in clause.
  - **focus with scalar contrast** (cf. also "true A'-movement of Frey 2006 and subs.).
- However, the movement is always optional even in these contexts.
- Both contexts require an accent, intonational peak, on the fronted constituent.

English "Topicalization"

- Felicitous in two English discourse contexts, both of which require a certain type of contrast and accent to appear on the fronted XP.

  (1) She's going to use three groups of mice. One, she'll feed them mouse chow, just the regular stuff they make for mice. Another she'll feed them veggies. And the third she'll feed junk food.

  (2) She was here two years. [checking transcript] Five semesters she was here.
  (Prince, 1999)

English Topicalization

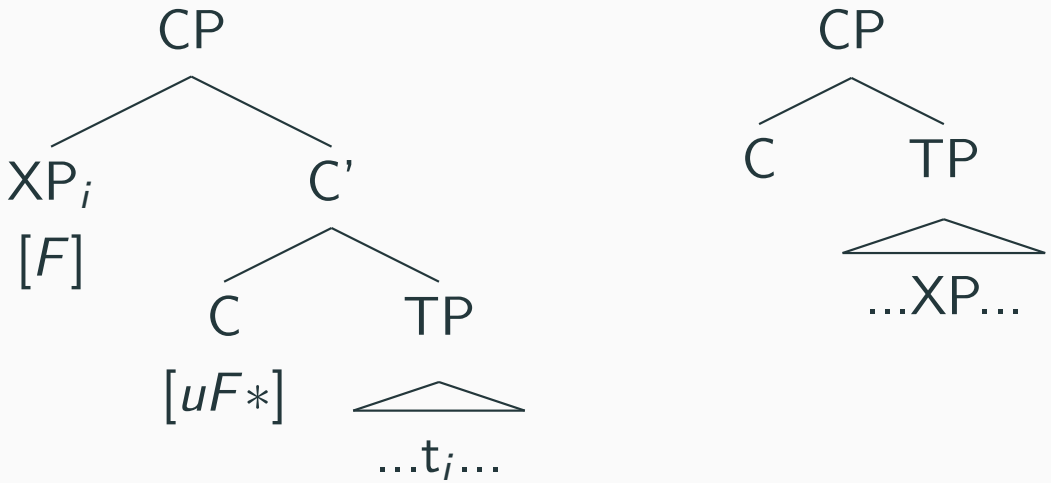
- As long as the accent pattern is kept constant, both orders are felicitous:

  (3) She's going to use three groups of mice. One, she'll feed them mouse chow, just the regular stuff they make for mice. Another she'll feed them veggies. And **the third** she'll feed **junk food**.

  (4) She's going to use three groups of mice. One, she'll feed them mouse chow, just the regular stuff they make for mice. Another she'll feed them veggies. And she'll feed **the third junk food**.

Sketch of Topicalization in Minimalism

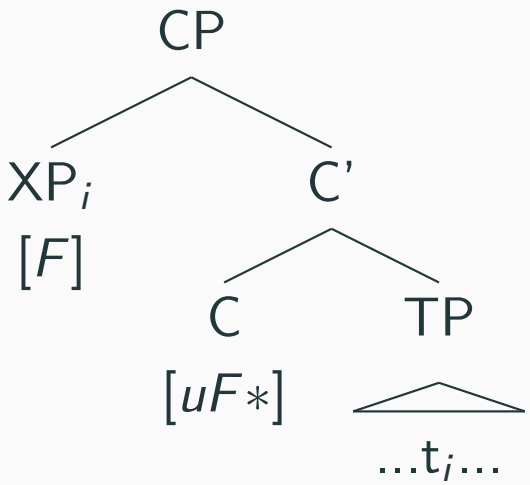
- Move is triggered by the feature content of some head.
- Given "Merge...preempts Move" (Chomsky, 2000), a feature cannot encode optional movement.
- Therefore, optional movement involves a choice (for the **Numeration**) between two variants of a functional head, out of an inventory of possible heads.
  - In line with "Borer-Chomsky Conjecture" (Borer, 1984), (Baker, 2008).

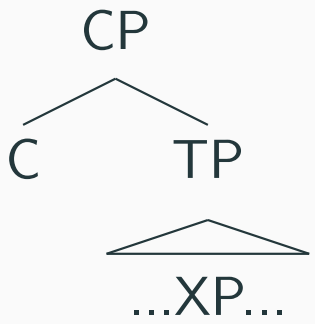


- This is the core case of morphosyntactic doublet (i.e. competing heads) described in (Kroch, 1994).

Sketch of Topicalization in Minimalism

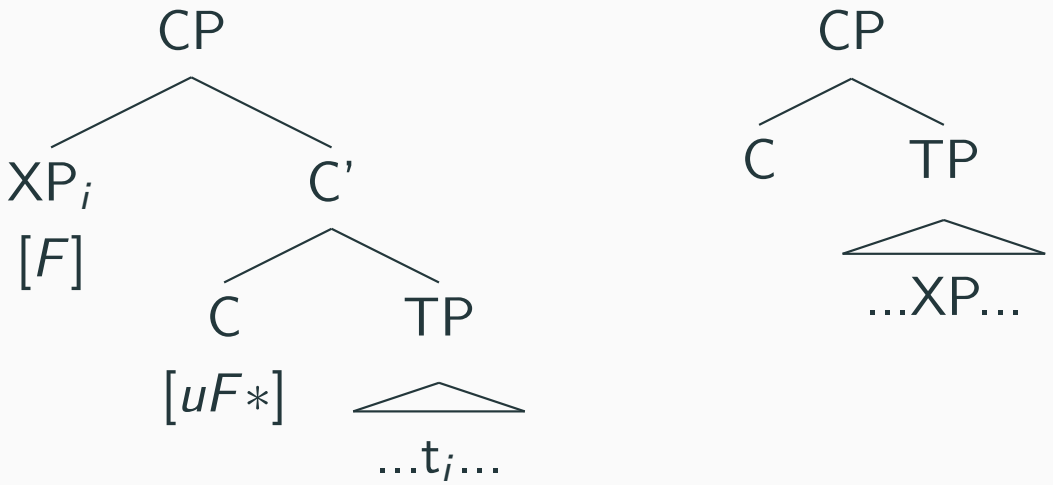
- To be slightly more precise...





Sketch of Topicalization in Minimalism

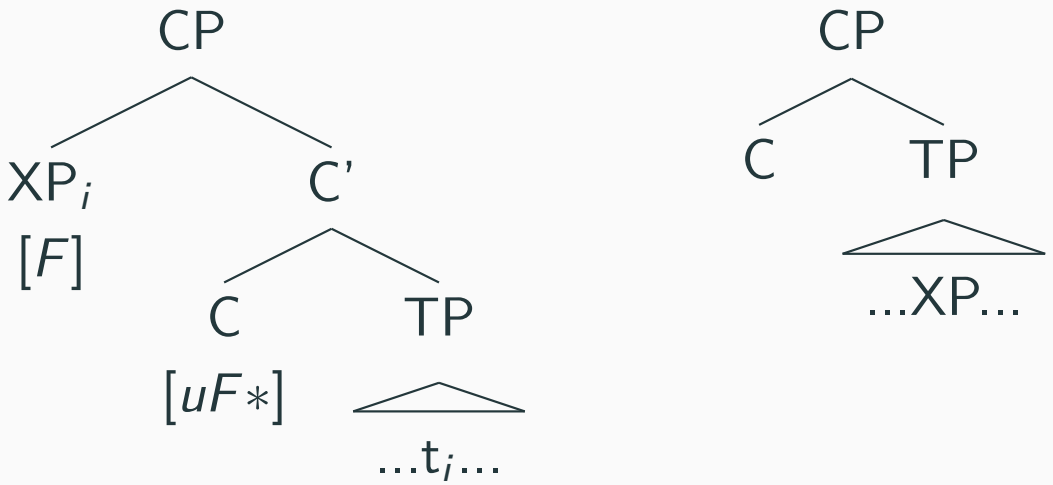
- To be slightly more precise...



- If an appropriate contrastive-top or contrastive-foc constituent exists, the language user can mark it with strong feature [F] and choose C[$uF*$], or choose [C].

Sketch of Topicalization in Minimalism

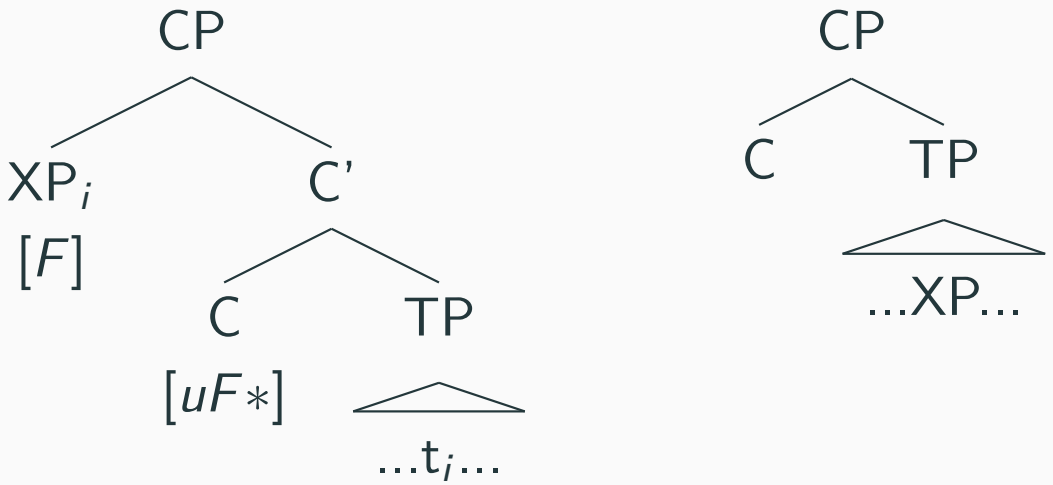
- To be slightly more precise...



- If an appropriate contrastive-top or contrastive-foc constituent exists, the language user can mark it with strong feature [F] and choose C[$uF*$], or choose [C].
- If the speaker uses C[$uF*$] and [F] with an inappropriate constituent, it's an infelicitous topicalization.

Sketch of Topicalization in Minimalism

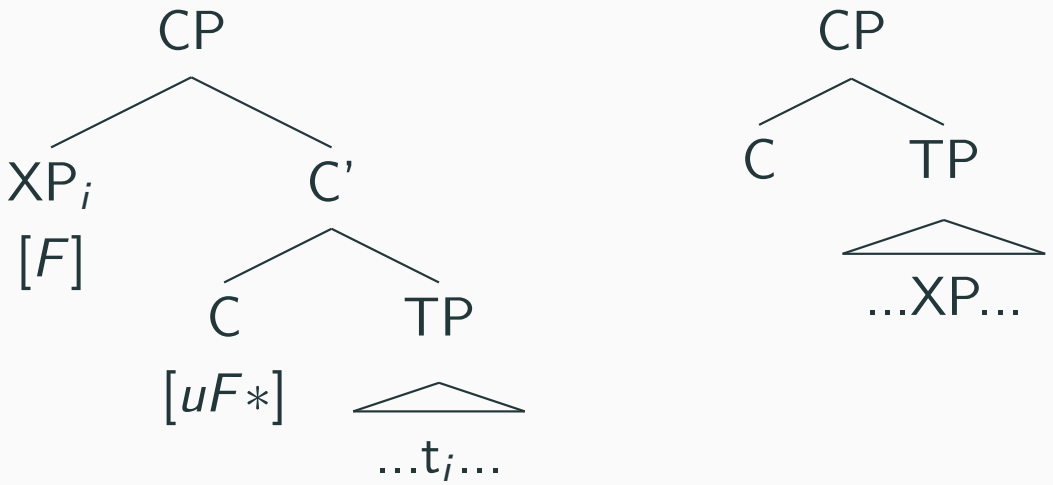
- To be slightly more precise...



- If an appropriate contrastive-top or contrastive-foc constituent exists, the language user can mark it with strong feature [F] and choose C[$uF*$], or choose [C].
- If the speaker uses C[$uF*$] and [F] with an inappropriate constituent, it's an infelicitous topicalization.
- The doublet exists when an appropriate contrastive constituent exists.

Sketch of Topicalization in Minimalism

- To be slightly more precise...



- If an appropriate contrastive-top or contrastive-foc constituent exists, the language user can mark it with strong feature [F] and choose C[$uF*$], or choose [C].
- If the speaker uses C[$uF*$] and [F] with an inappropriate constituent, it's an infelicitous topicalization.
- The doublet exists when an appropriate contrastive constituent exists.
- **Cost:** Default clausal prosody must be altered.

Topicalization as competing grammars

Topicalization and Accent Clash

Topicalization and Informational Uniformity

Methods:
testing information spread as a dimension of specialization

Conclusions

Topicalization and Accent Clash

- Work showing topicalization is sensitive to accent clash (Speyer, 2008, 2010).

Topicalization and Accent Clash

- Work showing topicalization is sensitive to accent clash (Speyer, 2008, 2010).

(9) Arable for Corn, you may wash with pale Straw-Colour
(WYLD-1725-2,114.263 in *Penn Parsed Corpus of Modern British English*, (Kroch et al., 2016))

(10) But this auspicious calm the savage Maximin at last disturbed.
(TURNER1-1799-1,45.214 in PPCMBE)

(11) ? Joel Viola likes (but Memel she doesn't).

(12) Joel, Viola likes (but Memel she doesn't).

From PPCMBE, nominal objects ($\chi^2 = 260$, $p < 2 \times 10^{-16}$):

	Fronted	In Situ	Prop. Fronted
Pron Sbj	631	20,071	0.031
Nom Sbj	119	16,808	0.0071

Topicalization and Accent Clash

- They do exist in the wild, but usually with pauses:

(13) All sins, except a sin against itself, Love should forgive.
(WILDE-1895-1,70.859 in PPCMBE)

(14) All lives, save loveless lives, true Love should pardon.
quad (WILDE-1895-1,70.860 in PPCMBE)

Topicalization and Accent Clash

- Speyer also showed that as V2 options are lost in English, direct object topicalization declines.
  - Note: not all apparent V2 orders have the same structure in ME and OE (Haeberli, 1999, 2000, 2002; Kroch & Taylor, 1997), ao.

(15) þone wæterscipe beworhte **se wisa cyning Salomon**
the conduit built the wise king Salomon
"The wise King Salomon constructed the conduit"
(Ælfric's Catholic Homilies, 10th c., from YCOE (Taylor, Warner, Pintzuk, & Beths, 2003)))

The accent clash account

- **Question:** what sense of accent clash includes two adjacent DPs, but excludes a lexical verb?
- It must be at a constituent level, and so fairly abstract.
- **Question:** do we expect the decline of topicalization to halt once V2 is lost, or for topicalization to specialize for the non-clash case?
- **Observation:** an account of the same facts in terms of informational spread – the Smooth Signal Redundancy Hypothesis (Shannon, 1948; Fenk & Fenk, 1980; Fenk-Oczlon, 2001; Aylett & Turk, 2004; Levy & Jaeger, 2007; Frank & Jaeger, 2008; Jaeger, 2010; Turk, 2010; Chingacham, Demberg, & Klakow, 2023; Wallenberg, Bailes, Cuskley, & Ingason, 2021)

  – makes the same predictions as to subject type and pauses, but also two more...

Topicalization as competing grammars

Topicalization and Accent Clash

Topicalization and Informational Uniformity

Methods:
testing information spread as a dimension of specialization

Conclusions

Topicalization and Informational Uniformity

Info Content: pronouns $<$ verbs $<$ nominals

- Potential **improvement in informational uniformity** is a continuous variable, **predicting slow change over time:**

Topicalization and Informational Uniformity

Info Content: pronouns $<$ verbs $<$ nominals

- Potential **improvement in informational uniformity** is a continuous variable, **predicting slow change over time:**
  - The two syntactic forms overlap in function (Prince, 1998), and therefore compete in use (Kroch, 1994).

Topicalization and Informational Uniformity

Info Content: pronouns < verbs < nominals

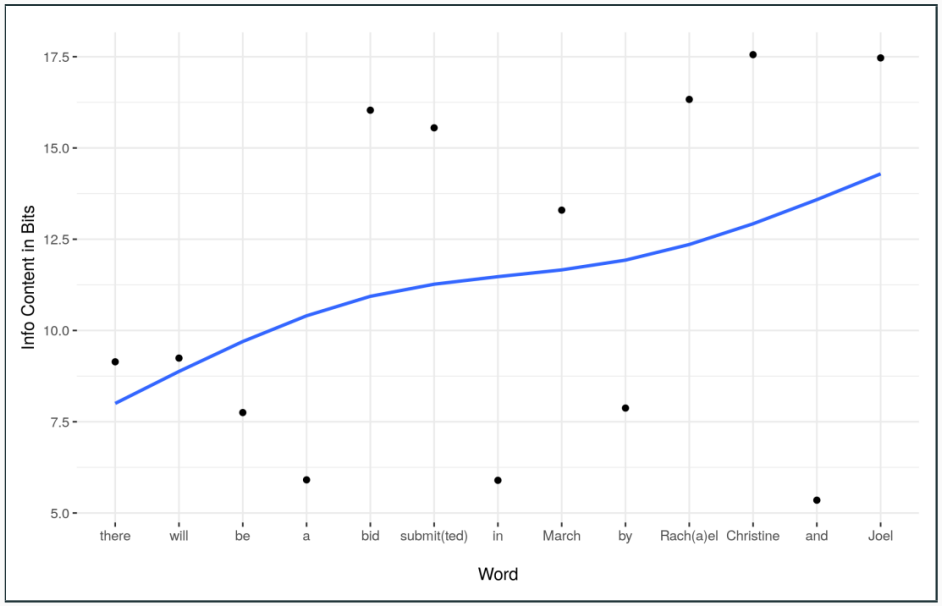
- Potential **improvement in informational uniformity** is a continuous variable, **predicting slow change over time:**
  - The two syntactic forms overlap in function (Prince, 1998), and therefore compete in use (Kroch, 1994).
  - The competition is mitigated by specialization, but along a continuous dimension, and so total specialization is impossible and slow change results (Wallenberg, 2016).

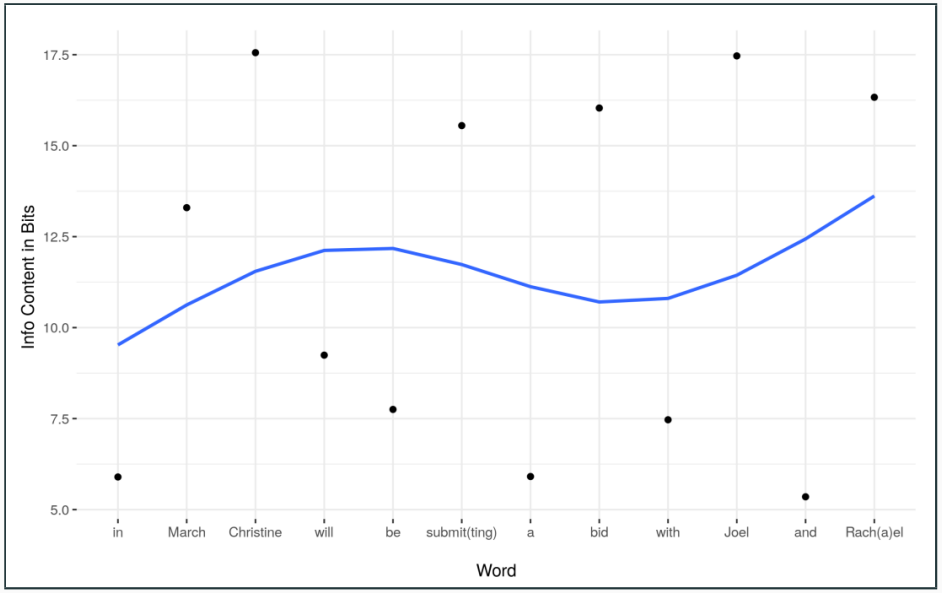
Uniformity of information distributions

Strings can vary in terms of how the information is spread or clustered. The **order of elements** in a sentence derives **more uniform** or **more asymmetric** distributions of information (Cuskley, Bailes, & Wallenberg, 2021).

Uniformity of information distributions

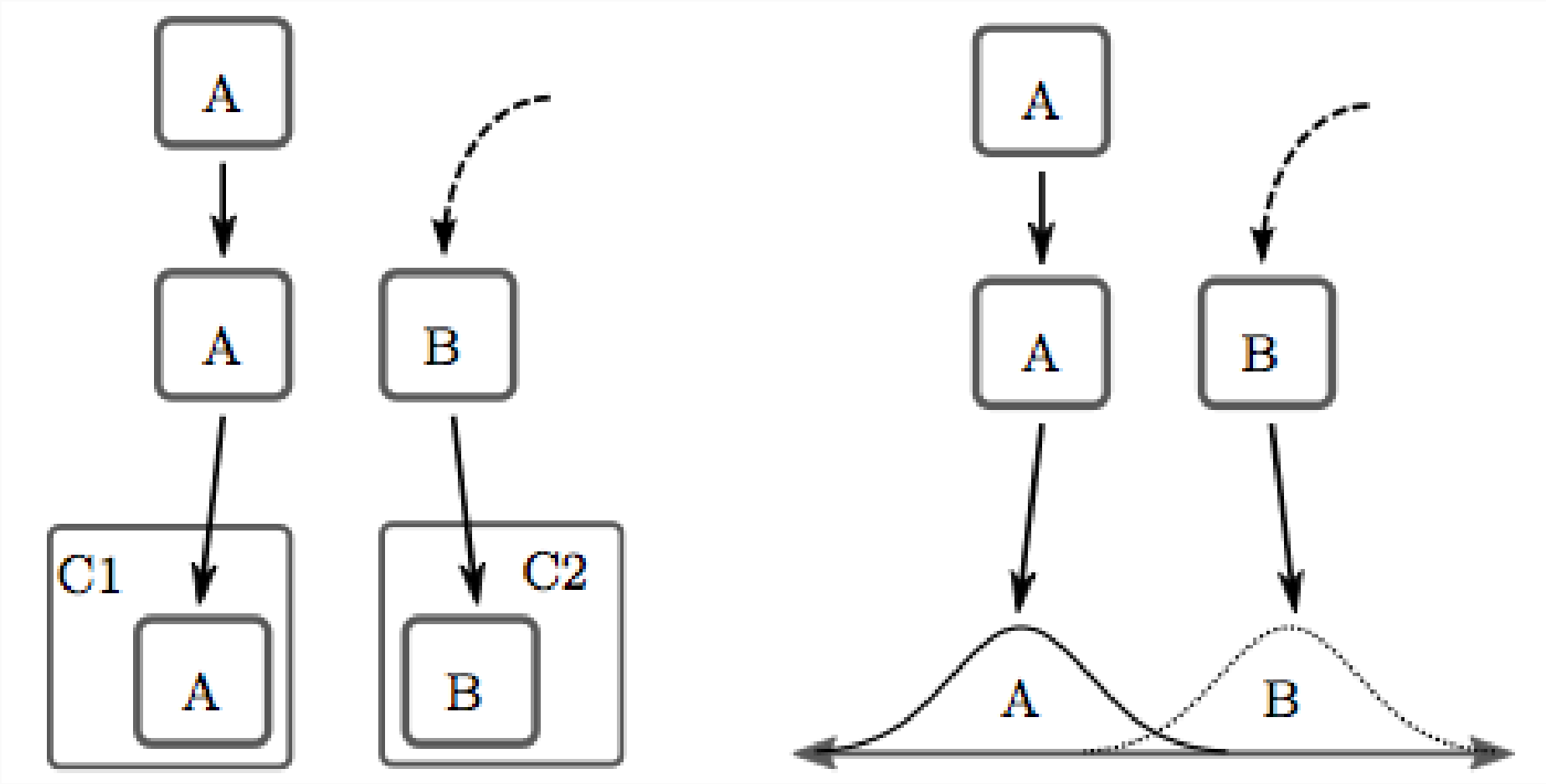
Strings can vary in terms of how the information is spread or clustered. The **order of elements** in a sentence derives **more uniform** or **more asymmetric** distributions of information (Cuskley et al., 2021).





We can measure how uniform an information distribution is with **Deviation of the Rolling Mean (DORM)** (Cuskley et al., 2021).

Hypothesis: specialization along continuous dimensions



(figure courtesy of J. Fruehwald)

Topicalization and Informational Uniformity

Predictions:

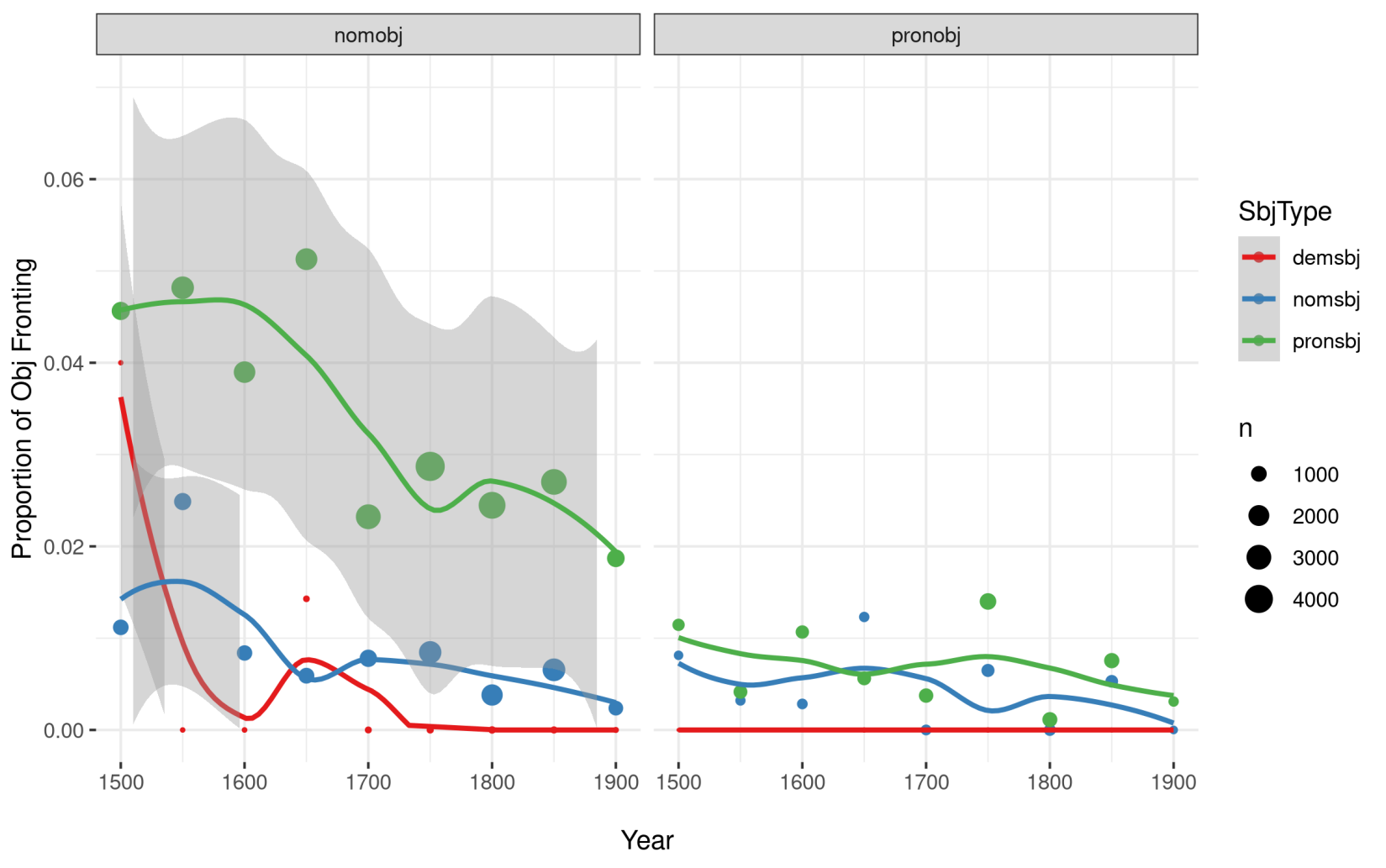
1. Slow change over time.
2. No specialization by subject type over time (i.e. CRE).
3. Topicalization should also be disfavored when Pron-Obj lands adjacent to Pron-Sbj.
4. Topicalizations should lead to lower calibrated DORMs.
5. We should see specialization by DORM-improvement over time.

Topicalization and Informational Uniformity

Predictions:

1. Slow change over time.
2. No specialization by subject type over time (i.e. CRE).
3. Topicalization should also be disfavored when Pron-Obj lands adjacent to Pron-Sbj.
4. ~~**Topicalizations should show lower calibrated DORMs.**~~
5. **?** We should see specialization by DORM-improvement over time. **?**

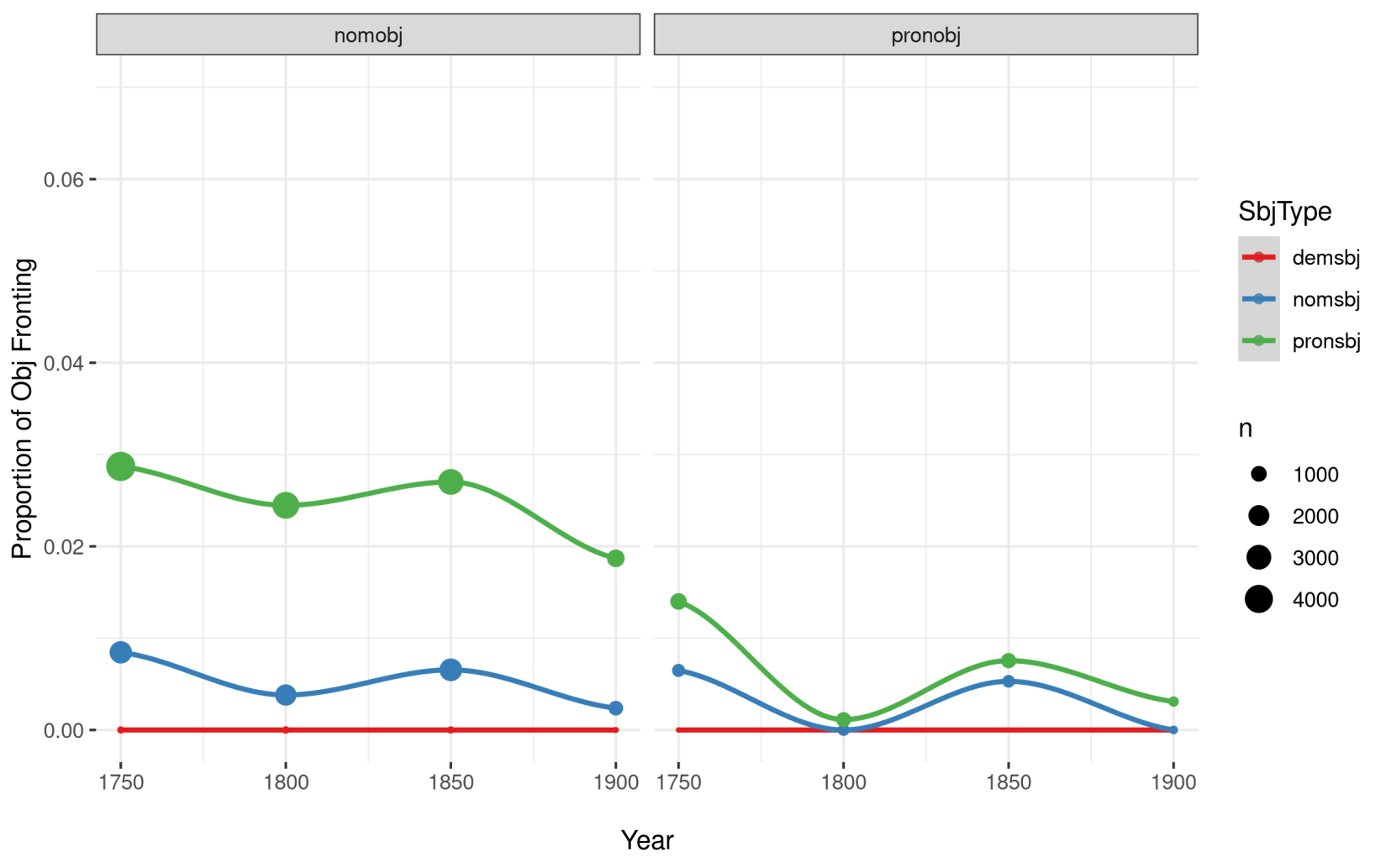
Object Topicalization, PPCEME/MBE, non-V2 clauses



(47,115 finite main clauses; dem Sbjs omitted)

Note: significant effects of **Year, Obj, Sbj**, sig **Obj:Sbj** interaction, but no sig Obj, Sbj interactions with Year (Constant Rate Effect).

Object Topicalization, PPCEME/MBE, non-V2 clauses



Note: significant effects of **Year, Obj, Sbj**, sig **Obj:Sbj** interaction, but no sig Obj, Sbj interactions with Year (Constant Rate Effect).

Methods: testing information spread as a dimension of specialization

- Above: CorpusSearch coding queries.

Methods: testing information spread as a dimension of specialization

- Above: CorpusSearch coding queries.
- Word frequencies derived from the ECCO portion (100m words) of The Penn-York Computer-annotated Corpus of a Large amount of English (PYCCLE; Ecay, 2015), converted to bits, plus PPCEME and PPCMBE (4.5m together).

Methods: testing information spread as a dimension of specialization

- Above: CorpusSearch coding queries.
- Word frequencies derived from the ECCO portion (100m words) of The Penn-York Computer-annotated Corpus of a Large amount of English (PYCCLE; Ecay, 2015), converted to bits, plus PPCEME and PPCMBE (4.5m together).
- Calibrated DORMS based on other possible orders of the string (UIDO).

Methods: testing information spread as a dimension of specialization

- Above: CorpusSearch coding queries.
- Word frequencies derived from the ECCO portion (100m words) of The Penn-York Computer-annotated Corpus of a Large amount of English (PYCCLE; Ecay, 2015), converted to bits, plus PPCEME and PPCMBE (4.5m together).
- Calibrated DORMS based on other possible orders of the string (UIDO).
- Used this word-bit frequency dictionary to create a Calibrated DORM-dictionary for every token ID in PPCEME and PPCMBE.

Methods: testing information spread as a dimension of specialization

- Above: CorpusSearch coding queries.
- Word frequencies derived from the ECCO portion (100m words) of The Penn-York Computer-annotated Corpus of a Large amount of English (PYCCLE; Ecay, 2015), converted to bits, plus PPCEME and PPCMBE (4.5m together).
- Calibrated DORMS based on other possible orders of the string (UIDO).
- Used this word-bit frequency dictionary to create a Calibrated DORM-dictionary for every token ID in PPCEME and PPCMBE.
- Appended Calibrated DORMs to output of CS coding queries.

DORM: Deviation of the Rolling Mean

en	eg	skal	sjá	yður	aftur
6.79	6.15	10.1	9.25	6.15	10.4

DORM: Deviation of the Rolling Mean

en		eg	skal	sjá	yður	aftur
6.79		**6.15**	10.1	9.25	6.15	10.4
	6.47					

DORM: Deviation of the Rolling Mean

en		eg		skal	sjá	yður	aftur
6.79		**6.15**		**10.1**	9.25	6.15	10.4
	6.47		**8.12**				

DORM: Deviation of the Rolling Mean

en			eg			skal			sjá	yður	aftur
6.79			6.15			**10.1**			**9.25**	6.15	10.4
	6.47			8.12			**9.67**				

DORM: Deviation of the Rolling Mean

en			eg			skal		sjá		yður	aftur
6.79			6.15			10.1		**9.25**		**6.15**	10.4
	6.47			8.12			9.67		**7.70**		

DORM: Deviation of the Rolling Mean

en		eg		skal		sjá		yður		aftur
6.79		6.15		10.1		9.25		**6.15**		**10.4**
	6.47		8.12		9.67		7.70		**8.29**	

DORM: Deviation of the Rolling Mean

en		eg		skal		sjá		yður		aftur
6.79		6.15		10.1		9.25		6.15		10.4
	6.47		8.12		9.67		7.70		8.29	

Sample variance of rolling means = 1.33 bits^2

low DORM $\rightarrow$ more uniform

high DORM $\rightarrow$ more lopsided

UIDO: Uniform information density optimization

- For a given array of values (e.g. information values of words in a sentence):
  1. The array of information content values is ordered greatest to least.
  2. Starting with the second and penultimate value in the array and moving inward, every other number is swapped, mixing up the large and small values.
  3. DORM is calculated for the resulting array.
  4. If this is lower than the original DORM, the array is kept. Otherwise, the algorithm proceeds with the original array.
  5. Every pair of numbers is swapped, checking to see whether the DORM is lowered after each swap. The lowest DORM is the UIDO.
- Calibrates DORMs by a baseline "best DORM", which should also capture **improvement in DORM**.

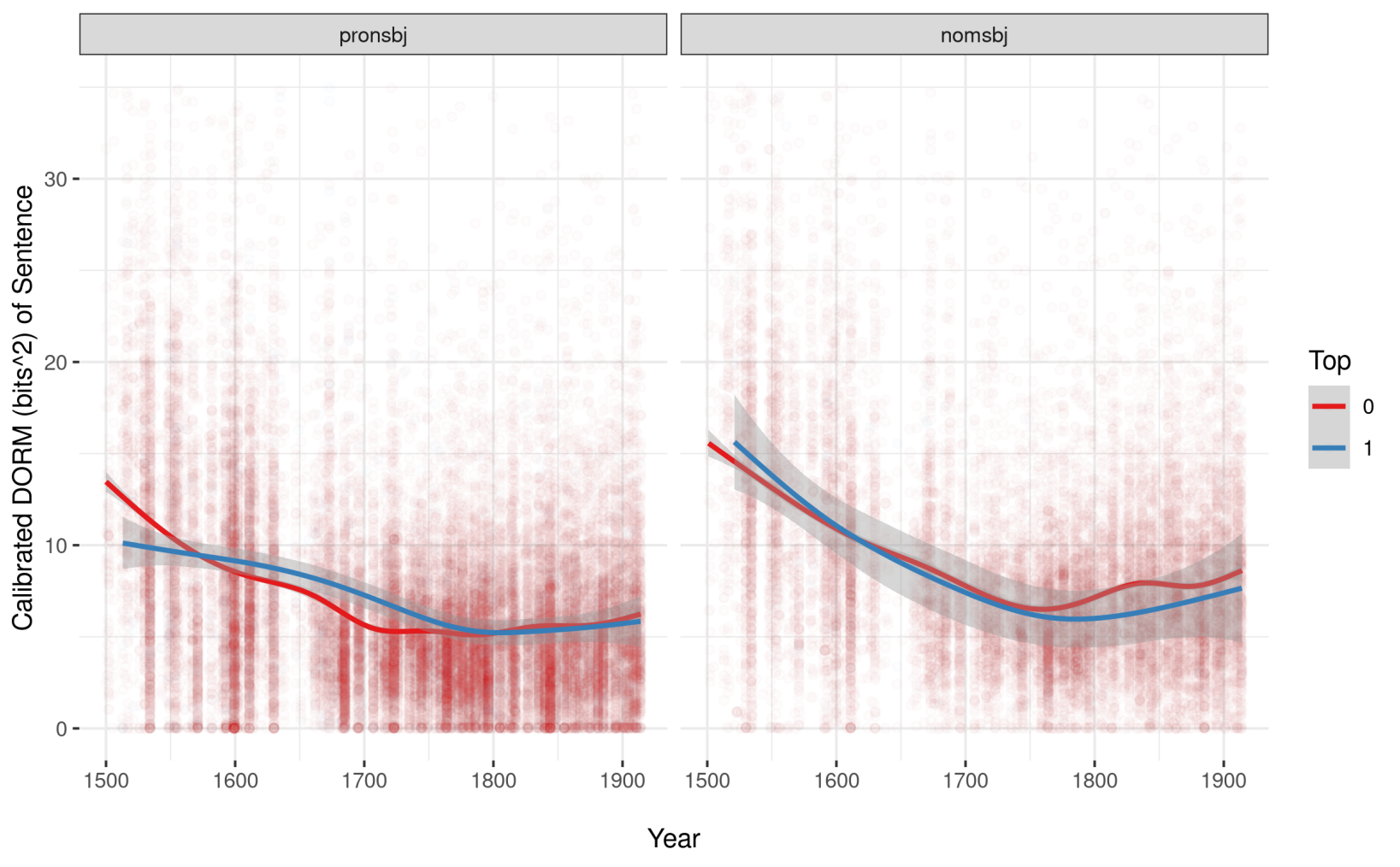
Calibrated DORM Worked Before

- Based on this account of information uniformity, (Wallenberg et al., 2021) predicted a previously undetected argument-type effect in the English and Icelandic OV to VO changes.

...but not super great this time: Results

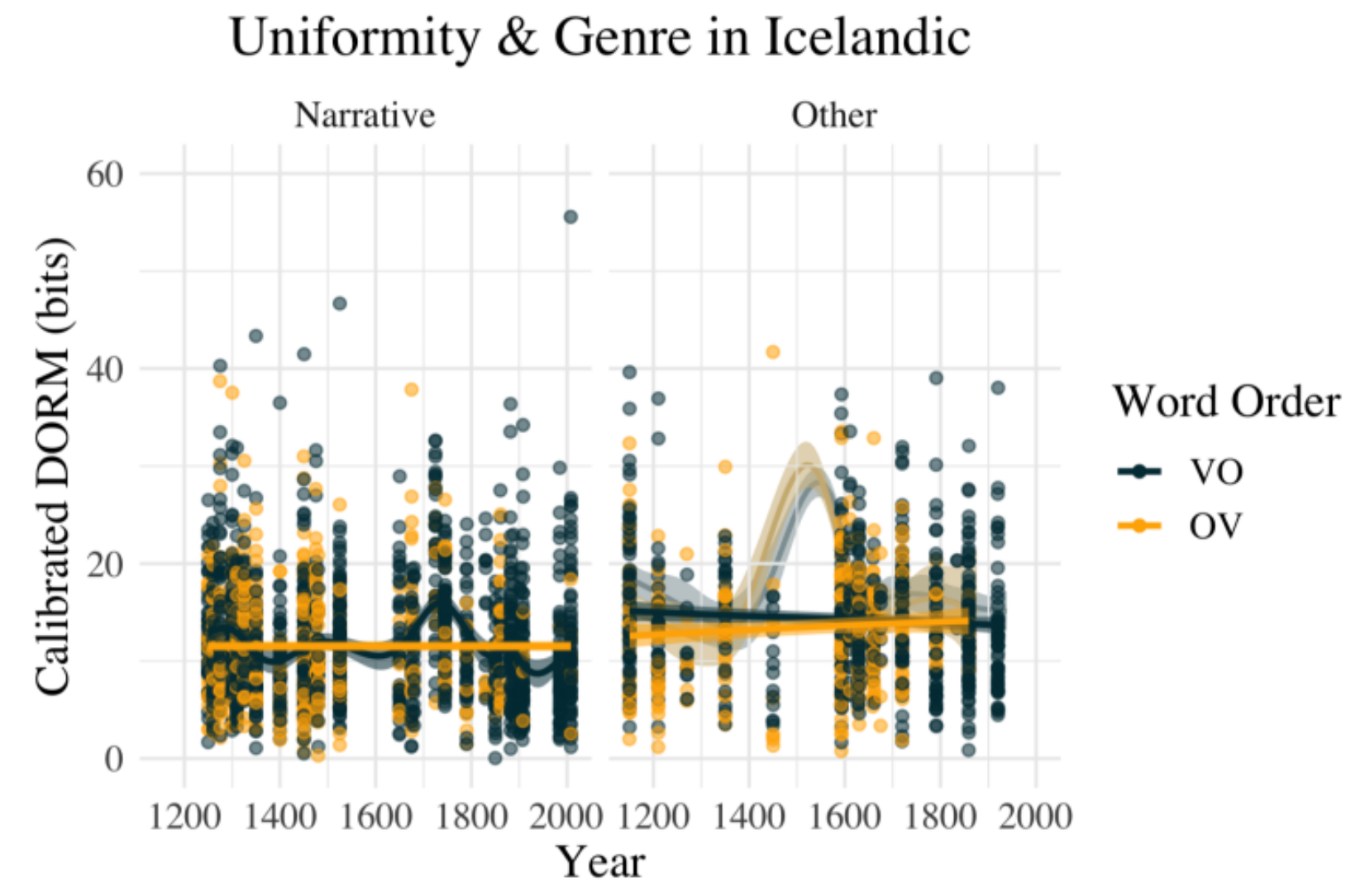
- Mixed-effects logistic regression predicting Top with SbjType, ObjType, Year, SbjType:ObjType, and calDORM showed **no significant main effect of calDORM** (β = -0.004, p = 0.51).
- Predicting Top with calDORM alone was borderline (p = 0.06).
- The SbjType:ObjType interaction **disfavors Top when both Sbj and Obj are pronouns** (β = -0.91, p = 0.004), as well as when both are nominals.

Results: Calibrated DORMs over Time



- Possible specialization in the nomsbj case: modeling DORM with 3-way Sbj:Top:Year interaction improves model fit with $p = 0.001$, but only AIC confirms: -10.

That Decline is Odd: Icelandic calDORMs over Time



- The average uniformity of sentences is constant across the history of Icelandic.

Topicalization as competing grammars

Topicalization and Accent Clash

Topicalization and Informational Uniformity

Methods:
testing information spread as a dimension of specialization

Conclusions

Conclusions and Suggestions

- The information uniformity account makes the same predictions as the accent clash account, plus a few more that were supported.

Conclusions and Suggestions

- The information uniformity account makes the same predictions as the accent clash account, plus a few more that were supported.
- Perhaps caIDORM is a continuous dimension of specialization, slowing the change in non-V2 topicalization.

Conclusions and Suggestions

- The information uniformity account makes the same predictions as the accent clash account, plus a few more that were supported.
- Perhaps caIDORM is a continuous dimension of specialization, slowing the change in non-V2 topicalization.
- However, the evidence does not show that topicalization leads to lower DORMs overall.

Conclusions and Suggestions

- The information uniformity account makes the same predictions as the accent clash account, plus a few more that were supported.
- Perhaps caIDORM is a continuous dimension of specialization, slowing the change in non-V2 topicalization.
- However, the evidence does not show that topicalization leads to lower DORMs overall.
- Perhaps there is too much noise in the frequency estimates or DORMs (i.e. use EEBO too).

Conclusions and Suggestions

- The information uniformity account makes the same predictions as the accent clash account, plus a few more that were supported.
- Perhaps caIDORM is a continuous dimension of specialization, slowing the change in non-V2 topicalization.
- However, the evidence does not show that topicalization leads to lower DORMs overall.
- Perhaps there is too much noise in the frequency estimates or DORMs (i.e. use EEBO too).
- Perhaps I need to compare to alternative orders directly (Voigtmann & Speyer, 2023).

Conclusions and Suggestions

- The information uniformity account makes the same predictions as the accent clash account, plus a few more that were supported.
- Perhaps calDORM is a continuous dimension of specialization, slowing the change in non-V2 topicalization.
- However, the evidence does not show that topicalization leads to lower DORMs overall.
- Perhaps there is too much noise in the frequency estimates or DORMs (i.e. use EEBO too).
- Perhaps I need to compare to alternative orders directly (Voigtmann & Speyer, 2023).
- Could prosodic cost be the selectional pressure driving the change?

Conclusions and Suggestions

- The information uniformity account makes the same predictions as the accent clash account, plus a few more that were supported.
- Perhaps calDORM is a continuous dimension of specialization, slowing the change in non-V2 topicalization.
- However, the evidence does not show that topicalization leads to lower DORMs overall.
- Perhaps there is too much noise in the frequency estimates or DORMs (i.e. use EEBO too).
- Perhaps I need to compare to alternative orders directly (Voigtmann & Speyer, 2023).
- Could prosodic cost be the selectional pressure driving the change?
- **I now regularly engage in, and hear of, ways of manipulating the parsed corpora that I couldn't imagine a decade ago.**

Acknowledgements, and thanks for listening!

Thank you to my collaborators, audiences at NYU, Saarbrucken, Manchester, Konstanz, Edinburgh, and the late great Anthony Kroch.

Questions?

References i

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech*, *47*(1), 31–56.

Baker, M. (2008). The macroparameter in a microparametric world. In T. Biberauer (Ed.), *The Limits of Syntactic Variation* (pp. 351–374). John Benjamins.

Borer, H. (1984). *Parametric syntax: case studies in Semitic and Romance languages*. Dordrecht: Foris Publications.

Chingacham, A., Demberg, V., & Klakow, D. (2023). A data-driven investigation of noise-adaptive utterance generation with linguistic modification. In *2022 ieee spoken language technology workshop (slt)* (pp. 353–360).

References ii

Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays on minimalist syntax in honor of howard lasnik* (pp. 89–155). Cambridge, Mass.: MIT Press.

Cuskley, C., Bailes, R., & Wallenberg, J. (2021). Noise resistance in communication: Quantifying uniformity and optimality. *Cognition*, *214*, 104754.

Ecay, A. (2015). *The Penn-York Computer-annotated Corpus of a Large amount of English based on the TCP (PYCCLE-TCP).* Public release 1. https://github.com/uoy-linguistics/pyccle.

Fenk, A., & Fenk, G. (1980). Konstanz im kurzzeitgedächtnis-konstanz im sprachlichen informationsfluß. *Zeitschrift für experimentelle und angewandte Psychologie*, *27*, 402.

Fenk-Oczlon, G. (2001). Familiarity, information flow, and linguistic form. *Typological Studies in Language*, *45*, 431–448.

References iii

Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 30).

Frey, W. (2006). Contrast and movement to the german prefield. In V. Molnár & S. Winkler (Eds.), *The architectur of focus.* Berlin: De Gruyter Mouton.

Haeberli, E. (1999). On the word order 'XP-subject' in the Germanic languages. *Journal of Comparative German Linguistics*, *3*(1), 1–36.

Haeberli, E. (2000). Adjuncts and the syntax of subjects in Old and Middle English. In S. Pintzuk, G. Tsoulas, & A. Warner (Eds.), *Diachronic syntax: Models and mechanisms* (pp. 109–131). Oxford: Oxford University Press.

References iv

Haeberli, E. (2002). Observations on the loss of verb second in the history of english. In *Studies in comparative germanic syntax: Proceedings from the 15th workshop on comparative germanic syntax* (pp. 245–272).

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology*, *61*(1), 23–62.

Kroch, A. S. (1994). Morphosyntactic variation. In K. B. et al (Ed.), *Papers from the 30th regional meeting of the chicago linguistics society: Parasession on variation and linguistic theory.*

Kroch, A. S., Santorini, B., & Diertani, A. (2016). *Penn Parsed Corpus of Modern British English 2nd edition, release 1.* (Size $\sim$ 2.8 million words.)

References v

Kroch, A. S., & Taylor, A. (1997). Verb movement in old and middle english: Dialect variation and language contact. In A. van Kemenade & N. Vincent (Eds.), *Parameters of morphosyntactic change.* Cambridge: Cambridge University Press.

Levy, R. P., & Jaeger, F. T. (2007). Speakers optimize information density through syntactic reduction. In *Advances in neural information processing systems* (pp. 849–856).

Prince, E. (1985). Fancy syntax and shared knowledge. *Journal of Pragmatics*, *9*(1), 65–81.

Prince, E. (1998). On the limits of syntax, with reference to left-dislocation and topicalization. *Syntax and semantics*, 281–302.

Prince, E. (1999). How not to mark topics: 'Topicalization' in English and Yiddish. In *Texas linguistics forum* (chap. 8). University of Texas, Austin: Citeseer.

References vi

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Speyer, A. (2008). *Topicalization and clash avoidance: On the interaction of prosody and syntax in the history of English with a few glimpses at German* (Unpublished doctoral dissertation). University of Pennsylvania.

Speyer, A. (2010). *Topicalization and stress clash avoidance in the history of english*. Berlin/New York: Mouton de Gruyter.

Taylor, A., Warner, A., Pintzuk, S., & Beths, F. (2003). *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. `http://www-users.york.ac.uk/~lang22/YcoeHome1.htm`.

Turk, A. (2010). Does prosodic constituency signal relative predictability? a smooth signal redundancy hypothesis. *Laboratory phonology*, *1*(2), 227–262.

References vii

Voigtmann, S., & Speyer, A. (2023). Where to place a phrase?: An informational and generative approach to phrasal extraposition. *Journal of Historical Syntax*, *7*(6-19), 1–48.

Wallenberg, J. C. (2016). Extraposition is disappearing. *Language*, *92*(4), e237–e256.

Wallenberg, J. C., Bailes, R., Cuskley, C., & Ingason, A. K. (2021). Smooth signals and syntactic change. *Languages*, *6*(2), 60.