Distinguishing change and stability – a quantitative study of Icelandic oblique subjects

Anton Karl Ingason Einar Freyr Sigurðsson
Joel C Wallenberg

University of Iceland

June 4th 2011, DIGS 13, University of Pennsylvania

Overview

1. Diachrony of oblique subjects
   - No change since 12th century (so OV to VO is not to blame)
2. Quantitative evidence
   - Stable extraction contexts
   - Stable use immediately following finite verb
   - Stable conjunction reduction
3. Analysis: Probing for gender features
   - Making Icelandic Bantu
4. Conclusion

Goals of the paper

- **Empirical:** Show that the status of oblique experiencers in Icelandic as syntactic subjects has remained stable since the 12th century, ruling out changes from this period as causes (notably the OV to VO change).
- **Methodological:** Show how the use of quantitative evidence can make up for lack of categorical diagnostics in historical texts.
- **Theoretical:** Make sense of the diachrony and synchronic system by proposing that the EPP is sensitive to Gender rather than Case or other ϕ-features in oblique subject languages.
- **Practical:** Demonstrate the use of parsed corpora in diachronic syntax, in particular IcePaHC (Icelandic Parsed Historical Corpus) (Wallenberg et al. 2011).

Oblique subjects in Modern Icelandic

- Consensus that oblique arguments can have subject status in Modern Icelandic (since Andrews 1976; Thráinsson 1979):

(1) **Mér** líkar ekki öskuskýið.
Me.DAT likes not ash.cloud.THE
'I don't like the ash cloud'

Diagnostics

- For example, dative experiencers can be PRO subjects in control infinitives in Modern Icelandic, unlike German (Zaenen et al 1985; Haider 2005, etc.) or Yiddish (Santorini 1989).

(2) Hún vonast til að **PRO** ganga vel einni.
She hopes for to DAT do well alone.DAT.
'She hopes to do well alone'

- Definitive diagnostics are more elusive in Old Icelandic, because they are low frequency and we are at the mercy of the written record.

Oblique subjects probably "existed" in Old Icelandic

- Rögnvaldsson (1996) provided arguments that oblique subjects were also attested in Old Icelandic.
- Argued from selected example sentences that oblique arguments in Old Icelandic showed behavior characteristic of modern oblique subjects, including conjunction reduction.
- Clear oblique PRO examples were lacking.
- Not clear if the examples cited are early indications of spread rather than a fully developed Modern Icelandic system.
- Not everyone was convinced (Faarlund 2003).
- The current paper tests Rögnvaldsson's claim of diachronic stability, and tests it with detailed large scale quantitative evidence.

Quantitative evidence from IcePaHC

- The most convincing evidence for oblique subjects in Modern Icelandic occurs at a low frequency in natural data so its absence from older written texts is simply uninformative.
- In addition to knock-down evidence, there are variable subject-object contrasts which produce quantitative patterns.
- To study the diachronic quantitative patterns in a way that can be replicated and challenged, we need a parsed corpus:
  - **IcePaHC** (Icelandic Parsed Historical Corpus) (Wallenberg et al. 2011)
  - Version 0.4, 440.000 words of hand-corrected phrase structure from 12th to 19th century inclusive (bigger v0.5 in July!)
  - LGPL license: Free download for research, profit, modification and/or redistribution. No registration required.
  - We encourage citing corpora like other published research
  - **www.linguist.is/icelandic_treebank/Download**
  - Comes with visual interface for Linux, Mac OS and Windows

Experiments

- 3 experiments:
  - Extraction contexts
  - Default postverbal position
  - Conjunction reduction
- Each experiment is designed to evaluate two quantitative hypotheses:
  - **Hyp. A: The obliques became subjects over time.** Oblique potential-subjects evolve so that a given quantitative behavior becomes more like the behavior of unambiguous nominative subjects (control group).
  - **Hyp. B: The obliques were subjects the whole time.** No change in relative quantitative patterning of obliques vs. nominative subjects.
- Separate sections for each experiment:
  - Data overview
  - Experimental design
  - Results

Exp I: Extraction contexts (data overview)

- Icelandic data:
  - Symmetrical V2 language (Thráinsson 1986, 2007)
  - Permits embedded topicalization
  - Position preceding the finite verb can contain either the subject or a fronted constituent, even in subordinate clauses.
- Extraction out of a clause degrades (or completely rules out) acceptability of embedded topicalization (Jónsson 1996, Angantýsson 2011):

(3) Ég spurði hvað hún gaf Jóni
I asked what.ACC she.NOM gave John.DAT
'I asked what she gave to John'

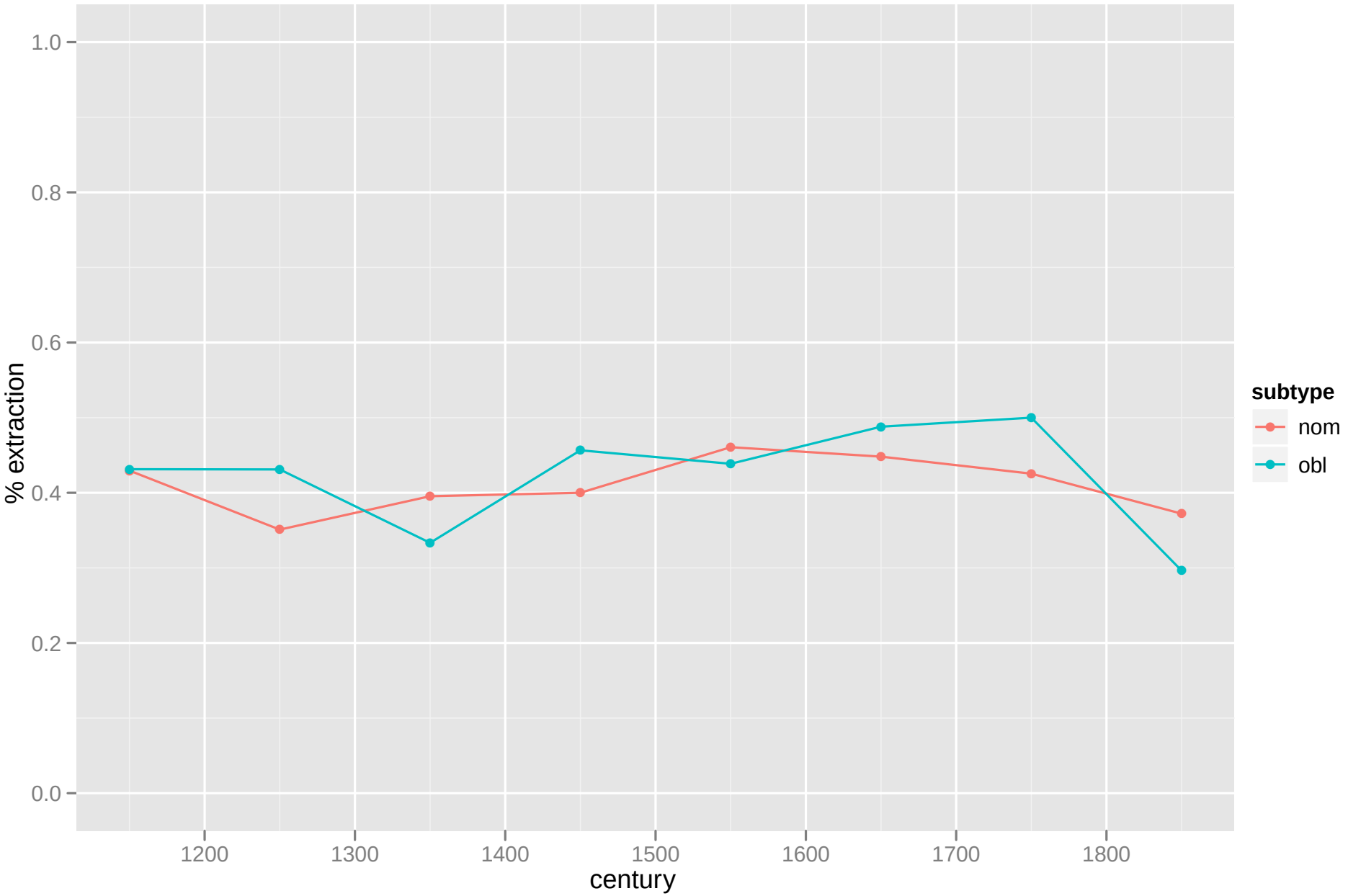
(4) *Ég spurði hvað Jóni gaf hún

(5) Ég spurði hvað Jóni var gefið
I asked what.NOM John.DAT was given
'I asked what John was given'

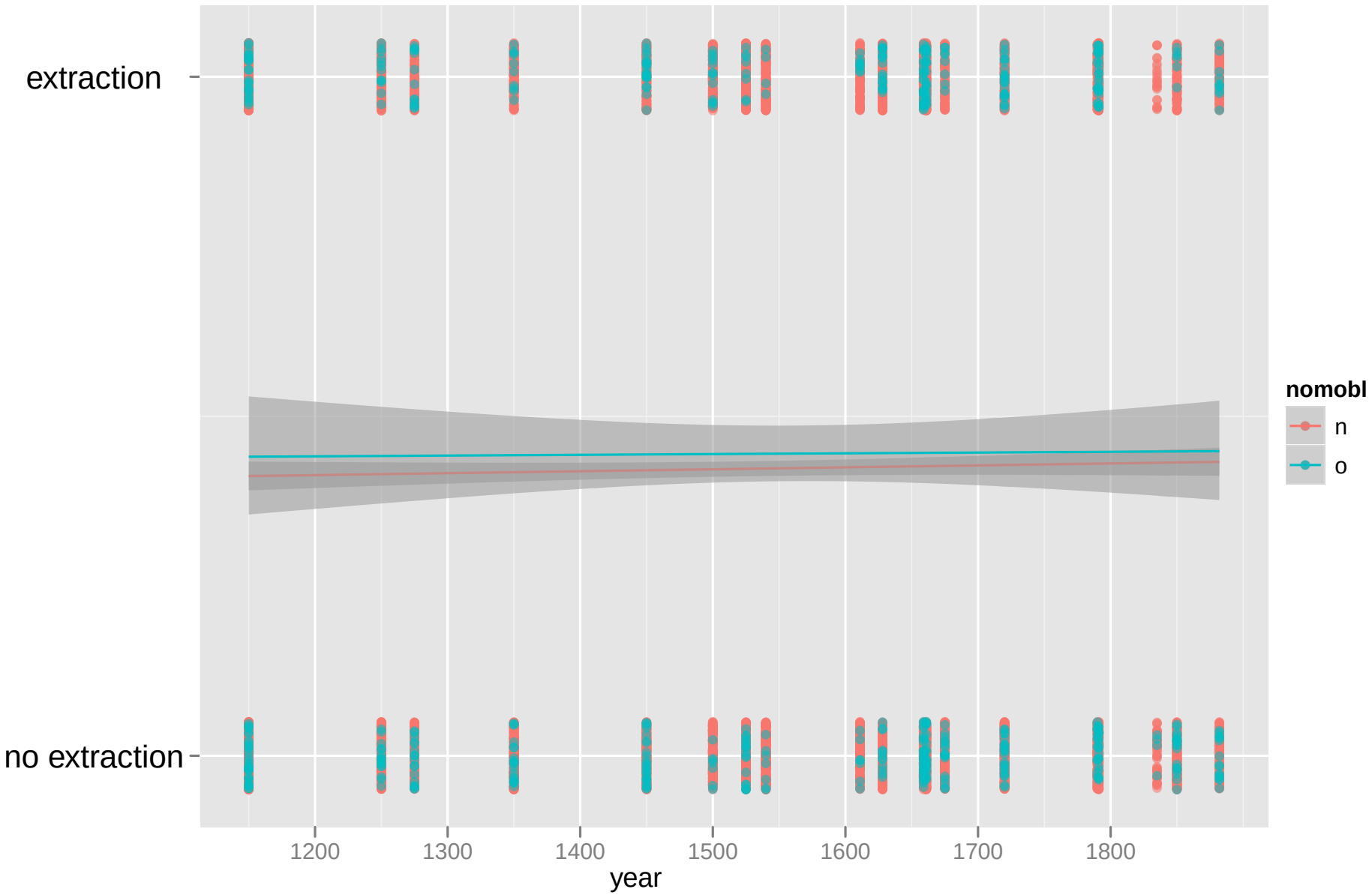
Exp I: Extraction contexts (experimental design)

- Queried for:
  - All subordinate clauses where initial element is either hypothesized oblique subject or a nominative subject
- Coded each clause for:
  - Clause containing an extraction vs. non-extraction
  - Year (and century)
- Different quantitative predictions:
  - **Hyp. A: The obliques became subjects over time**: Rise in frequency of gap-containing clauses with initial obliques, as this string becomes more and more grammatical (as the obliques become subjects).
  - **Hyp. B: The obliques were subjects the whole time**: Stable frequency of gappy clauses with initial obliques compared to the control sample with initial nominative subjects.
- We calculated the frequency by time period

Exp I: Extraction contexts (results) (n=8303)



Exp I: Extraction contexts (results) (n=8303)



Exp II: Default postverbal position (data overview)

- Icelandic data:
  - V2
  - Subjects follow the finite verb immediately by default when something else is topicalized

(6) Sjaldan líkar **málfræðingum** öskuský.
rarely like.3Sg linguists.DAT ash.clouds
'Rarely do linguists like ash clouds'

(7) *Sjaldan líkar öskuský **málfræðingum**.
rarely like.3Sg ash.clouds linguists.DAT
'Rarely do linguists like ash clouds'

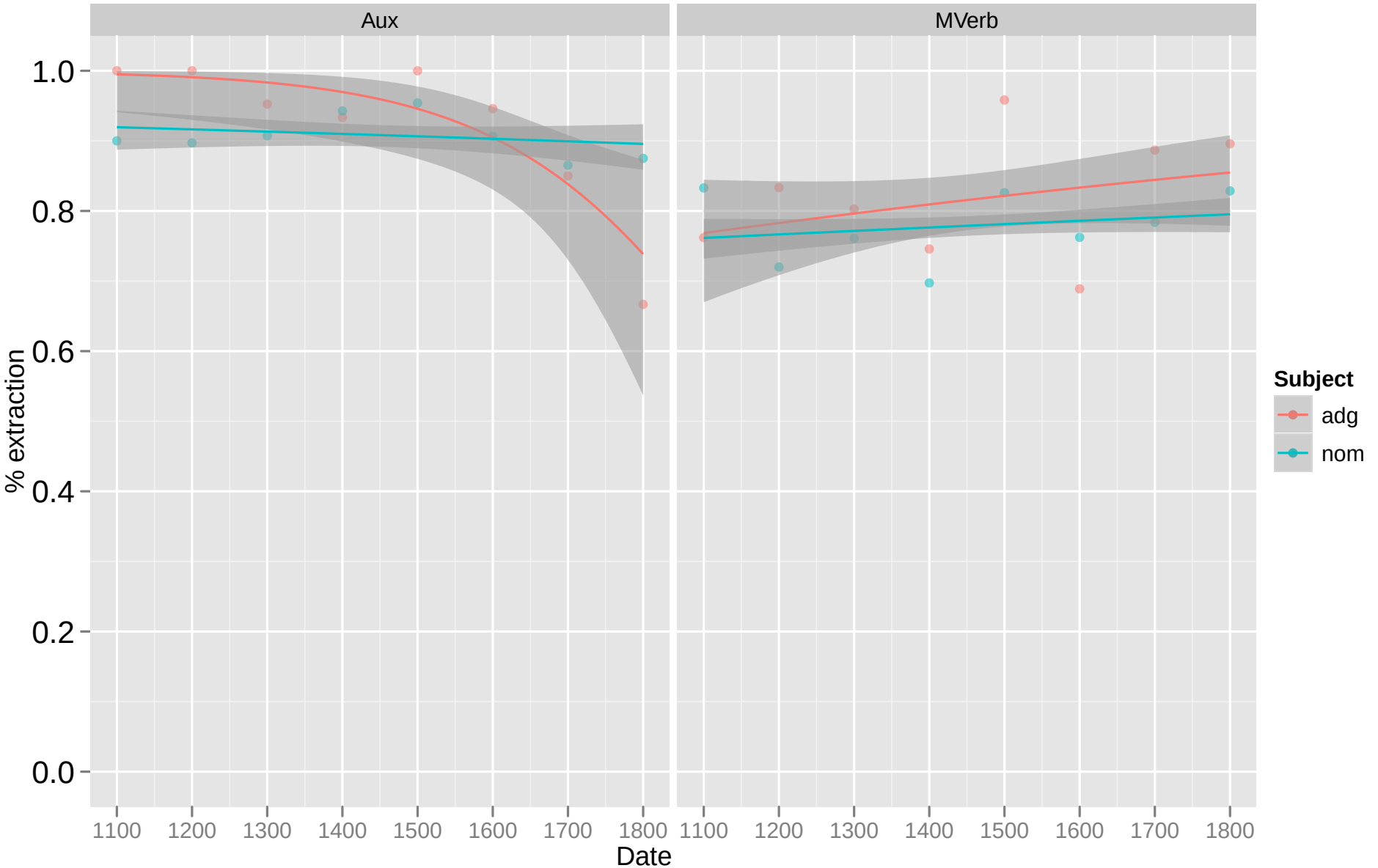
(8) Sjaldan líka **málfræðingar**
rarely facebook-like.3Pl linguists.NOM
Eyjafjallajökul.
Eyjafjallajokull.ACC
'Rarely do linguists facebook-like Eyjafjallajokull'

Exp II: Default postverbal position (experimental design)

- Queried for following word orders:
  - FiniteVerb - NomSubject - SomePhrase ← def. sbjpos.
  - FiniteVerb - SomePhrase - NomSubject
  - FiniteVerb - Oblique - SomePhrase ← def. sbjpos.
  - FiniteVerb - SomePhrase - Oblique
- Different quantitative predictions:
  - **Hyp. A: The obliques became subjects over time**: Rise in frequency of oblique arguments immediately following the finite verb compared to nominative control group.
  - **Hyp. B: The obliques were subjects the whole time**: Stable frequency of obliques in this position compared to nominatives.
- We calculated the frequency by century (12th to 19th inclusive)

Exp II: Default postverbal position (results)

Exp II: Default postverbal position (results)



Exp II: Default postverbal position (results)

- For clauses with auxiliaries, the only difference between the obliques and nominatives is in the opposite direction of the prediction that oblique subjects developed over time. (And there is sparse data for the last period, 15 clauses for the 1800s.)
- Logistic regression of the main verb data gives no significant difference between the position of nominatives and obliques over time (p=0.4077).
- In fact, it is not clear that there is a change at all over time: year as a predictor of subject position does not significantly improve the fit of the model (p=0.0726).
- In short: DPs in default postverbal position (with V2 inversion) do not present any evidence that the status of the obliques has changed.

Exp III: Conjunction reduction (data overview)

- Icelandic data:
  - The grammatical subject of a clause can be omitted under identity with the subject of the preceding clause
  - In many contexts, this is **only** possible for grammatical subjects

(9) Ég las bók og (mér) fannst hún góð
I read book and (me.DAT) found she.NOM good.NOM
'I read a book and (I) found it good'

(10) Jon kom og *(honum) gaf ég gjöf
John came and (him.DAT) gave I.NOM gift.ACC
'John came and I gave (him) a gift'

Exp III: Conjunction reduction (experimental design)

- Queried for:
  - All conjoined clauses with an initial overt or null hypothesized oblique subject or a nominative subject.
- Coded each clause for:
  - Initial element overt vs. null (reduced).
- Different quantitative predictions:
  - **Hyp. A: The obliques became subjects over time**: Rise in frequency of reduced initial obliques compared with frequency of reduction in nominative control group.
  - **Hyp. B: The obliques were subjects the whole time**: Initial obliques and nominative controls show same trend over time wrt frequency of reduction.
- We split the corpus into two periods century (pre-1540 and post-1540)

Exp. III: Conjunction reduction (results)

- There is an increase in overall conjunction reduction over time, but oblique reduction does not increase more than nominative.
- In fact, the frequency of oblique reduction changes less over time than that of nominative reduction (if it is different at all)

.

Time Period	Case	Reduced	Total	Frequency Reduced
pre-1540	obl	37	364	0.102
post-1540	obl	31	267	0.116
pre-1540	nom	3118	7184	0.434
pre-1540	nom	2660	5203	0.511

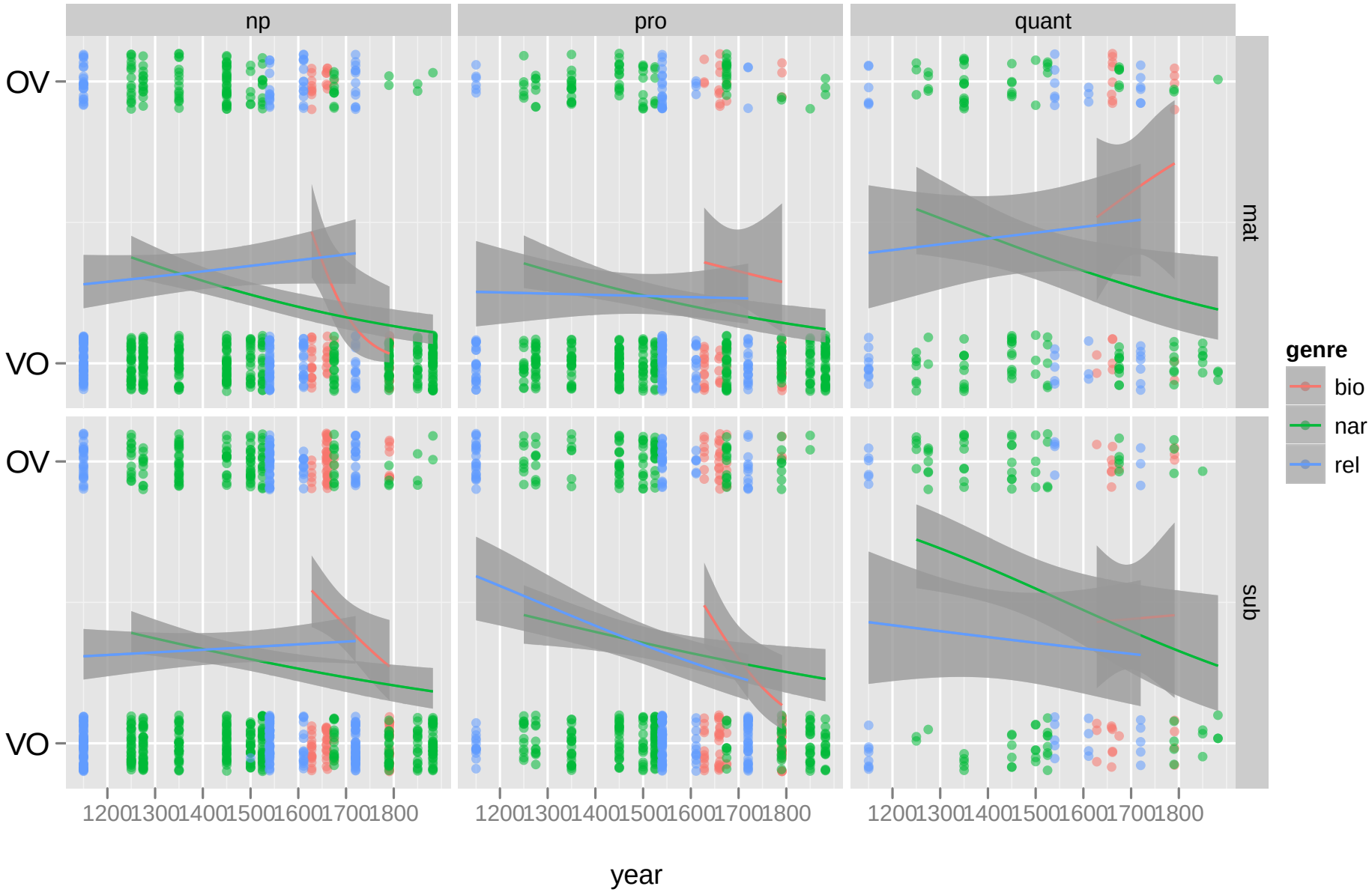
Exp. III: Conjunction reduction (results)

- Applying a loglinear model which assumes no interaction between Case, reduction, and time period yields a very good fit to the data ($G^2 = 0.384$ on 1 df, $p = 0.54$).
- This means that the effect of Case on the probability of reduction did not change over time.
- In fact, a simpler model which assumes conditional independence between Case of the subject and time period (controlling for reduction), also fits well and not significantly worse: $G^2 = 2.57$ on 2 df, $p = 0.28$.
- This means that the general frequency of hyp. oblique subjects in this sample, whether reduced or not, plausibly did not change over time.

From OV to VO in Icelandic

- We looked at the development from OV to VO (head-final *v*P to head-initial *v*P) to evaluate a possible link with the development of oblique subjects.
- Queried for:
  - All clauses with a finite auxiliary followed by a nonfinite lexical verb and and an overt object (n=2954).
- Coded each clause for:
  - Order of main verb and object
  - Object type (full np/pronoun/quantified)
  - Clause type (subordinate/matrix)
  - Year
  - Genre (biography, narrative, religious)

From OV to VO in Icelandic: 12th-19th century (n=2954)



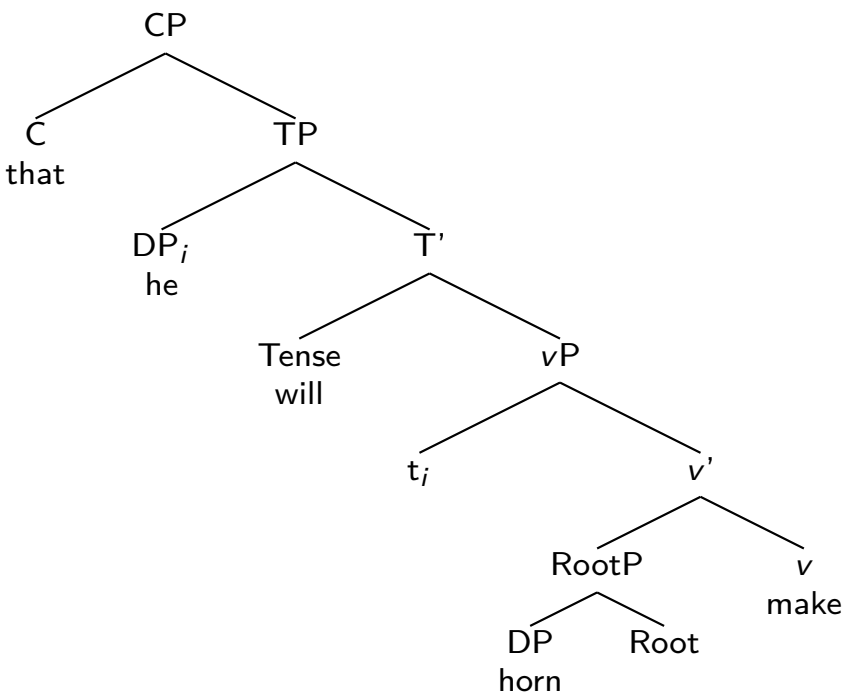
Development of Oblique Subjects

- How did they enter Icelandic?
  1. They didn't; they were always there (from PGmc or PIE) (Barðdal & Eythórsson 2005).
     The lack of oblique subject diagnostics in German, Yiddish, Dutch, and much of the rest of Europe makes this unlikely.
  2. A necessary consequence of OV to VO? **No.**
  3. A possible consequence of OV to VO? **No.**
  4. A necessary consequence of Tense-final to Tense-initial?
     No, because Yiddish changed to Tense-medial but does not have oblique PRO (Santorini 1989).
  5. A possible consequence of Tense-final to Tense-initial?

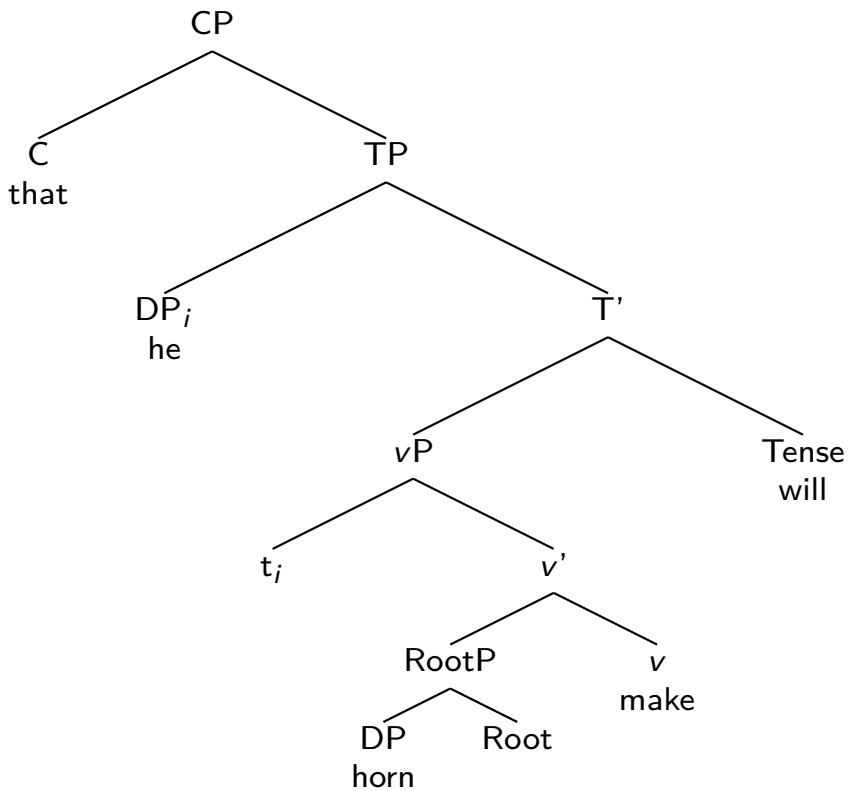
Why would the position of Tense affect the subject?

- Under a standard, head-parameterized phrase structure, there is no reason.

... that he **will** a horn make.



... that he a horn make **will**.

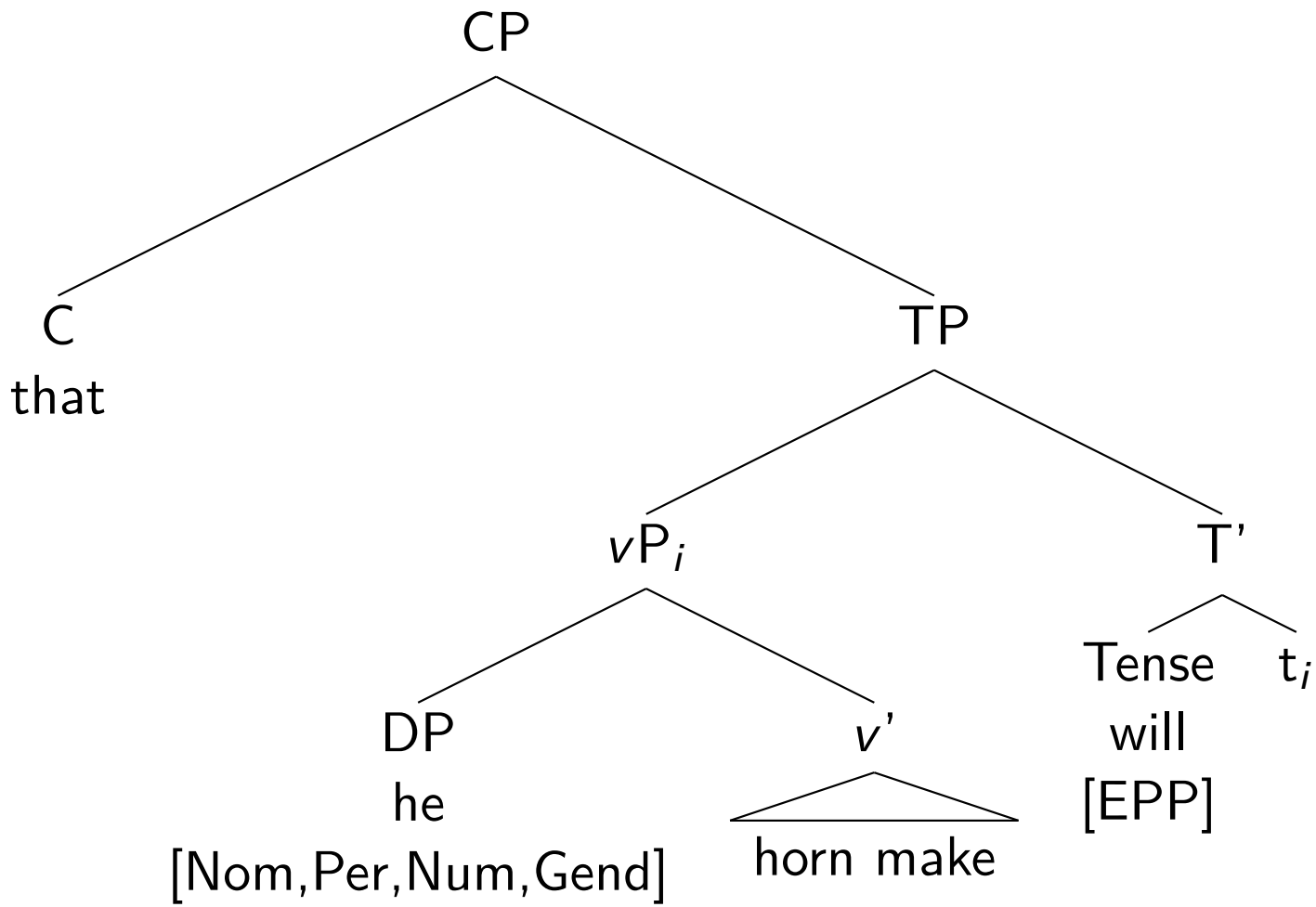


Why would the position of Tense affect the subject?

- Under an antisymmetric phrase structure of the variety found in Biberauer (2003), Biberauer & Roberts (2005), there is a reason.
- EPP satisfaction as in Biberauer & Richards (2003, 2004); see also Wallenberg (2010).
  - Tense probes in order to value some feature F.
  - If Tense is associated with an EPP feature, then an XP properly containing F must move to Spec(TP).
  - The XP may be larger or smaller, depending on the language; Tense-final languages pied-pipe *v*P to Spec(TP).
  - In some languages, Tense probes for ϕ-features and nominative Case (Lasnik 2005): such as German (pied-piping), English (non-pied-piping), or Yiddish (non-pied-piping).

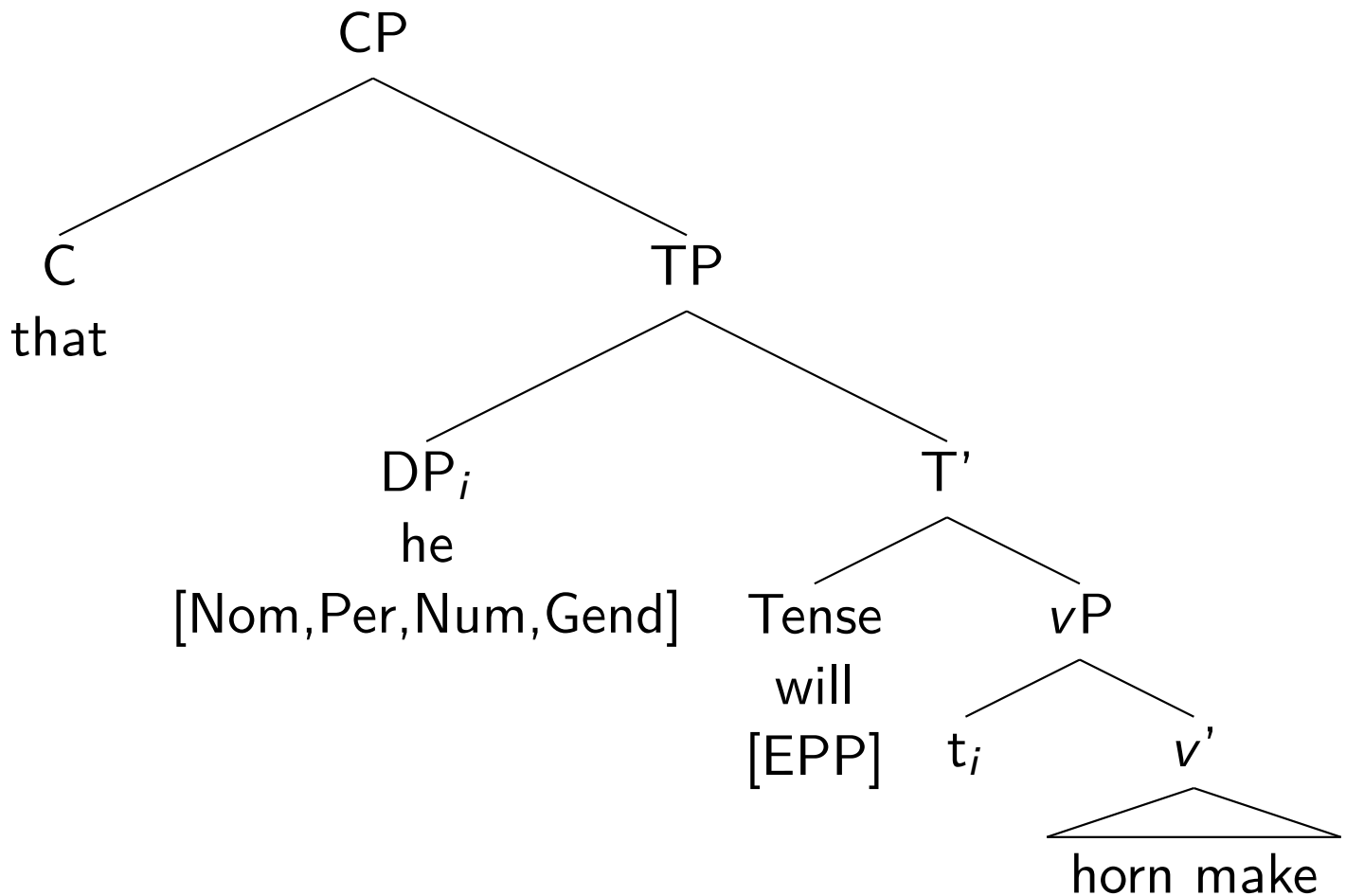
Why would the position of Tense affect the subject?

- Under an antisymmetric phrase structure, there is a reason, because a new subject position is open (see also Haider 2005).



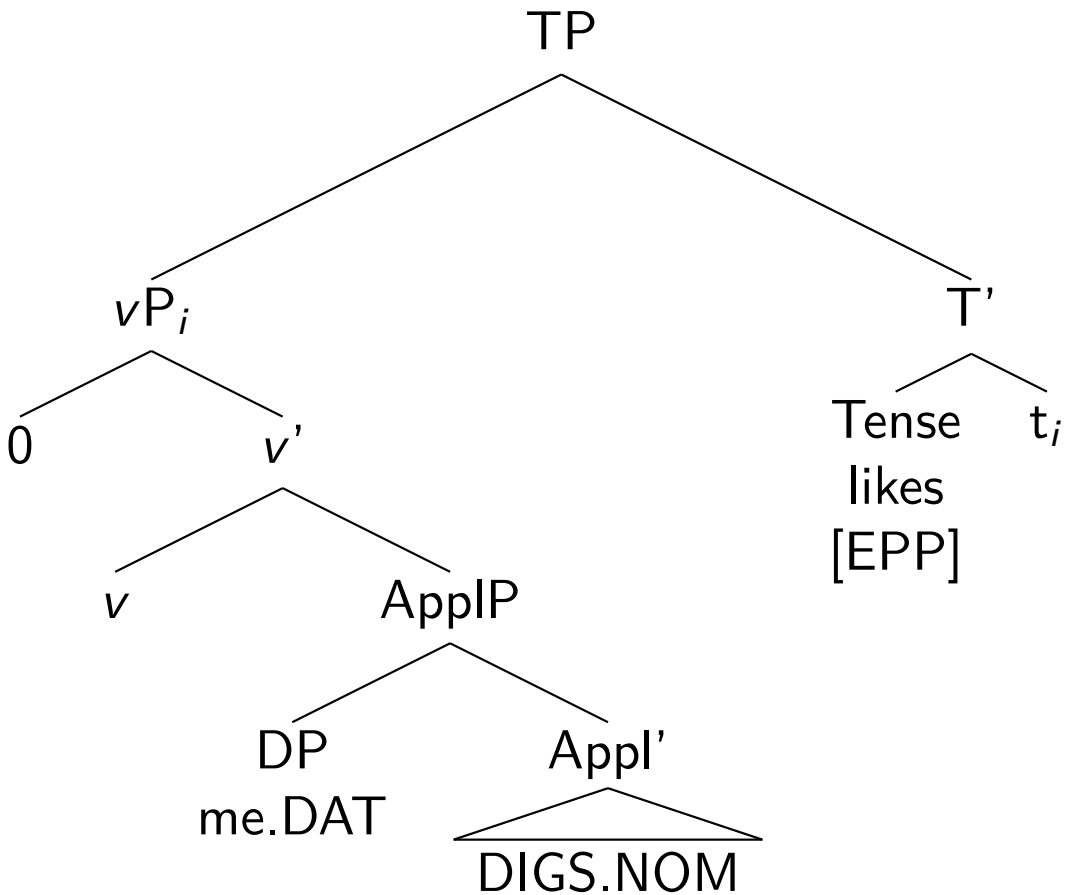
Why would the position of Tense affect the subject?

- Under an antisymmetric phrase structure, there is a reason, because a new subject position is open.



Experiencer Construction, Tense-Final

- Tense pied-pipes the *v*P when it attracts the nominative Theme. Note that this is opaque to the learner.

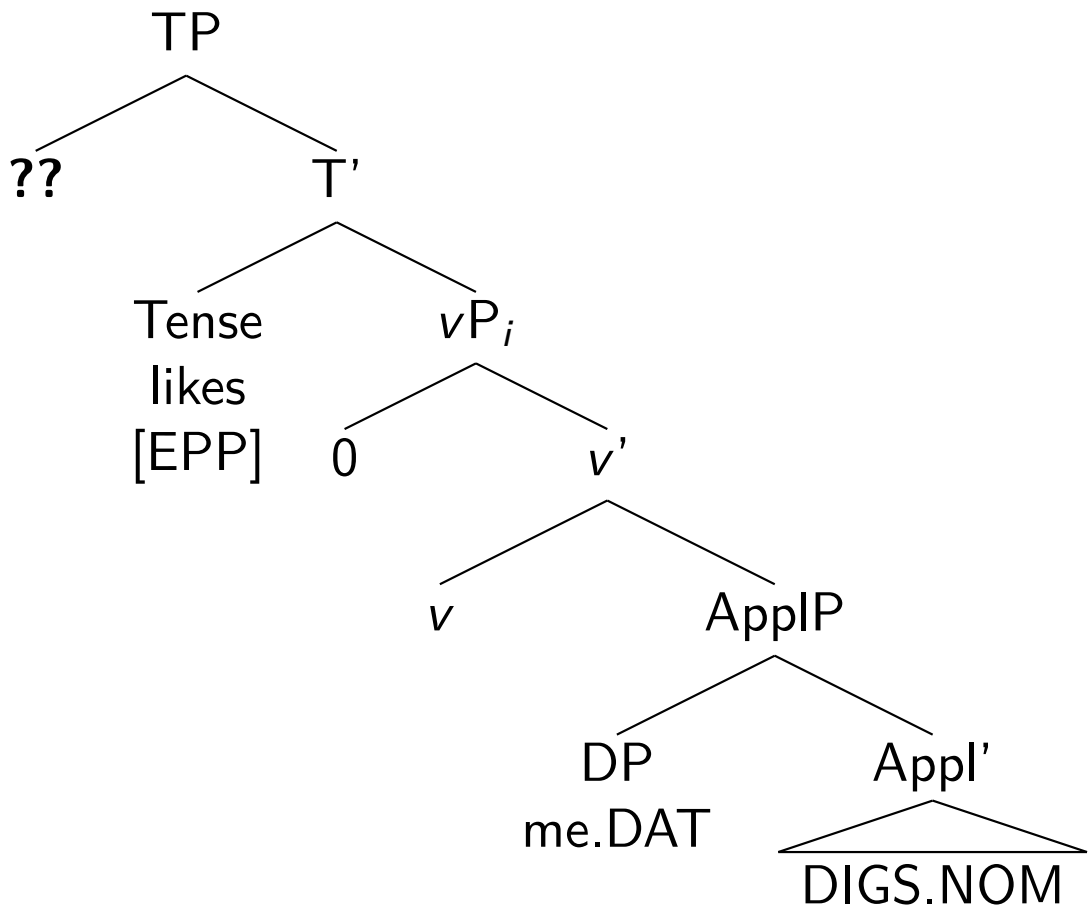


Experiencer Construction, Tense-Medial

- Pied-piping begins to be lost, and a new subject position is open. (Grammar competition ensues.)
- If the learner assumes that the EPP is sensitive to nominative Case and/or Person and Number, then modern Yiddish emerges: nominative Theme subject in Spec(TP).
- If the learner assumes that the EPP is sensitive to Gender, a different ϕ-feature, then modern Icelandic emerges: oblique Experiencer subject. (As well as oblique subjects in double object passives, etc.)
- Gender is the ϕ-feature which drives EPP-related movement in a number of Bantu varieties, e.g. Lubukusu and Lusaamia, so this is a perfectly plausible hypothesis (Carstens & Diercks 2010).

Experiencer Construction, Tense-Medial

- The closest DP with Gender is the Dative Experiencer. The closest DP with Nominative Case is the Theme. (Note that Person and Number agreement is dependent on nominative Case; see Bobaljik 2006)



Conclusion

- Oblique subjects in Icelandic have remained unchanged since the 12th century.
- They are not a consequence of VO, nor a necessary consequence of Tense-medial.
- Under a Kaynian architecture, oblique subjects are a potential consequence of Tense-medial.
- Under a pre-Kaynian architecture, they are not a consequence of anything.
- When definitive diagnostics are lacking, quantitative patterns can be used to distinguish between stability and change.
- Measuring quantitative patterns tells us more than a handful of categorical pieces of evidence could. The quantitative evidence can establish stability whereas individual examples might only be early indications of spread.