Word order change, case and expletives in the history of Icelandic

Hannah Booth[1] and Christin Schätzle[2]

[1]University of Manchester, [2]University of Konstanz

DiGS19
Stellenbosch University

8th September, 2017

Continuity and change in Icelandic

- ► Icelandic is said to be the most conservative Germanic language (Thráinsson 1996).

- ► **However, changes have been observed!**
  - ► 'freer' > less 'free' word order (Rögnvaldsson 1995)
  - ► decrease in V1 (Sigurðsson 1990)
  - ► increase in dative subjects (Barðdal 2011)
  - ► rise of expletives (Rögnvaldsson 2002)

- ► Overall, change in Icelandic, and in particular the interaction between changes, is still understudied.

- ► Existing studies mainly contrast Old Icelandic (1150-1350) with present-day language.

This Talk

- Corpus linguistic study using IcePaHC (historical treebank of Icelandic, 1150-2008; Wallenberg et al. 2011).
- Data visualization with HistoBankVis (Schätzle et al. 2017).

- Interaction between:
  - word order
  - expletives
  - dative subjects

- Evidence for the development of structure and positional licensing in Icelandic.

Icelandic Parsed Historical Corpus (IcePaHC)

- ▶ 12th to 21st century – all attested stages of Icelandic.
- ▶ 61 texts, 1 million words, different genres (not representative across centuries).
- ▶ annotation based on Penn Treebank style (Marcus et al. 1993).
- ▶ information about sentence types, constituents, word order, grammatical relations, tense, voice, and case.

Sample IcePaHC Annotation

```
(IP-MAT-SPE (NP-SBJ (PRO-D Mér-mér))
  (VBPI finnst-finna)
  (CP-ADV-SPE (WADVP-1 0)
     (C sem-sem)
     (IP-SUB-SPE (ADVP *T*-1)
             (NP-SBJ (PRO-N ég-ég))
             (BEPS sé-vera) (VBN sloppinn-sleppa)
          (PP (P úr-úr) (NP (NP-POS (ONE+Q-G einhvers-einhver)
             (N-G konar-konar)) (N-D fangelsi-fangelsi)))))
     (.  .-.))
  (ID 1882.TORFHILDUR.NAR-FIC,.603))
```

Icelandic V2

- ▶ Icelandic is a V2 language (Rögnvaldsson and Thráinsson 1990; Rögnvaldsson 1996)
  - → maximally one constituent in prefinite position

(1) a. Ég **gleymdi** þeim fljótt. **S-V-O**
         I.NOM forget.PST.1SG they.DAT quickly
         'I quickly forgot them.'

     b. Þeim **gleymdi** ég fljótt. **O-V-S**
         they.DAT forget.PST.1SG I.NOM quickly
         'Them I quickly forgot.'

     c. Fljótt **gleymdi** ég þeim. **X(P)-V-S**
         quickly forget.PST.1SG I.NOM they.DAT
         'Quickly I forgot them.'

Corpus Study I: Subject positions

- Positions for all subjects across IcePaHC.

Period	prefinite	postfinite	Total	% prefinite	χ^2
1150-1349	7045	6673	13718	51.4%	***
1350-1549	10091	8258	18349	55.0%	***
1550-1749	6076	5134	11210	54.2%	***
1750-1899	6490	4767	11257	57.6%	
1900-2008	7924	2937	10861	73.0%	***

- Almost **equal distribution of subject position** in the first time period.
- Subjects are increasingly realized in the prefinite position.
- **Large increase of prefinite subjects post-1900 (p<0.001***).**
- Development of designated subject position?

Icelandic V1

- ▶ Icelandic still allows for **V1 declaratives**, typical for older stages of Germanic (Sigurðsson 1990).

(2) **Gnæfði**   gaflinn   hátt yfir fjöruna
    tower.PST.3SG gable.the.NOM high over beach.the.ACC
    'The gable towered high over the beach' (Leysing, 1907)

V1 in Icelandic (Butt et al. 2014)

- Corpus study on V1 matrix declaratives in IcePaHC.
- V1 is attested throughout the history of Icelandic and is mainly confined to narrative texts.
- **But, V1 undergoes a marked decrease as of 1900!**

Period	V1	non V1	Total	% V1	χ^2
1150-1349	2829	10889	13718	20.6%	***
1350-1549	3656	14693	18349	19.9%	***
1550-1749	1654	9556	11210	14.8%	***
1750-1899	2072	9185	11257	18.4%	***
1900-2008	292	10569	10861	2.7%	***

Expletives in Icelandic

- **Possible explanation for V1 decrease:** rise of clause-initial expletives (Franco 2008, cf. Axel 2007 on historical German).

- **Modern Icelandic: V2 with expletive *það***

(3) **Það** var   töluverður   snjór
     EXPL be.PST.3SG considerable.NOM.SG snow.NOM.SG
     yfir öllu.
     over everything.DAT
     'There was a considerable amount of snow over everything.'
     (Ofsi, 2008)

- **Older Icelandic: V1 with no expletive**

(4) _   Var   þá gleði   mikil   í
     ØEXPL be.PST.3SG then joy.NOM.SG great.NOM.SG in
     kóngs   höll.
     king.GEN.SG hall'
     'There was then great joy in the king's hall.' (Jarlmann, 1480)

Corpus study II: Expletives and prefinite position

- Interaction between decrease of V1 and rise of expletive *það*.

Period	prefinite expl (V2)	no expl (V1)	Total	% expl	χ^2
1150-1349	16	153	169	9.5%	***
1350-1549	26	205	231	11.3%	***
1550-1749	13	87	100	13.0%	***
1750-1899	59	92	151	39.1%	
1900-2008	160	28	188	85.1%	***

- **Marked increase in expletives as of 1900.**
- Increase in expletive *það* in clause-initial prefinite position is a factor behind decrease in V1.

Explaining the decrease in V1

- **However**, not all V1 clauses in IcePaHC are constructions which could have had an expletive:

(5)   Sýndi   **drottinn**   mikla   miskunn
      show.PST.3SG lord.the.NOM   great.ACC   mercy.ACC
      vin   sínum   sankti   Georgíum
      friend.DAT   his.own.DAT   saint.DAT   George.DAT
      'The Lord showed great mercy to his friend St. George'
      (Georgius, 1525)

- The rise of expletive *það* can only be part of the story...

Explaining the decrease in V1

Expletive *það* is not a subject

(6)   a.   **Það** var   ekki minnst   á önnur dýr.
            EXPL be.PST.3SG NEG mentioned on other   animals
            'There was no mention of other animals.'

       b.   Ekki var   _   minnst   á önnur dýr.
            NEG be.PST.3SG ∅EXPL mentioned on other   animals
            'There was no mention of other animals.'

Swedish *det* 'subject expletive'

(7)   a.   **Det** dansades   i går.
            EXPL dance.PST.PASS yesterday
            'There was dancing yesterday.'

       b.   I går   dansades   **det**.
            yesterday dance.PST.PASS EXPL
            'Yesterday there was dancing.'

$\implies$ Against the idea of a designated subject position.

Verb position and information structure

What motivates clause-initial *það*?

- ▶ Older Icelandic: **flat structure**
    - ▶ Verb as information structural boundary, separating topic and comment in Germanic (cf. Hinterhölzl & Petrova 2010 on historical German)

        $\longrightarrow$ $\underbrace{Topic}$ V $\underbrace{Comment}$

    - ▶ V1 in topicless sentences (e.g. presentationals)

        $\longrightarrow$ V $\underbrace{\qquad}$ $\Longrightarrow$ whole clause in focus

- ▶ **Growth of structure** $\longrightarrow$ emergence of a fixed topic position in the history of Icelandic

- ▶ Expletive *það* emerges as a filler for topic position.
    - ▶ *það* is non-referential and unlikely to be a topic, contra Faarlund (1990) 'expletive topic'.
    - ▶ V1 decreases

IP

XP   I′

$$\left\{ \begin{array}{l} (\uparrow \text{TOPIC}) = \downarrow \\ (\uparrow \{\text{COMP}|\text{XCOMP}\}^* \text{ GF}) = \downarrow \end{array} \right.$$

I

$(\uparrow \text{SUBJ}) = \downarrow$

$$\left. \begin{array}{l} (\downarrow \text{EXPLETIVE}) =_c + \\ \neg(\uparrow \text{TOPIC}) \end{array} \right\}$$
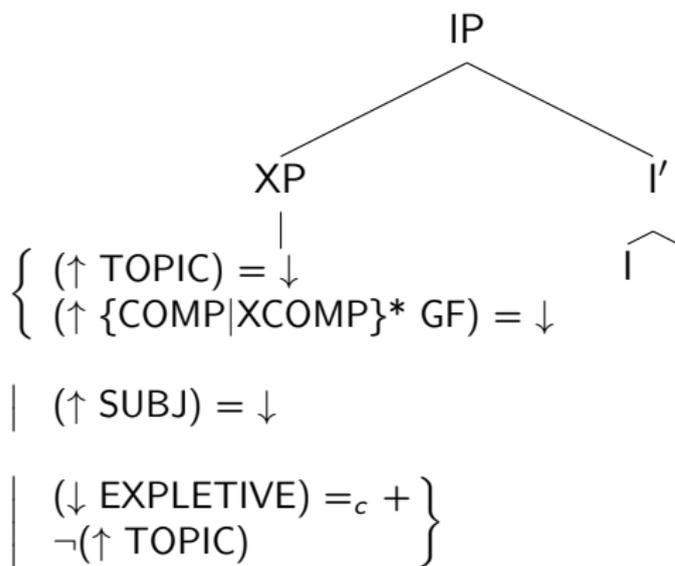
Meanwhile: Dative subjects in Icelandic

- ▶ Corpus study on dative subjects in IcePaHC (Schätzle et al. 2015).

(8)   Vel líkuðu   **goðrøði**   góð   røði.
      well like.PST.3PL Goðrøður.DAT good.NOM oars.NOM
      'Goðrøður (the good oarsman) liked good oars well.'
      (First grammar, 1150)

- ▶ On-going debate on whether dative subjects are a historical innovation or inherited.
- ▶ Dative subjects mainly occur together with experiencer/psych verbs or 'happenstance' verbs (e.g. see Barðdal 2011).
- ▶ Augmentation of IcePaHC's annotation scheme with verb class information.

- Diachronic distribution of subject case in IcePaHC.

Period	Nom	Dat	Acc	Gen	Total	% Dat	χ^2
1150-1349	13028	535	128	26	13718	3.9%	
1350-1549	17596	591	142	20	18349	3.2%	***
1550-1749	10676	417	106	11	11210	3.7%	
1750-1899	10733	428	80	16	11257	3.8%	
1900-2008	10113	626	115	7	10861	5.8%	***

- **Frequency of dative subjects increases as of 1900.**
- Dative subjects occur most often with **psych/experiencer predicates**.
- **Dative subjects become more systematically associated with experiencers and goals.**

Dative subjects and voice (Schätzle et al. 2015)

Period	active	middle	passive	Total	% middle	χ^2
1150-1349	397	66	72	535	12.3%	***
1350-1549	417	80	94	591	13.5%	***
1550-1749	239	69	109	417	16.6%	***
1750-1899	273	88	67	428	20.6%	
1900-2008	315	239	72	626	38.2%	***

- ▶ **Dative subjects are increasingly used with verbs carrying middle morphology**.
- ▶ Middles are typically agentless (Woods 2015).

$\Longrightarrow$ Dative case is becoming more systematically associated with lexical semantic conditions.

HistoBankVis (Schätzle et al. 2017)

- ▶ Generically applicable system for historical linguistic research.
- ▶ Flexible investigation of a potentially high number of interacting linguistic features stored in an SQL database.

Interactive Task-Based Feedback Loop

- ▶ Compact Matrix Visualization
  - ▶ Visualizes differences between selected dimensions across time
  - ▶ Measure of quality and "interestingness"
- ▶ Difference Histograms Visualization

Example: Subject case and word order
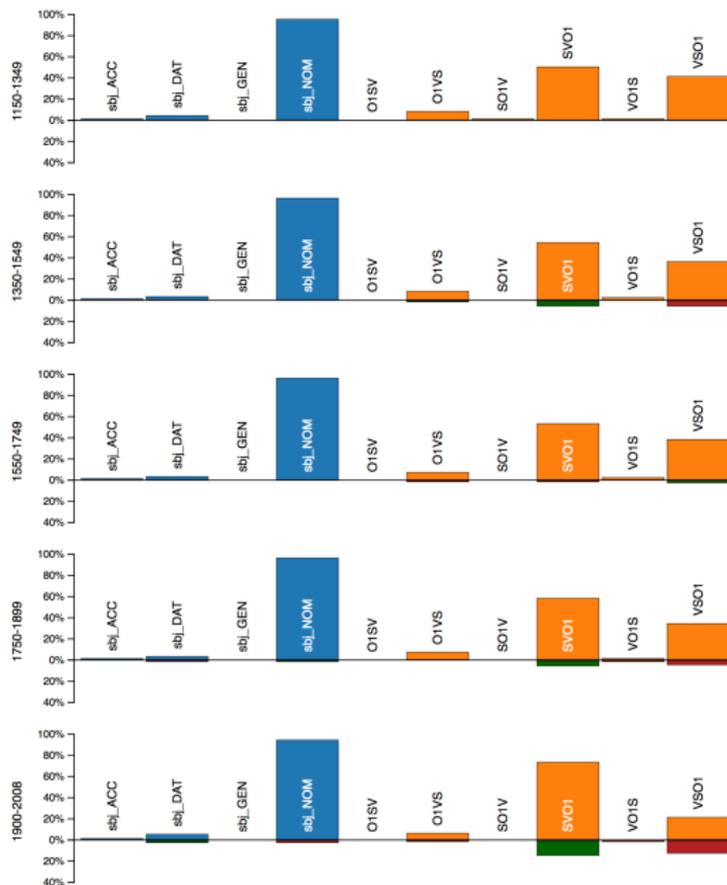
Corpus Study III: Dative subjects and subject position

- Subject positions for dative subjects across IcePaHC.

Period	prefin (Dat)	postfin (Dat)	Total	% prefin (Dat)	χ^2	% prefin (all)
1150-1349	131	404	535	24.5%	***	51.4%
1350-1549	126	465	591	21.3%	***	55.0%
1550-1749	119	298	417	28.5%	*	54.2%
1750-1899	151	277	428	35.3%		57.6%
1900-2008	353	273	626	56.4%	***	73.0%

- Dative subjects are preferably realized in the **postfinite position** in older stages of Icelandic.
- Prefinite dative subjects are increasing over time.
- **Significant increase of prefinite dative subjects after 1900; prefinite position becomes dominant.**

Corpus Study III: Dative subjects and subject position

- Dative subjects in V1 declaratives across IcePaHC.

Period	V1 (Dat)	non V1 (Dat)	Total	% V1 (Dat)	χ^2	% V1
1150-1349	173	362	535	32.3%	**	20.6%
1350-1549	254	337	591	43.0%	***	19.9%
1550-1749	106	311	417	25.4%		14.8%
1750-1899	126	302	428	29.4%		18.4%
1900-2008	20	606	626	3.2%	***	2.7%

- Larger tendency for dative subjects to occur in V1 constructions (i.e., postfinite) than for subjects overall.
- Decrease of V1 with dative subjects over time.
- **Very large drop of dative subjects in V1 as of 1900.**

Kiparsky's Linking Theory (1997)

- ▶ Theory of linking and case capturing the complex interrelation between
  - ▶ morphology
  - ▶ position
  - ▶ grammatical relations

- ▶ Unidirectionality of changes: loss of morphology implies fixed word order; converse is not true.
- ▶ Rise of positional licensing (rise of I) correlates with loss of morphology in historical English.
- ▶ In Modern Icelandic morphology is recessive, **position** is dominant.

Rise of Positional Licensing — LFG

- Overall the changes observed to Icelandic conform to a pattern in which functional structure is "added" over time (Börjars et al. 2016).
- Early Germanic had fairly free word order, with grammatical functions indicated by morphology.
- Topics tend to be clause initial.

```
                    S
         ┌──────────┼──────────┐
        XP         XP         VC
         │
   (↑ DF) = ↓
```

Rise of Positional Licensing — LFG
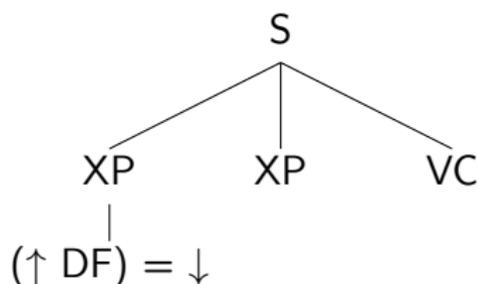
- Periphrastic tense/aspect arises, leading to an I.
- Finite verbs (I) partition a clause in terms of information-structural information (topic vs. comment, cf. Hinterhölzl & Petrova 2010).

IP
XP   I'
$(\uparrow DF) = \downarrow$ I   VP

Rise of Positional Licensing — LFG
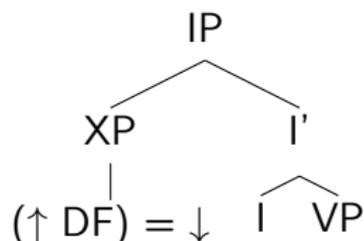
- Subjects tend to be topical and the SpecIP position becomes increasingly associated with subjects (current Icelandic).

$$
\begin{array}{c}
\text{IP} \\
\diagup \quad \diagdown \\
\text{XP} \qquad \text{I'} \\
| \qquad \diagup \diagdown \\
(\uparrow \text{DF}) = \downarrow \quad \text{I} \quad \text{VP} \\
(\uparrow \text{GF}) = \downarrow \\
\text{OR} \quad (\uparrow \text{SUBJ}) = \downarrow
\end{array}
$$

Rise of Positional Licensing — LFG
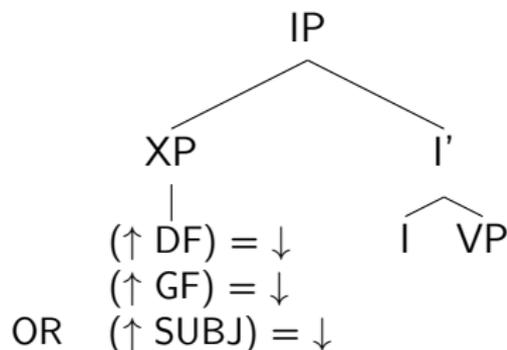
▶ Subjecthood and Information-Structure can be differentiated
  further.

```
                        CP
                   ╱         ╲
                XP              C'
                │           ╱       ╲
           (↑ DF) = ↓     C          IP
                                  ╱      ╲
                                XP        I'
                                │        ╱ ╲
                           (↑ SUBJ) = ↓  I   VP
```
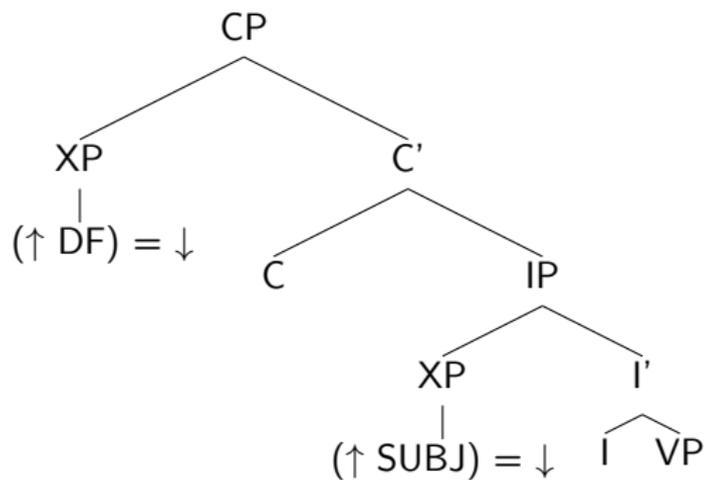
Hypothesis for Dative Subjects

- Over the history of Icelandic, the dative case becomes more clearly associated with experiencers.
- Experiencers are sentient and therefore make for better topics than stimuli.
- Dative experiencer arguments are increasingly associated with the subject position.

$$
\begin{array}{ccccc}
 & & \text{experiencer} & \text{stimulus} & \\
 & & | & | & \\
\text{Experiencer Verb} & < & \text{arg}_2 & \text{arg}_3 & > \\
 & & \text{[-r]} & \text{[+o]} & \\
 & & | & | & \\
 & & \text{subj} & \text{obj} & \\
 & & \text{dat} & \text{nom} &
\end{array}
$$
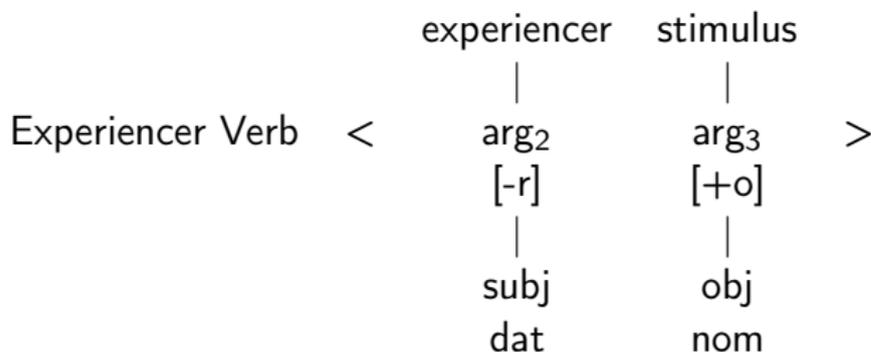
$\Longrightarrow$ Dative experiencers are more firmly linked to subjects than to objects.

Hypothesis for Dative Subjects

- ▶ SpecIP position becomes more firmly associated with topics.
- ▶ As a result, dative experiencers are also increasingly placed initially.
- ▶ The development of dative subjects **crucially** follows a prior identification of SpecIP as a general subject position.
- ▶ As non-canonical subjects they eventually follow/**conform to** the overall positional licensing developed in the language.

Conclusion

- ▶ Our corpus study provides evidence for the development of structure in the history of Icelandic; in particular for the rise of positional licensing.

- ▶ System becomes regularized over time to include a positional licensing for dative subjects.

- ▶ Against the idea of dative subjects as a stable, common Proto-Indo European inheritance.

- ▶ Complex interacting system of case, word order, lexical semantics (and information structure) in Icelandic.

$\implies$ Consideration of interaction between observed changes is key to understanding the diachrony.

Future work: Stylistic Fronting

- ▶ Stylistic Fronting and expletive *það* in complementary distribution (e.g. Holmberg 2000).

- ▶ **Stylistic Fronting**

(9) Ef **gengið**   er   eftir Laugaveginum...
     if walk.PASS.PTCP be.PRS along Laugavegur.DEF
     'If one walks along the Laugavegur...'

- ▶ **Expletive það**

(10) Ef **það** er   gengið   eftir Laugaveginum...
      if EXPL be.PRS walk.PASS.PTCP along Laugavegur.DEF
      'If one walks along the Laugavegur...'

- ▶ Previous claims: Stylistic Fronting decreases **at same time as** expletive increases (Hróarsdóttir 2000, Rögnvaldsson 1996).

Future work: Expletive construction types

- Our preliminary results show that expletive *það* develops at a quicker rate in presentational than in impersonal constructions.

- **Presentational: postfinite 'logical subject'**

(11)   **Það** rísu   upp **tveir**   **nýir**   **kaupmenn**.
     EXPL stand.PST up   two.NOM new.NOM merchants.NOM
     'There stood up two new merchants.'

- **Impersonal: genuinely subjectless**

(12)   Ef **það** er   gengið   eftir Laugaveginum...
     if EXPL be.PRS walk.PASS.PTCP along Laugavegur.DEF
     'If one walks along the Laugavegur...'

- Significant increase in *það* in impersonals **only as of 1900**.

Future work: OV/VO order

- **Older Icelandic**:   OV/VO variation
- **Modern Icelandic**:   almost exclusively VO

- Previous studies: significant decrease in OV in **19th century** (Hróarsdóttir 2000, Indriðason 1987).

- Kiparsky (1996) takes the Germanic shift from OV to VO to be connected to the rise of I.

- We intend to examine OV/VO word order in IcePaHC.

  $\longrightarrow$ How does this fit in with our previous observations?