

# Icelandic Parsed Historical Corpus (IcePaHC) Stability? Patterns that die and others that don't

Joel C. Wallenberg, Anton Karl Ingason  
Einar Freyr Sigurðsson & Eiríkur Rögnvaldsson

RILiVS Workshop – University of Iceland

October 9, 2010

# Overview

- 1 Introduction
- 2 DP-internal directionality
- 3 Passives
- 4 OV/VO
- 5 Conclusion

# The Icelandic Parsed Historical Corpus – IcePaHC

- A corpus of Icelandic (12th - 19th century)
  - Parsed for phrase structure
  - Mostly compatible with the Penn Corpora of Historical English (Kroch et al. 2004; Kroch & Taylor 2000; Kroch et al. 2010)
  - Compatible with CorpusSearch (Randall 2005)
- IcePaHC team:
  - Eiríkur Rögnvaldsson (PI)
  - Joel Wallenberg (PI)
  - Anton Karl Ingason
  - Einar Freyr Sigurðsson
  - Brynhildur Stefánsdóttir (BA research assistant)
  - Hulda Óladóttir (BA research assistant)
- Grants:
  - Icelandic Research Fund (RANNÍS) (#090662011)
  - U.S. National Science Foundation (NSF) (#OISE-0853114)
  - University of Iceland Research Fund (Rannsóknasjóður HÍ)

# The Icelandic Parsed Historical Corpus – IcePaHC

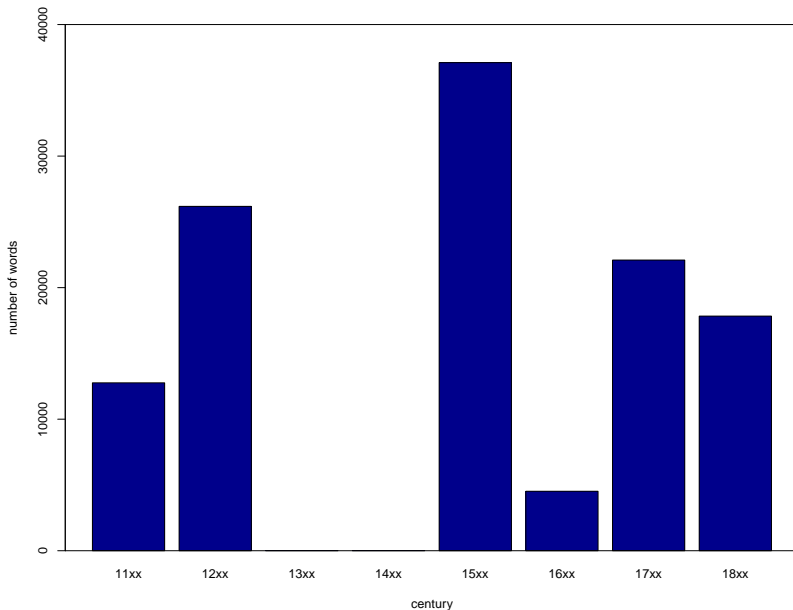
- Release dates every three months
  - IcePaHC, v0.1 (July 1st 2010, 31.000 words)
  - **IcePaHC, v0.2 (October 1st 2010, 120.000 words)**
  - IcePaHC, v0.3 (January 2011)
  - IcePaHC, v0.4 (April 2011)
  - IcePaHC, v1.0 (July 2011)

**Let us know if you find mistakes in the preview releases!**
- Free and open source policy
  - The corpus is LGPL-licensed
  - All development tools are free and open source
  - Freely available (including commercial use and derived works)
  - Available for download (no registration required)  
**[www.linguist.is/icelandic\\_treebank/Download](http://www.linguist.is/icelandic_treebank/Download)**
  - Should be cited **directly** if used in research

# Texts included in current release (IcePaHC, v0.2)

text	time	words
The First Grammatical Treatise (entire text)	12th century	4585
Íslensk hómiliubok (Icelandic book of homilies)	12th century	8179
Egils saga (theta fragment)	13th century	3459
Sturlunga saga	13th century	22719
New Testament's Gospel of John (entire text)	1540	20683
New Testament's Acts	1540	16421
Jón Indíafari's travelogue	1661	4521
Jón Steingrímsson's biography	1791	22097
Piltur og stúlka (novel by Jón Thoroddsen)	1850	17837
	<b>total:</b>	<b>120355</b>

# Words/century in IcePaHC 0.2



# Case study I

**DP-internal directionality**  
To go before or to go after

# DP-internal directionality

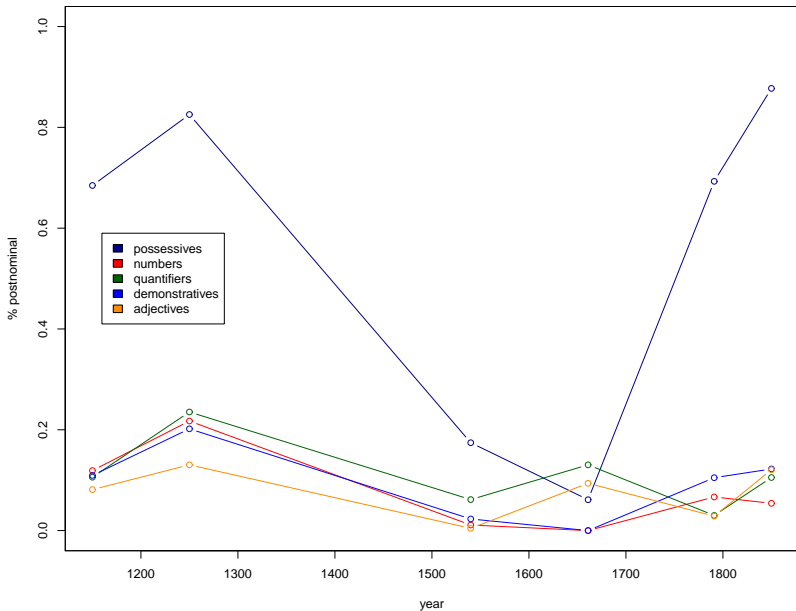
- Icelandic of all ages shows freedom in whether adjectives, quantifiers, numbers, possessives and demonstratives occur prenominally or postnominally

- (1) a. Ég á **góðar/allar/tvær/mínar/þessar** mýs  
 I own good/all/two/my/those mice  
 'I own some good mice' (etc.)
- b. Ég á mýs **góðar/allar/tvær/mínar/þessar**  
 I own mice good/all/two/my/those  
 'I own some good mice' (etc.)

- Can such a situation be stable?



## DP-internal directionality (% postnominal)



## Case study II

### Passives

To be a new or a not so new passive/construction

# The new passive/construction

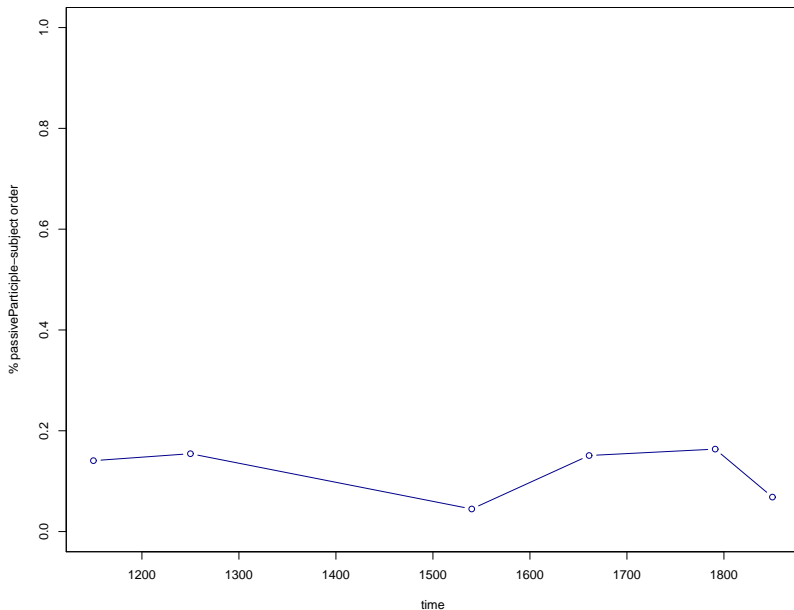
## New passive/construction:

- (2) Það var barið lítinn strákur  
it was beaten little.ACC boy.ACC  
'A little boy was beaten'

## Canonical passive:

- (3) a. Lítill strákur var barinn  
little.NOM boy.NOM was beaten  
'A little boy was beaten'
- b. Það var barinn lítill strákur  
it was beaten little.NOM boy.NOM  
'A little boy was beaten'

# Rate of subjects post passive participle



# Sources of ambiguity – low NOM/ACC DPs

## Sturlunga saga (13th century):

- (4) Var þá dæmt Sturlu **Staðarhólsland**  
Was then judged Sturla.DAT Staðarhólsland.NOM/ACC  
'Staðarhólsland was then ruled to be Sturla's'

IcePaHC 0.2: (ID 12XX.STURLUNGA.NAR-SAG,436.1625)

## Jón Steingrímsson's biography (18th century):

- (5) Þá var og úttekið **klaustrið**  
then was also out taken monastery.the.NOM/ACC  
'Then an assessment of the monastery was made'

IcePaHC 0.2: (ID 1791.JONSTEINGRIMS.BIO-AUT,109.214)

# Ambiguous NOM/ACC case on low DPs

Stable rate of ambiguity:

century	ambiguous	nom	total	% ambiguous
11xx	5	12	17	29%
12xx	5	15	20	25%
15xx	1	1	2	50%
16xx	2	1	3	67%
17xx	5	10	15	33%
18xx	1	2	3	33%

# Oldest examples of the new passive/construction

## 8 year old girl in Akureyri (1959):

- (6) Það var bólusett okkur  
it was inoculated us.ACC  
'We were inoculated'

(Maling & Sigurjónsdóttir 2002:129)

# The (not so) new passive/construction (18th century)

## Jón Steingrímsson's biography (18th century):

- (7) En þá kvöl, sem ég hafði að bera  
 But **that.ACC torture.ACC**, that I had to carry  
 af kitlum, sem ég hafði í yljum og tám, **verður**  
 from tickles, that I had in soles and toes, will.be  
**ei af mér útmálað**  
 not by me out painted  
 'But that torture, which I suffered from the ticklish feeling  
 in my soles and toes, will not be portrayed by me'

IcePaHC 0.2: (ID 1791.JONSTEINGRIMS.BIO-AUT,112.303)



## Case study III

**OV/VO**

The part when Joel will the talk continue

# From OV to VO

- We know that Old Icelandic showed some OV and some VO clauses, but we would like to have more detailed and quantitative information about the different possible OV and VO patterns, for different texts and various clause types.
- Modern Icelandic is VO, but there is much more research to be done concerning the time periods between the oldest and most recent stages of Icelandic syntax (see Þorbjörg Hróarsdóttir 2000 for an important first step in this direction).

# OV example with pronoun object

- (8) Kristur sjálfur hefir oss til boðið  
 Christ himself has us to offered  
 'which Christ himself has offered us'

```
( (IP-SUB (NP-OB1 *T*-4)
  (NP-SBJ (NPR-N Kristur-kristur)
    (NP-PRN (PRO-N sjálfur-sjálfur)))
  (HVPI hefir-hafa)
  (NP-OB2 (PRO-D oss-ég))
  (RP til-til)
  (VBN boðið-bjóða))
(ID 11XX.HOMILIUBOK.REL-SER,.6))
```

## VO example with pronoun object

- (9) ... nema maður vilji setja hann fyrir u  
 ... unless man wants put him for u  
 'unless one wants to use it for (the sound) u'

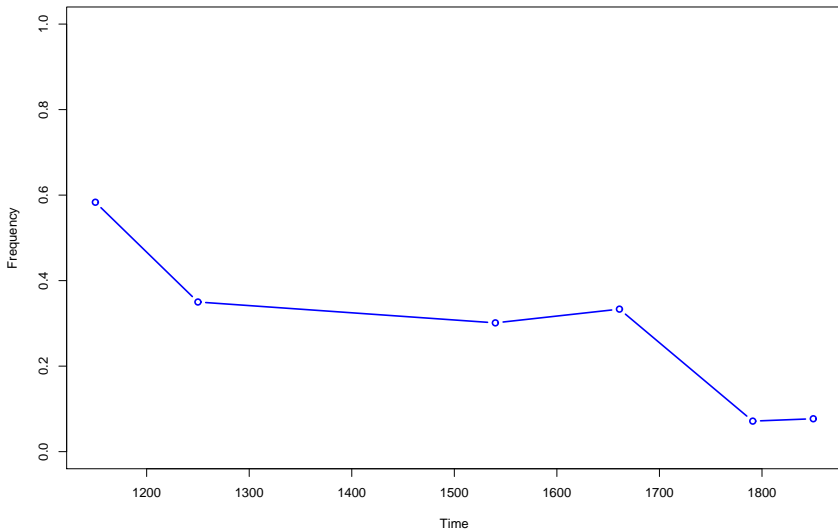
( (IP-SUB (NP-SBJ (N-N maður-maður))  
 (MDPS vilji-vilja)  
 (VB setja-setja)  
 (NP-OB1 (PRO-A hann-hann))  
 (PP (P fyrir-fyrir)  
 (NP (N-A u-u)))  
 (ADVP-TMP (ADV þá-þá)  
 (CP-REL (WADVP-1 0)  
 (C er-er)  
 (IP-SUB RMV:\*T\*-1\_hann-hann\_verður-verða...))))  
 (ID 11XX.FIRSTGRAMMAR.SCI-LIN,.136))

# OV example with NP and pronoun object

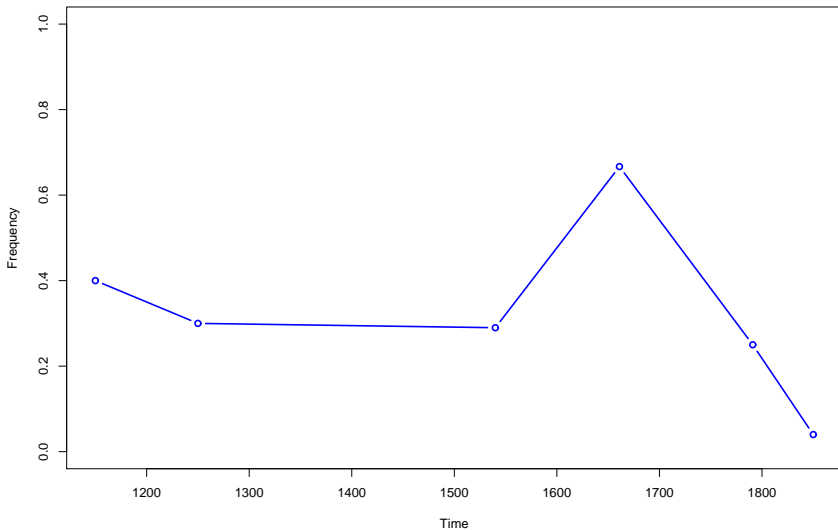
- (10) En er menn kenndu vildu öngvir honum mein gera  
 But when men knew wanted nobody him harm do  
 'When they recognized (him), nobody wanted to harm him'

( (IP-MAT (CONJ En-en)  
 (CP-ADV (C er-er)  
 (IP-SUB RMV:menn-maður\_kenndu-kenna...))  
 (MDDI vildu-vilja)  
 (NP-SBJ (Q-N öngvir-enginn))  
 (NP-OB2 (PRO-D honum-hann))  
 (NP-OB1 (N-A mein-mein))  
 (DO gera-gera))  
 (ID 12XX.STURLUNGA.NAR-SAG,449.2187))

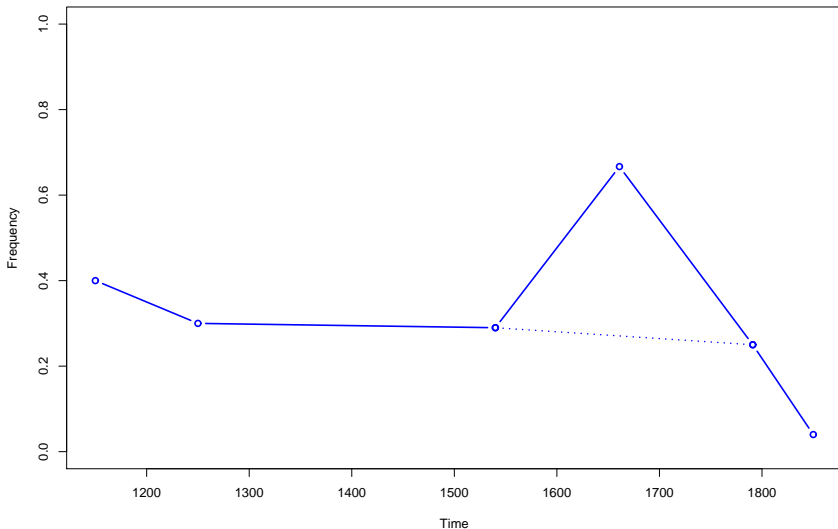
# Aux-O-V, pronoun objects, subordinate clauses



# Aux-O-V, pronoun objects, matrix clauses

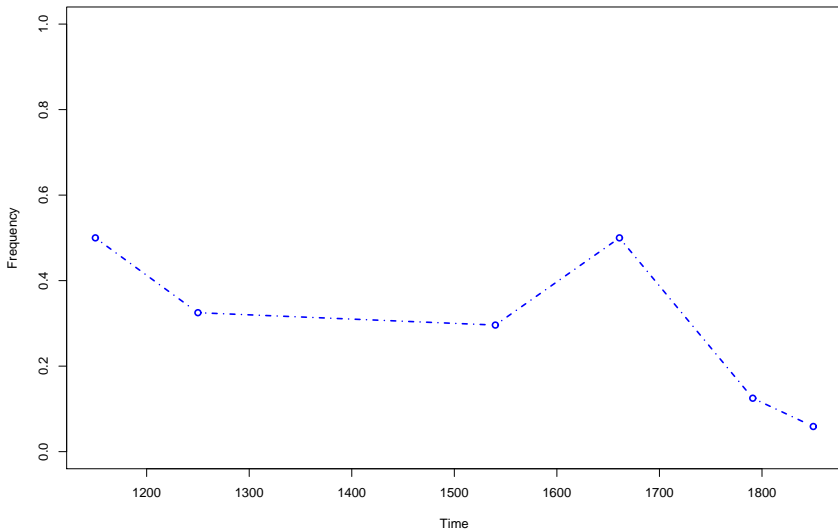


# Aux-O-V, pronoun objects, matrix clauses

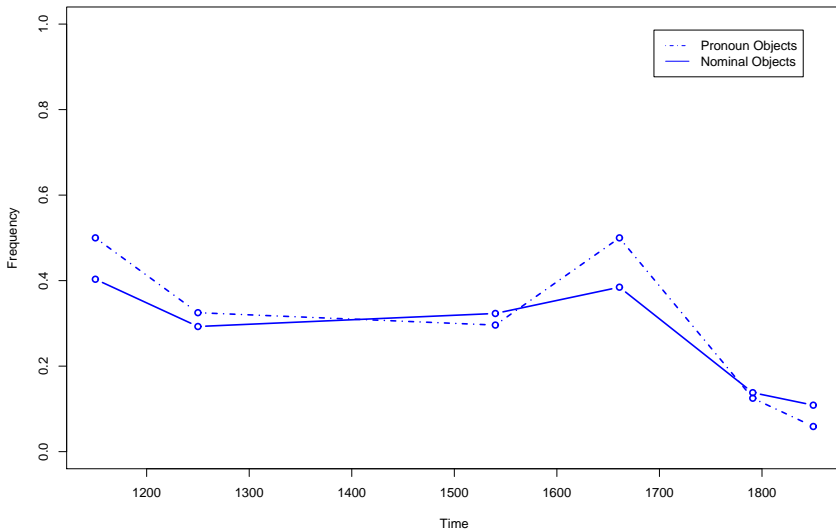




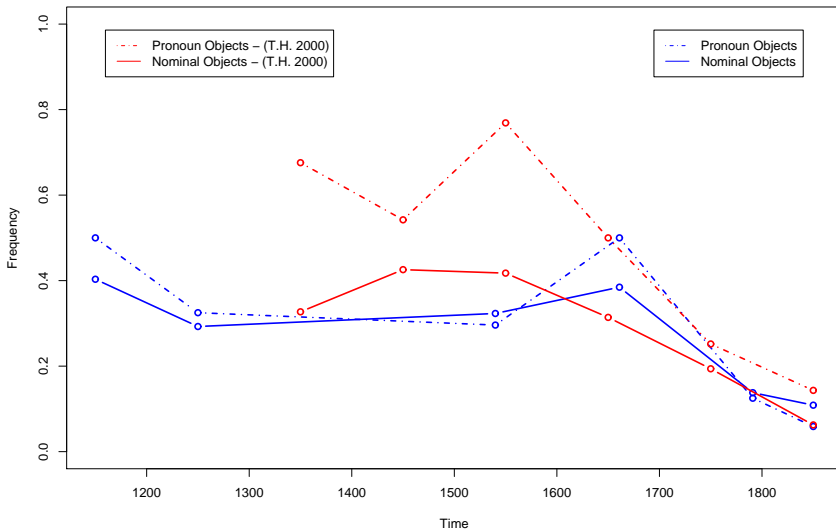
# Aux-O-V, all clauses, all non-quantified objects



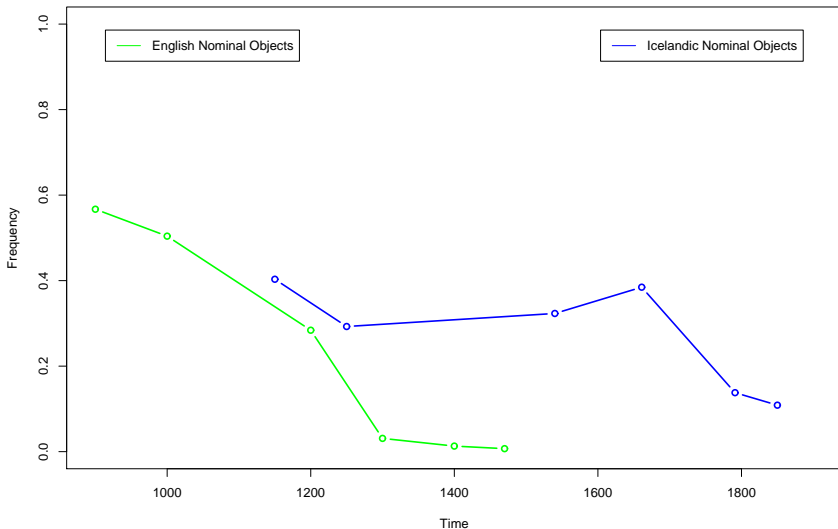
# Aux-O-V, all clauses, all non-quantified objects



# Comparison with Thorbjörg Hróarsdóttir (2000)



# Comparison with English (Pintzuk & Taylor 2006)



# Conclusion

- We are now beginning to have enough data to find meaningful diachronic results.
  - Diachronic stability (e.g. adnominal modifiers, passives).
  - Change over time (e.g. the headedness of  $vP$ ).
- The results should become even more interesting and statistically interpretable as we move closer to our goal of 1 million words.
  - More on the topics introduced here.
  - An inquiry into the relationship between the structure of  $vP$  and the existence of oblique subjects in modern Icelandic.
  - Further comparisons with historical English (YCOE, Penn Parsed Corpora of Historical English), Old/Middle French (Ottawa-Penn), Portuguese (Tycho Brahe), Early New High German (Caitlin Light) and Ancient Greek (Jana Beck).
- We also need more users for the corpus.