Dative Subjects: Historical Change Visualized

Christin Schätzle

Fachbereichskolloqium Sprachwissenschaft
University of Konstanz

December 21st, 2017

Thesis Overview

1. Introduction
2. Linguistic background
3. Visual analytics for historical linguistics
4. Dative subjects in Icelandic
5. Subject case and word order in Icelandic
6. Dative subjects in Marathi
7. Conclusion

Advisors: Miriam Butt and Frans Plank
External Advisor: Ashwini Deo, Ohio State University

- ▶ SFB-TRR 161 "Quantitative Methods for Visual Computing"
- ▶ **Project D02 "Evaluation Metrics for Visual Analytics in Linguistics"**
  - ▶ Language change in Germanic and Indo-Aryan
  - ▶ How useful are visual analytic approaches to linguistic data?
  - ▶ Which visual variables and representations are most effective for which kind of problem/type of data?

- ▶ Project A03: Identification of subspaces/patterns in larger amounts of high-dimensional data

⟹ Historical linguistic data is high-dimensional and contains subspaces (e.g., interacting factors, relevant time periods) which need to be identified and understood.

The diachrony of case

Empirical observation:

- ▶ Languages may lose their original case marking system and not innovate a new one (e.g., English).
- ▶ Languages may lose their original case marking system and innovate a novel system using new forms, replacing the old ones (e.g., Indo-Aryan: Marathi, Hindi/Urdu, Nepali, . . .).
- ▶ Languages may use the same case marking system over the centuries (e.g., Icelandic).

Questions:

- ▶ Why are case marking systems (not) lost?
- ▶ Why do languages innovate a new case marking system?

The diachrony of case

- Well-known trade-off between word order, case and/or agreement to mark grammatical relations (cf. Kiparsky 1987, 1988, 1997).
- Kiparsky (1997): Rise of positional licensing correlates with loss of morphology in historical English.

Icelandic . . .

- has retained a complex morphological case and agreement system, but also has a fairly fixed SVO word order!
- is famous for having non-nominative subjects, in particular **dative subjects** (Andrews 1976, Zaenen et al. 1985).

Dative subjects in Icelandic

- ▶ Dative subjects are attested throughout the history of Icelandic (Barðdal and Eythórsson 2009, Barðdal 2011).
- ▶ Dative subjects are mainly associated with experiencer/psych and happenstance predicates (Barðdal 2011).

(1) Vel líkuðu **goðrøði** góð røði.
     well like.PST.3PL Goðrøður.DAT good.NOM oars.NOM
     'Goðrøður (the good oarsman) liked good oars well.'

     (IcePaHC, Fyrsta málfræðiritgerðin, 1150)

The diachrony of dative subjects

On-going debate as to whether . . .

- ▶ dative subjects are a Proto Indo-European **inheritance** ('Oblique Subject Hypothesis')

  $\longrightarrow$ Evidence for continuity in Icelandic (Barðdal et al. 2012)

  or

- ▶ dative subjects are a historical **innovation** ('Object-to-Subject Hypothesis')

  $\longrightarrow$ Evidence from Indo-Aryan (Hock 1990, Deo 2003, Butt and Deo 2013)

The diachrony of dative subjects

- ▶ Old Indo-Aryan shows no evidence for dative subjects (Hock 1990).
- ▶ Loss of the original case system in Middle Indo-Aryan.
- ▶ Early New Indo-Aryan: Lexical semantic shifts of individual verbs lead to the emergence of dative experiencer subjects (Deo 2003, Butt and Deo 2013).

(2) na=enaṁ   dahati   pāvakaḥ.
     NEG=this-MAS-ACC-SG   burn-PRES-3-SG   fire-MAS-NOM-SG
     'The fire does not burn him (the soul).'[BG 2.23] – **Sanskrit**

(3) **mulī-lā**   āī-ca   rāgāvṇa
     girl-FEM-SG-DAT   mother-FEM-GEN-SG   scolding-NEU-NOM-SG
     ḍāj-ta
     trouble-PRES-NEU-SG
     'The mother's scolding torments the girl.' – **Marathi**

(Deo 2003, 6)

Continuity and change

- The Icelandic attestation only goes back to the 12th century.

- This is about when dative subjects begin to be possible in Indo-Aryan in the first place.

- Moreover, the distribution of dative subjects is changing in present-day Icelandic: **'Dative Sickness/Substitution'**
  - Accusative experiencer subjects are systematically replaced by dative subjects.
  - Increasing systematic association of dative case with experiencer semantics (Smith 1996, Jónsson 2003).

(4) **Mig** langar   að fara.   (5) **Mér** langar   að fara.
     I.ACC long.PRES to go   **I.DAT** long.PRES to go
     'I long to go.'   'I long to go.'

(Smith 1996, 22)

How stable is the distribution of dative subjects in the history of Icelandic?

Thesis Overview

1. Introduction
2. Linguistic background
3. Visual analytics for historical linguistics
4. Dative subjects in Icelandic
5. **Subject case and word order in Icelandic**
6. Dative subjects in Marathi
7. Conclusion

Advisors: Miriam Butt and Frans Plank
External Advisor: Ashwini Deo, Ohio State University

Continuity and change

- ▶ Icelandic is said to be the most conservative Germanic language (Thráinsson 1994).

- ▶ **However, changes have been observed!**
  - ▶ 'freer' > less 'free' word order (Rögnvaldsson 1995)
  - ▶ OV/VO variation > exclusively VO (Hróarsdóttir 2000)
  - ▶ decrease in V1 (Sigurðsson 1990, Butt et al. 2014)
  - ▶ increase in dative subjects/dative substitution (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; Wallenberg et al. 2011).
- ▶ Data visualization with HistoBankVis (Schätzle et al. 2017).

- ▶ Interaction between:
    - ▶ **dative subjects**
    - ▶ **word order**
    - ▶ expletives (Hannah Booth, University of Manchester)

- ▶ Evidence for the development of structure and positional licensing in Icelandic.
- ▶ Evidence against the Proto Indo-European inheritance of dative subjects.

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))
```

Visual Analytics for Historical Linguistics

> **Problem:** Diachronic investigations involve understanding highly complex interactions between various linguistic and extra-linguistic features and structures.
>
> $\longrightarrow$ Meaningful patterns are difficult to see in the forest of numbers.

Emmanuelle Moureaux 'Forest of Numbers'

Visual Analytics for Historical Linguistics

Visual Analytics

- ▶ "Analyze first, show the important, zoom, filter and analyze further, details on demand" (Keim et al., 2008)
- ▶ Compact presentation of large amounts of data
- ▶ Different levels of detail on demand (interactivity)
- ▶ Exploratory and confirmatory data analysis
- ▶ Iterative process of hypothesis testing and generation

HistoBankVis: Visualizing language change

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

Data Processing → Filter & Dimension Selection → Define Time Periods → Visualization

Interactive Task-Based Feedback Loop

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

HistoBankVis: Word order and subject case

DEMO
http://subva.dbvis.de/histobankvis-v1.0/#/

Summary of Results

- **Prefinite position** becomes the **preferred subject position** in the history of Icelandic.

- 19th century is a major key turning point.

- Dative subjects lag behind in being realized in a particular position.

- Frequency of dative subjects increases as of 1900.

- Dative subjects become more systematically associated with experiencers and goals (Chapter 4).

- **Manchester collaboration**: Decrease of V1 and rise of expletives are connected to the observed changes (and also with each other).

$\Longrightarrow$ Evidence for the development of structure and the rise of positional licensing in Icelandic.
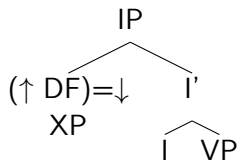
Rise of positional licensing

- ▶ Kiparsky (1995): Germanic languages developed structure and functional categories not present in Indo-European ancestor.
- ▶ Growth of structure and the development of functional categories in Icelandic noun phrases (Börjars et al. 2016).
- ▶ Early Germanic had fairly free word order, with grammatical functions indicated by case morphology.
- ▶ Flat tree in which word order is used to signal information-structural content (cf. Urdu/Hindi, Butt and King 2004).

$$
\begin{array}{c}
\text{S} \\
\end{array}
$$

(↑ DF)=↓   XP   VC
   XP
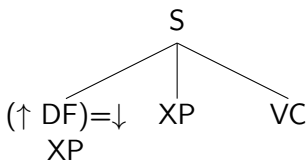
Rise of Positional Licensing

- ▶ Periphrastic tense/aspect arises, leading to an I (cf. Old English, Kiparsky 1997).
- ▶ Finite verbs (I) partition a clause in terms of information-structural information (topic vs. comment, cf. Hinterhölzl & Petrova 2010).
- ▶ V1 in topicless sentences (e.g. presentationals).

```
              IP
             /  \
   (↑ DF)=↓      I'
      XP        /  \
               I    VP
```

Rise of Positional Licensing

- ▶ Blueprint for clausal structure in current Icelandic proposed by Sells (2001, 2005).
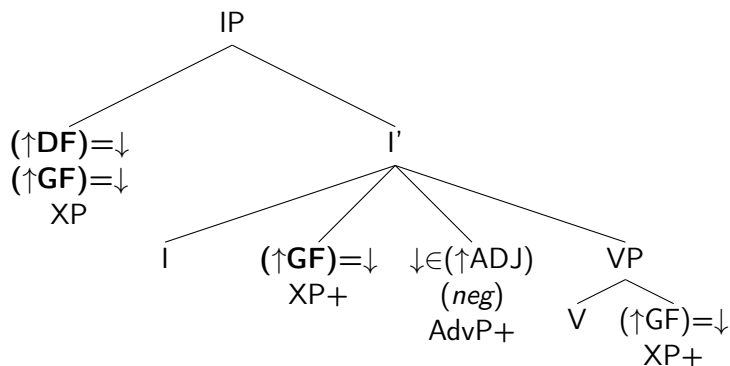
Rise of Positional Licensing

- ▶ Prefinite position in SpecIP is associated with a discourse function (i.e., topic).
- ▶ Subjects tend to be topical and the SpecIP position becomes increasingly associated with subjects.
- ▶ Subjects can occur in the immediately postfinite position when the prefinite position is occupied.

- ▶ **Problem:** Expletive *það* occurs in the prefinite position and is neither a topic nor a subject (contra Sells 2001, 2005)!
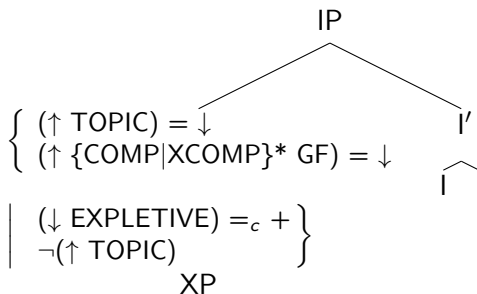
What motivates clause-initial *það*?

Transitive Expletive Constructions

(6) **Það** hafa [**margir jólasveinar**]
EXPL have.PST.3PL many.NOM Christmas-trolls.NOM
borðað búðing.
eat.PST.PTCP pudding.ACC
'Many Christmas trolls have eaten pudding.'
(Bobaljik and Jonas, 1996, 209)

Core observation: Expletive *það* licenses a clause in which there is no topic (Rögnvaldsson and Thráinsson 1990).

Rise of Positional Licensing

- ▶ SpecIP can be a topic position, preferably hosting subjects.
- ▶ SpecIP can be filled by expletive *það*, stating that the sentence is topicless.
- ▶ Not filling the SpecIP position leads to (topicless) V1 structures.

$$
\begin{array}{c}
\text{IP}\\
\diagup \qquad \diagdown\\
\left\{ \begin{array}{l} (\uparrow \text{TOPIC}) = \downarrow \\ (\uparrow \{\text{COMP}|\text{XCOMP}\}^* \text{ GF}) = \downarrow \end{array} \right. \qquad \text{I}'\\
\qquad\qquad\qquad\qquad \widehat{\text{I}}\\
\left. \begin{array}{l} (\downarrow \text{EXPLETIVE}) =_c + \\ \neg(\uparrow \text{TOPIC}) \end{array} \right\}\\
\text{XP}
\end{array}
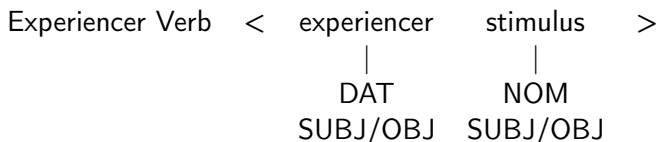$$

Dative subjects and positional licensing

- 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.
- **Significant increase of prefinite dative subjects after 1900; prefinite position becomes dominant.**

Dative subjects and positional licensing

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

$$
\begin{array}{ccccc}
\text{Experiencer Verb} & < & \text{experiencer} & \text{stimulus} & > \\
 & & | & | & \\
 & & \text{DAT} & \text{NOM} & \\
 & & \text{SUBJ/OBJ} & \text{SUBJ/OBJ} &
\end{array}
$$

$\Longrightarrow$ Dative experiencers become more firmly linked to subjects than to objects.

Dative subjects and positional licensing

- ▶ Dative experiencers are not prototypical subjects and may retain some object properties.
- ▶ SpecIP position becomes more firmly associated with topics.
- ▶ As a result, dative experiencers are also increasingly placed initially.
- ▶ Over time, dative experiencers become more firmly established as subjects.
- ▶ The prefinite (topic) position becomes the preferred subject position over the history of Icelandic.
- ▶ As non-canonical subjects, the dative experiencers eventually **conform to the overall positional licensing** developed in the language.

Conclusion

Linguistic Insights

- ▶ The 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.
  - ▶ Position ↔ grammatical functions ↔ lexical semantics
  - ▶ Information structure ↔ word order

Conclusion

HistoBankVis

- ▶ HistoBankVis immensely facilitates the analysis of historical linguistic data.
- ▶ Combination of knowledge-based and data-driven modeling.
- ▶ HistoBankVis bridges the gap between annotated values, statistical analysis and the actual underlying data.
- ▶ Each analysis step is accessible via a single identification URL.
  - ▶ Storage of multiple perspectives on the data
  - ▶ Support of collaborative research
- ▶ System can be applied to any Penn Treebank-style corpus or well-structured data set.

Thank you!
Questions?

The Icelandic Yule Lads

Stekkjastaur (Sheep-Cote Clod)
Harasses sheep, impaired by his stiff peg-legs

Giljagaur (Gully Gawk)
Hides in gullies, waiting for an opportunity to sneak into the cowshed and steal milk.

Stúfur (Stubby)
Abnormally short. Steals pans to eat the crust left on them

Þvörusleikir (Spoon-Licker)
Steals Þvörur (a type of a wooden spoon) to lick. Is extremely thin due to malnutrition.

Pottaskefill Pot-Licker
Steals leftovers from pots

Askasleikir Bowl-Licker
Hides under beds waiting for someone to put down their 'askur' (a type of bowl), which he then steals

Hurðaskellir Door-Slammer
Likes to slam doors, especially during the night

Skyrgámur Skyr-Gobbler
A Yule Lad with an affinity for skyr

Bjúgnakrækir Sausage-Swiper
Would hide in the rafters and snatch sausages that were being smoked

Gluggagægir Window-Peeper
A voyeur who would look through windows in search of things to steal

Gáttaþefur Doorway-Sniffer
Has an abnormally large nose and an acute sense of smell which he uses to locate laufabrauð

Ketkrókur Meat-Hook
Uses a hook to steal meat

Kertasníkir Candle-Stealer
Follows children in order to steal their candles (which in those days was made of tallow and thus edible)

Grýla
The mother of the Yule Lads. Icelandic parents did scare their children from misbehaving by telling them that Grýla could come and abduct them

Leppalúði
The husband of Grýla. Not that evil, but a lazy one.