Visualizing Language Change: Dative Subjects in Icelandic

**Christin Schätzle & Dominik Sacha**

University of Konstanz

VisLR II workshop
LREC 2016

May 23, 2016

Introduction

- Diachronic visualization of multidimensional language data

- Corpus study of the interaction among dative subjects, lexical semantic verb classes and voice

- Icelandic Parsed Historical Corpus (IcePaHC; Wallenberg et al., 2011)

- Adaption of V1-visualization developed by Butt et al. (2014)

- 'Overview first – details on demand' (Keim et al., 2008; Schneiderman, 1996)
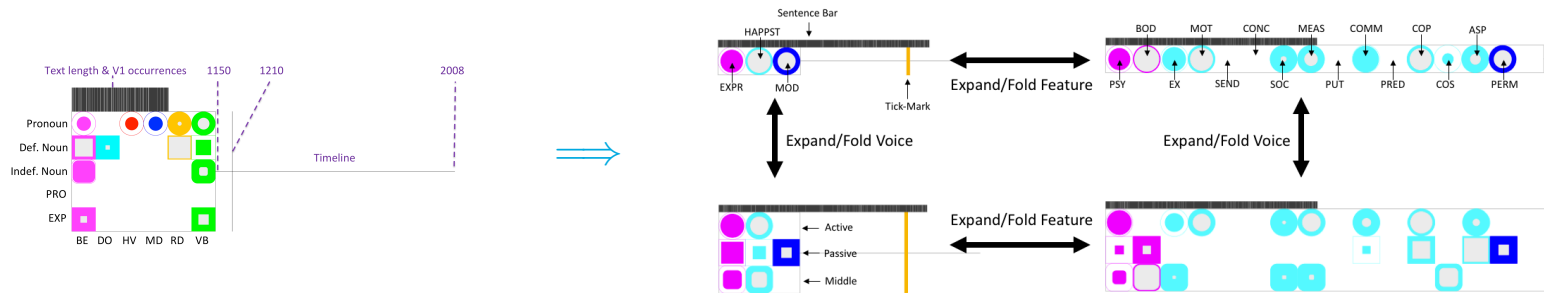
The Diachrony of Dative Subjects

- Dative subjects exist in a multitude of Indo-European languages.
- Debate as to whether...

dative subjects are a Proto Indo-European Inheritance (Oblique Subject Hypothesis)
⟶ evidence for continuity in Icelandic (e.g. Barðdal et al. 2012)

or

dative subjects are a historical innovation (Object-to-Subject Hypothesis)
⟶ evidence from Indo-Aryan (e.g. Butt and Deo 2013)

Case in Icelandic

- Icelandic has morphologically distinct expressions for nominative, accusative, dative and genitive case.
- The standard case-marking pattern is nominative on subjects, accusative on direct objects, dative on indirect objects and genitive on possessors.
- Icelandic is famous for having non-nominative subjects, in particular dative subjects (see, e.g. Zaenen et al. 1985).
- Also, nominative objects are possible.

Dative Subjects in Icelandic

- Dative subjects are attested throughout the history of Icelandic (see, e.g. Barðdal and Eythórsson 2009).

- Dative subjects in Icelandic are mainly associated with experiencer/psych and happenstance predicates (see, e.g. Barðdal 2011)

  (1)   Vel líkuðu   **goðrøði**   góð   røði.
  well like.PAST.3.PL **PROP.DAT.SG** good.NOM.PL oar.NOM.PL

  'Goðrøði (the good oarsman) liked good oars well.'

  (Fyrsta málfræðiritgerðin, 1150)

- . . .but there are changes!

Dative Substitution/ Dative Sickness

- Change in progress: accusatives are original, datives the innovation

(2)	**Mig** langar að fara.	(3)	**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)		(Smith 1996, 22)

- Change began in the latter part of the 19th century (Barðdal 2011)
- Datives are systematically associated with either goals or experiencers in Modern Icelandic (Smith 1996).

$\Longrightarrow$ Lexical semantic factors play a role in the determination of dative subjects.

Voice – Passive

- Dative subjects appear in active, passive and middle constructions.
- Passivization:
    - *vera* 'to be' + past participle
    - External argument is demoted
    - Accusative objects are realized as nominative subjects.

    (4) einhver barði **strákana** í skolanum
        somebody.NOM hit.PAST.3.SG **the.boy.ACC.PL** in the.school.DAT.SG
        'Somebody hit the boys in school.'
        (Thráinsson 1994,177)

    (5) **strákarnir** voru   barðir   í skolanum
        the.boy.NOM.PL be.PAST.3.PL hit.PPART.M.NOM.PL in the.school.DAT.SG
        'The boys were hit in school.'
        (Thráinsson 1994,177)

Voice – Passive

- However, dative objects preserve their case marking under passivization and are realized as dative subjects:

(6)   Skipstjòrinn   sökkti   **skipinu**.
the.captain.NOM.SG sink.PAST.3.SG **the.ship.DAT.SG**

'The captain sank the ship.'

(Zaenan and Maling, 1984, 141)

(7)   **Skipinu**   var   sökkt   af skipstjóranum.
the.ship.DAT.SG be.PAST.3.SG hit.PPART.N.NOM.SG by the.captain.DAT.SG

'The ship was sunk by the captain.'

(Zaenan and Maling, 1984, 142)

Voice – Middle

- Middle formation:
    - Middle voice is marked via the suffix *-st* (see, e.g. Anderson, 1990).
    - Accusative objects are realized as nominative subjects (similar to passivization).

    - **Dative objects: themes/patients are realized as nominative subjects, but benefactives/goals see case preservation!**

Voice – Middle

- Dative themes/patients are realized as nominative subjects.

(8)   Ég   hellti   **mjólkinni**   niður.
I.NOM spill.PAST.1.SG **the.milk.DAT.SG** down
'I spilled the milk down.'
(Sigurðsson, 1989, 265)

(9)   **Mjólkin**   helltist   niður.
the.milk.NOM.SG spill.PAST.MID down
'The milk spilled down.'
(Sigurðsson, 1989, 265)

Voice – Middle

- Dative benefactives/goals are realized as dative subjects.

   (10) Pétur   bauð   mér vinnu.
       Peter.NOM.SG offer.PAST.3.SG I.**DAT** job.ACC.SG
       'Peter offered me a job.'
       (Sigurðsson, 1989, 260)

   (11) Mér bauðst   vinna.
       I.**DAT** offer.PAST.MID job.NOM.SG
       'I got the opportunity to get a job.'
       (Sigurðsson, 1989, 260)

⟹ Further indicator for a lexical semantic approach to case marking in Icelandic

Corpus Study

- Barðdal and colleagues claim Icelandic has been stable with respect to dative subjects.

- The earliest attested stages of Icelandic stem from the 12th century.

- This is about when the New Indo-Aryan languages began to develop a new case marking system and dative subjects emerged (see, e.g. Butt & Deo 2013).

- Case alternations in middles and dative substitution in Icelandic

$\Longrightarrow$ The distribution of dative subjects across semantic verb classes and voice in Icelandic should change or at least vary diachronically!

Corpus Study

Icelandic Parsed Historical Corpus (IcePaHC)

- 12th to 21st century – annotates dative subjects from the earliest stages
- 60 texts, 1 million words
- different genres, not representative across centuries
- annotated according to syntactic annotation of the Penn Treebank (Marcus et al. 1993)
- provides information about grammatical relations and case

Corpus Study

- Automatic extraction of verbs inducing dative subjects (>4000)

- Addition of a lexical semantic annotation layer of verb classes for dative subject predicates based on Levin (1993) and Barðdal et al. (2012)

```
(IP-MAT-SPE (NP-SBJ (PRO-D Mér-mér))
        (VBPI-PSY 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))
```

- Division of data into time stages as per Haugen (1984)

- χ^2 to test whether our observed distributions differ from what could be expected

Some Results

Time	active	passive	middle	χ^2
1150-1350	64.4%	18.2%	17.4%	***
1350-1550	66.8%	17.5%	15.7%	***
1550-1750	46.1%	28.8%	25.1%	***
1750-1900	53.1%	20.8%	26.2%	
1900-present	43.2%	14.3%	42.5%	***
all	55.3%	19.5%	25.5%	

Table: Diachronic distribution of dative subject predicates by voice (* p<0.05, ** p<0.01, *** p<0.001)

- Actives:
  - Dative subjects most often occur in active constructions, but they show an overall decrease.
  - Mostly psych predicates, e.g. *líka* 'to like'

Some Results

Time	active	passive	middle	χ^2
1150-1350	64.4%	18.2%	17.4%	***
1350-1550	66.8%	17.5%	15.7%	***
1550-1750	46.1%	28.8%	25.1%	***
1750-1900	53.1%	20.8%	26.2%	
1900-present	43.2%	14.3%	42.5%	***
all	55.3%	19.5%	25.5%	

Table: Diachronic distribution of dative subject predicates by voice (* $p<0.05$, ** $p<0.01$, *** $p<0.001$)

- Passives:
    - Passives with dative subjects appear less often and they also decrease.
    - Mainly associated with verbs of communciation, e.g. *tylkinna* 'announce', and change of possession, e.g. *kaupa* 'buy'

Some Results

Time	active	passive	middle	χ^2
1150-1350	64.4%	18.2%	17.4%	***
1350-1550	66.8%	17.5%	15.7%	***
1550-1750	46.1%	28.8%	25.1%	***
1750-1900	53.1%	20.8%	26.2%	
1900-present	43.2%	14.3%	42.5%	***
all	55.3%	19.5%	25.5%	

Table: Diachronic distribution of dative subject predicates by voice (* $p<0.05$, ** $p<0.01$, *** $p<0.001$)

- Middles:
  - Middles with dative subjects are on the increase over the whole time span.
  - Most often psych predicates, e.g. *leiðast* 'be bored'
- Deviating percentages in the third time stage can be attributed to a genre effect (see Butt et al. 2014).

Problems/Challenges

- Significant changes with respect to voice could be identified, analysis of verb classes displays a more tedious task → 15 different verb classes across three different voices
- Generation of multiple data tables with various different characteristics
- Meaningful patterns may stay undetected.
- Some of the relative frequencies are based on very few occurences of the actual observation.
- Only binary comparisons are possible.
- Limited access to actual interactions

$\implies$ Visualize the data in order to be able to account for all dependencies and inter-actions of factors

    May 23, 2016   VisLR II - LREC 2016   Christin Schätzle & Dominik Sacha

Visual Analytics

- Compact presentation of large amounts of data
- Enables an exploratory and confirmatory data analysis
- Fosters the generation and validation of hypotheses which may lead to often unexpected insights
- Data is analyzed as is, no specific assumptions needed in advance
- The visual design is driven to support the analyst's tasks best.

Diachronic visualization of V1 (Butt et al. 2014)

- V1 (Verb First) word order in IcePaHC

- Glyph visualization allowing for a comparative diachronic evaluation

- 'Overview first – details on demand'
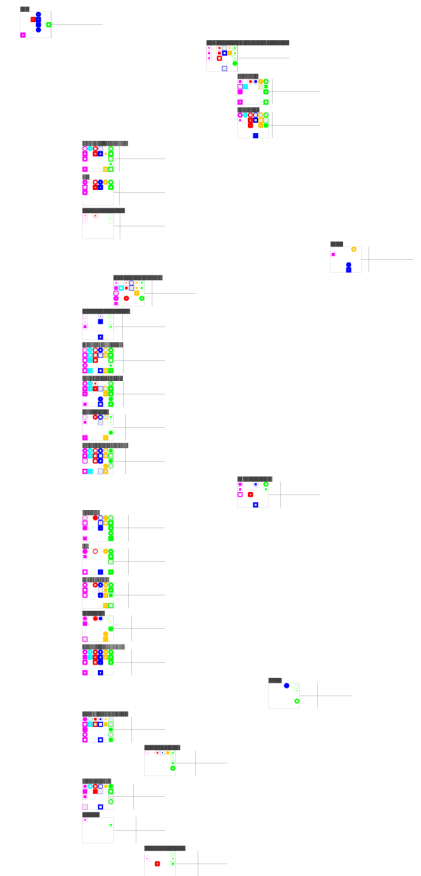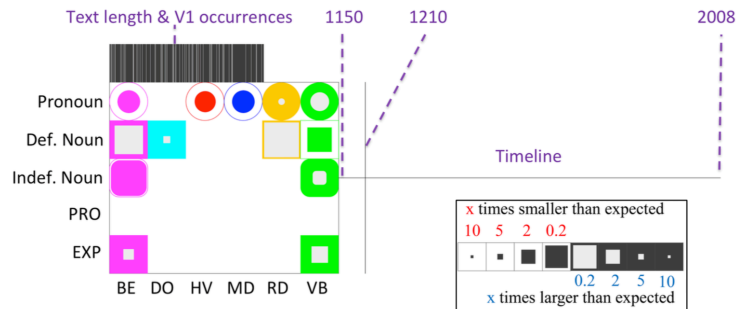
Diachronic Visualization of Dative Subjects

- Visualization adopts the general design of Butt et al. (2014)

- Aggregation of verb classes into a higher class categorization as given by Barðdal (2011) and Barðdal et al. (2012)

- One line of features without further distinctions

- Detailed features can be shown on demand in order to cope with the high number of features
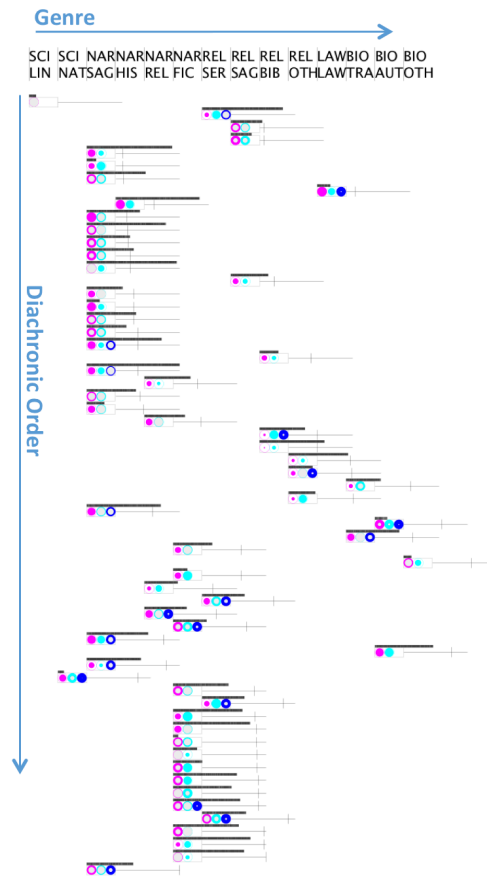
# Diachronic Visualization of Dative Subjects

Label	Verb Class	Encoding
EXPR	Experiencer verbs	Magenta
PSY	Psych verbs	Magenta
BOD	Body verbs	Magenta
HAPP	Happenstance verbs	Light Blue
EX	Existence verbs	Light Blue
MOT	Motion verbs	Light Blue
SEND	Verbs of sending	Light Blue
CONC	Verbs of concealment	Light Blue
SOC	Verbs of social interaction	Light Blue
MEAS	Measure verbs	Light Blue
PUT	Put verbs	Light Blue
COMM	Communication verbs	Light Blue
PRED	Predicative complements	Light Blue
COP	Change of possession	Light Blue
COS	Change of state	Light Blue
ASP	Aspectual verbs	Light Blue
MOD	Modality verbs	Dark Blue
PERM	Permission verbs	Dark Blue

Table: Visual encodings of verb classes

Voice	Encoding
Active	Circle
Passive	Rectangle
Middle	Rounded Rectangle

Table: Visual encodings of voice

Diachronic Visualization of Dative Subjects
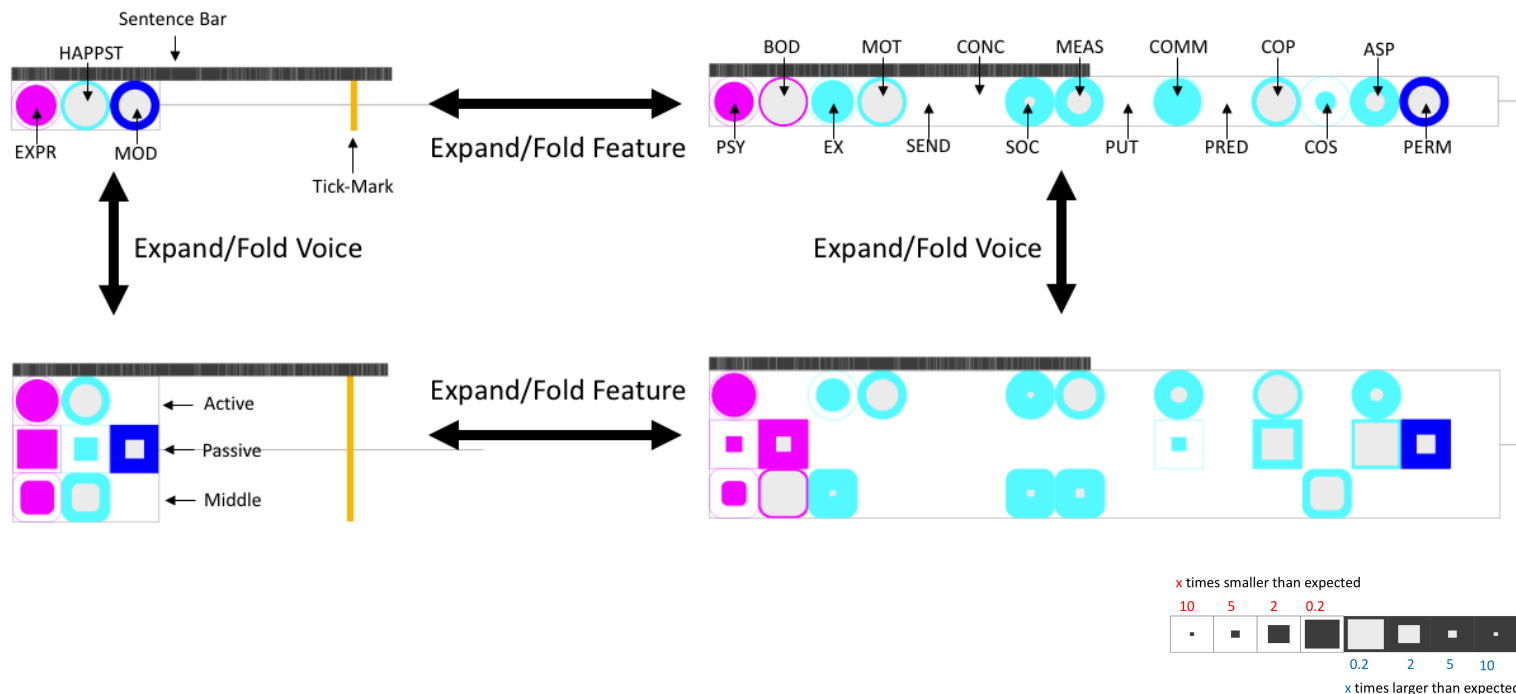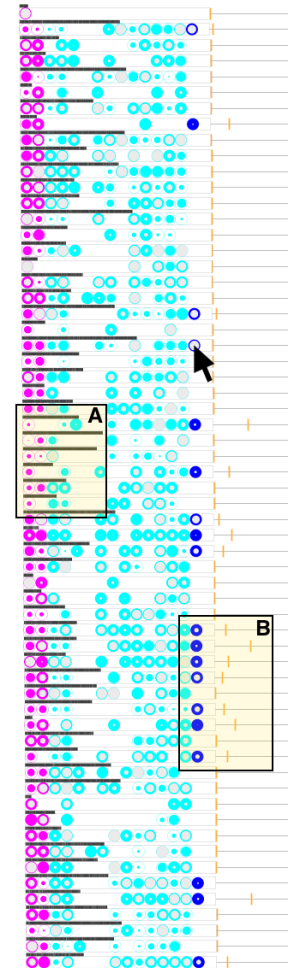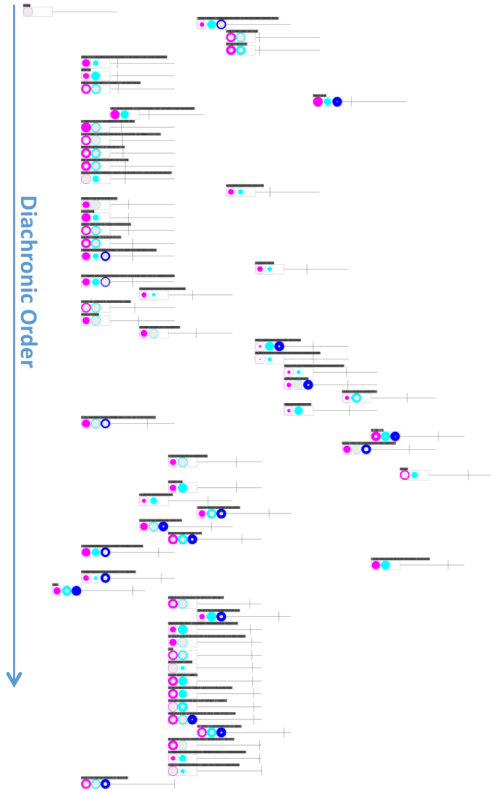
Results

Dative subjects …

- are common throughout the history of Icelandic

- are mainly associated with experiencer predicates and, to a lesser extent, with happenstance predicates throughout the corpus

- appear in all three voices with experiencer and happenstance predicates

- occur rarely with verbs of modality (permission) in the first half of the corpus, but appear more often in the latter part

- with verbs of modality most often occured in passives and more rarely in middles, but never with actives

- are increasingly used with experiencer predicates from the end of the 19th century on which correlates with an increasing use of middle verbs

Results

- However, in the visual analysis of the more fine-grained verb classes a large set of visual cues has to be distinguished and processed → at a glance identification of salient patterns is merely impossible
- Visualization clearly identifies a genre effect (see Butt et al., 2014)

$\implies$ The distribution of dative subjects in Icelandic has been changing over the past millenium.

$\implies$ These changes point towards an increasing systematic association of dative case with experiencer arguments and against the Oblique Subject Hypothesis.

Conclusion

- Our visualization represents a useful and powerful tool for the analysis of multidimensional and complex language data.
- The visualization helps to uncover and understand multidimensional interactions of factors conditioning diachronic language change.
- Future work:
    - improve the visualization of the lower verb classes
    - revise lexical semantic annotation scheme making it more likely to find significant patterns

Thank you!
Questions?

Acknowledgements

DFG Deutsche Forschungsgemeinschaft

Some References

Anderson, S. R. (1990). The grammar of Icelandic verbs in -st. In J. Maling et al., editors, *Modern Icelandic Syntax*, pages 235—273. Academic Press, San Diego.

Barðdal, J., Smitherman, T., Bjarnadóttir, V., Danesi, S., Jenset, G. B., and McGillivray, B. (2012). Reconstructing constructional semantics: The dative subject construction in Old Norse-Icelandic, Latin, Ancient Greek, Old Russian and Old Lithuanian. *Studies in Language*, 36(3):511–547.

Barðdal, J. (2011). The rise of dative substitution in the history of Icelandic: A diachronic construction grammar account. *Lingua*, 121(1):60–79.

Butt, M. and Deo, A. (2013). A historical perspective on dative subjects in Indo-Aryan. Paper presented at the LFG13 Conference.

Butt, M., Bögel, T., Kotcheva, K., Schätzle, C., Rohrdantz, C., Sacha, D., Dehe, N., and Keim, D. (2014). V1 in Icelandic: A multifactorial visualization of historical data. In *Proceedings of VisLR: Visualization as added value in the development, use and evaluation of Language Resources*, Workshop at the 9th edition of the Language Resources and Evaluation Conference (LREC 2014), Reykjavik, Iceland.

Keim, D., Andrienko, G., Fekete, J.-D., Górg, C., Kohlhammer, J., and Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In A. Kerren, et al., editors, *Information Visualization*, pages 154–175. Springer, Berlin.

Levin, B. (1993). *English Verb Classes and Alternation. A Preliminary Investigation*. Chicago: Universtiy of Chicago Press.

Marcus, M. P., Marcinkiewicz, M. A., and Santorini, B. (1993). Building a large annotated corpus of English: the penn treebank. *Computational Linguistics*, 19(2):313– 330.

Computational Linguistics Sigurðsson, H. A. (1989). *Verbal Syntax and Case in Icelandic: A Comparative GB Approach*. Ph.D. thesis, Lund University. Smith, H. (1996). Restrictiveness in Case Theory. Cambridge University Press, Cambridge.

Wallenberg, J. C., Ingason, A. K., Sigurðsson, E. F., and Rögnvaldsson, E. (2011). *Icelandic Parced Historical Corpus (IcePaHC)*.

Zaenen, A. and Maling, J. (1984). Unaccusative, passive, and quirky case. In M. Cobler, et al., editors, *Proceedings of the Third West Coast Conference on Formal Linguistics (WCCFL)*, pages 317–329.

Zaenen, A., Maling, J., and Thráinsson, H. (1985). Case and Grammatical Functions. *The Icelandic Passive. Natural Language and Linguistic Theory*, 3(4):441—-483.