Dative Subjects: Historical Change Visualized

Doctoral thesis submitted for the degree of Doctor of Philosophy (Dr.phil.)

presented by

Christin Schätzle

at the

Universität Konstanz

Faculty of Humanities

Department of Linguistics

Konstanz, 2018
Date of the oral examination: December 12th, 2018

First referee: Prof. Dr. Miriam Butt

Second referee: Prof. Dr. George Walkden

Third referee: Prof. Dr. Annie Zaenen
Abstract

The Icelandic case system presents an interesting linguistic puzzle. Languages tend to use either word order, case and/or agreement to signal grammatical relations (Kiparsky 1987, 1988, 1997). Icelandic is atypical in this respect as it has a rather rigid word order, but also retained a rich morphological case system over the centuries. Moreover, non-nominative subjects exist in the language, with in particular the synchronic existence of dative subjects being well-established (Andrews 1976, Zenen et al. 1985). From a diachronic perspective, dative subjects have also attracted a good deal of research, specifically with respect to the question about whether dative subjects are a common Proto-Indo-European feature or whether they are a more recent historical innovation (see, e.g., Haspelmath 2001, Barðdal and Eythórsson 2009, Barðdal et al. 2012).

In this thesis, I investigate factors conditioning the diachronic occurrence of dative subjects in the Icelandic Parsed Historical Corpus (IcePaHC, Wallenberg et al. 2011) to provide a window of understanding of the complex system licensing grammatical relations in the language, contributing to the discussion which evolved around the historical origin of dative subjects. As method of investigation, I utilize novel visualization techniques coming from the field of Visual Analytics (Keim et al. 2008). The investigations presented in this thesis show that dative subjects are part of a complex interlinked system in which case, word order, grammatical relations, lexical semantics and event structure interact in the mapping of arguments to grammatical relations. For one, I provide my findings with respect to the interaction between dative subjects, thematic roles, event structure and voice in IcePaHC, showing that the distribution of dative subjects has been changing in the history of Icelandic, in particular with respect to an increasingly systematic association between dative subjects and experiencer semantics. This correlates with an increasing use of verbs carrying middle morphology, which have been lexicalized as
stative experiencer predicates with a dative subject over time. I furthermore present
an investigation of the interaction between subject case and word order which ex-
amines the interrelation between dative case, subject positions, and verb placement
in IcePaHC. This investigation provides evidence for the diachronic development of
structure and the rise of positional licensing in the language (in line with Kiparsky
1997); developments in which dative subjects consistently lag behind.

For the theoretical analysis of the historical developments observed in IcePaHC,
I present a novel linking theory couched in the Lexical-Functional Grammar (LFG)
framework in this thesis. My linking theory builds on the enhancements of LFG’s
Lexical Mapping Theory by Zaenen (1993) and Kibort (2014) with respect to lex-
cical semantic entailments and argument positions, separating out lexical semantics
from structural positions. As core component of the linking system, I implement
a reference frame in the form of Talmy’s (1978) figure-ground division, which
functions as mediator between word order, lexical semantics, and event structure.
Grammatical relations are linked to arguments via a set of lexical semantic entail-
ments which follow from the event structure, the reference frame, and the sentence
of arguments, associating grammatical relations with particular structural positions.
Event structure is encoded in the linking system via the event participants assumed
in Ramchand’s (2008) event-decompositional framework of the first-phase syntax
and is taken to license case marking in Icelandic as has been suggested by Sveno-
nius (2002). Overall, the linking analysis of the diachronic corpus data shows that
the licensing conditions for case and grammatical relations have been changing over
time, which questions the inheritance of a stable and monolithic dative subject con-
struction from earlier language stages.
Zusammenfassung


Acknowledgements

This thesis would not have been possible without the support and encouragement of many people.

First of all, I would like to thank my supervisor, Miriam Butt. Miriam hired me as a student assistant back in 2012, and has shaped my academic career ever since. She drew my interest towards using computational linguistic methods to investigate the diachrony of case and moreover planted the idea in my head to combine historical linguistic research with visual analytics. I am particularly thankful that she has provided me with permanent job opportunities during my doctoral studies, which enabled me to focus on my research without having to worry about funding. This moreover gave me the possibility to present my work at different venues around the world. She provided valuable feedback on various aspects of this thesis, discussed linking with me many times and managed to eventually turn me into a ‘real’ linguist.

I would also like to thank George Walkden who agreed to be my co-advisor and regularly commented on my work. I moreover thank Annie Zaenen for providing valuable feedback on this thesis and for being part of my dissertation committee. Gillian Ramchand discussed dative subject verb classes with me, furthering my understanding of the Icelandic data. I also want to thank Frans Plank, who encouraged me to do a PhD after I finished my master’s thesis and has supported my research ever since.

I want to thank Ashwini Deo, who agreed to host me as a visiting scholar for three months at Yale in 2015. I learned a lot from her, advanced as a linguist and had the chance to present my research to the linguistics department at Yale. Many thanks also go to the members of the Language and Brain lab at Yale, who took care of me during my stay. Thank you, Sara, Martin, Jun, Yao-Ying, Andy, and Maria for organizing pot-lucks, taking me to the movies, and showing me New Haven and the surroundings.
In the end of 2016, Miriam sent me to Manchester to meet Kersti Börjars and her PhD student Hannah Booth, who I was supposed to work with. What felt like an ‘arranged marriage’ in the beginning turned into a very productive and successful collaboration. Hannah helped me understanding pieces of data that I was clueless about and furthered my knowledge of Old Icelandic. Moreover, she made a great travel companion when we explored Texas together after ICHL. Thank y’all, Hannah!

I am very thankful to my colleagues and friends from my working group and the linguistics department in Konstanz. Annette Hautli-Janisz, Tina Bögel, Sebastian Sulger, and Melanie Seiss helped me organizing my PhD and always had an open ear. Katerina Kalouli suffered with me on many long office days, brought me cake and chocolate whenever I was in need and was always there for me. Jessica Zipf, Farhat Jabeen and Saira Bano always managed to cheer me up when I felt stressed. Thank you Mark-Matthias Zymla for the regular Friday afternoon visits, suffering together with me from the process of writing a dissertation. Thanks also go to Alexandra Rehn for helping me last-minute with the thesis submission process.

I am particularly grateful to my friends and family, who always believed that I could do this and supported me in every respect. They provided the necessary amount of diversion, but also gave me enough freedom and space to achieve my goals. Many thanks go in particular to my ‘Handballerinas’, the best team one could wish for.

Last, I want to thank the most important person in my life. Simon, I could never have done this without you!
Contents

1 Introduction 1

2 Background 7
  2.1 Introduction ................................................. 7
  2.2 Icelandic .................................................. 8
  2.3 Non-canonical case marking ................................. 13
  2.4 Dative subjects in Modern Icelandic ....................... 16
  2.5 The diachrony of dative subjects ......................... 24
    2.5.1 The Object-to-Subject Hypothesis .................... 24
    2.5.2 The Oblique Subject Hypothesis ..................... 28
  2.6 Dative subjects and lexical semantics .................... 31
    2.6.1 Lexical semantic verb classes ....................... 31
    2.6.2 Passivization and middle formation .................. 35
  2.7 Dative case and the structure of events ................. 39
  2.8 The Icelandic Parsed Historical Corpus (IcePaHC) ....... 43
    2.8.1 Data processing ...................................... 45
    2.8.2 The IcePaHC dataset ................................ 50
    2.8.3 A note on data curation ............................... 50
  2.9 Visual Analytics for historical linguistics (HistLingVis) 52
    2.9.1 Challenges for historical linguistic research ....... 54
    2.9.2 Diachronic linguistic visualizations ................. 56
    2.9.3 Designing visualizations .............................. 60
  2.10 Summary and conclusion .................................... 63

3 Linking Theories 65
  3.1 Introduction .............................................. 65
3.2 Lexical Mapping Theory ........................................... 66
   3.2.1 Lexical-Functional Grammar ................................. 66
   3.2.2 C-structure and f-structure ................................. 67
   3.2.3 A-structure and thematic roles ............................ 69
   3.2.4 Quirky case and the Icelandic Association Principles .... 70
   3.2.5 Feature decomposition ..................................... 72
   3.2.6 Mapping principles and wellformedness conditions ....... 74
   3.2.7 Argument alternations .................................... 75
   3.2.8 Argument positions ....................................... 78
   3.2.9 Proto-Role entailments .................................... 81

3.3 Kiparsky’s linking theory ....................................... 85
   3.3.1 Semantic Form and argument structure ..................... 85
   3.3.2 Feature decomposition ..................................... 86
      3.3.2.1 Abstract Case ........................................ 86
      3.3.2.2 Morphosyntactic Case ............................... 88
   3.3.3 Quirky case and argument scrambling ...................... 90

3.4 The first-phase syntax and event semantics ................... 92
   3.4.1 Dynamic predicates ....................................... 96
   3.4.2 Stative predicates ....................................... 99

3.5 Summary and conclusion ....................................... 101

4 Dative subjects and event semantics .............................. 103
   4.1 Introduction .................................................. 103
   4.2 Event structure and dative case ............................. 105
   4.3 Corpus study I: The diachronic distribution of dative subjects ... 107
      4.3.1 Subject and object case ................................. 108
      4.3.2 Dative subjects and voice ............................... 113
   4.4 Dative Substitution, lexicalization and middle formation ...... 117
   4.5 Annotating event structure ................................... 122
      4.5.1 Event structure classification ........................... 122
         4.5.1.1 Class I – Stative predicates ..................... 124
         4.5.1.2 Class II – Transfer verbs .......................... 125
         4.5.1.3 Class III – Transitions ............................ 127
         4.5.1.4 Class IV – Scalar changes ......................... 129
      4.5.2 Event structure annotation ............................... 130
## Contents

4.6  Corpus study II: Dative subjects and event structure ........................................ 133
   4.6.1  Glyph visualization ......................................................................................... 133
   4.6.2  Dative subjects and event structure ................................................................. 140
   4.6.3  Dative subjects, event structure and voice ..................................................... 143
4.7  Summary and conclusion ....................................................................................... 149

5  Dative subjects and the rise of positional licensing .................................................. 153
   5.1  Introduction ........................................................................................................... 153
   5.2  Corpus study: The diachrony of subject case and word order .............................. 156
      5.2.1  The HistoBankVis system .................................................................................. 156
      5.2.1.1  System implementation .............................................................................. 157
      5.2.1.2  Iterative multilayer approach .................................................................. 158
      5.2.1.3  Visualizing change over time ................................................................. 160
      5.2.2  Subject case and word order .......................................................................... 166
      5.2.3  Dative subjects, word order and voice ......................................................... 168
      5.2.4  Case and subject positions .............................................................................. 170
      5.2.5  Subject case and V1 ...................................................................................... 173
   5.3  V1 and expletives ................................................................................................. 177
      5.3.1  The diachrony of expletive \( það \) .................................................................... 177
      5.3.2  V1, expletives and subject positions .............................................................. 179
   5.4  The rise of positional licensing – an LFG analysis ................................................. 181
      5.4.1  V1 and information structure ....................................................................... 182
      5.4.2  Expletives and positional licensing ............................................................... 183
   5.5  Summary and conclusion ...................................................................................... 186

6  Linking events, case and grammatical relations in Icelandic .......................................... 189
   6.1  Introduction ........................................................................................................... 189
   6.2  Linking system ..................................................................................................... 191
      6.2.1  Event participants and argument positions ................................................... 192
      6.2.2  Reference frame ............................................................................................ 194
      6.2.3  Sentience and animacy .................................................................................. 196
      6.2.4  Lexical semantic entailments ....................................................................... 197
         6.2.4.1  The Asymmetrical Object Parameter ...................................................... 200
         6.2.4.2  Oblique arguments ................................................................................. 201
      6.2.5  Passivization ................................................................................................. 203
      6.2.6  Middle formation ......................................................................................... 204
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2.7</td>
<td>Summary</td>
<td>207</td>
</tr>
<tr>
<td>6.3</td>
<td>Event structure and case</td>
<td>208</td>
</tr>
<tr>
<td>6.4</td>
<td>Dative subjects and the rise of positional licensing</td>
<td>210</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Experiencer subjects and locative inversion</td>
<td>210</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Dative experiencers and positional licensing</td>
<td>215</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Case versus position</td>
<td>216</td>
</tr>
<tr>
<td>6.5</td>
<td>Lexicalized experiencer and raising predicates</td>
<td>217</td>
</tr>
<tr>
<td>6.5.1</td>
<td>Reanalysis via middle formation</td>
<td>218</td>
</tr>
<tr>
<td>6.5.2</td>
<td>Secondary predication and raising</td>
<td>221</td>
</tr>
<tr>
<td>6.6</td>
<td>Dative goals and middle formation</td>
<td>223</td>
</tr>
<tr>
<td>6.6.1</td>
<td>Passivization with transfer verbs</td>
<td>225</td>
</tr>
<tr>
<td>6.6.2</td>
<td>Middle formation with transfer verbs</td>
<td>226</td>
</tr>
<tr>
<td>6.6.3</td>
<td>Transfer verbs and stative experiencer predicates</td>
<td>228</td>
</tr>
<tr>
<td>6.6.4</td>
<td>Interim summary</td>
<td>229</td>
</tr>
<tr>
<td>6.7</td>
<td>Dative themes and event structure</td>
<td>230</td>
</tr>
<tr>
<td>6.7.1</td>
<td>Caused and accompanied motion</td>
<td>230</td>
</tr>
<tr>
<td>6.7.2</td>
<td>Dative themes and passivization</td>
<td>233</td>
</tr>
<tr>
<td>6.7.3</td>
<td>Dative themes and middle formation</td>
<td>234</td>
</tr>
<tr>
<td>6.7.4</td>
<td>Unaccusative predicates</td>
<td>235</td>
</tr>
<tr>
<td>6.7.5</td>
<td>Diachronic perspective</td>
<td>237</td>
</tr>
<tr>
<td>6.7.6</td>
<td>Experiencer subjects and scalar changes</td>
<td>238</td>
</tr>
<tr>
<td>6.8</td>
<td>Summary and conclusion</td>
<td>239</td>
</tr>
</tbody>
</table>

## 7 Summary and conclusion

Appendix

241

Appendix

247
List of Tables

2.1 Inflectional paradigms of Icelandic nouns (Thráinsson 1994, 153). . . 9
2.2 Predicate type frequencies across different subject cases in two corpora
  of Old Icelandic and Modern Icelandic (Barðdal 2011, 73). . . . . . . 34
2.3 Diachronic distributions of definite and indefinite NPs across different
  word orders in the history of Icelandic (Hróarsdóttir 2000, 136). . . . 55

3.1 Verb classes in the first-phase syntax (taken from Ramchand 2008, 108).101

4.1 Diachronic distribution of subject case in IcePaHC. . . . . . . . . . 108
4.2 Diachronic distribution of direct object case in IcePaHC. . . . . . 109
4.3 Diachronic distribution of indirect object case in IcePaHC. . . . . . 110
4.4 Object case distribution in dative subject clauses from IcePaHC. . . 112
4.5 Subject case distribution for active constructions in IcePaHC. . . . 114
4.6 Subject case distribution for passive constructions in IcePaHC. . . . 114
4.7 Subject case distribution for middle constructions in IcePaHC. . . . 115
4.8 Diachronic distribution of dative subject predicates by voice in IcePaHC.115
4.9 Type and token frequencies of dative subject predicates in IcePaHC. 118
4.10 Event structure classification for dative subject predicates in IcePaHC.124
4.11 Annotation labels for event classes with respect to dative subject pred-
  icates in IcePaHC. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 131
4.12 Diachronic distribution of event classes with respect to dative subject
  predicates in IcePaHC. . . . . . . . . . . . . . . . . . . . . . . . . . . 141
4.13 Diachronic distribution of voice with stative predicates taking a dative
  subject in IcePaHC. . . . . . . . . . . . . . . . . . . . . . . . . . . . 144
4.14 Diachronic distribution of voice with transfer verbs taking a dative
  subject in IcePaHC. . . . . . . . . . . . . . . . . . . . . . . . . . . . 145
4.15 Diachronic distribution of voice with transition verbs taking a dative subject in IcePaHC. 146
4.16 Diachronic distribution of voice with verbs describing scalar changes which take a dative subject in IcePaHC. 148

5.1 Diachronic distribution of subject positions in IcePaHC. 170
5.2 Diachronic distribution of subject position for dative subjects in IcePaHC. 173
5.3 Diachronic distribution of V1 matrix declaratives in IcePaHC. 174
5.4 Diachronic distribution of dative subjects in V1 declaratives in IcePaHC. 174
5.5 Diachronic distribution of prefinite expletives in IcePaHC. 178

6.1 Diachronic distribution of subject positions in IcePaHC. 215
6.2 Diachronic distribution of subject position for dative subjects in IcePaHC. 216
List of Figures
2.1

Semantic Map Model for verb classes instantiating the Dative Subject Construction in Old Norse-Icelandic, Ancient Greek, Latin, Old

33

2.2

Sample annotation for a sentence from IcePaHC. . . . . . . . . . . .

44

2.3

IcePaHC annotation for an infinitival construction with the infinitive
marker að ‘to’. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

2.4

IcePaHC annotation for a V1 declarative with an empty expletive
subject. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

2.5

49

Dataset extracted from IcePaHC for the diachronic investigation of
factors conditioning dative subjects in Icelandic. . . . . . . . . . . . .

2.6

47

50

Top: Scatterplot visualization of word contexts and their diachronic
distribution across contextual senses. Bottom: Visualization of the
temporal sense developments for to browse and to surf. Figures taken
from Rohrdantz et al. (2011, 307f.). . . . . . . . . . . . . . . . . . . .

2.7

57

Structured Parallel Coordinates visualization of modal meanings in
scientific texts in the 1970/80s and 2000s. Figure taken from Lyding
et al. (2012, 47). . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

2.8

58

High-resolution screenshot of the World’s Languages Explorer showing the main components. Left: sunburst visualization of linguistic
features, top-right: a node-link diagram displaying the language genealogy of the languages in question, bottom-right: a map providing
information about the geographical distribution of the respective languages. The example at hand consists of 19 language features for
27 Indo-European languages that were automatically extracted from
parallel Bible texts. Figure taken from Rohrdantz et al. (2012a). . .

xvii

59


3.1 C-structure and f-structure representation of *Simon ate cake*. 68

4.1 Event class annotation of sentences containing a dative subject predicate in IcePaHC. 131

4.2 Different glyph representations for texts from IcePaHC with respect to dative subjects, verb classes and voice. 135

4.3 Scale for the quantification of text feature deviations. 136

4.4 The sentences which are involved in the calculation of a feature interaction can be accessed via mouse-over in the glyph visualization. 137

4.5 Glyph representations for each text from IcePaHC are positioned among each other on the vertical axis based on their diachronic order and are aligned on the horizontal axis with respect to their genre. Genre labels are shown on top and can be read as columns (SCI scientific texts, NAR narratives, REL religious texts, LAW law texts, and BIO biographies). 138

4.6 Glyph visualization in the temporal comparison layout. The tick-mark indicates the relative occurrence frequency of the selected/hoovered feature which enables the temporal comparison of the feature value. Moreover, the temporal comparison layout allows for the emergence of visual patterns showing diachronic differences, see, e.g., the patterns A and B. 139

4.7 Glyph visualization in the temporal comparison layout showing detailed verb classes. 142

4.8 Glyph visualization in the temporal comparison layout expanded for voice showing event classes (left) and more detailed verb classes (right) in texts shortly before and after 1900. 144

5.1 Iterative workflow of HistoBankVis. 158

5.2 Filtering component of HistoBankVis. 159

5.3 Compact matrix visualization showing differences between ranges via $\chi^2$ (left) or Euclidean distance (middle) and a bar chart displaying record counts per range (right). 161

5.4 Difference histograms for the distribution of subject case and word order in transitive sentences from IcePaHC. 163

5.5 Structured Parallel Coordinates visualization of co-occurrence frequencies developed by Culy et al. (2011b). 164
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.6</td>
<td>Dimension interaction for voice and word order in transitive sentences with a dative subject from 1150–1349</td>
<td>165</td>
</tr>
<tr>
<td>5.7</td>
<td>Dimension interactions for subject case and word order from 1150–1349 (top) and 1900–2008 (bottom)</td>
<td>167</td>
</tr>
<tr>
<td>5.8</td>
<td>Dimension interaction for voice and word order in transitive sentences with a dative subject from 1750–1899 (top) and 1900–2008 (bottom)</td>
<td>169</td>
</tr>
<tr>
<td>5.9</td>
<td>Difference histograms for subject case and position in IcePaHC.</td>
<td>171</td>
</tr>
<tr>
<td>5.10</td>
<td>Dimension interaction for subject case and position from 1150–1349 (left) and 1900–2008 (right)</td>
<td>172</td>
</tr>
<tr>
<td>5.11</td>
<td>Difference histograms for subject case and V1 in IcePaHC.</td>
<td>175</td>
</tr>
<tr>
<td>5.12</td>
<td>Dimension interactions for subject case and V1 from 1750–1899 (top) and 1900–2008 (bottom)</td>
<td>176</td>
</tr>
</tbody>
</table>
List of Abbreviations

In this thesis, the morphological glossing of examples generally follows the Leipzig Glossing Rules (Lehmann 1982, Croft 2003). However, unless otherwise stated, I retained the original gloss of examples taken from other authors. Additionally, the abbreviations listed in the following are used in this thesis:

AVM  Attribute-value matrix
CSD  Coordinate Subject Deletion
EXPL Expletive
GF   Grammatical function
HistLingVis Visual Analytics for historical linguistics
IcePaHC Icelandic Parsed Historical Corpus
INIT Initiator
LFG  Lexical-Functional Grammar
MID Middle morphology
OBJθ/2OBJ Secondary object
OBL/OBLθ Oblique
RES Resultee
RH   Rheme
sby  somebody
SF   Semantic Form
SPC  Structured Parallel Coordinates
sth  something
SUBJ Subject
UND  Undergoer
V1   verb-first
V2   verb-second
XCOMP Open clausal complement
∅EXPL Empty expletive
∅PRO Empty pronoun
Chapter 1

Introduction

Case is generally acknowledged to be a system which establishes the semantic relation which holds between arguments of a clause and the syntactic relationship they bear to their heads (Blake 2001, Butt 2006). Already in ancient times, at least going back to the Indian grammar tradition (6th century BCE), case was interpreted as expressing thematic roles such as actor and undergoer, indicating the relationships which hold between arguments of a clause (Butt and King 2004, Butt 2006). Modern theories of syntax preserved the view that case serves as the expressive means of thematic roles and moreover understand grammatical relations to be mediating between the lexical semantics of a verb and the case marking of its arguments (see, e.g., Fillmore 1968 and, for an overview of theories, Butt 2006). In the Government-Binding grammar tradition, case is divided into structural Case and inherent Case (Chomsky 1981): Structural Case is assigned configurationally and strictly binds nominative case to subjects and accusative to objects in accusative languages, whereas inherent Case is linked to particular thematic roles. There are moreover instances of quirky case marking, stipulated idiosyncratically by individual lexical items, which are not motivated by thematic roles (Butt 2006).

Cross-linguistically, languages tend to use either word order, case and/or agreement to mark grammatical relations (see, e.g., Kiparsky 1987, 1988, 1997). Icelandic, the most conservative Germanic language, is atypical in this respect, employing a fixed word order while having a rich case morphology together with complex agreement patterns. The Icelandic case system is moreover exceptional in that non-nominative subjects, and in particular dative subjects, are common in the language (Andrews 1976, Zaenen et al. 1985). Although often referred to as ‘quirky’ subjects,
Chapter 1. Introduction

Dative subjects are systematically associated with experiencer semantics in Icelandic (Smith 1996, Jónsson 2003, Barðdal 2011). In general, the Icelandic dative case correlates regularly with goal roles, including experiencers, recipients and beneficiaries, and moreover occurs frequently on theme objects of motion verbs (Maling 2001, 2002, Jónsson 2003). Yet, dative case assignment cannot entirely be motivated on the basis of thematic roles in Icelandic. For one, under middle formation, dative goal arguments retain their case marking, whereas dative theme objects are realized as nominative subjects instead. Moreover, while dative subjects tend to be experiencers, experiencer subjects are most often marked nominative in Icelandic (e.g., see Smith 1994, 1996, Jónsson 2003). Given that dative case assignment in Icelandic is neither structural nor truly inherent, i.e., lexical, nor purely idiosyncratic, Svenonius (2002, 2006) opts for an alternative explanation of dative case marking in Icelandic and proposes that event structure is the key component of the Icelandic case marking system.

In this thesis, I investigate the interaction between dative subjects, lexical semantics, event structure, voice, and word order in a thorough corpus study, using the Icelandic Parsed Historical Corpus (IcePaHC, Wallenberg et al. 2011), in order to shed more light on the function of case marking in the complex system which licenses grammatical relations in Icelandic. This thesis moreover contributes to the on-going discussion on whether dative subjects are inherited from a common Proto-Indo-European language stage or are in fact a more recent historical innovation. The so-called ‘Oblique Subject Hypothesis’ (see, e.g., Barðdal and Eythórsson 2003, 2009, Barðdal et al. 2012) takes dative subjects to be a common Proto-Indo-European feature, mainly drawing on the continuous existence of a monolithic dative subject construction throughout the Icelandic diachrony. The Oblique Subject Hypothesis challenges the more traditional ‘Object-to-Subject Hypothesis’ (e.g., Cole et al. 1980, Haspelmath 2001) which generally assumes that dative subjects are the result of the gradual reanalysis of former objects. Evidence for the Object-to-Subject Hypothesis can be found in Indo-Aryan, a related branch of Indo-European. While no evidence for the existence of dative subjects can be adduced for Old Indo-Aryan (Hock 1990), former objects were gradually reanalyzed as dative subjects during the New Indo-Aryan period (Deo 2003, Butt and Deo 2013), a process connected to lexical semantic changes of individual verbs, even after the original case system of Old Indo-Aryan eroded away during the Middle Indo-Aryan period. The Indo-Aryan evidence furthermore contests the validity of Icelandic as evidence for the Oblique
Subject Hypothesis, given that the Icelandic attestation only goes back to the 12th century which is about when dative subjects emerged in Indo-Aryan.

The methods employed for data analysis in the corpus investigations are a combination of standard corpus linguistic methods with novel methods coming from the field of Visual Analytics (Keim et al. 2008). The linguistic relevance of the findings obtained by means of the visualizations used in the corpus study particularly highlights the potential which Visual Analytics has for historical linguistic research, being able to account for various complex interactions of linguistic features contained in the data, regardless of whether the interactions were previously anticipated or unknown.

The overall picture which emerges from the corpus study is that the distribution of dative subjects has been changing over time. This argues against the inheritance of a monolithic dative subject construction from an earlier language stage, questioning the Oblique Subject Hypothesis. For one, the systematic association between dative subjects and experiencer semantics is increasing over time. However, this increase cannot be correlated with Dative Substitution as has been suggested in the literature (see, e.g., Barðdal 2011), but correlates with an increase of dative subjects together with verbs carrying middle morphology. I show that the middle forms in question have been lexicalized as stative experiencer predicates with a dative subject over time, with the diachronic reanalysis being effectuated by middle formation. Moreover, the data provides evidence for the diachronic development of a designated, clause-initial topic position in SpecIP which led to the rise of positional licensing in the history of Icelandic (in line with Kiparsky 1997). Dative subjects consistently lag behind in being realized in a particular position, which indicates that they were not licensed as subjects straightaway.

In order to provide a theoretical analysis of my empirical findings, I propose a novel linking theory which is based on the Lexical Mapping Theory, the linking theory established within the Lexical-Functional Grammar (LFG) framework (see, e.g., Bresnan and Kanerva 1989), drawing on Zaenen’s (1993) enhancement of the theory with respect to lexical semantics and Kibort’s extension with respect to argument positions. Accounting for the role of event structure in linking case to grammatical relations in Icelandic, I draw on Svenonius’ (2002) insights, but incorporate event structure in the form of the event participants proposed by Ramchand (2008) in her first-phase syntax approach to event decomposition, which allows for a more sophisticated analysis of the interaction between event structure and case. I furthermore assume a reference frame as per the FIGURE-GROUND division by Talmy (1978) in my
Chapter 1. Introduction

linking theory which represents position and information structure. Moreover, from a
diachronic perspective, I build on Kiparsky’s (1997) insights that positional licensing
can become dominant in a language over time, overriding the inherent features
associated with dative case marking in Icelandic.

This thesis shows that case marking in Icelandic is part of a complex system in
which event structure, lexical semantics, and word order interact with one another in
the licensing of grammatical relations via information structure, tying grammatical
relations to specific positions and particular lexical semantics. Moreover, I show that
the licensing conditions for dative subjects have been changing in the history of Ice-
landic which increasingly associates dative subjects with particular event structural
configurations. With stative experiencer predicates and middles of ditransitive verbs
with a goal argument, dative subjects are realized as state holders and denote the
figure argument, while a nominative rheme object refers to the ground. These
constructions are increasingly used in the corpus in the period post-1900, driving
the overall diachronic increase of dative subjects observed in IcePaHC.

The thesis proceeds as follows: Chapter 2 provides the crucial background in-
f ormation for this thesis, introducing the basic syntactic features of Icelandic, non-
canonical case marking and dative subjects. This includes a description of the gram-
matical properties of dative subjects in Modern Icelandic and discussions of the
Object-to-Subject Hypothesis and the Oblique Subject Hypothesis. Moreover, the
lexical semantic conditions which were associated with dative subjects in Icelandic
by the previous literature are introduced and Svenonius’ (2002) event structural ap-
proach to case marking is presented. The last part of the chapter introduces IcePaHC
and Visual Analytics for historical linguistics (HistLingVis).

Chapter 3 introduces the linking theories which are relevant for the linking theory
developed in this thesis. Therefore, LFG’s standard Lexical Mapping Theory and the
enhancements proposed by Zaenen (1993) and Kibort (2014) with respect to lexical
semantics and argument positions are introduced in this chapter. Moreover, I detail
Kiparsky’s linking theory (Kiparsky 1997) which accounts for the diachronic trade-
off between word order and case morphology for licensing grammatical relations.
Chapter 3 ends with an introduction of the first-phase syntax approach by Ramchand
(2008) which is employed for the analysis of event structure in this thesis.

Chapters 4 and 5 present the corpus investigations conducted in the thesis. In
Chapter 4, I investigate the diachronic interrelation between dative subjects, voice,
lexical semantics and event structure in IcePaHC using the glyph visualization de-
developed by Schätzle and Sacha (2016). Chapter 5 presents a corpus study examining the interaction between subject case and word order in IcePaHC which employs the HistoBankVis system (Schätzle et al. 2017) for data analysis.

Chapter 6 provides the theoretical analysis of the empirical findings from the corpus studies presented in Chapters 4 and 5. After introducing the linking theory developed for the analysis, I show how case interacts with event structure in the system which licenses grammatical relations in Icelandic. In particular, I account for the increase of dative subjects with stative experiencer predicates and the diachronic reanalysis of dynamic process verbs as dative subject predicates via middle formation. Moreover, I provide an analysis of the licensing of dative case on goal and theme objects, accounting for the preservation of dative case on goal arguments under middle formation on the basis of event structure. In Chapter 7, I conclude the thesis.
Chapter 2

Background

2.1 Introduction

This chapter introduces the relevant background information for this thesis. First, an introduction to Icelandic and the language’s basic syntactic features is given. Icelandic syntax represents an interesting object of study as Icelandic has both a fixed word order and a rich case morphology. This is cross-linguistically atypical. In this thesis, I thoroughly investigate factors conditioning dative case marking on subjects in the history of Icelandic to provide a window of understanding of the correlation between word order and case marking. I therefore begin with an introduction to non-canonical case marking in Section 2.3 which is followed by a description of the syntactic properties of dative subjects in Modern Icelandic in Section 2.4. Although the synchronic existence of dative subjects is generally acknowledged, their historical origin is highly controversial. The major point of debate concentrates on two competing hypotheses discussed in Section 2.5: the Object-to-Subject Hypothesis and the Oblique Subject Hypothesis. Next, the lexical semantic conditions and thematic roles which correlate with dative subjects in Icelandic are presented in Section 2.6. This is followed by an introduction to the event semantics approach to dative case marking in Icelandic taken by Svenonius (2002) in Section 2.7, which paves the way for the theoretical analysis of the data presented in this thesis. The historical data adduced in this thesis stems from the Icelandic Parsed Historical Corpus (Wallenberg et al. 2011), which is described in Section 2.8. Finally, Section 2.9 introduces the concept of Visual Analytics for historical linguistics (HistLingVis), presenting the methods employed for data analysis in this thesis.
2.2 Icelandic

Icelandic is a Northern Germanic language which is spoken by the approximately 330,000 inhabitants of Iceland. The ancestor language of Icelandic is Old Norse, brought to the island around the late ninth and early tenth century by the first settlers coming from Norway (Thráinsson 1994). The diachrony of Icelandic is generally divided into two major time periods, Old Icelandic and Modern Icelandic, which describe periods before and after 1540 CE respectively, with 1540 designating the publication year of the first Icelandic translation of the New Testament (Thráinsson 1994). From a linguistic perspective however, these two periods are often assumed to only differ marginally from one another. Due to Iceland’s geographical isolation, a strong literary tradition and a comparably small population, Icelandic has changed considerably less than the other Scandinavian languages and is generally considered to be the most conservative Germanic language. The First Grammatical Treatise (“Fyrsta málfræðirútgerðin”) is an Old Icelandic manuscript which provides a detailed description of the Old Icelandic sound system. The descriptions in this manuscript offer valuable clues about several phonological changes in Icelandic, for example the loss of distinctive vowel length, nasality and the diphthongization of long vowels (Thráinsson 1994).

Icelandic has maintained a rich case morphology over the centuries and distinguishes between four different cases: nominative, accusative, dative and genitive. Case endings differ across the three gender classes masculine, feminine and neuter and can be further divided into two inflectional classes, weak and strong. Nouns belong into the weak inflectional class if all of their singular inflections end in a vowel. Strong nouns in turn have their genitive singular as well as nominative plural inflection ending in a consonant. Table 2.1 shows examples of Icelandic nouns and their corresponding inflectional paradigms classified into weak and strong across the three genders.

The default case marking pattern in Icelandic is nominative on subjects, accusative on direct objects and dative case on indirect objects, while genitives typically mark possession (see, e.g. Thráinsson 2007). Example (1) shows the default case marking for transitive predicates, i.e., nominative on the subject and accusative on the object, and a prototypical genitive possessor. An example for the default case marking pattern for ditransitives is given in (2). The default case marking for ditransitives is nominative case on the subject, dative on the indirect object and accusative
on the direct object. Furthermore, the finite verb agrees with the nominative subject in person and number.

(1) Anna las bókina Jóns.
Anna.NOM read.pst.3sg book.the.acc John.gen
‘Anna read John’s book.’

(2) Anna gaf Jóni bókina.
Anna.NOM give.pst.3sg John.dat book.the.acc
‘Anna gave John the book.’

<table>
<thead>
<tr>
<th></th>
<th>Masculine</th>
<th>Feminine</th>
<th>Neuter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong 1</td>
<td>‘horse’</td>
<td>‘time’</td>
<td>‘table’</td>
</tr>
<tr>
<td>Weak 1</td>
<td>‘needle’</td>
<td>‘tongue’</td>
<td>‘eye’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom.</td>
<td>hest-ur</td>
<td>hest</td>
<td>hest-i</td>
<td>hest-s</td>
</tr>
<tr>
<td>Acc.</td>
<td>tím-i</td>
<td>tím-a</td>
<td>tím-a</td>
<td>tím-a</td>
</tr>
<tr>
<td>Dat.</td>
<td>nál</td>
<td>nál</td>
<td>nál-ar</td>
<td>nál-ar</td>
</tr>
<tr>
<td>Gen.</td>
<td>tung-a</td>
<td>tung-u</td>
<td>tung-u</td>
<td>tung-u</td>
</tr>
<tr>
<td>Plural</td>
<td>borð</td>
<td>borð-i</td>
<td>borð-s</td>
<td>aug-a</td>
</tr>
<tr>
<td>Nom.</td>
<td>hest-ar</td>
<td>hest-a</td>
<td>hest-um</td>
<td>hest-a</td>
</tr>
<tr>
<td>Acc.</td>
<td>tím-ar</td>
<td>tím-a</td>
<td>tím-um</td>
<td>tím-a</td>
</tr>
<tr>
<td>Dat.</td>
<td>nál-ar</td>
<td>nál-ar</td>
<td>nál-um</td>
<td>nál-a</td>
</tr>
<tr>
<td>Gen.</td>
<td>tung-ur</td>
<td>tung-ur</td>
<td>tung-um</td>
<td>tung-na</td>
</tr>
<tr>
<td></td>
<td>borð</td>
<td>borð</td>
<td>borð-um</td>
<td>aug-um</td>
</tr>
<tr>
<td></td>
<td>aug-u</td>
<td>aug-a</td>
<td>aug-um</td>
<td>aug-na</td>
</tr>
</tbody>
</table>

Table 2.1: Inflectional paradigms of Icelandic nouns (Thráinsson 1994, 153).

There is no one-to-one relationship between morphological case and grammatical relations in Icelandic (Thráinsson 2007). While nominative is the canonical case for subjects, Icelandic exhibits a fairly large amount of predicates that take non-nominative subjects, with in particular the synchronic existence of dative subjects being well-established (Andrews 1976, Zaenen et al. 1985; see also Section 2.4 for more details) and their diachronic continuity being widely accepted (see, e.g., Rögnvaldsson 1991, Barðdal and Eythórsson 2003). For example, (3) shows a sentence in which the verb batna ‘recover, get better’ takes a dative subject in combination with a nominative object.

(3) Barninu batnáði veikin.
the-child (DAT) bettered (NOM) bettered the-disease (NOM)
‘The child recovered from the disease’ (Andrews 1990, 169)
Chapter 2. Background

While accusative is the standard object case, nominative is the default case for objects co-occurring with non-nominative subjects (Thráinsson 2007). Finite verbs do not agree with a non-nominative subject in Icelandic, but occur in the third person singular instead. Agreement in number with a nominative object is generally possible, see (4-b), but not obligatory, see (4-a).

(4) a. Barninu batnaði veiknar.
    the.child.DAT better.PST.3SG the.diseases.NOM
    ‘The child recovered from the diseases.’

b. Barninu bótnuðu veiknar.
    the.child.DAT better.PST.3PL the.diseases.NOM
    ‘The child recovered from the diseases.’

Apart from dative subjects and the standard nominative subject case-marking, Icelandic also employs accusative and genitive subjects, albeit to a lesser extent. Barðdal (2011) counts about 700 different verb types which take a dative subject in Icelandic, but only 180-200 verbs with accusative subjects and merely 10-15 different verb types exist that can have a genitive subject. Examples for an accusative and a genitive subject are given in (5-a) and (5-b) respectively. As with dative subjects, agreement is either neutral (third person singular) or optional with a nominative object.

(5) a. Skessuna vant mat.
    the-giantess (ACC) lacks food (ACC)
    ‘The giantess lacks food.’

b. Vindsins gætir ekki.
    the.wind (GEN) matters not
    ‘The wind does not matter, is negligible.’

(Andrews 1990, 169)

Example (5-a) additionally shows that accusative objects can co-occur with accusative subjects. With nominative subjects, dative and genitive objects are also possible, but are more marked and occur less often than accusative objects (Thráinsson 1994). Thráinsson (2007, 167) summarizes the possible case patterns for Icelandic transitive verbs as given in (6) where the non-existing combinations are marked by an asterisk and an overstrike. N, A, D, G stand for nominative, accusative, dative and genitive respectively. The very rare or exceptional patterns are enclosed in parentheses.
Ditransitive predicates always have a nominative subject and most often an accusative direct object in combination with an indirect, dative marked, object (NDA) as was shown in (2). Other possible, but less frequent, ditransitive patterns are NAD, NDG, NDD, NAG and NAA (cf. Thráinsson 2007).

Due to the variety of case marking patterns, identifying grammatical relations on the basis of case marking alone is difficult in Icelandic. Thráinsson (1994) concludes that word order plays a more important role in identifying grammatical relations: noun phrases with subject properties always occur leftmost and indirect objects precede direct ones, at least in the unmarked, default word order. The basic word order pattern in finite declarative clauses is SVO (Subject-Verb-Object) in Icelandic (Zaenen et al. 1985, Thráinsson 2007). Icelandic moreover is a V2 (verb-second) language which allows for maximally one constituent in the prefinite position (see Røgnvaldsson and Thráinsson 1990, Røgnvaldsson 1996), see (7).

(7) a. Ég gleymdí þeim fljótt.
   I.NOM forget.PST.1SG they.DAT quickly
   ‘I quickly forgot them.’

b. Peim gleymdí ég fljótt.
   they.DAT forget.PST.1SG I.NOM quickly
   ‘Them I quickly forgot.’

c. Fljótt gleymdí ég þeim.
   quickly forget.PST.1SG I.NOM they.DAT
   ‘Quickly I forgot them.’

d. *Peim ég gleymdí fljótt.
   they.DAT I.NOM quickly forget.PST.1SG
   ‘Them I quickly forgot.’
Whenever a non-subject is fronted to initial sentence position, the subject has to immediately follow the verb, which remains in second position, as shown in example (7-b) in which the object from (7-a) is fronted.\(^3\)

V1 (verb first) word order is generally known to signal yes/no questions or imperatives in Germanic, see the Icelandic examples in (8) and (9) respectively. Icelandic is exceptional in that it regularly allows for V1 constructions in matrix declarative sentences with the subject immediately following the verb, see (10), a construction typical for older stages of Germanic, but uncommon in the modern languages (see, e.g., Axel 2007 on V1 in Old German).

(8) Hefur Anna lesið bókina?
  have.PRS.3SG Anna.NOM read.PST.PTCP book.the.NOM
  ‘Has Anna read the book?’

(9) Far-ðu burt!
  go.IMP-you away
  ‘Go away!’

(10) Kom Ólafur seint heim.
  came Olaf late home.
  ‘Olaf came home late.’ (Sigurðsson 1990, 41)

In sum, Icelandic is a language with a number of different word order patterns. Yet, the word order variations in Icelandic are structured, adhering to specific rules, which results in a fairly fixed word order. Atypically for a language with a fixed word order, Icelandic sports a rich morphological case system in combination with verbal agreement. Crosslinguistically, languages tend to use either word order, case and/or agreement to mark grammatical relations (Kiparsky 1987, 1988, 1997). In this respect, Icelandic constitutes an interesting object of study, given that case seems to be somehow redundant in determining grammatical relations. This thesis investigates factors conditioning the occurrence of dative subjects in Icelandic from a diachronic as well as synchronic perspective to provide an understanding of the evidently complex system which associates arguments with grammatical relations and the function of case marking therein. The next section presents an overview about canonical case alignment patterns, providing fundamental knowledge on non-canonical subject case marking.

\(^3\)A restricted class of adverbs (e.g., auðvitað ‘obviously’ and líklega/sennilega ‘probably’) can be placed between the clause-initial subject and the finite verb, resulting in a V3 order (cf. Thráinsson 2007).
2.3 Non-canonical case marking

Several Indo-European languages exhibit non-canonical case marking of their core arguments. Dixon defines the core arguments A, S, and O as the universal semantic-syntactic primitives inherent to all languages which “serve as intermediaries between meaning and grammatical marking” (Dixon 1994, 25). S denotes the single core argument of an intransitive clause, i.e., the subject. A and O are the two core arguments of a transitive clause, with A designating the subject and O the object of a transitive predicate. Moreover, there is a general tendency for A and O to be assigned to a verb’s arguments based on the lexical semantic content of the verb (Dixon 1994). The initiator or controller of the activity expressed by a verb, i.e., the ‘agent’ role, is prototypically mapped onto the A function. O typically denotes the participant affected by the activity, i.e., the ‘patient’ role (Dixon 1994).

Case-marking languages can, in most instances, be divided into two patterns with respect to the relational organization of the three basic syntactic relations A, S, and O, see (11), taken from Dixon (1994, 9).

\[
\begin{align*}
\text{nominative} & \quad \begin{cases} 
A \quad \text{ergative} \\
S 
\end{cases} \\
\text{absolutive} & \quad \begin{cases} 
O 
\end{cases}
\end{align*}
\]

Accusative languages mark A and S with nominative case as opposed to the O argument which usually receives accusative case marking. Examples for case marking in German, an accusative language, are given in (12), where the subject is always nominative. Moreover, the transitive object in (12-a) is marked accusative. Nominative case is often unmarked in accusative languages and contrasts with accusative which is marked overtly.

\[
\begin{align*}
(12) & \quad \begin{align*}
a. & \quad \text{Der Junge isst Kuchen.} \\
& \quad \text{the.NOM boy.NOM eat.PRS.3SG cake.ACC} \\
& \quad \text{‘The boy eats cake.’} \\
b. & \quad \text{Der Junge schwimmt.} \\
& \quad \text{the.NOM boy.NOM swim.PRS.3SG} \\
& \quad \text{‘The boy swims.’}
\end{align*}
\]

Ergative languages on the other hand mark S and O with absolutive case, but mark A differently via ergative case (see also Comrie 1989 and Plank 1979 on ergativity). In ergative languages, absolutive case is usually unmarked. Dyirbal, a Pama-Nyungan language spoken in Australia (Dixon 1979, 1994), is an ergative language. Examples for Dyribal are given in (13). Sentences (13-a) and (13-b) both have absolutive marked intransitive subjects. The transitive example from (13-c) contains the unmarked (absolutive) object ŋuma ‘father’, while the subject yabu ‘mother’ receives ergative marking via the case clitic -ŋgu. Example (13-d) contrasts with this by marking yabu and ŋuma inversely: In (13-d), yabu corresponds to the O function and remains unmarked, while ŋuma receives ergative marking (ŋuma-ŋgu) and corresponds to A.

(13) a. ŋuma banaga-nŋu
    father.ABS retrun-NONFUT
    ‘Father returned.’

b. yabu banaga-nŋu
    mother.ABS retrun-NONFUT
    ‘Mother returned.’

c. ŋuma yabu-ŋgu bura-n
    father.ABS mother-ERG see-NONFUT
    ‘Mother saw father.’

d. yabu ŋuma-ŋgu bura-n
    mother.ABS father-ERG see-NONFUT
    ‘Father saw mother.’

(Dixon 1994, p.10)

There are a number of languages which have mixed alignment systems and show both ergative and accusative alignments, each under certain circumstances. Alignment splits are conditioned by several different factors, e.g., lexical semantics, tense, aspect or mood, or the grammatical status of a clause (cf. Dixon 1994). Marathi, an Indo-Aryan language, for example is a split-ergative language which shows ergative alignment when the verb of the clause has perfective morphology, see (14-a), and accusative alignment in imperfective clauses, see (14-b).\textsuperscript{4}

\textsuperscript{4}In Marathi, nominative case designates the unmarked case in both alignments and thus the nominative object in sentence (14-a) is conceptually equivalent to an absolutive object.
2.3. Non-canonical case marking

(14) a. simhā=ne hattī mār-lā
   lion.OBL=M.ERG elephant.M.NOM kill-PRF.M.3SG
   ‘The lion killed an elephant.’

b. simha hattī=lā mār-to.
   lion.M.NOM elephant.OBL=M.ACC kill-PRS.M.3SG
   ‘The lion kills the elephant.’

Some languages deviate from their canonical alignment patterns in that they mark A/S and O differently with respect to the transitivity status of the clause involved (Hopper and Thompson 1980). Low transitivity surfaces for example in clauses which describe non-volitional or stative events that have a non-affected object, or none at all. That is, non-canonical case marking typically arises in clauses with a low transitivity that do neither have a prototypical agent nor a prototypical patient. Transitivity furthermore refers to the degree of individuation of NPs. Some languages make use of non-canonical case marking to indicate the status of referentiality, specificity and/or animacy of subjects and objects. For instance, Urdu/Hindi, another Indo-Aryan language with a split-ergative system conditioned by verbal aspect, deviates from the canonical alignments by showing alternation patterns in which a difference in case marking on one of the NPs signals a difference in semantic interpretation between otherwise identical sentences (see Butt and King 1991, 2004). These alternations may occur on subjects as well as objects in the form of Differential Case Marking.

Differential Object Marking is conditioned by the specificity or referentiality of the NP in Urdu/Hindi. For example, the sentences in (15) differ in that sentence (a) has an unmarked (nominative) object NP which yields a generic interpretation of the noun ‘giraffe’ in contrast to (b) where accusative case indicates reference to a particular giraffe.

(15) a. nadya=ne jiraf dek-h-na hε
    Nadja.F.Sg=Erg giraffe.M.Sg.Nom see-Inf.M.Sg be.Pres.3.Sg
    ‘Nadja wants to see a giraffe/giraffes.’

b. nadya=ne jiraf=ko dek-h-na hε
    Nadja.F.Sg=Erg giraffe.M.Sg=Acc see-Inf.M.Sg be.Pres.3.Sg
    ‘Nadja wants to see the giraffe.’

(Butt and King 2004, 3)
An example for Differential Subject Marking in Urdu/Hindi is given in (16), where the subject in (16-b) is marked with dative case, deviating from the canonical ergative case marking. This difference has an impact on the semantics of the clause: The ergative subject gives rise to a reading in which the participant has control over the action, while the dative subject expresses the obligation of the participant to perform the action (Butt and King 2004).

(16) 
\[
\text{(16a) } \quad \text{nadya=ne} \quad \text{zu} \quad \text{ja-na} \quad \text{h} \\
\quad \text{Nadja.F.Sg=Erg} \quad \text{zoo.M.Sg.Obl} \quad \text{go-Inf.M.Sg} \quad \text{be.Pres.3.Sg} \\
\quad \text{‘Nadja wants to go to the zoo.’} \\
\text{(16b) } \quad \text{nadya=ko} \quad \text{zu} \quad \text{ja-na} \quad \text{h} \\
\quad \text{Nadja.F.Sg=Dat} \quad \text{zoo.M.Sg.Obl} \quad \text{go-Inf.M.Sg} \quad \text{be.Pres.3.Sg} \\
\quad \text{‘Nadja has to go to the zoo.’} \\
\]

(Butt and King 2004, 2)

Non-canonical subjects, and in particular dative subjects, exist in a multitude of Indo-European languages, including, for example, Icelandic and Marathi besides Urdu/Hindi. Icelandic is particularly famous for having non-canonical subjects, i.e., non-nominative or so-called ‘oblique’ subjects, with the synchronic existence of dative subjects being well established (cf. Andrews 1976, Zaenen et al. 1985). The next section details the grammatical properties of non-nominative subjects in Modern Icelandic, distinguishing dative subjects from objects.

### 2.4 Dative subjects in Modern Icelandic

Building on seminal work by Andrews (1976, 1982a,b), Thráinsson (1979), Maling (1980) and Zaenen (1985), Zaenen et al. (1985) were among the first to provide an exhaustive set of tests which distinguish subjects from topicalized objects in Modern Icelandic. As it turned out, non-nominative subjects, including dative subjects, pattern with nominative subjects regarding the following properties: raising, reflexivization, topicalization and subject-verb inversion, extraction, indefinite subject postponing, subject ellipsis, and infinitive complements. These properties and the corresponding illustrative examples as given by Zaenen et al. (1985) are presented below.

**Raising** In Icelandic, embedded subjects can be raised into matrix object position, see example (17-b), in which the subject of the embedded verb *sakna* ‘miss’ is raised
2.4. Dative subjects in Modern Icelandic

into the object position of the matrix verb *telja* ‘believe’. The verb *sakna* generally takes a nominative subject and a genitive object as in (17-a), while *telja* has a nominative subject and an accusative object. Thus, the subject NP of *sakna* receives accusative marking in (17-b) as it is raised to the matrix object position.\(^5\) In example (17-c), the genitive object is topicalized, i.e., placed into the preverbal position. However, the topicalized constituent (the object) does not undergo raising in (17-d).

\[(17)\]
\[
a. \text{Guðrún saknar Haraldar.} \\
\text{Guðrún.NOM miss.PRS.3SG Harald.GEN} \\
\text{‘Guðrún misses Harald.’}
\]
\[
b. \text{Ég taldi Guðrúnu í barnaskap mínum} \\
\text{L.NOM believe.PST.1SG Guðrún.ACC in foolishness.DAT my.DAT} \\
sakna Haraldar. \\
\text{miss.INF Harald.GEN} \\
\text{‘I believed Guðrún to miss Harald in my foolishness.’}
\]
\[
c. \text{Haraldar saknar Guðrún.} \\
\text{Harald.GEN miss.PRS Guðrún.NOM} \\
\text{‘Harald, Guðrún misses.’}
\]
\[
d. *\text{Ég taldi} \{\text{Haraldar/Harald}\} \text{ sakna} \\
\text{L.NOM believe.PST.1SG \{Harald.GEN/Harald.ACC\} miss.INF} \\
\{\text{Guðrún/Guðrúnu}\}. \\
\{\text{Guðrún.NOM/Guðrún.ACC}\} \\
\text{‘I believe Harald, Guðrún to miss.’}
\]

\[(Zaenen\ et\ al.\ 1985,\ 448;\ gloss\ modified)\]

Unlike topicalized objects, oblique subjects can undergo raising, see (18) and (19) in which the dative subject of *þykja* ‘think’ is raised to the matrix object position, with the dative case being retained. Consequently, Zaenen et al. (1985) conclude that irrespective of the surface case-marking, all, but only, grammatical subjects can raise.

\[(18)\]
\[
\text{Henni hefur alltaf þótt Ólafur leiðinlegur.} \\
\text{she.DAT have.PRS.3SG always think.PTCP. PST Ólafur.NOM boring.NOM} \\
\text{‘She has always thought Olaf was boring.’}
\]

\[(Zaenen\ et\ al.\ 1985,\ 447;\ gloss\ modified)\]

\(^5\)The adverbial *í barnaskap mínum* ‘in my foolishness’ provides further evidence for *Guðrúnu* being raised into matrix object position as the adverbial belongs to the matrix clause and is situated between *Guðrúnu* and the infinitive complement *sakna Haraldar* (see Zaenen et al. 1985, 448).
Chapter 2. Background

(19) Ég tel henni hafa alltaf þótt
    I.NOM believe.PST.1SG she.DAT have.INF always think.PTCP.PST
    Ólafur leiðinlegur.
    Ólafur.NOM boring.NOM
‘I believe her to have always thought Olaf was boring.’
    (Zaenen et al. 1985, 449; gloss modified)

**Reflexivization** Zaenen et al. (1985) moreover found that grammatical subjects contrast with objects with respect to reflexivization in Icelandic. Grammatical subjects control obligatory reflexivization for all speakers of Icelandic, whereas objects at best control reflexives optionally. Zaenen et al. (1985) provide judgements for those speakers who allow subject-controlled reflexivization only. These judgements are given below, with boldface indicating the intended coreference.

(20) a. **Sigga** barði mig með dúkkunni
    Sigga.NOM hit.PRS.3SG I.ACC with doll.the.DAT
    sinni/*hennar.
    her.DAT/*she GEN([-REFL])
    ‘Sigga hit me with her doll.’

b. Êg barði **Siggu** með dúkkunni
    I.NOM hit.PRS.3SG Sigga.ACC with doll.the.DAT
    hennar/*sinni.
    she.GEN/*her DAT([+]REFL])
    ‘I hit Sigga with her doll.’

c. **Siggu** barði ég með dúkkunni
    Sigga.ACC hit.PRS.3SG I.NOM with doll.the.DAT
    hennar/*sinni.
    she.GEN/*her.DAT([+]REFL])
    ‘Sigga, I hit with her doll.’
    (Zaenen et al. 1985, 449f.; gloss modified)

The judgements in (20) show that nominative subjects trigger obligatory reflexivization, while objects do not. Moreover, dative subjects control reflexivization, see (21) and can thus not be topicalized objects.

(21) a. **Henni** þykir bróðir **sinn/**hennar
    she.DAT think.PRS.3SG brother.NOM her.NOM/*she GEN([-REFL])
    leiðinlegur.
    boring.NOM
    ‘She thinks her brother is boring.’
2.4. Dative subjects in Modern Icelandic

b. Hverjum þykir sinn fugl
   everyone.DAT think.PRS.3SG his.NOM(〔+REFL〕) bird.NOM
   fagur. (Proverb)
   beautiful.NOM
   ‘Everyone thinks his bird is beautiful.’
   (Zaenen et al. 1985, 450; gloss modified)

Topicalization and subject-verb inversion  Icelandic is a V2 language (see Section 2.2). Hence, whenever a non-subject constituent is fronted to initial position for topicalization, the verb occurs in second position with the subject immediately following it. In (22), the nominative subject Ólafur ‘Ólaf’ has to appear immediately after the verb for the sentences to be grammatical. In addition, once the object has been topicalized, no further constituent can be fronted.

(22) a. Refinn skaut Ólafur með þessari byssu.
    fox.the.ACC shoot.PST.3SG Ólafur.NOM with this.DAT shotgun.DAT
    ‘The fox, Ólafur shot with this shotgun.’
   b. *Með þessari byssu skaut refinn Ólafur.
      with this.DAT shotgun.DAT shoot.PST.3SG fox.the.ACC Ólafur.NOM
      ‘With his shotgun, the fox, Ólafur shot.’
      (Zaenen et al. 1985, 450; gloss modified)

Topicalization can also be applied to sentences with an oblique subject, see example (23) in which the dative subject inverts with the finite verb when something else (the object) is topicalized.

(23) Ólafur hefur henni alltaf þótt leiðinlegur.
    Ólafur.NOM have.PRS.3SG she.DAT always think.PTCP.PST boring.NOM
    ‘Ólafur, she has always thought was boring.’
    (Zaenen et al. 1985, 451; gloss modified)

Extraction  Most Icelandic speakers do not allow for topicalization in binding domains, e.g., in the extraction domain of a wh-question (Zaenen et al. 1985), see (24-d) which displays the ungrammatical topicalized version of (24-b). Yet, these speakers generally allow for topicalization in embedded clauses introduced by the complementizer að ‘that’, see (24-c) as opposed to (24-a).
(24) a. Jón telur að María hafi
   Jón.NOM believe.PRS.3SG that María.NOM have.SBJV.PRS.3SG
   kysst Harald í gær.
   kiss.PTCP.PST Harald.ACC yesterday
   'Jón believes that María has kissed Harald yesterday.'

b. Hvenær telur Jón að María
   when believe.PRS.3SG Jón.NOM that María.NOM
   hafi kysst Harald?
   have.SBJV.PRS.3SG kiss.PTCP.PST Harald.ACC
   'When believes Jón that María has kissed Harald?'

c. Jón telur að Harald hafi
   Jón.NOM believe.PRS.3SG that Harald.ACC have.SBJV.PRS.3SG
   María kysst í gær. (Topicalization)
   María.NOM kiss.PTCP.PST yesterday
   'Jón believes that Harald, María has kissed yesterday.'

d. *Hvenær telur Jón að Haraldi
   when believe.PRS.3SG Jón.NOM that Harald.ACC
   hafi María kysst?
   have.SBJV.PRS.3SG María.NOM kiss.PTCP.PST
   'When does Jon believe that Harold, Mary kissed?'

   (Zaenen et al. 1985, 451; gloss modified)

Once more, dative subjects differ from topicalized NPs in that they pattern with
nominative subjects, see (25).

(25) a. Jón telur að henni hafi alltaf
   Jón.NOM believe.PRS.3SG that she.DAT have.SBJV.PRS.3SG always
   þótt Ölafur leiðinlegur.
   think.PTCP.PST Ölafur.NOM boring.NOM
   'Jón believes that she has always thought Ólafur was boring.'

b. Hvenær telur Jón að henni hafi
   when believe.PRS.3SG Jón.NOM that she.DAT have.SBJV.PRS.3SG
   þótt Ölafur leiðinlegur?
   think.PTCP.PST Ölafur.NOM boring.NOM
   'When believes Jón that she has thought Ólafur was boring?'

c. Jón telur að Ólafur hafi
   Jón.NOM believe.PRS.3SG that Ölafur.NOM have.SBJV.PRS.3SG
   henni alltaf þótt leiðinlegur. (Topicalization)
   she.DAT always think.PTCP.PST boring.NOM
   'Jón believes that Ólafur, she has always thought was boring.'
2.4. Dative subjects in Modern Icelandic

Indefinite Subject Postposing  In Icelandic, indefinite subjects can be postposed, either when the expletive það is inserted into the clause-initial position, see (26-a), or when another constituent is topicalized. Indefinite subject postposing, however, is impossible when a non-subject is in first position, as shown in (26-c) and (26-d).

(26)  

a. það hefur þjófur stolið hjólinu mínú.  
expl have.prs.3sg thief.nom steal.pTCP.pst bicycle.dat my.dat  
‘There has a thief stolen my bicycle.’

b. Hjóli hefur þjófurinn stolið.  
bicycle.dat have.prs.3sg thief.nom steal.pTCP.pst  
‘A bicycle, the thief has stolen.’

c. *það hefur hjóli þjófurinn stolið.  
expl have.prs.3sg bicycle.dat thief.nom steal.pTCP.pst  
‘There has a bicycle, the thief stolen.’

d. *það hefur hjóli stolið þjófurinn.  
expl have.prs.3sg bicycle.dat steal.pTCP.pst thief.nom  
‘There has a bicycle, the thief stolen.’  

(Zaenen et al. 1985, 452; gloss modified)

Again, dative subjects pattern like nominative subjects and contrast with the topicalized constituents in (26), see (27).

(27)  

a. það hefur einherjum þótt Ólafur  
expl have.prs.3sg someone.dat think.pTCP.pst Ólafur.nom  
leíðinlegur.  
boring.nom  
‘There has someone thought Ólafur was boring.’

b. Ólafur hefur einherjum þótt leíðinlegur.  
Ólafur.nom have.prs.3sg someone.dat think.pTCP.pst boring.nom  
‘Ólafur, someone has thought was boring.’

d. *Hvenær telur Jón að Ólafur  
when believe.prs.3sg Jón.nom that Ólafur.nom  
hafi henni þótt leíðinlegur?  
have.sbjv.prs.3sg she.dat think.pTCP.pst boring.nom  
‘When believes Jón that Ólafur, she has thought was boring?’  

(Zaenen et al. 1985, 452; gloss modified)
c. *Það hefur Ólafur einherjum þótt
expl have.prs.3sg Ólafur.nom someone.dat think.ptcp.pst
leiðinlegur.
boring.nom
‘There has Ólafur, someone thought was boring.’
(Zaenen et al. 1985, 453; gloss modified)

**Subject Ellipsis** Subject ellipsis distinguishes between non-subjects and subjects in Modern Icelandic (Zaenen et al. 1985). Only subjects can be deleted in coordinated clauses under identity with the subject of the preceding main clause. Objects however do not allow for deletion under coordination, neither under identity with an object nor with a subject, see (28). Note that examples (28-c) and (28-d) are only grammatical if the verb *graða* carries the intransitive reading ‘to dig’ instead of the transitive reading ‘to bury’ (Zaenen et al. 1985).

(28) a. Þeir fluttu líkið og þeir grófu
they.nom move.pst.3pl corpse.the.acc and they.nom bury.pst.3pl
það.
it.acc
‘They moved the corpse and they buried it.’

b. Þeir fluttu líkið og _ grófu
they.nom move.pst.3pl corpse.the.acc and pro bury.pst.3pl
það. (Subj-Subj)
it.acc
‘They moved the corpse and buried it.’

c. ≠ Þeir fluttu líkið og þeir grófu
they.nom move.pst.3pl corpse.the.acc and they.nom bury.pst.3pl
og
‘They moved the corpse and digged.’

d. ≠ Líkið var flutt og þeir
corpse.the.nom be.pst.3sg move.ptcp.pass and they.nom
grófu
‘The corpse was moved and they digged.’
2.4. Dative subjects in Modern Icelandic

Dative subjects may also be deleted under identity with the subject of the preceding conjunct clause, even if the preceding subject carries nominative case, see (29).

(29) Hann segist vera duglegur, en _ finnst verkefnið of þungt. (Subj-Subj)
   ‘He says of himself to be diligent, but _ finds the homework too hard.’
   (Zaenen et al. 1985, 454; gloss modified)

Infinitive Complements In Icelandic, only subjects coreferent with the matrix subject can be deleted when they are embedded in an infinitive complement, i.e. only subjects are the target of EQUI-NP-Deletion (Zaenen et al. 1985). Furthermore, only subjects can be understood as arbitrary or anaphorically controlled pronouns and be not overtly realized at the same time. This holds for nominative as well as non-nominative subjects, see (30) and (31) respectively.

(30) a. Ég vonast til að fara heim.
   ‘I hope to go home.’

   b. Að fara heim snemma er óvenjulegt.
   ‘To go home early is unusual.’
   (Zaenen et al. 1985, 454; gloss modified)

(31) a. Mig vantar peninga.
   ‘I need money.’

   b. Ég vonast til _ að vanta ekki peninga.
   ‘I hope _ to not lack money.’
c. Að vanta peninga er allt af algengt.
   to lack.INF money.NOM be.PRS.3SG all.too common.NOM
   ‘To lack money is all too common.’

(Zaenen et al. 1985, 454f)

Summing up the results of these tests, non-nominative subjects, including dative subjects, syntactically pattern like nominative subjects and can clearly be distinguished from objects in Modern Icelandic. The only properties that distinguish them from regular, nominative subjects are their non-canonical case marking and their inability to trigger verb agreement. While the synchronic existence of dative subjects has been well-established for Icelandic, their historical origin constitutes a major point of debate, attracting a good deal of research in recent years. The debate which has evolved around the diachrony of dative subjects is discussed in the next section.

2.5 The diachrony of dative subjects

Although dative subjects are common in many modern Indo-European languages, how and when they came into existence has been the subject of fierce debates in recent years. These debates mainly concentrate on two competing narratives: (i) the so-called Oblique Subject or Semantic Alignment Hypothesis and (ii) the Object-to-Subject Hypothesis. The Oblique Subject Hypothesis takes dative subjects to be a common Proto-Indo-European inheritance (e.g., see Barðdal and Eythórsson 2009, Barðdal et al. 2012), while the Object-to-Subject Hypothesis assumes that dative subjects are a historical innovation. The latter is the more traditional hypothesis and generally understands dative subjects to be the result of the gradual reanalysis of former objects (e.g., Cole et al. 1980, Haspelmath 2001), a theory which has undergone major criticisms recently.

2.5.1 The Object-to-Subject Hypothesis

The Object-to-Subject Hypothesis is supported by empirical facts from Indo-Aryan. While there is no evidence for the existence of dative subjects in Old Indo-Aryan (Hock 1990), there is evidence for objects being reanalyzed as dative subjects during several later stages of Indo-Aryan (Deo 2003, Butt and Deo 2013). Moreover, the Old Indo-Aryan case system eroded away over the course of Middle Indo-Aryan which
led to the loss of the original case system. In the New Indo-Aryan period, new case markers were drawn into the system from around 1200 on (cf. Butt and Deo 2013), which coincides with the emergence of dative subjects.

Marathi as a New Indo-Aryan language exhibits a comparably large number of predicates taking a dative experiencer subject (Deo 2003). These predicates belong to the class of psychological predicates and include verbs of knowledge, belief, desire, perception, and mental or physical states (cf. Dhongde and Wali 2009), see, for example, the predicate āvād ‘like’ in (32) which occurs together with a dative subject and a nominative object.

(32) lili-la babu awād-t-o.
Lili-DAT Babu-NOM-MSG like-IMPF-MSG
‘Lili likes Babu.’ (Dhongde and Wali 2009, 187)

Deo (2003) shows that the emergence of dative subjects in Marathi correlates with an increasing systematic association of dative case with experiencer arguments. Marathi experiencer predicates are the result of lexical semantic shifts of individual verbs which led to changes in their argument structure from Old Marathi, i.e., early New Indo-Aryan, on. Concomitantly, the case marking of experiencers changed to dative case (Deo 2003). Deo (2003) moreover shows that these changes emerged from three distinct types of Sanskrit sources, which are reflected in the classes of modern Marathi experiencer predicates.

The first type of source are ‘know predicates’, i.e., verbs involving mental activities (Deo 2003). These predicates take a nominative experiencer subject and an accusative theme object in Sanskrit. From Old Marathi on, these verbs undergo a shift in argument realization which is extended verb by verb, creating a class of verbs which optionally take either dative or nominative subjects. The different argument realizations correlate with a lexical semantic change, e.g., from saṃ-jñā ‘know’ to samaj ‘understand’, giving rise to different readings: The more agentive experiencer reading is only available when the subject is marked nominative, see (33-a), otherwise the experiencer is non-agentive as in (33-b).

(33) a. kanyā pāṭha-m saṃjñā-ti
girl.F.SG.NOM lesson.M.SG-ACC know-PRS.3.SG
‘The girl knows the lesson.’ (Sanskrit)

---

6The glosses in examples (33) and (34) were revised by Ashwini Deo.
A second type of source is exemplified by a set of transitive causative verbs, the ‘burn predicates’, which have either a psych or a physical reading in Sanskrit (Deo 2003). These typically have a nominative marked cause subject and an accusative object experiencer. In Marathi, only the psych interpretation persists for the descending predicates with the former accusative patient object being reanalyzed as a dative experiencer subject. Examples (34-a) and (34-b) illustrate that the Sanskrit predicate *dah* has both, the non-psych reading ‘burn’ and the psych reading ‘torment’, while the Marathi version *dāj*, as given in (34-c), only retained the psych interpretation.

(34) a. na=ena-m daha-ti pāvaka-h.
   NEG=this.M.SG.ACC burn-IMPF.PRS.3.SG fire.M.SG.NOM
   ‘The fire does not burn him (the soul).’ [BG 2.23]
   (Sanskrit – non-psych)

b. haṁś-ānāṁ vacana-m yat-tu tad mā-m
   daha-ti pārthiva
   burn-IMPF.PRS.3.SG king.M.SG.VOC
   ‘O king, those words of the swans torment me.’ (Sanskrit – psych)

c. mulī=lā āī=ca rāgāvṇa
   girl.F.SG.OBL=DAT mother.F.SG.OBL=GEN.N.SG scolding.N.SG.NOM
   dāj-ta
   trouble-IMPF.PRS.N.3.SG
   ‘The mother’s scolding torments the girl.’ (Marathi – psych)
   (Deo 2003, 6; gloss modified)

The third type of source are ‘go predicates’, Sanskrit intransitive verbs with an internal theme argument, e.g., *gam* ‘go’ (Deo 2003). These predicates undergo a lexical semantic change by which they license an additional dative-marked experiencer argument, from meaning ‘go’ to ‘like’. The dative experiencers were realized as oblique arguments in Sanskrit and only became subjects in Marathi (Deo 2003).

The diachronic acquisition of subjecthood has generally been described as a gradual development in the existing literature (e.g., see Cole et al. 1980), with the acquisition of behavioral subject properties preceding the acquisition of coding properties,
2.5. The diachrony of dative subjects

i.e., nominative case and verb agreement. Allen (1995) provides further evidence for the gradual development of subjecthood with her extensive study on the diachrony of experiencer verbs and grammatical relations in English. Old English (7th to 11th century CE) had a large number of experiencer verbs with dative case marking on the experiencer, e.g., *ofhreowan* ‘cause/feel pity for something’ and *lician* ‘to cause/feel pleasure’ (Allen 1995), as shown in examples (35) and (36) respectively.

(35) ... him ofhreow þæs mannæ
    him-DAT caused-pity the-GEN man-DAT
    ‘There was pity in him for the man’
    or: ‘He felt sorry for the man.’
    (Æ1c.Th.I.p.192.16, taken from Allen 1995, 68)

(36) ac gode ne licode na heora geleafleast, ne heora
    but God-DAT not liked not their faithlessness-NOM, nor their
    ceorung, ac asende him to fyr
    grumbling-NOM but sent them to fire
    ‘but God did not like their unbelief or their grumbling, but sent fire to them’
    ((COE), ÆHom 21 68, taken from Allen 1995, 114f.)

Allen (1995) shows that some of the dative experiencer arguments behaved like subjects at the Old English stage with respect to coordinate subject deletion (CSD), but only when the experiencer was preposed, i.e., preceded a nominative theme argument, see (36). The nominative theme of these verbs could also control CSD, but again, only if it was preposed. Yet, the Old English experiencer predicates generally showed a larger preference to have object experiencers. Old English word order has been rather free, generally allowing for object fronting (cf. Allen 1995). For Middle English (11th to 16th century BE), more evidence accounting for the subjecthood of the dative experiencers exists as the nominative theme arguments of experiencer verbs consistently failed to trigger subject-verb-agreement and sporadic agreement with dative experiencers started to appear (Allen 1995). The experiencer arguments then eventually accomplished ‘full subject status’ by receiving canonical nominative subject case marking before the English case system was lost.

This is in line with Haspelmath (2001) who argues for the gradual development of dative subjects in Maltese, a Semitic language which has been in close contact with Romance and Germanic languages, i.e., Sicilian, Italian, and English, over the centuries. Haspelmath (2001) takes Maltese verbs like *irnexxlu* ‘succeed’ originally to be dative-object verbs, see the reconstructed example in (37).
The dative experiencer construction as in (37), however, is ungrammatical in contemporary Maltese. The experiencer argument instead has to precede the verb and the verb has to additionally agree with the dative experiencer in person, number and gender via a suffixed indirect-object agreement marker, see (38).

(38) L-it-tifla t-itfa’ il-ballun.
    to-the-girl succeed.PRF.to.her [she-throw.IP the-ball]
    ‘The girl managed to throw the ball.’ (Haspelmath 2001, 78)

Haspelmath (2001) suggests this as evidence for the gradual acquisition of subject status of the preverbal experiencer over time. A further indicator of the experiencer argument eventually developing subject status is the optional realization of the preverbal experiencer as a nominative argument, as shown in (39).

(39) It-tifla t-itfa’ il-ballun.
    the-girl succeed.PRF-to.her [she-throw.IP the-ball]
    ‘The girl managed to throw the ball.’ (Haspelmath 2001, 78)

Crosslinguistically, there is a clear tendency for dative experiencers to acquire subjecthood over time, leading to non-canonically marked subjects. Haspelmath (2001) describes the general mechanism as follows: Experiencers generally refer to a definite human participant and are thus increasingly placed in topic position. As most human topics are subjects, the experiencer gradually assimilates and acquires morphosyntactic subject behavior. Despite all this evidence on hand, the Object-to-Subject Hypothesis has been challenged by the more recent Oblique Subject Hypothesis.

### 2.5.2 The Oblique Subject Hypothesis

The Oblique Subject Hypothesis generally draws on the continuous existence of dative subjects in Icelandic, the most conservative Germanic language. Dative subjects are common in present-day Icelandic (Andrews 1976, Zaenen et al. 1985) and can be attested at older stages of the language, see e.g., (40) for an Old Icelandic example as provided by Barðdal and Eythórsson (2003), in which the experiencer verb þykja ‘seem’ has a dative subject.
2.5. The diachrony of dative subjects

(40) og þótti honum sem fóstra sínum mundi mein að verða
and seemed him as foster-father self.-dat would harm to become
‘and it seemed to him as if his foster father would be harmed’

(Ljósvetninga saga; taken from Barðdal and Eythórsson 2003, 442)

Barðdal and Eythórsson (2003) show that most of the subjecthood diagnostics established for dative subjects in Modern Icelandic (cf. Section 2.4), including syntactic position, long distance reflexivization, raising, and controlling infinitives, generally hold for the relevant dative arguments in Old Icelandic (see also Barðdal and Eythórsson 2012).

On the basis of the pervasiveness and stability of dative subjects in the history of Icelandic, Barðdal and Eythórsson (2009) furthermore claim that dative subjects must have already existed in earlier stages of Indo-European. Barðdal and Eythórsson (2009) take dative subjects to be a substantial part of the Icelandic core grammar, given its high frequency in Modern Icelandic of around 700 existing dative subject predicates. The existing predicates moreover belong to lexical semantic verb classes that are generally low on the transitivity scale, an observation which holds cross-linguistically. This in turn speaks for the Icelandic dative subjects to be lexicalized to a large extent (Barðdal and Eythórsson 2009). Moreover, dative subjects are not productive in the history of Icelandic, as the type frequency of verbs taking a dative subject has gone down from Old to Modern Icelandic as Barðdal and Eythórsson (2009) report (from 72 types in a 20,000 words corpus to 48 types in a compatible modern corpus). Barðdal and Eythórsson (2009) thus conclude that Icelandic inherited a monolithic dative subject construction from earlier stages of Indo-European which was already deeply rooted in Old Icelandic. Furthermore, Barðdal and Eythórsson (2009) claim that structures containing at least ‘subject-like obliques’ are to be found in all the ancient and archaic Indo-European languages and should therefore be reconstructed for their proto-language, Proto-Indo-European (cf. Barðdal and Eythórsson 2009).

Further support for the Oblique Subject Hypothesis comes from a comparative typological study conducted by Barðdal et al. (2012). Barðdal et al. (2012) compared predicates that instantiate the ‘Dative Subject Construction’, i.e., predicates whose leftmost argument of their argument-structure is in the dative case, across several languages from different chronological layers of Indo-European, namely Old (Norse-) Icelandic, Ancient Greek, Early/Classical Latin, Old Russian, and Old Lithuanian. Barðdal et al. (2012) found that the Dative Subject Construction co-occurs consis-
tently with similar lexical semantic verb classes across the different language branches and thus conclude that the construction must have existed in a language stage prior to these languages, possibly (and at least) reflecting a common West-Indo-European language stage.

However, the Indo-Aryan evidence questions the inheritance of a semantically stable, monolithic dative subject construction from Proto-Indo-European. Deo (2003) showed that in Marathi, lexical semantic shifts of individual verbs have effectuated argument structure changes, leading to the reanalysis of these verbs as taking a dative subject (see Section 2.5.1). Moreover, the Icelandic attestation only goes back to the 12th century which is around when dative subjects began to be possible in Indo-Aryan in the first place (Deo 2003, Butt and Deo 2013), without any evidence for dative subjects in older stages of the language. Lexical semantic change is a common phenomenon of Indo-European (see, e.g., Traugott 2005 for an overview) and Icelandic is generally no exception to this matter. Icelandic is undergoing a change with respect to subject case marking connected to lexical semantics, which can be traced back to the 19th century and is currently still in progress (Svavarsdóttir 1982, Smith 1996, Jónsson 2003, Barðdal 2011). This change in progress has been dubbed as Dative Sickness (þágufallssýki) or Dative Substitution and describes the systematic replacement of accusative experiencer subjects by datives, see, e.g., (41).

(41) a. Mig langar að fara.
   I.ACC long.PRS to go.INF
   ‘I long to go.’

   b. Mér langar að fara.
   I.DAT long.PRS to go.INF
   ‘I long to go.’

(Smith 1996, 22)

Dative substitution has been taken to be driven by the increasing systematic association of dative case with experiencer semantics (Smith 1996, Jónsson 2003). Although dative case marking can in general be correlated with particular lexical semantics in Icelandic, the exact lexical semantics associated with dative subjects are yet unclear. The next section illustrates the lexical semantic factors which have been suggested to condition dative subjects in Icelandic by the previous literature.
2.6 Dative subjects and lexical semantics

Dative case in Icelandic was initially analyzed as an instance of quirky case marking, stipulated by the lexical entries of individual verbs (cf. Zaenen et al. 1985). However, Zaenen et al. (1985) also noticed that the Icelandic datives are mostly restricted to arguments which denote specific thematic roles, i.e., goal and theme arguments. Maling (2001, 2002) revisited this observation and showed that the mappings between dative case, objects and thematic roles in Icelandic are more regular than had been previously acknowledged. In particular, there seems to be a strong systematic correlation between goal roles, including experiencers, recipients and beneficiaries, and dative case marking. Moreover, theme objects which undergo movements also tend to be marked dative (Maling 2002). With respect to subject case marking in Icelandic, Jónsson (2003) showed that non-nominative case is unavailable for agentive subject arguments. Dative subjects have been broadly associated with two major lexical semantic verb classes (Barðdal 2011): (i) experiencer predicates, and (ii) happenstance predicates, which can also be seen as involving some kind of experiencer/goal. The studies examining the diachronic development of these lexical semantic verb classes with respect to dative subjects are detailed in the following.

2.6.1 Lexical semantic verb classes

Barðdal et al. (2012) identified a set of 49 narrowly-defined lexical semantic verb classes belonging to 14 different higher class categories which occur together with predicates in the Dative Subject Construction, i.e., a verbal construction whose first or leftmost argument at argument-structure is marked dative, across several ancient Indo-European languages, including Old Icelandic. Apart from providing evidence for the Oblique Subject Hypothesis (cf. Section 2.5), Barðdal et al. (2012) present the lexical semantic distribution of the Dative Subject Construction in Old Icelandic. Their data was gathered electronically as well as manually from various resources such as databases, lexicons and reference grammars and also through the inspection of textual resources. The 14 established verb classes are listed below, including some Old Icelandic examples as given by Barðdal et al. (2012). Verbs of possession were not found for Old Icelandic, but are listed for the sake of completeness as they occur

---

\[\text{Except for the Latin resources which are described in McGillivray (2010), no detailed descriptions are given about the data sources in Barðdal et al. (2012). The reader is referred to the webpage www.uib.no/noncancase for access to the full database, which unfortunately has no content.}\]
in all of the other investigated languages, i.e., Ancient Greek, Latin, Old Russian
and Old Lithuanian. Verbs of happening are assumed to occur in the Dative Subject
Construction in Old Icelandic by Barðdal et al. (2012), but no examples were given.

- **Verbs denoting Emotions**: falla í geð ‘like’, leiðast ‘dislike, be bored’, leng-
  jast ‘long’, bjóða þekt ‘feel good’, finnast ‘feel, think’, ofþjóða ‘have fear/agony’,
  renna í skap ‘become angry’, leiðast ‘be bored’, léttað ‘feel relieve’, angra ‘be
  bothered, regret’, sárna ‘cause pain’, skipa máli ‘find (un)important for sby’

- **Verbs denoting Attitudes**: duga ‘be sufficient, suffice’, verða gagn að ‘have
  use of sth’, sama ‘suit, become, be proper for sby’

- **Verbs of Gain**: aflast ‘benefit’, bjóðast ‘be offered’, vaxa fjaðrir ‘grow’

- **Verbs of Bodily States**: hitna ‘feel warm’, batna ‘get better (health)’, sortna
  fyrir augum ‘lose consciousness’, svelgjast á ‘go down the wrong way (food)’,
  blæða ‘bleed’

- **Verbs of Cognition**: boða hugur ‘suspect’, hugsast ‘occur to one’s mind’,
  gangast hugur við ‘change one’s mind’, hnykkja við ‘be amazed’, fyrnast ‘forget’

- **Verbs of Perception**: heyrast ‘hear’, spyrjast ‘hear news’

- **Verbs of Speaking**: kveðast ‘say’, mælast ‘speak well’

- **Verbs of Success**: fara fram ‘make progress’, farast vel ‘fare well’, ganga ‘do
  well’, takast ‘succeed, manage’

- **Verbs of Hindrance**: dveljast ‘get a hindrance, stay’, gefask yfir ‘fail, do
  wrong’, missa ‘slip, stagger’

- **Verbs of Ontological States**: kippa í kynið ‘turn into one’s kin’, lenda
  saman ‘come in collision with’, vera farið ‘be in a certain way’, muna ‘make a
  difference’, vera gefið ‘have an ability’

- **Verbs of Happening**

- **Verbs of Modality**: beru ‘be obliged’, bila ‘lack’, vera þörfr á ‘need’

- **Verbs of Evidentiality**: virðast ‘seem’, birtast ‘appear’

- **Verbs of Possession**
In earlier work, Barðdal (2004, 2008) proposed similar verb classes for dative subject predicates in Icelandic and subsumed them under the two semantic categories ‘experience-based predicates’ and ‘happenstance predicates’. These categories are picked up in Barðdal et al. (2012). The category of experience-based predicates consists of verbs of emotions, bodily states, cognition, attitudes and perception. Happenstance predicates denote non-volitional, often accidental events and are verbs of gain, success, happening, hindrance, ontological states, speaking, and possession. Verbs of modality and evidentiality are not classified as belonging into one of these categories by Barðdal et al. (2012).

Verbs belonging into the umbrella categories can be found across all the five languages investigated by Barðdal et al. (2012) with only a few language-specific deviations, see the semantic map in Figure 2.1. All verb classes, except for verbs of speaking, are present in at least four of the five languages in question, with the great majority of classes being found in all investigated language branches. This suggests a high semantic coherency and stability of the Dative Subject Construction in early stages of Indo-European which argues in favor of the Proto-Indo-European inheritance of dative subjects according to Barðdal et al. (2012).

Based on the same categorization, Barðdal (2011) shows in a comparative study of experience-based and happenstance predicates in Old and Modern Icelandic that
although experience-based predicates already make up the majority of the dative subject construction in Old Icelandic, the type frequency of happenstance predicates is further decreasing towards the Modern Icelandic period. Óðardal’s investigation builds upon two corpora consisting of four different genres: Icelandic fiction, translated fiction, biographies and memoirs, and non-fictive texts. The Old Icelandic texts date from approximately 1200 CE and comprise mainly Sagas, while the Modern Icelandic corpus contains texts from the 1980s (see also Óðardal 2001). For each corpus, 500 words were extracted from each text with 10 texts per genre, resulting in 20,000 running words for each time stage respectively.

Table 2.2 shows the verb type frequencies for the possible four subject cases which Óðardal (2011) found in the two corpora. According to this investigation, the frequency of dative subject predicates has gone down from 72 in Old Icelandic to only 48 types in MI. These numbers contain a small proportion of passive constructions. When only considering active constructions, there are 66 dative subject predicates in Old Icelandic and 33 in MI which amounts to a marked overall reduction of the construction’s type frequency of about 50%.

<table>
<thead>
<tr>
<th>Subject Case</th>
<th>Old Icelandic</th>
<th>Modern Icelandic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Nom</td>
<td>299</td>
<td>76.6</td>
</tr>
<tr>
<td>Acc</td>
<td>12</td>
<td>3.1</td>
</tr>
<tr>
<td>Dat</td>
<td>72</td>
<td>18.5</td>
</tr>
<tr>
<td>Gen</td>
<td>7</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>390</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2.2: Predicate type frequencies across different subject cases in two corpora of Old Icelandic and Modern Icelandic (Óðardal 2011, 73).

The 66 Old Icelandic dative subject predicates are equally distributed across experience-based and happenstance predicates. This distribution, however, shifts towards 76% experience-based and 24% happenstance predicates in the Modern Icelandic corpus with the experience-based predicates mainly denoting verbs of cognition and emotion in the modern texts. Óðardal (2011) takes her findings to be the result of Dative Substitution with mainly cognition/emotion verbs attracting former accusative subject predicates, while happenstance predicates change to default nominative subject marking.

Overall, the existing studies show that lexical verbal semantics are conditioning the occurrence of dative subjects throughout the Icelandic diachrony. Moreover,
the distribution of dative subjects changes over time, with an increasing correlation between dative case and experiencer arguments. More evidence for the interrelation between lexical semantic factors and dative subjects comes from morphosyntactic operations such as passivization and middle formation, where case is only retained when certain lexical semantic conditions apply as exemplified below.

2.6.2 Passivization and middle formation

Icelandic distinguishes between three voices: active, passive and middle. If a canonical transitive sentence is passivized, the nominative subject is demoted while the accusative object is promoted and realized as a nominative subject in the passive, see (42). Passives are formed periphrastically via the auxiliaries *vera* ‘be’ and *verða* ‘become’ in conjunction with a past participle (see, e.g., Thráinsson 2007). In the passive, the auxiliaries as well as the participle agree with the nominative subject. The auxiliaries show agreement for person and number and the participle agrees with the nominative in number, gender and case. In addition, the agent argument, i.e., the former active nominative subject, can be optionally realized as a prepositional construction together with the preposition *af* ‘by’ which takes a dative complement, see example (43-b).

(42) a. einhver barði strákana í skólanum
   somebody.NOM hit.PST.3SG boys.the.ACC in school.the.DAT
   ‘Someboy hit the boys in school’

   b. strákarnir voru barðir í skólanum
   boys.the.NOM be.PST.3PL hit.PASS.PTCP.M.PL.NOM in school.the.DAT
   ‘The boys were hit in school’

   (Thráinsson 1994, 177, gloss modified)

Apart from manifesting the existence of dative subjects in Modern Icelandic, Zaenen et al. (1985) have shown that passivization gives rise to dative subjects. In contrast to canonical transitives, verbs that take a dative object preserve their case marking under passivization. Hence, the dative object of the verb *sökkva* ‘sink’ in the active construction given in (43-a) surfaces as a dative subject in the corresponding passive, see (43-b). The finite verb shows no agreement with the dative subject, but occurs in third person singular by default. The past participle appears in the neuter nominative-accusative form which is equivalent to the Icelandic uninflected past participle, sometimes called the supine (cf. Andrews 1990, Thráinsson 2007).
Another valency changing operation which gives rise to dative subjects in Icelandic is middle formation. Icelandic middles are systematically derived from their transitive counterparts via the suffix -st (Sigurðsson 1989, Anderson 1990, Wood 2015), see (44) for an example pair. As with passives, the object role is promoted to subject under middle formation. However, the external subject role is not merely demoted as in passives, but eliminated with middles (see, e.g., Sigurðsson 1989).

The ending -st was historically derived from the reflexive pronoun sik during the Old Nordic period (cf. Ottósson 1992, Wood 2015). In contemporary Icelandic, -st predicates still denote reflexives, but there are also reciprocal, anticausatives and more generic middle readings for verbs carrying -st, see the examples in (45) which Wood cites for the different usages (Wood 2015, 62). The distribution of -st-verbs is rather complex and not all Icelandic verbs carrying the suffix -st are automatically generic middles. However, the majority of the -st marked verbs are compatible with cross-linguistically identified middle readings (cf. inter alia Kemmer 1993, Kaufmann 2007, Wood 2015) and I refer to all -st-marked predicates as middles throughout my thesis.
2.6. Dative subjects and lexical semantics

b. Jón dulbjóst sem prestur.
   John.NOM disguised-ST as priest
   ‘John disguised himself as a priest.’
   (Reflexive)
   (Jónsson 2005, 400)

c. Glugginn opnaðist af sjálfu sér.
   window.the.NOM opened-ST by itself
   ‘The window opened by itself.’
   (Anticausative)
   (Sigurðsson 1989, 268)

d. Rafmagnsbílar seljast (vel) hér.
   electric.cars.NOM sell-ST (well) here
   ‘Electric cars sell well here.’
   (Generic Middle)

In this thesis, I follow Kaufmann’s (2007) analysis of middles, which describes middle voice as essentially encoding the non-canonical control properties of a verb’s arguments. Kaufmann (2007) shows that the non-canonical control properties result from the structural absence of the agent argument which is, however, still available on a pragmatic or inferential level. The structural absence of an agent is exemplified by the fact that the Icelandic middles are incompatible with agentive modifiers, purpose clauses and by-phrases, but allow for the non-agentive phrase af sjálfu sér ‘by itself/automatically’ (cf. Sigurðsson 1989, Wood 2015). This starkly contrasts with passives, which still have an understood external argument (Wood 2015), see example (46) as opposed to (47).

(46) a. Rúðunni var splundrað {viljandi/ *af sjálfu sér}.
   window.the.DAT was shattered {intentionally/ *by itself}
   ‘The window was shattered on purpose.’

b. Rúðunni var splundrað (af rænigjumum).
   window.the.DAT was shattered (by robbers.the.DAT)
   ‘The window was shattered by the robbers.’

c. Rúðunni var splundrað (til þess að gera hann reiðan).
   window.the.DAT was shattered (for it to make him mad)
   The window was shattered in order to make him mad.’
   (Wood 2015, 67)

(47) a. Rúðan splundraðist {*viljandi/ af sjálfu sér}.
   window.the.NOM shattered-ST {*intentionally/ by itself}
   ‘The window shattered by itself.’
b. **Rúðan** splundraðist (*af rænigjumun*). window.the.NOM shattered-ST (*by robbers.the.DAT*)
‘The window shattered (*by the robbers*).’

c. **Rúðan** splundraðist (*til þess að gera hann reiðan*). window.the.NOM shattered-ST (*for it to make him mad*)
The window shattered (*in order to make him mad*).’

(Wood 2015, 67)

A further difference between the passive and the middle constructions in (46) and (47) concerns the occurrence of dative case on the subject. Dative case is only preserved under middle formation when the object is a thematic goal as in (48). When the object is a dative theme, case is not preserved, see (49) (and (47)). Middles do not agree with dative subjects, along the lines of what was shown before for actives and passives.

(48) a. Pétur bauð mér vinnu. Peter offered me(D) a job(A)
b. Mér bauðst vinna. me(D) offered a job(N)
‘I got the opportunity to get a job.’

(Sigurðsson 1989, 260)

(49) a. Ég hellti mjólkinni niður. I spilled the milk(D) down
b. Mjólin helltist niður. the milk(N) spilled down

(Sigurðsson 1989, 265)

The empirical facts about passivization and middle formation show that voice is a further conditioning factor for the occurrence of dative subjects in Icelandic. However, while middle formation provides further evidence that dative case marking and experiencer/goal semantics are interrelated, dative case and theme arguments can not be pulled together in the same way. Jónsson (2003) concludes that dative goals are predictable from lexical semantics, but dative themes must be assigned idiosyncratically. Yet, Svenonius (2002, 2006) opts for an approach which factors in event semantic considerations to provide a more generic analysis of case marking in Icelandic. The event semantic factors conditioning case marking in Icelandic as suggested by Svenonius are detailed in the following section.
2.7 Dative case and the structure of events

Svenonius (2002) proposes a novel analysis for case in Icelandic which is based on event semantic considerations. His theory is inspired by the cross-linguistic observation that accusative case marking often correlates with a certain aspect of interpretation, in particular, the measuring out of an event (see, e.g., Krifka 1992, Verkuyl 1993, Ramchand 1997). Svenonius (2002) describes the measuring out of an event as consisting of two components: the temporal run-time of an event and the physical extent of some participant (‘the measurer’) or the degree to which the measurer has a property. Svenonius follows previous work on verbal decomposition (e.g., Hale and Keyser 1993, Kratzer 1994, 1996, Harley 1995) and decomposes transitive verbs into two components, the two verbal heads \( v \) and \( V \), whereas \( v \) generally introduces the external argument. \( v \) and \( V \) each introduce subevents construed as parts of a single event which is temporally indivisible. When the temporal relation between the introduced subevents is one of complete overlap, accusative case is assigned with respect to the following principles (Svenonius 2002, 201):

\[
\begin{align*}
\text{(50) a. Certain syntactic elements, e.g., } v \text{ and } V, & \text{ introduce event variables in } \\
& \text{their semantic representations (Davidson 1967). Others do not.} \\
\text{b. Event variables introduced within a syntactic constituent } \alpha & \text{ may represent a complex event } x \text{ consisting of two (or more) subevents } y \text{ and } z \\
& \text{(and } \ldots \text{).} \\
\text{c. If the event } x & \text{ consists of subevents } y \text{ and } z, \text{ then } y \text{ and } z \text{ are related } \\
& \text{temporally.} \\
\text{d. If the temporal relation of } y \text{ and } z & \text{ is one of total overlap, then accusative} \\
& \text{case is licensed in } \alpha. \\
\text{e. Aspectual features of } y \text{ and } z & \text{ may force or prevent total overlap.}
\end{align*}
\]

In Svenonius’ approach, accusative case assignment depends on Aktionsart features, i.e., the temporal signature of the verb. Moreover, case is not identified by the identity of individual verbal heads, but depends on specific combinations of verbal heads. There are various configurations in which a \( v \)-\( V \) combination fails to assign accusative. Svenonius (2002, 209) postulates that Icelandic verbs which take dative objects introduce subevents which do not have the same temporal extension, either because of the temporal profile of \( v \) or \( V \):
In a syntactic context $\alpha$ representing an event $x$ composed of subevents $y$ and $z$, dative case is licensed in $\alpha$ iff the temporal relationship of $y$ and $z$ is not total overlap.

In explaining his theory, Svenonius (2002) first discusses verbs of perception and cognition which generally take accusative objects in Icelandic when used transitively, e.g., sjá ‘see’ or heyra ‘hear’. According to Svenonius’ intuition, these verbs tally well with an account in which a temporal isomorphism between the subevents introduced by $v$ and $V$ produces accusative case: The seer and the seen are both equally long participating in an event of seeing. If either of the participants is removed, the event is no longer sustained. This contrasts with verbs which take experiencer subjects which generally describe the experience of a sensation in relation to the object of $V$. The object of $V$ however could disappear hypothetically, while the experience continues. Yet, Svenonius (2002) remains rather vague about the licensing of dative case on subjects and provides no straightforward analysis.

With respect to motion verbs, Svenonius (2002) observes that object case assignment depends on the extent to which the motion is accompanied throughout the event by a causer, i.e., the degree to which the subevents introduced by $v$ and $V$ overlap temporally. For example, verbs of ballistic motion, such as kasta ‘throw, fling, hurl’ or henda ‘throw away, discard’, contain a subevent introduced by $V$ which describes a smoothly flowing inertial movement of, e.g., a ball that has been thrown. This subevent does not completely overlap with the causative subevent of $v$ as the causer does not accompany the moving object until its endpoint and such verbs thus take dative objects. Furthermore, verbs in which the movement of the object is independent of the action of an agent or causer have dative objects in Icelandic, e.g., dreypa ‘drip’ and sökkva ‘sink’. This contrasts with verbs of caused motion, e.g., drauga ‘pull, drag, draw’, which take accusative objects because the causing force accompanies the object throughout the whole motion event which renders the subevents introduced by $v$ and $V$ inseparable. Moreover, verbs that typically take affected objects such as brenna ‘burn’ assign accusative case to their objects as the causer/agent’s participation in the event is conceived as co-temporaneous with the participant that is undergoing the event. Svenonius (2002) further proves his point by providing the minimal pairs given in (52) which illustrate that object case marking for certain Icelandic verbs depends on whether the object undergoes an ‘unaccom-
panied' ballistic motion event or whether the object is fully affected by the event described by the verb.

(52) a. *skjóta fuglinn* ‘shoot the bird’ (ACC)
    b. *skjóta kúlunni* ‘shoot the bullet’ (DAT)
    c. *skutla hvalinn* ‘harpoon the whale’ (ACC)
    d. *skutla skutlinum* ‘throw the harpoon’ (DAT)

Another group of verbs which standardly take dative objects in Icelandic, and also cross-linguistically (see, e.g., Blake 2001), are what Svenonius (2002) dubs *help* verbs. Examples of verbs belonging into this class are *hjálpa* ‘help’, *trúa* ‘trust, have faith in’ and *þakka* ‘thank’. Dative case marking on the objects of such verbs is often characterized in terms of the general animacy of the object. Svenonius (2002) however notes that the class of *help* verbs does allow for inanimate dative arguments synchronically, see example (53), which in turn suggests that other factors are at play.

(53) Vaxtalækkun hjálpar efnahaginum/*efnahaginn.
    interest.rate.cut.NOM helps the.economy.DAT/the.economy.ACC

An interest rate cut helps the economy. (Svenonius 2002, 213)

Svenonius (2002) suggests that the aspectual signature of V cannot overlap completely with that of the external argument in *v* for *help* verbs and thus dative case is licensed. The aspectual signature of V stems from the lexicalization of *help* verbs as being internally caused with an internal argument in V that has its own, independent, trajectory.

Dative objects moreover alternate with accusatives in a class of verbs when the object is an animate beneficiary (‘beneficiary alternations’, cf. Svenonius 2002), see example (54-a) vs. (54-b). According to Svenonius (2002), the presence of an animate object yields that V has the temporal signature of the ‘internally caused’ type which is unavailable for inanimate objects. The signature of V in turn leads to the temporal mismatch between *v* and V and dative case is licensed.

(54) a. Kristín greiddi hárið.
    Kristín combed the.hair.ACC
    Kristín combed her hair.
Kristín greiddi Jóni.
Kristin combed Jon's hair.

(Svenonius 2002, 216)

Dative case is moreover common on goal arguments of Icelandic ditransitive predicates, whereas accusative is licensed on the theme argument. In order to account for the presence of three arguments, Svenonius (2002) borrows the label RP for ‘Result Phrase’ from Ramchand’s (2002) decomposition of the verb phrase which occurs in addition to vP and VP, see the structure for the ditransitive verb gefa ‘give’ in (55), taken from Svenonius (2002, 218). In Ramchand’s event decomposition framework, which has been fully fledged as the first-phase syntax in Ramchand (2008), events can generally be decomposed into three parts, which allows for a more sophisticated and detailed analysis of the interrelation between event structure and case overall. Therefore, I make use of Ramchand’s event decomposition in this thesis, which is discussed in detail in Section 3.4 of Chapter 3, instead of arguing on the basis of vP and VP.

\[
\text{gefa ‘give’ } [vP \text{ Agent cause } [VP \text{ Theme go } [RP \text{ TO Goal }]]]
\]

Svenonius assumes that the dative on the goal argument is licensed within RP via a temporal mismatch between V and R, or even in a higher structural position. The accusative, on the other hand, is licensed via the complete temporal overlap between v and V. This account also holds for ditransitives with two dative marked objects such as lofa ‘promise’ in which there are three subevents involved of which no two subevents completely overlap. Moreover, the stative predicate kosta ‘cost’ takes two accusative objects which fits into an analysis in which the three subevents overlap completely.

In more recent work, Svenonius (2006) employs his theory to explain the differences observed in the Icelandic passive and middle with respect to dative case preservation. His approach is based on two assumptions. First, dative case is determined lower in the decomposed verbal structure than accusative, given the lexical semantic constraints on the Icelandic dative case distribution. Second, since middles do not imply the existence of an external argument, the middle is licensed lower in the verbal structure than the passive. With respect to the lexical decomposition of the verb phrase, Svenonius heavily draws on the system for event decomposition
proposed by Ramchand (2002, 2006). Moreover, his analysis of passives and middles relies on minimalist ideas of feature-checking and cyclic spell-out and assumes a Voice head which interacts in a complex fashion in order to license dative case.

The core observation of the theory outlined in both Svenonius (2002) and Svenonius (2006) is that dative case cannot be purely idiosyncratic in Icelandic, but also that case marking in Icelandic is not straightforwardly determined on the basis of thematic roles. Rather, case is licensed by a combination of lexical and event-related, i.e., ‘aktionsartal’, semantics in the Icelandic verb phrase. While Svenonius’ basic insights are valuable, his notion of overlap remains vague in terms of the formalism employed. Therefore, although I draw on his core observations in this thesis, I follow a Lexical-Functional Grammar (LFG) approach to argument structure which separates out lexical semantics from syntactic structure and incorporate the event decomposition framework proposed by Ramchand (2008) in order to provide a detailed account of the interaction between case marking, voice, and event structure in Icelandic.

While this section provided further background information on factors conditioning dative case marking in Icelandic, the next section introduces the historical data which forms the main empirical basis for this thesis.

2.8 The Icelandic Parsed Historical Corpus (IcePaHC)

The historical data in this thesis is based on the Icelandic Parsed Historical Corpus (IcePaHC, Wallenberg et al. 2011, Rögnvaldsson et al. 2012). IcePaHC consists of 61 text extracts with around one million words which date from the 12th to the 21st century. Most pre-existing studies examining syntactic change in Icelandic amassed their evidence by comparing data from the Old Icelandic Sagas (ca. 1150–1350 CE) with data from the modern language, leaving the diachronic detail of the intervening periods unclear. The texts in IcePaHC cover all attested stages of Icelandic, with roughly 100000 words per century, which enables access to an unusually comprehensive level of diachronic detail.

IcePaHC contains texts from different genres, i.e., narratives, religious texts, biographies, legal texts and scientific texts. However, there is a genre issue in IcePaHC in that certain genres are over-represented and others are under-represented in individual periods. While the majority of texts in the corpus are narratives, including the Old Icelandic Sagas and modern fiction, the texts from the 16th century con-
tained in IcePaHC are mainly religious texts and biographies are predominant in the 17th century.

The corpus is syntactically annotated according to the Penn Treebank-style (Marcus et al. 1993) which includes the sophisticated annotation of case and grammatical relations, ideal for a corpus linguistic study of dative subjects. The corpus annotation furthermore specifies clause types, constituent order, noun types (proper nouns, empty and overt pronouns, and expletives, etc.), verb types (modals, main verbs, ‘have’, ‘be’, and ‘become’), tense, and voice. Figure 2.2 shows a sample annotation of IcePaHC.

(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))

Figure 2.2: Sample annotation for a sentence from IcePaHC.

The sample annotation in Figure 2.2 shows a matrix declarative IP with a clause-initial, pronominal, dative subject (mér ‘me’) followed by an inflected form of the verb finna, meaning ‘find, feel, think, seem’, in present tense. Each sentence in IcePaHC contains a sentence ID which is given at the bottom of each sentence, see Figure 2.2. The sentence ID provides information about the year date of the text in which the corresponding sentence occurred (the first four digits), the text name and genre of the text. When citing an example from IcePaHC in this thesis, I specify the full sentence ID of the corresponding sentence so that the annotations can be retrieved in the corpus, see, e.g., (56) which corresponds to the sample annotation from Figure 2.2. Unless otherwise stated, I adhere to the information given by the IcePaHC annotation with respect to the morphological glossing of examples from the corpus throughout this thesis.
This thesis investigates factors conditioning the occurrence of dative subjects in the history of Icelandic. In this chapter, I have introduced the factors which are known to correlate with dative case marking in Icelandic, i.e., voice, lexical semantics, and event semantics. Moreover, dative subjects are examined with respect to word order to provide an understanding of the interaction between case marking, word order and grammatical relations in Icelandic. Investigating these factors in IcePaHC necessitates the extraction of the relevant information from the IcePaHC annotation. The data processing and extracted data are detailed in the following.

2.8.1 Data processing

My investigations are based on the matrix declarative clauses from IcePaHC in which subjects are distinctly annotated for case, see, e.g., the dative case marked subject in the sample annotation given in Figure 2.2, excluding ambiguous or non-marked subjects such as foreign words and annotations which were marked as uncertain. These clauses were identified using my own Perl scripts which are furthermore used for the extraction of the relevant data. In total, 65,394 matrix declarative clauses with case-marked subjects form the basis of my investigations. Besides subject case marking, information about object case marking, if available, was extracted from the corpus for each of the 65,394 matrix declarative clauses. Moreover, I extracted the verbs and verb types involved in the clause as well as information about voice. The verb tags in IcePaHC differentiate between different verb types including forms of vera ‘be’ (BE), geru ‘do’ (DO), hafa ‘have’ (HV), verða ‘become’ (RD), modal verbs (MD) and main verbs (VB). In order to account for instances in which an infinite verb form licenses the case marking of the matrix clause, matrix sentences containing particular types of verbs, e.g., auxiliaries and modals, had to be handled in a specific manner which is exemplified below.
Bare infinitives When a finite auxiliary or modal was present in the sentence, I took the non-finite verb, whether participle or infinitive, as being in charge of the case frame and argument structure. Note however that in Icelandic only epistemic modals show properties similar to those of auxiliaries (Thráinsson 2007). The verbs tagged as modals in IcePaHC consist of epistemic and root modals and modals are often ambiguous with respect to these readings (see, e.g., Platzack 1979 for the distinction between epistemic and root modals). In general, IcePaHC annotates the following verbs as modals (MD): mega ‘may’, munu ‘will’, skulu ‘shall’, vilja ‘will’, geta ‘be able to’ and fá ‘be able to’. Apart from the modal verb geta ‘be able to’, which takes a participial complement, the modals listed generally take a bare infinitive as complement. While root modals always take a nominative subject, independent of their infinitival complement, epistemic modals can take a non-nominative subject assigned to it by the infinitive or participle (Thráinsson 2007). For example, in (57-a), the modal verb vilja takes an accusative subject licensed by the predicate vanta ‘lack’, but can only be interpreted as epistemic in the sense of ‘will’, while the root meaning ‘want’ as shown in (57-b) is unavailable.

(57) a. Harald vill oft vanta peninga.
   Haraldur.ACC will.PRS.3SG frequently lack.INF money:NOM
   ‘Haraldur frequently tends to lack money.’

   (Thráinsson 2007, 426; gloss modified)

b. Haraldur vill selja bókina.
   Haraldur.NOM want.PRS.3SG sell.INF book.the.ACC
   ‘Haraldur wants to sell the book.’

   (Thráinsson 2007, 424; gloss modified)

I tested whether modals have a significant impact on the occurrence of a particular case frame via $\chi^2$-tests and found that the distribution of case with modals does not differ significantly from the overall picture (which is presented in Chapter 4). Therefore, in constructions with a modal and a bare infinitive, or participle in the case of geta ‘be able to’, I extracted the relevant information about the infinite verb form, i.e., lemma and verb type, in conjunction with the finite modal verb from the corpus and related the infinite verb to the case frame of the matrix clause in the resulting dataset (see Section 2.8.2 for a description of the dataset).
Infinitival clauses  IcePaHC only annotates modals that take a bare vb (main verb infinitive) as MD. However, there are a number of verbs that have an (epistemic) modal meaning when taking an infinitive with the infinitive marker að ‘to’, e.g., ætla ‘intend, need’ (cf. Thráinsson 2007, 422), see (58), which were not distinctly annotated as modals. Besides ætla ‘intend, need’, these verbs include eiga ‘ought (to)’, hljóta ‘must’, kunna ‘can’, verða ‘must’, and þurfa ‘need’ (cf. Thráinsson 2007, 422). The epistemic modals can also take non-nominative subjects licensed by the að-infinitive, as shown in (58), where the infinitive líka ‘like’ assigns dative case to the subject.

(58) Haraldi ætlar að líka vel í Stuttgart.
Haraldur.DAT intend.PRS.3SG to like.INF well in Stuttgart
‘It looks like Haraldur will like it in Stuttgart.’

(Thráinsson 2007, 426; gloss modified)

In IcePaHC, að-infinitives are annotated in the form of an IP-INF, i.e., as an infinitival clause, subordinated to the matrix IP-clause, see Figure 2.3 and the corresponding sentence given in example (59).

```
( (IP-MAT (CONJ og-og)
 (VBDI ætluðu-ætla)
 (NP-SBJ (PRD-N þau-það))
 (IP-INF (TO að-að) (VB sigla-sigla))
 (...-.))
 (ID 1350.FINNBOGI.NAR-SAG,657.1852))
```

Figure 2.3: IcePaHC annotation for an infinitival construction with the infinitive marker að ‘to’.

(59) og ætluðu þau að sigla.
and intend.PST.3PL they.NOM to sail.INF
‘and they intended to sail away.’

(IcePaHC, 1350.FINNBOGI.NAR-SAG,657.1852)

As the infinitive in the að-clause can in general license the case pattern of the modal verb, information about the infinitives was extracted and related to the case frame of the respective matrix clause in the dataset, in analogy to the extraction of the relevant bits of information from modal constructions containing bare infinitives.
The modals taking an að-infinitive were moreover included when testing whether modals interfere significantly with case marking, and were found to have no effect on the overall distributions.

Furthermore, a number of aspectual verbs take an að-infinitive in the form of a subordinated IP-INF in IcePaHC which cannot assign a thematic role to their subject position allowing for licensing of non-nominative subjects by the non-finite verb in their infinitival complement (Thráinsson 2007, 428). The aspectual verbs in question are byrja ‘begin’, fara ‘begin’, hætta ‘stop’, taka ‘begin’, vera ‘be’ (i.e., the progressive) and vera búinn ‘be finished, be done’, see, e.g., (60-a) for an example in which byrja ‘begin’ has an accusative subject licensed by klæja ‘itch’ and the progressive construction with vera ‘be’ in (60-b), where the infinitive kólna ‘get colder’ licenses the dative subject.

(60) a. Mig byrjaði að klæja í þetta í gær.
   l.acc begin.pst.3sg to itch.inf in this.acc yesterday
   ‘I began to itch in this yesterday.’

b. Honum var að kólna.
   he.dat be.pst.3sg to get.colder.inf
   ‘He was getting colder.’

(Thráinsson 2007, 428; gloss modified)

Similar to the handling of modal verbs, with the aspectual verbs in question, information about the infinitive was recorded together with the case frame of the matrix clause in the resulting dataset. Again, the presence of the aspectual verbs had no significant impact on the overall distribution of case patterns according to a \( \chi^2 \)-test.

**Voice** Given that voice has been identified as a conditioning factor for the occurrence of dative subjects in Icelandic (see Section 2.6.2), information about the voice of a clause has been extracted on the basis of the inflected verb forms and the verb tags involved. Passive participles are tagged as VAN in the corpus and could thus be readily identified by the annotation itself. Middle verbs have been identified via their -st-ending. IcePaHC only annotates middle forms with their corresponding middle infinitive lemma if the meaning of the middle considerably differs from the meaning of the corresponding non -st-verb. Thus, the endings of the verb forms themselves had to be matched in my search.
2.8. The Icelandic Parsed Historical Corpus (IcePaHC)

**Word order** Studying the interaction between subject case and word order in IcePaHC demands the extraction of information about the possible word order patterns. To this end, I extracted information about verb placement, subject position and the general order of subject, verb, indirect and direct object. With respect to subject position, I gathered information on whether the subject occurred before or after the finite verb, i.e., in the pre- or postfinite position. As regards verb placement, I documented whether the sentence is in V1 (verb-first) order or not. Only identifying sentences with a verb in initial position however is not doing justice to the full range of V1 declaratives in the matrix sentences from IcePaHC. V1 structures may contain an empty subject, e.g., an empty expletive or *pro*-dropped argument, which is placed into the prefinite position by definition via the IcePaHC annotation. Thus, V1 structures with a null subject surface as V2 structures in IcePaHC, see, e.g., the annotation of the sentence in example (61).\(^8\)

![Figure 2.4: IcePaHC annotation for a V1 declarative with an empty expletive subject.](image)

( (IP-MAT (NP-SBJ-1 *exp*)
  (BEDI Var-vera)
  (NP-1 (Q-N fátt-fár)
    (NP-POS (NS-G manna-maður)))
  (ADVP-LOC (ADV heima-heima))
  (.-.-))
  (ID 1350.FINNBOGI.NAR-SAG,655.1696))

Moreover, V1 clauses may begin with a conjunction, e.g., og ‘and’, in which case the verb immediately follows the conjunction. These factors were all taken into account in the Perl script for data extraction.

\[^{8}\text{The subject in Figure 2.4 is not explicitly marked for case. The IcePaHC annotation only marks empty pronouns for non-nominative case. Unmarked instances such as the one in Figure 2.4 represent nominative constituents. Thus, although not marked explicitly, the annotation provides distinct information about subject case marking in sentences involving an empty pronoun.}\]
2.8.2 The IcePaHC dataset

The extracted information was used to generate a well-structured dataset which forms the foundation for the frequency calculations in my investigations. Figure 2.5 shows a snapshot of the dataset. In the dataset, the extracted data features of a sentence are mapped onto its sentence ID, encoding additional information about the text it occurs in as well as the text’s genre and the corresponding year date. In sum, the extracted features are the verb licensing the case frame and the corresponding verb tag, the voice of the matrix clause, subject and object case marking, and word order, including subject position and verb placement. Moreover, I recorded whether a modal or aspectual verb was present in the matrix declarative clause, when a non-finite verb form is taken to be licensing the case frame. The dataset was furthermore enriched with lexical and event semantic information which is detailed in Chapter 4.

<table>
<thead>
<tr>
<th>ID</th>
<th>VERB</th>
<th>VERB TYPE</th>
<th>MODAL/ASPECT</th>
<th>VOICE</th>
<th>SUB_CASE</th>
<th>OBJ1_CASE</th>
<th>OBJ2_CASE</th>
<th>WORD_ORDER</th>
<th>SUBJ_POSITION</th>
<th>V1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150.FIRSTGRAMMAR.SCI-UN_1</td>
<td>seta</td>
<td>VB</td>
<td>active</td>
<td>subj NOM</td>
<td>subj ACC</td>
<td></td>
<td></td>
<td>VSO1</td>
<td>postfinite</td>
<td>no</td>
</tr>
<tr>
<td>1150.FIRSTGRAMMAR.SCI-UN_2</td>
<td>seta</td>
<td>VB</td>
<td>active</td>
<td>subj NOM</td>
<td>subj ACC</td>
<td></td>
<td></td>
<td>V15V</td>
<td>postfinite</td>
<td>no</td>
</tr>
<tr>
<td>1150.FIRSTGRAMMAR.SCI-UN_3</td>
<td>hafa</td>
<td>HV</td>
<td>passive</td>
<td>subj NOM</td>
<td>subj ACC</td>
<td></td>
<td></td>
<td>VSO1</td>
<td>prefinite</td>
<td>no</td>
</tr>
<tr>
<td>1150.FIRSTGRAMMAR.SCI-UN_4</td>
<td>hafa</td>
<td>HV</td>
<td>passive</td>
<td>subj NOM</td>
<td>subj ACC</td>
<td></td>
<td></td>
<td>VSO1</td>
<td>postfinite</td>
<td>yes</td>
</tr>
<tr>
<td>1150.FIRSTGRAMMAR.SCI-UN_5</td>
<td>hafa</td>
<td>HV</td>
<td>passive</td>
<td>subj GEN</td>
<td>-</td>
<td>V5</td>
<td>postfinite</td>
<td>V5</td>
<td>postfinite</td>
<td>no</td>
</tr>
<tr>
<td>1150.FIRSTGRAMMAR.SCI-UN_6</td>
<td>hafa</td>
<td>HV</td>
<td>passive</td>
<td>subj NOM</td>
<td>-</td>
<td>V5</td>
<td>postfinite</td>
<td>V5</td>
<td>postfinite</td>
<td>no</td>
</tr>
<tr>
<td>1150.FIRSTGRAMMAR.SCI-UN_7</td>
<td>hafa</td>
<td>HV</td>
<td>passive</td>
<td>subj NOM</td>
<td>-</td>
<td>V5</td>
<td>postfinite</td>
<td>V5</td>
<td>postfinite</td>
<td>no</td>
</tr>
<tr>
<td>1150.FIRSTGRAMMAR.SCI-UN_8</td>
<td>hafa</td>
<td>HV</td>
<td>passive</td>
<td>subj NOM</td>
<td>-</td>
<td>V5</td>
<td>postfinite</td>
<td>V5</td>
<td>postfinite</td>
<td>no</td>
</tr>
<tr>
<td>1150.FIRSTGRAMMAR.SCI-UN_9</td>
<td>hafa</td>
<td>HV</td>
<td>passive</td>
<td>subj NOM</td>
<td>-</td>
<td>V5</td>
<td>postfinite</td>
<td>V5</td>
<td>postfinite</td>
<td>no</td>
</tr>
<tr>
<td>1150.FIRSTGRAMMAR.SCI-UN_10</td>
<td>hafa</td>
<td>HV</td>
<td>passive</td>
<td>subj NOM</td>
<td>-</td>
<td>V5</td>
<td>postfinite</td>
<td>V5</td>
<td>postfinite</td>
<td>no</td>
</tr>
<tr>
<td>1150.FIRSTGRAMMAR.SCI-UN_11</td>
<td>hafa</td>
<td>HV</td>
<td>passive</td>
<td>subj NOM</td>
<td>-</td>
<td>V5</td>
<td>postfinite</td>
<td>V5</td>
<td>postfinite</td>
<td>no</td>
</tr>
</tbody>
</table>

Figure 2.5: Dataset extracted from IcePaHC for the diachronic investigation of factors conditioning dative subjects in Icelandic.

For the analysis of the frequency calculations performed on the basis of the dataset, the data was divided into time periods as defined per Haugen (1984) for the historical development of Scandinavian. These periods group the IcePaHC data into the following five time stages: 1150-1349, 1350-1549, 1550-1749, 1750-1899, and 1900-2008 CE. I furthermore conducted $\chi^2$-tests of homogeneity to examine whether the observed distributions in the individual time stages differed from what could be expected given the distributions of the respective constructions in the whole corpus.

2.8.3 A note on data curation

During my investigations, I encountered a number of annotation mistakes concerning subject case in IcePaHC. For the purpose of this thesis, I corrected around 150 instances of erroneously annotated dative subject clauses. The annotation mistakes mainly consisted of the spurious annotation of objects as subjects, nominative proper
nouns as datives and instances of false interpretations of the pronoun þér. þér is ambiguous between a reading as honorific form of address in nominative case which was frequently used in Old Icelandic and the 3rd person singular dative form of the personal pronoun þu ‘you’.

The corrections were performed in my own dataset in order to keep consistent with future released versions of IcePaHC. The erroneous patterns were corrected manually in my dataset. I furthermore documented all the corrections carefully in order to be able to keep track of possible bias during my investigations. I moreover came across other instances of spurious annotations and discovered ungrammatical constructions in IcePaHC. These, for example, included nominative or dative subjects which were spuriously annotated as accusatives with middles as well as passives. These constructions were also corrected. The figures that I present throughout my thesis for the IcePaHC data are therefore calculated on the basis of my corrected dataset and may for this reason differ from previous studies concerning the same corpus material. The most recent version of the dataset can be investigated and downloaded via the HistoBankVis website, see Section 5.2.1 in Chapter 5 for more information.

This section presented the historical data underlying the investigations in this thesis. The next section introduces the methods employed for data analysis. Besides calculating occurrence frequencies and statistical significance, complex feature interactions contained in the multidimensional IcePaHC dataset are analyzed using methods coming from the field of Visual Analytics.

\footnote{Many thanks go to Joan Maling and Jóhannes Gisli Jónsson who reviewed the Schätzle et al. (2015) paper and pointed us towards some of the erroneous annotations by critically questioning frequency calculations which we gave for various constructions in IcePaHC. I also owe many thanks to Hannah Booth who went over the corrections I made regarding the dative subject predicates and commented on each single one of them. The erroneous annotations identified together with Joan Maling and Jóhannes Gisli Jónsson were forwarded to the corpus developers and I moreover intend to report back to them on the remaining errors in the near future.}

\footnote{http://histobankvis.dbvis.de}
2.9 Visual Analytics for historical linguistics  
(HistLingVis)

Historical linguistic research is corpus-based by nature. In recent years, diachronic corpora have increasingly been made available to the historical linguistic research community in a digitized form, often including linguistically sophisticated annotations. With the increasing availability of digitized text material, historical linguistic researchers have also availed themselves more and more of quantitative and statistical methods for data analysis. Well-established methods are the calculation of (co-)occurrence frequencies, correlations and dispersion statistics, but also more elaborated methods, e.g., clustering, have found their way into the historical linguistic toolbox (see, e.g., Hilpert and Gries 2016).

Linguistic data is inherently multidimensional, with complex interactions between different linguistic features and structures being the norm rather than the exception. Historical linguistic change typically is the result of such complex interactions. The core remit of historical linguistic work is to identify a language change and to understand how different relevant factors have interacted with each other across time to effectuate the change. Statistical calculations are an extremely useful means for the quantification of changes of individual linguistic structures, but they are per se not suited for the uncovering and understanding of complex interactions between various linguistic features over time. Statistical tests are suited to either confirm or to reject a given hypothesis. A central issue of historical linguistics is that the precise factors leading to a change are often unknown or at least highly debated among researchers. Without a priori knowledge about potential interactions contained in the data, using statistical methods is not feasible.

In this thesis, I use methods coming from the field of Visual Analytics (Keim et al. 2008) for the analysis of historical change, illustrating their power and efficacy for historical linguistic research. Visual Analytics is defined as “the science of analytical reasoning facilitated by interactive visual interfaces” (Thomas and Cook 2005, 28), coupling automated algorithmic analyses with interactive visual components to enable a human-informed discourse (Thomas and Cook 2005, Keim et al. 2008). Visual analytic methods contrast with methods coming from the longer-standing field of Information Visualization which directly transforms mostly numerical data into visualizations, but build on insights from the field. Accordingly, Shneiderman’s (1996) Information Seeking Mantra for Information Visualization ‘Overview first,
zoom and filter, then details on demand’ has been transformed into the Visual Analytics Mantra ‘Analyze first, show the important, zoom, filter and analyze further, details on demand’ (Keim et al. 2008) which integrates the human into the analysis loop. The aim of Visual Analytics is to saliently present interesting and significant correlations between data dimensions, leading to the visual emergence of significant patterns. Visual analytic tools and techniques are generally suited for information synthesis and are moreover meant to help deriving insights from large amounts of dynamic, ambiguous, and often conflicting data (Thomas and Cook 2005); data properties which are also characteristic of linguistic data. In terms of historical linguistics, this facilitates the identification of language change and relevant interacting factors, furthering the understanding of the diachronic data.

Over the past decade, sophisticated visualization methods as developed within the field of computer science have been introduced to work on a small but growing range of linguistic problems. Beginning with seminal work on the visualization of natural language processing techniques as pursued by Collins and colleagues (e.g., Collins 2007, Collins et al. 2007, Collins 2010), linguistic visualizations have moved from visualizing word categories (Honkela et al. 1995, Honkela 1997) to representing interactive visual analytic systems for the analysis of more complex linguistic structures. Visual analytic approaches to linguistic data range from the visualization of word concordances (Culy and Lyding 2010, 2011), syntactic structure (Culy et al. 2011a, 2012, Meurer et al. 2016), the cross-linguistic distribution of phonological and phonotactic patterns, e.g. vowel harmony (Rohrdantz et al. 2010, Mayer et al. 2010a) and similar place avoidance (Mayer et al. 2010b), the cross-linguistic comparison of linguistic features in context of their genealogical and geo-spatial distribution (Rohrdantz et al. 2012a, Mayer et al. 2014), complex predication patterns (Butt et al. 2012, Lamprecht et al. 2013), and pitch contours (Sacha et al. 2015, Asano et al. 2016), to discourse analysis (Zhao et al. 2012, Gold et al. 2016, El-Assady et al. 2016, Hautli-Janisz and El-Assady 2017, El-Assady et al. 2017).

Although the majority of diachronic studies still employ purely statistical methods for data analysis, the interest in incorporating visualizations in historical linguistic investigations has been growing (see, e.g., Culy et al. 2011b, Lyding et al. 2012, Theron and Fontanillo 2015 and Section 2.9.2). However, only in rare cases, visualizations have been specifically designed for the analysis of diachronic linguistic data. This section motivates the application of Visual Analytics to historical linguistic investigations and provides an overview about the few existing diachronic
linguistic visualizations. From the existing examples for diachronic linguistic visualizations and my own experiences, I deduce a number of design principles for a successful diachronic linguistic visualization which are presented as design space in Section 2.9.3.

### 2.9.1 Challenges for historical linguistic research

Historical linguistics is almost exclusively text-based and the available amount of digitized text corpora for historical linguistic research has been increasing over recent years. Diachronic corpora comprise large linguistically unannotated collections of historical texts from several Indo-European languages such as e.g., the Bibliotheca Augustana,\(^{11}\) TITUS (Thesaurus Indogermanischer Text- und Sprachmaterialien),\(^{12}\) and GRETIL (Göttingen Register of Electronic Texts in Indian Languages),\(^{13}\) but also increasingly include linguistically annotated corpora. The large digital collections and corpora generally offer the amount of textual data needed for applying unsupervised Natural Language Processing (NLP) techniques, but are not able to capture the amount of linguistic detail necessary for the analysis of complex linguistic phenomena and structures.

Linguistically annotated corpora are usually smaller in size because they have undergone a manual annotation process in addition to an automatic preprocessing. Albeit being time-consuming, the manual annotation procedure allows for a sophisticated linguistic annotation, often including a deep syntactic analysis of hierarchies and dependencies between phrase structure constituents. Such structural information is prototypically stored, i.e., *banked*, in so-called treebanks. For instance, corpora annotated in the Penn Treebank-style (Marcus et al. 1993) are treebanks which include an annotation for syntactic hierarchies and dependencies. Examples are the Penn Parsed Corpora of Historical English (Kroch and Taylor 2000, Kroch et al. 2004, 2010), the Icelandic Parsed Historical Corpus (Wallenberg et al. 2011), and the Heliand Parsed Database (Walkden 2015). Other treebanks containing annotations for syntactic dependencies are the Latin Dependency Treebank (Bamman and Cane 2006, Bamman et al. 2007), the Prague Dependency Treebank (Hajič 1998) and PROIEL (Pragmatic Resources of Indo-European, Haug and Jøhndal 2008).

---

\(^{11}\)https://www.hs-augsburg.de/~harsch/augustana.html  
\(^{12}\)http://titus.uni-frankfurt.de/indexd.htm  
\(^{13}\)http://gretil.sub.uni-goettingen.de/
The standard corpus linguistic approach to a quantitative analysis of diachronic data incorporates the usage of specialized programming languages for text processing and statistical analysis, e.g., Python, Perl and R (Baayen 2008, Bird et al. 2009, Christiansen et al. 2012). Scripts are usually employed for the automatic extraction of relevant linguistic patterns based on annotated values and for the calculation of co-occurrence frequencies and statistical significances across different temporal episodes. In pursuing such an analysis, a multitude of high-dimensional data tables with different features and characteristics are generated for the diachronic analysis of a single phenomenon. For example, Table 2.3 represents a prototypical historical linguistic dataset, but still shows just a small subset of the many different tables utilized for the comprehensive diachronic analysis of OV (Object-Verb) word order in Icelandic conducted by Hróarsdóttir (2000).

<table>
<thead>
<tr>
<th></th>
<th>Indefinite NPs</th>
<th>Definite NPs</th>
<th>NPs as proper names</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OV</td>
<td>VO</td>
<td>% OV</td>
</tr>
<tr>
<td>14th century</td>
<td>28</td>
<td>33</td>
<td>45.9%</td>
</tr>
<tr>
<td>15th century</td>
<td>23</td>
<td>30</td>
<td>43.4%</td>
</tr>
<tr>
<td>16th century</td>
<td>15</td>
<td>28</td>
<td>34.9%</td>
</tr>
<tr>
<td>17th century</td>
<td>28</td>
<td>59</td>
<td>32.2%</td>
</tr>
<tr>
<td>18th century</td>
<td>6</td>
<td>28</td>
<td>17.6%</td>
</tr>
<tr>
<td>19th century</td>
<td>34</td>
<td>425</td>
<td>7.4%</td>
</tr>
<tr>
<td></td>
<td>134</td>
<td>603</td>
<td>18.2%</td>
</tr>
</tbody>
</table>

Table 2.3: Diachronic distributions of definite and indefinite NPs across different word orders in the history of Icelandic (Hróarsdóttir 2000, 136).

Finding significant patterns and feature interactions across such tables requires the pair-wise comparison of the relevant bits of information in the form of numbers. This is by no means a trivial task as numbers computed for various feature combinations from different data dimensions have to be compared across multiple data tables of varying size, while factoring in a temporal component. Moreover, a well-known problem of historical linguistic and corpus-based research is data sparsity. Thus, statistical significances are often calculated on the basis of only few occurrences of the actual observation, derogating the statistical measurements and conclusions. Based on this, meaningful patterns may be lost in the forest of numbers, while irrelevant patterns are likely to surface as significant. Furthermore, statistical calculations entail the definition of fixed parameters, e.g., in the form of fixed temporal episodes. Yet, interesting patterns may stay hidden when an analyst chooses epochs that are
either too fine or too coarse grained for the analysis. In general, statistical analyses are suited to either confirm or reject previously anticipated knowledge. This can be appropriate in order to verify a priori hypotheses and does provide insights into whether a particular factor is significant or not. However, under this scenario, statistical analyses are limited to the investigation of more or less evident hypotheses about the data. Given that the precise factors for historical linguistic change are often elusive and hypotheses are usually tentative, a researcher may have to conduct several different analyses and experiment with different combinations of feature interactions to eventually find significant patterns. This is not only time-consuming, but also makes the resulting data difficult to navigate.

Visual analytic methods allow one to push beyond what is possible with traditional corpus linguistic methods and statistical analysis. By granting an interactive and exploratory access to a large and complex dataset, the researcher can interact with the data freely and directly. Potentially interesting and significant correlations contained in the data are presented saliently, irrespective of specific and predefined hypotheses about the dataset. This, in combination with an explorative visual analysis, may lead to the identification of unknown patterns in the data, generating novel and unexpected insights, instead of merely conforming or disconfirming the expected.

### 2.9.2 Diachronic linguistic visualizations

Visualizations tailored to the analysis of historical linguistic data are rare, but the existing approaches highlight the potential of Visual Analytics for historical linguistic research. For example, Rohrdantz et al. (2011) developed a novel Visual Analytics approach for the analysis of semantic change by visually modeling the diachronic developments of word contexts and their senses, extracted via topic modeling from a large diachronic newspaper corpus. The visualization provides two views of the data which allow for a detailed inspection of word contexts in the form of scatterplots, see, e.g., Figure 2.6-top, and the diachronic investigation of word sense developments, see Figure 2.6-bottom. Another visualization which provides for a diachronic analysis of lexical semantic content based on topic modeling is the pixel visualization designed by Rohrdantz et al. (2012b) for the investigation of the cross-linguistic as well as historical spread of new suffixes and their meaning throughout mass media (see also Rohrdantz 2014).

Further examples for historical linguistic visualizations are the approach taken by Lyding et al. (2012) for the investigation of diachronic changes in the use of modal
verbs across different registers of academic discourse using Structured Parallel Coordinates (Culy et al. 2011b) as shown in Figure 2.7, and diachronlex diagrams (Theron and Fontanillo 2015) which visualize the diachronic evolution of word meanings in historical dictionaries as parallel coordinate plots.

Figure 2.6: Top: Scatterplot visualization of word contexts and their diachronic distribution across contextual senses. Bottom: Visualization of the temporal sense developments for *to browse* and *to surf*. Figures taken from Rohrdantz et al. (2011, 307f.).
The World’s Languages Explorer is a visualization tool which allows for the cross-linguistic comparison of linguistic features (Rohrdantz et al. 2012a, Mayer et al. 2014). In addition to the comparison of features across languages, the World’s Language Explorer provides for the exploration of interrelations and (dis)similarities among features within subtrees of a language’s genealogy as well as for the investigation of areal (contact) influences by using the sunburst visualization technique as shown in Figure 2.8. Although the World’s Languages Explorer is mainly typological in nature, the visualization allows a researcher to draw historical linguistic conclusions by exploring genealogical information related to linguistic features in comparison to their areal distribution.

Within the context of the research conducted for this thesis, I developed two novel visualization systems in collaboration with Visual Analytics experts from the computer science department at the University of Konstanz for the analysis of his-
torical change. The first system is a glyph visualization (Butt et al. 2014, Schätzle and Sacha 2016) which allows for a visual analysis of the interactions of several pre-defined syntactic factors in the history of Icelandic, using data from IcePaHC. The glyph visualization was initially designed for the diachronic analysis of verb placement in Icelandic and has been extended to cope with a larger number of potentially interacting factors for the diachronic analysis of dative subjects.

![Image of the World’s Languages Explorer](image)

Figure 2.8: High-resolution screenshot of the World’s Languages Explorer showing the main components. Left: sunburst visualization of linguistic features, top-right: a node-link diagram displaying the language genealogy of the languages in question, bottom-right: a map providing information about the geographical distribution of the respective languages. The example at hand consists of 19 language features for 27 Indo-European languages that were automatically extracted from parallel Bible texts. Figure taken from Rohrdantz et al. (2012a).

While building on the experiences gathered by the glyph visualization, the second system, HistoBankVis (Schätzle et al. 2017), aims at providing a more elegant and generic solution for the analysis of a large amount of possibly interacting data dimensions, suitable for any kind of historical linguistic research question. The glyph visualization is introduced in Chapter 4, where it complements a complex analysis of factors conditioning the occurrence of dative subjects in the history of Icelandic. The HistoBankVis system is presented and used in Chapter 5 for a thorough and
detailed investigation of the diachronic interaction between word order changes and subject case marking in IcePaHC.

Based on our experiences in designing visualizations for historical linguistic research, my colleagues and I developed a generalized design space for diachronic visualizations (Hautli-Janisz et al. to appear) because we found that the design parameters for a diachronic Visual Analytics system heavily depend on the type of available historical data and the historical linguistic research question related to it. The generalized design space and the established guidelines are illustrated in the following section and provide an understanding of the design choices underlying the visualization systems used for the historical linguistic investigations in this thesis.

2.9.3 Designing visualizations

The process of designing visualizations for research is both structured and creative. Creating a visualization involves finding optimal solutions for mapping data dimensions, which may be either numerical, ordinal, or categorical, to visual representations, i.e., visual variables such as color, position, shape, size, orientation, value, texture, brightness and motion (Bertin 1983). However, the number of visual variables is limited and not every visual variable is suited to represent every type of data dimension. The visual variable color is often used for categorical data, whereas position is commonly used for representing numerical data. Yet, a good choice in one visualization might turn out to be a bad choice in another visualization.

Visualizations aim at fostering the emergence of visual patterns which in turn point to relevant hidden structures contained in the raw data. Thus, a visualization designer has to anticipate what kinds of patterns might be of interest for the data analyst and choose the necessary means to make these patterns stand out visibly. The design effort is both iterative and collaborative, merging knowledge with data-driven modeling in order to provide the most optimal overview of and access to the data. For linguistic visualizations, this usually requires input from both domains: After the linguist has found or generated a suitable corpus for the research task, relevant data dimensions are extracted and brought into a format which can be processed by the visualization. The Visual Analytics expert needs to establish design parameters that represent the data best, e.g., scatterplots, glyphs, or histograms, and find statistical measures to uncover significant patterns tailored to the analysis task and the analyst’s need. Moreover, appropriate interaction techniques granting an exploratory access to the data have to be implemented.
The main task of diachronic visualizations is to support a researcher in understanding how different data dimensions develop along a temporal axis, i.e., the time dimension. Most design decisions thus depend on the characteristics of the time dimension and how the time dimension relates to the other data dimensions involved. Moreover, whether potential correlations among different data dimensions are of interest has to be taken into account. The available design space which results from these parameters is elaborated on in the following.

**Time dimension** The time dimension is inherent to historical linguistic data as investigations generally focus on uncovering and understanding change over time. The design of the time dimension depends on the following data characteristics:

- **Time resolution:** In diachronic investigations, each data point, e.g., a document or sentence, can be considered to be a time-stamped observation of language use. The time resolution plays an important role for the visualization design which depends on whether developments across years, decades or centuries are investigated. The time resolution is either given by the data or implied by the research task.

- **Distribution of observations over time:** Diachronic data usually consists of less data objects for the longer-standing past than for more recent time stages. Plotting such data along a linearly scaled timeline may have the effect that some epochs are only scarcely populated, while others suffer from overplotting. To prevent these effects, it might be a good choice to offer an analysis of the data points as time sequences or aggregate data points in the form of fixed or variable time frames instead.

- **Amount of observations:** The overall amount of available data points also influences the design choices. If a dataset is rather small, each data object might be visualized individually and given its own visual representation within the visualization. However, if the data consists of tens of thousands of data points or more, aggregating data objects might yet again be the better solution in order to arrive at a meaningful visualization.

**Data dimensions under investigation** Identifying patterns of linguistic change and deriving hypotheses about which factors potentially caused the change is the main purpose of diachronic linguistic research. The visualization has to support
a researcher in finding these patterns of change across different data dimensions and visually represent them in the best possible way. The data dimensions may be extracted from different types of resources, i.e., linguistically annotated corpora or large raw text collections, and be either manually edited or directly computed from the raw data. In the following, the different kinds of available data dimensions are explained:

- Manually created: These data dimensions result from manual annotations performed by a domain expert. A manual annotation generally allows for the precise and accurate annotation of quite complex facts and relations. However, manual annotations are time-consuming and complicated. The amount of the resulting data is thus necessarily limited. Moreover, errors may be caused by human unsystematicity.

- Manually revised: These data dimensions result from annotations which are generated automatically, but additionally undergo a careful manual revision by a domain expert. The resulting annotation might be less complex than with fully manually created annotations, but they are generally of good quality. Moreover, the semi-automated process allows for the annotation of a larger dataset.

- Predefined computed: These data dimensions are computed automatically from the raw data. They are generally used as is as their quality is known or expect to be good enough for the research task. Predefined computed data dimensions are, for example, counts of occurrence frequencies of a certain, predefined, linguistic phenomenon in different contexts. In contrast to manually edited data dimensions, very large datasets can be made use of, but the phenomena under investigation are usually less complex. Furthermore, systematic biases may occur due to processing errors.

- Open computed: These data dimensions are also computed from the raw data. However, the structure and interpretation of the resulting dimension is unknown beforehand as the data is not categorized by means of fixed parameters, but grouped automatically, e.g., via a clustering algorithm. The task is then to understand the emerging clusters with respect to why the data are clustered into particular groups and how the clusters develop over time. Open
computed data dimensions may thus lead to novel and unexpected insights, providing access to developments contained in a very large dataset.

**Correlations of data dimensions** Detecting change in one data dimension can yield important insights into the diachrony of a language, confirming hypotheses about whether specific constructions changed over time or remained rather stable. Moreover, novel hypotheses can be generated on the basis of the identified change. Yet, language change is generally correlated with complex interactions between features from different data dimensions. In some cases, a researcher cannot anticipate which data dimensions are correlated with a change. Visual Analytics supports a researcher in this respect by providing an easy interactive approach to experimenting with correlations of different dimensions, some of which may prove to be significant. Finding and understanding correlations between data dimensions may then again lead to the formulation of new hypotheses on the nature of the historical change. Furthermore, potentials for model improvements might be revealed and can be implemented through feedback loops, leading to better automated analyses and an improved data foundation for research.

**2.10 Summary and conclusion**

This chapter has introduced the crucial background for the investigation of factors conditioning the occurrence of dative subjects in the history of Icelandic. After having provided the language-specific background on Icelandic and a brief introduction to non-canonical case marking, the syntactic subjunctive properties of dative subjects have been established for Modern Icelandic, showing that the dative arguments in question qualify as subjects. This was followed by a detailed discussion about the controversy evolving around the historical origin of dative subjects, presenting the Oblique Subject Hypothesis in comparison with the Object-to-Subject Hypothesis. Moreover, I have presented the lexical and event semantic conditions which correlate with dative case marking in Icelandic to set the stage for my diachronic investigations. The last part introduced the corpus data which is analyzed in this thesis and Visual Analytics for historical linguistics as method for data analysis.

Several conclusions can be drawn from this introductory chapter: The distribution of dative subjects has been changing in Icelandic in that experiencer subjects are increasingly marked with dative case (‘Dative Substitution’). Overall, dative
case correlates systematically with thematic goals in Icelandic. Moreover, passivization and middle formation give rise to dative subjects in Icelandic. However, dative arguments differ with respect to middle formation in that case is only retained with goal arguments, but lost with dative themes. Thus, dative case marking can not be explained solely on the basis of thematic roles. As an alternative solution, Svenonius (2002) provides an event structural analysis of dative case marking in Icelandic, accounting for dative case-marked goals and themes. Yet, the precise event semantic conditions which apply to dative experiencer subjects remain elusive in Svenonius’ account. Investigating dative subjects in conjunction with the Oblique Subject Hypothesis, which is the prevalent theory in the Icelandic literature regarding the historical origin of dative subjects, necessitates an analysis of word order: Syntactic position is a main criterion for the identification of dative subjects in historical Icelandic (e.g., see Barðdal and Eythórsson 2003) and changes in word order have been identified to be a factor behind the reanalysis of datives as subjects in historical English (Allen 1995) and other Indo-European languages (e.g., see Haspelmath 2001).

On these grounds, I investigate and analyze the diachronic distribution of dative subjects in Icelandic with respect to the factors voice, lexical semantics, event semantics, and word order in the following chapters, showing that case marking in Icelandic is part of a complex and interacting system. Before detailing the investigations and providing a theoretical analysis, the relevant linguistic theories for linking arguments to grammatical relations are introduced in the next chapter.
Chapter 3

Linking Theories

3.1 Introduction

This chapter introduces the linking theories which are crucial for providing a theoretical account of the complex interrelation between word order, grammatical functions, lexical semantics, event structure and case marking being at play in Icelandic. Maling describes linking as “the problem of discovering and explaining the regularities which govern the syntactic realization of a verb’s arguments” (Maling 2001, 420). Linking generally refers to the mapping of predicate-argument structures to a syntactic representation, explaining the relationship between lexical semantics, argument structure, case marking, grammatical relations and syntactic structures (Butt 2006). However, the existing theories differ with respect to the particularities employed and no linking theory exists which includes all the relevant features. Therefore, I propose a novel linking theory for Icelandic which is presented in Chapter 6 of this thesis, accounting for the complex system which links case to grammatical relations in the language. The linking theory developed in this thesis is mainly based on LFG’s standard Lexical Mapping Theory (Bresnan and Kanerva 1989), but draws on the enhancements of the theory provided by Zaenen (1993) with respect to lexical semantic entailments and the extended version of Lexical Mapping Theory by Kibort (2014) with respect to argument positions. The Lexical Mapping Theory, including Zaenen’s and Kibort’s ideas, is detailed in Section 3.2 of the present chapter. My analysis furthermore builds on the linking theory developed by Kiparsky (1997). Whereas the standard Lexical Mapping Theory does not account for case marking in linking, Kiparsky’s theory employs the three linkers case, agreement, and posi-
tion in order to be able to account for the diachronic trade-off between word order and case morphology in licensing grammatical relations. Kiparsky’s linking theory is presented in Section 3.3. The theory presented last in this chapter is the first-phase syntax approach by Ramchand (2008) which is not a linking theory per se, but an event decomposition framework which relates syntactic arguments to event structure at the syntax-semantics interface (Section 3.4). Ramchand’s event decomposition is relevant for the linking theory established in this thesis as the first-phase syntax provides a detailed and sophisticated approach for the analysis of the complex interaction between event structure and case marking in Icelandic. Section 3.5 summarizes the relevant features of the linking theories and concludes the chapter.

3.2 Lexical Mapping Theory

Lexical Mapping Theory is the theory of linking developed within the framework of Lexical-Functional Grammar (LFG). This section provides a detailed description of the Lexical Mapping Theory, including an introduction to the basic features of LFG to fully grasp the concept behind its theory of linking.

3.2.1 Lexical-Functional Grammar

LFG originally surfaced as a theory of syntax in the late 1970s and early 1980s in order to address problems and inconsistencies of Transformational Grammar (e.g., see Chomsky 1965). The development of LFG as a generative, but non-transformational theory was motivated by Joan Bresnan’s concern that grammatical transformations were “psychologically unrealistic” (Bresnan 1978, 2), not conforming to the experimental evidence from the existing speech perception and production studies at the time (e.g., Fodor et al. 1974). Together with Ron Kaplan, Bresnan provided the first fully-fledged formalism of LFG in Bresnan and Kaplan (1982), presenting a psychologically feasible as well as computationally implementable linguistic theory, resulting from Bresnan’s research on the viability of Transformational Grammar (Bresnan 1977, 1978) and Kaplan’s research on computational and psycholinguistic modeling (Kaplan 1972, 1973).

In Bresnan and Kaplan (1982), the mapping between semantic predicate-argument relationships and syntactic structure is depicted as the central descriptive problem that any syntactic theory has to deal with. Addressing this problem, LFG defines the complex interrelation between predicate-argument relations and surface structure in
two stages. Instead of assuming that syntactic transformations generate different
surface structures for the same predicate-argument relation, e.g., for different verbal
diatheses, lexical entries are used to define the relations between semantic argu-
ments and surface grammatical functions within LFG (Bresnan and Kaplan 1982,
Dalrymple 2001). Lexical entries are then inserted into the syntax via specific rules.
Moreover, LFG is functional in that grammatical functions are inherently part, i.e.,
primitives, of the theory (cf. Dalrymple 2001). In general, LFG operates on a mod-
ular architecture of language in which parallel levels of linguistic organization are
interrelated via functional constraints (Dalrymple 2001). With respect to syntax,
LFG differentiates between two levels of representation (Bresnan and Kaplan 1982):
the c(onstituent)-structure and the f(unctional)-structure.

3.2.2 C-structure and f-structure

C-structures represent the linear surface order of words and the hierarchical orga-
nization of a sentence’s constituents in the form of conventional phrase structure
trees. F-structures represent the abstract functional syntactic organization of a
sentence. They encode the sentence’s predicate-argument structure and the cor-
responding surface grammatical relations in the form of attribute-value matrices.
An attribute-value matrix is a set of pairs in square brackets which consists of an
attribute and a value: [ATTRIBUTE value]. Values are mapped to attributes in a one-
to-one correspondence, but may contain a ‘nested’ f-structure, describing a further
level of embedding, for example, [ATTRIBUTE [ATTRIBUTE value]]. In general, the
top level of embedding in the f-structure is reserved for information which is part of
the main predicate or specifies the part of the sentence which has functional scope
over the clause in question. However, f-structures do not impose any linear order
on the grammatical relations involved. Figure 3.1 shows a sample c- and f-structure
representation for the English sentence Simon ate cake.

The first row of the f-structure in Figure 3.1 depicts the sentence’s predicate-
argument structure which is inserted from the lexicon. The lexical entry of the main
predicate ate is shown in (1). In the lexical entry, information about the predicate’s
subcategorization frame is specified under PRED, providing the information that the
predicate eat is transitive and thus requires a subject and an object. Tense is also
specified in the lexical entry and is represented at the top level of the f-structure,
having scope over the whole sentence.
Chapter 3. Linking Theories

Figure 3.1: C-structure and f-structure representation of Simon ate cake.

\[ \text{(1) \hspace{1cm} ate} \hspace{1cm} V \hspace{1cm} (\uparrow \text{PRED}) = \text{‘eat(\text{SUBJ,OBJ})’} \]
\[ \hspace{1cm} (\uparrow \text{TENSE}) = \text{past} \]

The functional information contained in the lexical entry is made available to f- and c-structure via a functional projection (ϕ-projection), indicated by the up-arrows in (1). For example, \((\uparrow \text{TENSE}) = \text{past}\) in (1) determines that the attribute-value pair \((\text{tense, past})\) will appear at the top-level of f-structure. This corresponds to the mother node of the predicate’s lexical entry in the c-structure, i.e., V, which is indicated by the arrows pointing up towards V. The subject Simon and the object cake are both encoded as nested structures in the f-structure in Figure 3.1, providing further lexical information (such as e.g., number, gender, person). Functional projections generally relate the individual parts from the c- and f-structure to one another. For example, one ϕ-projection in Figure 3.1 relates the NP cake at c-structure to the functional information corresponding to the object at f-structure, while the other displayed ϕ-projection indicates that the whole f-structure is under the IP-node at c-structure.

The structural separation of functional dependencies from surface form via c- and f-structure allows for a language-independent representation of functional information and language-specific differences with respect to surface phrase structure. The syntactic representations of c- and f-structure are moreover complemented with a further level of representation at the syntax-semantics interface: a(rgument)-structure. A-structure is the locus of assigning grammatical relations to thematic roles, passing on information from the lexicon and c-structure to the f-structure.
3.2.3 A-structure and thematic roles

In the modular LFG architecture, a-structure functions as the “interface between the semantics and syntax of predicators” (Bresnan 2001, 304), and mediates between the lexical semantics of a verb and its syntactic representation:

\[(2) \quad \text{lexical semantics} \downarrow \quad \text{a-structure} \downarrow \quad \text{syntactic structure}\]

A-structure provides information about predicate-argument relationships in terms of thematic roles. For example, the a-structure of the predicate chase as shown in (3) contains the information that the first argument of chase is an agent, whereas the second is a patient.

\[(3) \quad \text{chase} < \text{agent patient}>\]

The thematic roles at a-structure follow a specific linear order according to their relative prominence on the thematic role hierarchy (4) (e.g., see Kiparsky 1987, Bresnan and Kanerva 1989) in which the agent is specified as the most prominent role:

\[(4) \quad \text{Thematic Role Hierarchy}\]

\[\text{agent} > \text{beneficiary} > \text{recipient/experiencer} > \text{instrument} > \text{theme/patient} > \text{location}\]

Functional structures, i.e., grammatical relations, are projected from the thematic roles at a-structure to skeletal f-structures via general principles as formulated within the Lexical Mapping Theory (Bresnan and Kanerva 1989, Bresnan 2001). The standard Lexical Mapping Theory of LFG grew out of the loosely formulated association principles for thematic roles, grammatical relations, and case marking in Icelandic presented in Zaenen et al. (1985) which are illustrated in the following.
3.2.4 Quirky case and the Icelandic Association Principles

Zaenen et al. (1985) defined the Icelandic Association Principles as given in (5) in order to account for the abundance of different case marking patterns existing in Icelandic (cf. Section 2.2). The association principles were meant to capture the complex relationship between thematic roles and grammatical functions in Icelandic and include, contrary to standard Lexical Mapping Theory, principles for case marking.

(5) Icelandic Association Principles

a. If there is only one thematic role, it is assigned to subj; if there are two, they are assigned to subj and obj; if there are three, they are assigned to subj, obj, 2obj. (Universal)
b. Agents are linked to subj. (Universal)
c. Case-marked themes are assigned to the lowest available gf. (Language Specific)
d. Default Case Marking: the highest available gf is assigned nom case, the next highest acc. (Universal)

(Zaenen et al. 1985, 467)

The association principles make predictions about cross-linguistic regularities which are taken to be universal. For example, principle (5-a) reflects the general inventory of grammatical relations, while principle (5-b) corresponds to the observation that agents are associated with subjects across languages. The third universal principle (5-d) corresponds to the observation that languages tend to have the default case marking pattern of nominative on subjects and accusative on objects. However, languages differ with respect to the relation between thematic roles and case. In Icelandic, the language specific principle (5-c) that case-marked themes are assigned to the lowest available grammatical function, with subj>obj>2obj, applies before all other principles. Case-marked themes are taken to be the result of idiosyncratic or quirky case marking, being idiosyncratically assigned by a lexical item. In Zaenen et al. (1985), idiosyncratic case is generally associated with a particular thematic role and is furthermore assigned before thematic roles are mapped to grammatical functions.

As an illustratory example for the association principles, Zaenen et al. (1985) apply the principles to the analysis of the predicate óska ‘wish’. The verb óska
‘wish’ can be used as a ditransitive predicate, see (6-a), or a transitive predicate, see (6-b), depending on whether the goal argument is present or not. The linking schemes for both versions of óska are displayed in (7).

(6)  
   a. Þu hefur óskað henni þess.
       you have wished her(D) this(G)
   b. Þu hefur óskað þess.

(Zaenen et al. 1985, 470)

(7)  
    óska < agent  theme   (goal) >
    [+gen]    [+dat]
   a. SUBJ  2OBJ  OBJ
   b. SUBJ  OBJ

(Zaenen et al. 1985, 471)

The genitive case marking on the theme argument and the dative on the goal argument in (7) are analyzed as instances of idiosyncratic case marking by Zaenen et al. (1985). The language specific principle (5-c) applies to idiosyncratically marked themes and links the genitive theme to the lowest available grammatical function which is the secondary object (2OBJ) in the ditransitive scenario and the primary object (OBJ) in the transitive scenario. This principle moreover prevents that the genitive argument is mapped onto OBJ with ditransitive óska, which would happen if principle (5-a) would apply first. After the genitive argument has been assigned to the lowest available function, principle (5-a) assigns the remaining arguments to the available grammatical functions. Nominative case on the subject is assigned via default case marking. However, the idiosyncratic dative case marking of the goal argument, if present, precludes the default accusative case assignment.

This analysis tallies well with the case marking patterns which arise under passivization with óska. In classical LFG, passives are related to their active counterpart via a lexical redundancy rule. This rule is given in (8) and demotes the subject under passivization, while the object is promoted to the subject of the passive. In Icelandic, the former agentive subject may optionally be realized in a prepositional phrase introduced with af ‘by’ which is indicated by AF-OBJ/∅ in (8).

(8)  
   a. SUBJ ⇔ AF-OBJ/∅
   b. OBJ ⇔ SUBJ

(Zaenen et al. 1985, 463)
Chapter 3. Linking Theories

(9) a. Þess var óskað (*henni).
   this(G) was wished (her(D))
   b. Henni var óskað þess.
   (Zaenen et al. 1985, 471)

The passivization rules predict that only an OBJ can passivize. When oska ‘wish’ is used transitively, the genitive theme object undergoes passivization and becomes the subject, see (9-a) respectively. In the ditransitive version however, only the dative goal argument can be passivized because the theme argument is linked to 2OBJ. In both scenarios, case is retained in the passive which gives rise to idiosyncratically assigned non-nominative case on subjects.

Although the standard Lexical Mapping Theory is superficially different from these initially formulated association principles, the underlying spirit of the approach has not changed. Lexical Mapping Theory still relies on specific principles which associate thematic roles with grammatical functions, but decomposes grammatical relations and thematic roles into basic features.

3.2.5 Feature decomposition

The Lexical Mapping Theory employs a classification of grammatical relations into a set of binary features which constrain the mapping of thematic roles to grammatical relations. One of these features is [$\pm r$ (restricted)] which specifies whether a grammatical function or thematic role is sensitive to semantic restrictions. The other feature is [$\pm o$ (objective)] and indicates whether a thematic role is likely to be linked to an object-like grammatical function or not (cf. Butt 2006). With respect to grammatical relations, the classification results in the matrix given in (10).

(10) Feature Decomposition of Grammatical Relations

<table>
<thead>
<tr>
<th></th>
<th>$-r$</th>
<th>$+r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-o$</td>
<td>SUBJ</td>
<td>OBL$_{\theta}$</td>
</tr>
<tr>
<td>$+o$</td>
<td>OBJ</td>
<td>OBJ$_{\theta}$</td>
</tr>
</tbody>
</table>

(Bresnan 2001, 308)

Subjects and objects are not restricted to specific thematic roles and are [$-r$]. Semantically restricted functions ([$+r$]), i.e. secondary objects OBJ$_{\theta}$ and oblique arguments OBL$_{\theta}$, are the ones which intuitively correspond to fixed thematic roles
cross-linguistically (see Bresnan and Kanerva 1989). The $\theta$ is used as a placeholder for any kind of thematic role. Whereas the feature $[+o]$ designates functions that are object-like, $[-o]$ refers to non-objective functions which complement intransitive categories, e.g. nouns and adjectives (Bresnan and Kanerva 1989).

Grammatical functions are hierarchically ordered according to their degree of markedness (cf. Bresnan and Zaenen 1990, Bresnan 2001). Negative features designate the unmarked feature values, while positive features are marked. Thus, in the following markedness hierarchy, $\text{SUBJ}([-r,-o])$ is the least and $\text{OBJ}_\theta ([+r,+o])$ the most marked function (Bresnan 2001, 309):

$$\text{(11) Partial Ordering of Argument Functions}$$
$$\text{SUBJ} > \text{OBJ}, \text{OBL}_\theta > \text{OBJ}_\theta$$

In order to enable the mapping between thematic roles and grammatical functions, the feature values are used for an intrinsic classification of thematic roles (e.g., see Bresnan and Zaenen 1990):

$$\text{(12) Feature Classification of Thematic Roles}$$
patientlike roles: $\theta$
$[-r]$
secondary patientlike roles: $\theta$
$[+o]$
other roles: $\theta$
$[-o]$

The feature specification of thematic roles is held rather general to be able to account for cross-linguistic differences of linking. Example (13) shows the possible linking patterns of thematic roles to grammatical relations for the predicate $\textit{chase}$ via the set of features specified above.

$$\text{(13) chase < agent patient >}$$
$$\text{subject/obl} \quad \text{subject/obj}$$
The agent role in (13) is classified as $[-o]$, i.e., not patientlike, in contrast to the patient role which is $[-r]$. However, the thematic roles in (13) can in principle be
linked to more than one grammatical function: \([-o]\) refers to a SUBJ as well as an OBL, while \([-r]\) can refer to both a SUBJ and an OBJ. In order to ensure that each thematic role is linked to a single grammatical function, mapping principles and wellformedness conditions which further constrain and regulate the assignment of grammatical relations to thematic roles have been established within the Lexical Mapping Theory.

3.2.6 Mapping principles and wellformedness conditions

The mapping of thematic roles from a-structure to syntactic functions at f-structure is constrained by the set of the following basic syntactic principles (Bresnan and Zaenen 1990, 51):

\[(14) \text{Mapping Principles}\]

a. Subject roles:
   (i) \(\hat{\theta}\) is mapped onto SUBJ when initial in the a-structure; \([-o]\) otherwise:
   (ii) \(\theta\) is mapped onto SUBJ.
   \([-r]\)

b. Other roles are mapped onto the lowest compatible function on the markedness hierarchy (11).

The first part of the mapping principles specifies the subject role: The most prominent role (referred to as \(\hat{\theta}\)) with respect to the Thematic Role Hierarchy, which is at the same time classified as \([-o]\), is mapped to SUBJ. Otherwise, the subject function is assigned to a non-agentive unrestricted role (\([-r]\)) (cf. Bresnan 2001). Once the subject role is linked, all other roles are assigned to the lowest compatible function on the markedness hierarchy. The wellformedness conditions in (15) ensure that only one subject is assigned per predicator and that each thematic role at a-structure is assigned to a single unique function (see Bresnan and Zaenen 1990).

\[(15) \text{Wellformedness Conditions}\]

a. The Subject Condition: Every predicator must have a subject.

b. Function-Argument Bi-uniqueness: Each a-structure role must be associated with a unique function, and conversely.
The mapping principles and wellformedness conditions control, for example, the mapping of the agent role of *chase*, which is the most prominent role classified as $[-o]$, to one unique syntactic function, i.e., the subject. The other thematic role at the a-structure of *chase*, i.e., the patient, is assigned to the lowest compatible syntactic function carrying the feature $[-r]$ and is, finally, linked to the object function, see (16).

(16)  
\[
\begin{array}{ccc}
\text{chase} & \text{agent} & \text{patient} \\
[-o] & [-r] \\
\text{subj} & \text{obj}
\end{array}
\]

The mapping principles also account for argument changing operations by which thematic roles are added or suppressed (cf. Bresnan and Zaenen 1990). In passives for example, the highest thematic role, i.e., the subject, is suppressed and no longer available for linking ($\emptyset$), see (17) which shows the passivized version of *chase*. If there is no initial $[-o]$ role at a-structure, the $[-r]$ role, i.e., the patient and former object, is promoted and linked to the subject.

(17)  
\[
\begin{array}{ccc}
\text{chase} & \text{agent} & \text{patient} \\
[-o] & [-r] \\
\emptyset & \text{subj}
\end{array}
\]

Linking theories generally avoid a direct one-to-one mapping from thematic roles to grammatical relations in order to be able to account for cross-linguistically recurring argument alternations (Butt 2006). An example for such an alternation is locative inversion in Chichewa, a phenomenon which has been thoroughly investigated within LFG, and eventually led to the first fully-fledged formulation of the Lexical Mapping Theory in Bresnan and Kanerva (1989).

### 3.2.7 Argument alternations

Chichewa is a Bantu language, spoken in East Central Africa (Bresnan and Kanerva 1989). In locative inversion, a locative phrase is preposed, compare the inverted structure in (18-b) with the uninverted form in (18-a). The inverted and uninverted forms do not occur in free variation. Locative inversion has a specific discourse
function, i.e., presentational focus, by which the location is focused (Bresnan and Kanerva 1989).

   7-well  7SB-be 17-3-village
   ‘The well is in the village.’

   17-3-village 17SB-be 7-well
   ‘In the village is a well.’

(Bresnan and Kanerva 1989, 2)

Although the arguments in the uninverted form and the locative inversion share the same thematic roles, the arguments are mapped to different grammatical relations in each counterpart. Bresnan and Kanerva (1989) show that the subject from the uninverted form, i.e., ‘well’ in (18-a), is not the subject of the inverted counterpart in (18-b). Instead, the location, i.e. ‘village’, is the subject in the locative inversion.

Evidence for the subject status of the inverted location is provided, inter alia, on the basis of subject-verb agreement. Chichewa uses a complex noun class system for subject-verb agreement. The noun classes are indicated via numbers in (18). In the inverted form, the verb agrees with ‘village’ via the class 17 marker, while in the uninverted form, agreement is with ‘well’ via the class 7 marker instead.

The Lexical Mapping Theory allows for an elegant analysis of locative inversion and the associated dual linking possibilities of thematic roles. For example, the predicate *khâla* ‘remain’ has two thematic roles, theme and location, which may undergo locative inversion (Bresnan and Kanerva 1989). The more general linking pattern is the uninverted form in which the theme argument is linked to subj and the location is realized as an obl, see (19). Linking in (19) proceeds as follows: The thematic roles are classified via the intrinsic feature classification given in (12). The theme argument is classified as $[-r]$ and the location is $[-o]$. By default, the location is $[+r]$ as locatives are usually oblique arguments. Thus, the location is linked to obl$_{loc}$. The feature specification of the theme argument allows for it to be linked to a subj or an obj alternatively. However, the wellformedness conditions specify that there must be a subject, ruling out the obj option.
In locative inversion, the location is linked to subj, while the theme argument becomes the obj. The subject linking of the location is possible because of the special focus context which assigns the [−r] feature to the the argument (Bresnan and Kanerva 1989, Butt 2006). As an [−o, −r] argument, the location can only be assigned to subj, whereas the theme argument is linked to obj. If the theme would be assigned to subj, the wellformedness conditions would be violated.

Although the standard Lexical Mapping Theory provides for a neat analysis of locative inversion, several revised versions of Lexical Mapping Theory exist (e.g., by Alsina 1996, Ackerman 1992, Butt 1998, Zaenen 1993, Kibort 2014), addressing the general dissatisfaction of the linguistic community with respect to the assumption of a thematic role hierarchy. For instance, Levin and Rappaport Hovav (2005) show that the existing thematic role hierarchies differ significantly with respect to the ranking of roles and with respect to the range of phenomena they can account for. On these grounds, no thematic role hierarchy exists that is able to satisfactorily represent universal constructs of argument realizations.

In the following, I introduce Kibort’s (2014) extended Lexical Mapping Theory which pleads for the distinction between argument positions and thematic roles at a-structure, in order to create a better and cleaner formalism for morphosyntactic operations, e.g., passivization and locative inversion, without having to assume a thematic role hierarchy (see also Kibort 2007, 2008, 2013 for earlier formulations of the theory).
3.2.8 Argument positions

Restoring the early LFG dissociation between argument positions and thematic roles (see, e.g., Bresnan 1982), Kibort assumes that an independent tier of representation, i.e., the argument positions, mediates the mapping between thematic roles and grammatical functions, allowing for semantic participants to change order and re-associate with different argument positions for non-default mappings (Kibort 2007). Argument positions are fixed at a-structure in order to represent the relative syntactic prominence of the arguments of the predicate specified by its subcategorization frame (Kibort 2014). Each argument position comes with a particular fixed specification using the standard Lexical Mapping Theory’s binary feature decomposition:

\[
< \text{arg}_1 \text{arg}_2 \text{arg}_3 \ldots \text{arg}_4 \ldots \text{arg}_n >
\]

\[
[+o/−r] [−r] [+o] [−o] [−o]
\]

(Kibort 2014, 266)

Kibort reinterprets the features, going away from the original semantic intuition of Lexical Mapping Theory, in order to match with the traditional linguistic division of syntactic functions into subject and complements and the distinction between core arguments and oblique arguments (Kibort 2013, 2014):

\[
[+o]\text{ complements} \quad (\text{‘internal arguments’ of the predicate})
\]

\[
[−o]\text{ non-complements} \quad (\text{the ‘external’ argument and oblique arguments})
\]

\[
[−r]\text{ core arguments} \quad (\text{subject and object only})
\]

\[
[+r]\text{ non-core arguments} \quad (\text{all arguments except subject and object})
\]

(Kibort 2014, 266)

With this classification, the order of arguments shown in (21) moreover corresponds to LFG’s relational hierarchy of syntactic functions (Bresnan 2001, 96):

\[
\text{SUBJ} > \text{OBJ} > \text{OBJ}_θ > \text{OBL}_θ > \text{COMPL} > \text{ADJUNCT}
\]

The first argument position is generally associated with the subject function, the second argument position then refers to an OBJ and the third argument can be an OBJ_θ. Obliques can be associated with arguments from the fourth argument position on. Moreover, grammatical functions are mapped to the argument positions in accord with the following principle (Kibort 2014, 267):
3.2. Lexical Mapping Theory

(24) **Mapping Principle**

The ordered arguments are mapped onto the highest (i.e. least marked) compatible function on the markedness hierarchy.

For example, (25) illustrates the a-structure of the predicate *cook* as used in the sentence *Both parents cooked supper for the children* in which the first argument is associated with SUBJ, the second with OBJ, and the OBL is assigned to the fourth argument position.

(25)  


cook  <  arg₁  arg₂  arg₄  >  

\[[-o] [-r] [-o]\]

SUBJ OBJ OBLθ

Morphosyntactic operations, e.g., passivization, are taken care of at the level of argument-to-function mapping because they interfere with the ‘default’ mapping, but leave the lexical semantic interpretation of the predicate unaffected (Kibort 2007, 2014). A morphosyntactic operation restricts an argument in that it increases an argument’s markedness by adding a [+r] or [+o] specification to the syntactic feature classification. Kibort (2014) provides the following set of morphosyntactic operations:

(26)  

a. adding the [+r] specification to a [−o] argument;
b. adding the [+o] specification to a [−r] argument; and
c. adding the [+r] specification to a [+o] argument.

(Kibort 2014, 267)

Kibort assumes passivization to be a morphosyntactic operation which restricts an external argument ([−o]) by adding the [+r] specification, see (27). The [+r] specification restricts an argument to be linked to non-core arguments only and the agent argument can no longer be mapped to the subject role. However, the agent argument is still structurally available for linking and is realized as an OBLθ.

(27)  

verb\_passive  <  arg₁  arg₂  >  

\[[-o] [-r] [+r]\]

OBLθ  SUBJ
In contrast to passivization, locative inversion is analyzed as a morphosyntactic operation which restricts the first $[−r]$ argument by adding the $[+o]$ argument (Kibort 2007), see (28). This results in the first argument being linked to the OBJ function. If the second argument selects a non-core $[−o]$ argument, the Mapping Principle associates the second argument with the SUBJ function.

\[
\text{verb}_{\text{locative inversion}} < \text{arg}_1 \text{arg}_4 > \\
[−r] [−o] \\
[+o] \\
\text{OBJ} \quad \text{SUBJ}
\]

With regard to mapping semantic participants to arguments, Kibort (2014) flags atomic thematic roles as inadequate for the semantic representation of arguments because theoretical accounts, even within the LFG framework, are generally at odds with each other, failing to agree on the definition and content of thematic roles. Instead, Kibort (2014) draws on insights from Zaenen (1993) that the lexical meaning of a verb encodes some semantically definable characteristics which interfere with the mapping of arguments. Zaenen incorporates Dowty’s (1991) Proto-Role entailments into LFG’s linking theory, dispensing with thematic roles and the thematic role hierarchy. Yet, while Zaenen (1993) integrates the semantic entailments neatly into linking by using the binary feature classification of the Lexical Mapping Theory, Kibort (2014) employs a complicated system of numerical markers to map semantic participants to argument positions. Therefore, although I make use of Kibort’s argument positions in this thesis, the theoretical analysis presented in Chapter 6 builds on Zaenen’s original account of incorporating Proto-Role entailments into Lexical Mapping Theory as an alternative to using discrete thematic roles, which is detailed in the following.
3.2.9 Proto-Role entailments

In Dutch, auxiliary selection is a syntactic criterion to distinguish unaccusative from unergative verbs. While unergative verbs, e.g., ‘phone’ in (29-a), select for the auxiliary *hebben* ‘have’, unaccusative verbs select for *zijn* ‘be’, e.g., ‘die’ in (29-b). Zaenen (1993) conducted a detailed study of the interaction between syntax and lexical semantics in Dutch, showing that the auxiliary distinction does not solely brand unaccusativity, but marks more generalizable semantic properties which go beyond intransitive verbs.

\[(29)\]
\begin{align*}
\text{a. } & \text{ De jongen heeft getelefoneerd. (with *hebben*)} \\
& \text{The boy has phoned.}
\end{align*}
\[\text{b. } \text{In dat ziekenhuis zijn (er) veel patienten gestorven. (with *zijn*)} \\
& \text{In that hospital have (there) died many patients.}
\]
\[(\text{Zaenen 1993, 131})\]

Zaenen (1993) found that selection of the auxiliary *hebben* implies control over the action, whereas *zijn* correlates with affected arguments and arguments undergoing a change of state, designating a telic event. This distinction is exemplified by the transitive psych verbs ‘please’ and ‘irritate’ which differ with respect to auxiliary selection. The verb ‘please’ occurs with the auxiliary *zijn* ‘be’, whereas ‘irritate’ selects *hebben* ‘have’ and implies control over the action, compare the examples in (30).

\[(30)\]
\begin{align*}
\text{a. } & \text{Hij heeft me jarenlang geirriteerd. (with *hebben*)} \\
& \text{He has irritated me for years.}
\end{align*}
\begin{align*}
\text{b. } & \text{Dat is me jarenlang goed bevallen. (with *zijn*)} \\
& \text{that has pleased me well for years.}
\end{align*}
\[(\text{Zaenen 1993, 144})\]

In order to account for these lexical semantic generalizations, Zaenen (1993) incorporates Dowty’s (1991) Proto-Roles into linking theory. Instead of working with an ill-defined set of thematic role labels, Dowty’s Proto-Role approach uses just two roles, Proto-Agent and Proto-Patient, which are characterized by a list of lexical semantic entailments, see (31) and (32) respectively (culled from Dowty 1991, 572–574 and Butt 2006, 98f.).\footnote{Working primarily on English, Dowty’s Proto-Role entailments do not take case into account.}

\[(31)\]
\begin{align*}
\text{1. } & \text{A schema for Proto-Agent entailments:} \\
& \text{Proto-Agent entails having a Proximal Location, a Distal Location, a Benefactive, an Agentive, a Benefactive Agent, an Agentive Benefactive, and an Agentive Benefactive Agent.}
\end{align*}
\[(32)\]
\begin{align*}
\text{1. } & \text{A schema for Proto-Patient entailments:} \\
& \text{Proto-Patient entails being the Benefactive of another entity, being the Agentive of another entity,} \\
& \text{being the Benefactive Agent of another entity, being the Agentive Benefactive of another entity,} \\
& \text{being the Agentive Benefactive Agent of another entity, being the Benefactive of another entity,} \\
& \text{being the Agentive Benefactive Agent of another entity,} \\
& \text{and being the Agentive Benefactive Agent of another entity.}
\end{align*}
Chapter 3. Linking Theories

(31) **Proto-Agent entailments**
   a. Volitional involvement in the event or state
      Example: *John* in *John is ignoring Mary.*
   b. Sentience/perception
      Example: *John* in *John sees/fears Mary.*
   c. Causing an event or change of state in another participant
      Example: *unemployment* in *Unemployment causes delinquency.*
   d. Movement (relative to the position of another participant)
      Example: *tumbleweed* in *The rolling tumbleweed passed the rock.*
   e. Independent existence
      Example: *John* in *John needs a new car.*

(32) **Proto-Patient entailments**
   a. Change of state
      Examples: *mistake* in *John made a mistake.*, *error* in *John erased the error.*
   b. Incremental theme
      Example: *apple* in *John ate the apple.*
   c. Causally affected by another participant
      Example: *Mary* in *John kicked Mary.*
   d. Stationary relative to another participant
      Example: *rock* in *The rolling tumbleweed passed the rock.*
   e. Existence not independent of event
      Example: *house* in *John built a house.*

For an argument to qualify as a Proto-Agent or Proto-Patient, it does not necessarily have to entail all the listed properties, but at least one. The distinction between Proto-Agents and Proto-Patients is context-sensitive and drawn on a clause by clause basis. With respect to argument selection, Dowty’s approach proceeds according to the following principle (Dowty 1991, 576).

---

In order to be able to account for languages that have overt case marking, Primus (1999, 2002) reformulates Dowty’s Proto-Role entailments and introduces a third, ‘Proto-Recipient’, role.
3.2. Lexical Mapping Theory

(33) **Argument Selection Principle**
In predicates with grammatical subject and object, the argument for which the predicate entails the greatest number of Proto-Agent properties will be lexicalized as the subject of the predicate: the argument having the greatest number of Proto-Patient entailments will be lexicalized as the direct object.

Zaenen implements Dowty’s argument selection principle and the Proto-Role entailments using the binary feature classification system of standard Lexical Mapping Theory (Zaenen 1993, 150, 152):

(34) **Feature Classification**
1. If a participant has more patient properties than agent properties, it is marked $-r$.
2. If a participant has more agent properties than patient properties, it is marked $-o$.
3. If a participant has an equal number of properties, it is marked $-r$.
4. If a participant has neither agent nor patient properties, it is marked $-o$.

In Zaenen’s approach, the $[\pm o,r]$ marked arguments are directly linked to grammatical functions in keeping with the hierarchy of grammatical functions, see (35) (Zaenen 1993, 151), instead of having to define thematic roles and assuming a hierarchical organization of such roles.

(35) **Association Principles**
order the participants as follows according to their intrinsic markings:
$-o < -r < +o < +r$

order of the GR [grammatical relations] as follows:
$\text{subj} < \text{obj} < \text{obj}_\theta < \text{(obl)}$

Starting from the left, associate the leftmost participant with the leftmost GR it is compatible with.

With respect to auxiliary selection, differences between unaccusative and unergative verbs can now be analyzed straightforwardly: Unaccusative verbs, e.g., ‘die’, have a single argument which standardly has more patient than agent properties and is thus classified as $[-r]$. Therefore, the argument is linked to subj. Unerga-
tives on the other hand, e.g., ‘phone’, have a single argument with more agent than patient properties which is [−o] and also linked to SUBJ. Zaenen (1993) assumes that the selection of the ‘be’ auxiliary is sensitive to subjects with the [−r] feature which explains the auxiliary distinction with respect to unaccusativity. This analysis moreover extends to the alternation observed with the transitive psych predicates in (30). The predicates in (30) both have two arguments. The difference in auxiliary selection can be explained on the basis of the features assigned to the first argument of each predicate. With ‘irritate’, the first argument, i.e., ‘he’, qualifies as a Proto-Agent, thus receiving the [−o] feature. The first argument of ‘please’ however has no Proto-Agent properties and is classified as [−r]. The feature classification leads to its realization as a subj which in turn triggers the selection of the auxiliary zijn instead of hebben.

To conclude the discussion of linking in LFG, the Lexical Mapping Theory is able to account for a variety of different linguistic phenomena on the basis of its binary feature classification and structured mapping principles. Case is not inherently part of the standard Lexical Mapping Theory, but the early linking principles formulated within the LFG framework by Zaenen et al. (1985) account for case marking in Icelandic by assuming that idiosyncratic case marking constrains the linking principles, while default case marking is associated with grammatical functions after linking has taken place. The separation of argument positions and thematic roles suggested by Kibort (2014) presents an elaborate enhancement of the Lexical Mapping Theory with respect to the analysis of argument alternations and other morphosyntactic operations involving a-structure changes. By incorporating Proto-Role entailments into Lexical Mapping Theory, Zaenen (1993) dispenses with thematic roles and the thematic role hierarchy in order to be able to account for the lexical semantic properties of arguments in a more generalizable fashion.

LFG’s standard Lexical Mapping Theory does not account for the role which case marking plays in linking. In contrast, Kiparsky’s (1997) linking theory, which is detailed in the next section, implements case directly. Moreover, the theory accounts for the complex interaction between position, case marking, and grammatical relations in Icelandic (see, e.g., Kiparsky 1997) and provides fundamental ideas for the diachronic account of dative subjects in Icelandic presented in Chapter 6 of this thesis.
3.3 Kiparsky’s linking theory

Over the years, Kiparsky developed an elaborate theory of linking which is able to capture complex interrelations between inflectional morphology, positional constraints and grammatical relations (see Kiparsky 1987, 1988, 1997, 2001). In his theory, case, agreement, and position are seen as equally privileged licensers of grammatical function which may interact with one another in a complex fashion. The theory is inspired by the cross-linguistic observation that there is a diachronic correlation between the loss of inflectional morphology and the development of a rigid word order in Germanic. Kiparsky (1997) takes this correlation to be a unidirectional implication: lack of inflectional morphology implies a fixed word order, but the converse is not true. As an example, Kiparsky (1997) shows that in Old English, position and case functioned as a partly joint, partly independent system to license grammatical relations. However, over time, English lost its morphological case marking, while the functional category I was established firmly, leading to the rise of positional licensing. Icelandic differs interestingly in that the language retained its morphological case system, but employs a fairly fixed word order. Kiparsky’s analysis for licensing grammatical relations in Icelandic presupposes that morphological marking has not been lost in Icelandic, but became recessive in determining grammatical relations. Instead, position has become the dominant licensing constraint. Kiparsky’s linking theory and his approach to the licensing of grammatical relations in Icelandic are presented in the following sections.

3.3.1 Semantic Form and argument structure

Kiparsky’s theory is based on the assumption that syntactic argument structure is projected from semantic content (see, e.g., Dowty 1979, Jackendoff 1983, Foley and Van Valin 1984, Givón 1984). More specifically, Kiparsky employs a level of Semantic Form (SF) in the sense of Bierwisch (Bierwisch 1983, 1986, Bierwisch and Schreuder 1992) which decomposes a predicate into semantic primitives and variables representing the predicate’s arguments, see (36).

\[
\text{(36) } \text{show: } \lambda x \lambda y \lambda z \ [x \text{ cause } \lambda\text{can } [y \text{ see } z]]
\]

The SF in (36) represents the verb *show* which is decomposed into the three semantic primitives *cause*, *can*, and *see* and has three arguments, *x*, *y* and *z*. This representation describes *show* as an event in which an *x* causes another event by
which it is possible for \( y \) to see \( z \). The semantic content of the arguments in the form of thematic roles is extracted via lambda (\( \lambda \)) abstraction over the semantic primitives, resulting in the argument structure left to the SF (\( \lambda z \lambda y \lambda x \)). For example, being the first argument of \texttt{cause}, \( x \) can be interpreted as the thematic ‘agent’ of \texttt{show}. A variable’s depth of embedding moreover represents the corresponding thematic role’s rank on the thematic hierarchy, i.e., the one from (4), in that the less a variable is embedded, the higher it is ranked. This follows from the idea that the order of \( \lambda \)-abstraction reflects semantic depth. In (36), \( x \) is saturated first as it is the least embedded argument. As the least embedded argument, \( x \) represents the highest thematic role which can in turn be linked to the most prominent syntactic function, i.e., the subject.

### 3.3.2 Feature decomposition

In order to link the thematic roles to syntax, Kiparsky introduces the cross-classifying relational case features \( \{\pm H(ighest) \ R(ole)\} \) and \( \{\pm L(owest) \ R(ole)\} \). The case features are important at all levels of grammatical structure. At the level of morphology, the case features classify case and agreement morphemes which are passed on to their stems. At SF and argument structure, Abstract Case features are assigned to thematic roles based on their ranks on the thematic hierarchy and identify their grammatical functions. At morphosyntax, Morphosyntactic Case features are assigned to arguments via inflectional case, agreement morphemes and their structural argument position.

#### 3.3.2.1 Abstract Case

Abstract Case features are assigned to thematic roles according to their relative position on the thematic role hierarchy (Kiparsky 1997, 18):

\begin{align*}
(37) & \quad \text{a. } [+HR] \text{ is assigned to the highest role.} \\
& \quad \text{b. } [+LR] \text{ is assigned to the lowest role.}
\end{align*}

For example, (38) shows the assigned Abstract Case features of the three ordered thematic roles of the verb \texttt{show}. The least embedded role \( x \) is highest on the thematic hierarchy and thus receives the \([+HR]\) feature, whereas \( z \) is \([+LR]\). The middle role of a three-place predicate receives neither \([+HR]\) nor \([+LR]\) (Kiparsky 1997). Being
3.3. Kiparsky’s linking theory

Intrinsically relational, a predicate can have at most one [+HR] role and at most one [+LR] role.

\( \lambda z \) \([+LR]\) \( \lambda y \) \([+HR]\) \( x \text{ cause} \) \( y \text{ can} \) \( z \text{ see} \)

Kiparsky defines an inventory of four Abstract Cases representing the basic grammatical relations which are assigned to thematic roles in the argument structure. The four Abstract Cases along with their feature specifications are shown in (39) (see Kiparsky 1997, 2001).

<table>
<thead>
<tr>
<th>Abstract Case</th>
<th>Feature Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transitive Subject</td>
<td>[+HR, −LR]</td>
</tr>
<tr>
<td>Intransitive Subject</td>
<td>[+HR, +LR]</td>
</tr>
<tr>
<td>Higher (Indirect) Object</td>
<td>[−HR, −LR]/[ ]</td>
</tr>
<tr>
<td>Lower (Direct) Object</td>
<td>[−HR, +LR]</td>
</tr>
</tbody>
</table>

In (38), the \( x \) argument picks out the subject function on the basis of its feature specification. The \( z \) argument corresponds to the lowest role, but is not [+HR], and is linked to the object function. With its empty feature specification, \( y \) is assigned to the indirect object.

According to Kiparsky (1997), Abstract Case features not only allow for the identification of grammatical relations, but also allow for the appropriate representation of valency-changing operations, e.g., passivization. Kiparsky (1997) assumes that valency changing operations are triggered by verbal morphology in the lexicon. Passive affixes demote the [+HR] argument, i.e., the subject, and render the argument ineligible for structural case. This is symbolized by the absence of a bracket around the respective (empty) feature matrix, see (40). Concomitantly, the next highest feature receives [+HR] and becomes the subject of the passive construction.

\( \lambda x \) \([+HR]\) \( \lambda y \) \([+LR]\)

Although Abstract Case features relate grammatical functions to thematic roles, Abstract Case is not a licensing property per se, but ‘only’ constrains Morphosyntactic Case via a set of features (Kiparsky 1997). Instead, the assignment of a thematic...
role to an argument is licensed by the successful unification of case features at all levels, and more specifically, by the unification of Abstract with Morphosyntactic Case features.

### 3.3.2.2 Morphosyntactic Case

Morphosyntactic Case feature complexes are assigned to arguments by the morphological features of their heads, agreement relations, and the structural licensing position they occupy. Kiparsky introduces the following inventory of Morphosyntactic Case features which license structural case (Kiparsky 1997, 20):

<table>
<thead>
<tr>
<th>Structural Case</th>
<th>Morphosyntactic Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative (including “absolutive”)</td>
<td>[ ]</td>
</tr>
<tr>
<td>Accusative</td>
<td>[−HR]</td>
</tr>
<tr>
<td>Genitive, ergative</td>
<td>[−LR]</td>
</tr>
<tr>
<td>Dative, partitive</td>
<td>[−HR,−LR]</td>
</tr>
</tbody>
</table>

Morphosyntactic case features are generally described by negative values, i.e., [−HR]/[−LR], to not restrict a case or position to a specific thematic role. Nominative case is compatible with any kind of thematic role cross-linguistically and can in principle be assigned to any argument, thus receiving the empty feature specification. Accusative case on the other hand is [−HR] because accusatives tend to mark objects and non-agentive roles cross-linguistically. Furthermore, dative and partitive are in general characterized as neither being the highest nor the lowest role and are assigned to indirect objects. Genitive and ergative can not be assigned to the lowest role as they generally occur together with a further, lower-ranked argument.

The Morphosyntactic Case features are moreover used to classify agreement patterns and position. Nominative is specified as the unmarked case, accounting for the cross-linguistic generality that subject agreement is typically assigned to the nominative argument. Other types of agreements are also possible. For instance, object agreement in the form of accusative agreement is classified as [-HR].

With respect to internal argument positions, the Morphosyntactic Case features are assigned as follows (Kiparsky 1997, 20):

(42) **Positional Licensing**

a. Complement positions are [−HR].

b. Non-final complement positions are [−LR].
Specifier positions remain featureless because they are the positional equivalents of nominative case (cf. Kiparsky 1997). Kiparsky generally assumes that the basic licensing position for subjects is in SpecIP.  

Arguments which have been assigned Morphosyntactic Case are associated with thematic roles, which have been marked via Abstract Case, in accord with the following principles (Kiparsky 1997, 20):

\[(43)\]
\[\text{a. Unification: } \text{Associated feature matrices must be non-distinct (one must not have a plus value where the other has a minus value).}\]
\[\text{b. Specificity (Blocking, ‘Elsewhere’): Specific rules and morphemes block general rules and morphemes in the shared contexts.}\]

The combined effect of Unification and Specificity is that thematic roles are linked to the argument with the most specific Morphosyntactic Case features compatible with the thematic role’s Abstract Case. Unification ensures that entities from different case levels, i.e., Abstract and Morphosyntactic Case, are only associated with one another when their feature specifications are non-contradictory. Specificity allows for the application of more specific rules and language-specific constraints before more general and universal principles which may even be blocked. Combining the Abstract Case features of \textit{show} as given in (38) with Morphosyntactic Case results in the linking scheme shown in (44).

\[(44)\]
\[
\begin{align*}
\text{Thematic roles with Abstract Case} & : \quad \lambda x \quad [| +HR ] \quad | \quad \lambda y \quad [| ] \quad | \quad \lambda z \quad [| +LR ] \\
\text{Morphosyntactic Case} & : \quad [| -HR ] \quad [| -LR ] \quad [-HR ]
\end{align*}
\]

\textit{show}:

\text{English is a language which has only very little morphology, but relies on positional constraints mediated by subject agreement to mark grammatical relations. The empty Morphosyntactic Case feature specification of } x \text{ in (44) results from the SpecIP position and subject agreement in English which are unified with the Abstract Case features of } x, \text{ licensing the linking of } x \text{ (the ‘agent’) to the subject function.}

---

\textsuperscript{3}Kiparsky (2001) also accounts for subject licensing in SpecVP in order to explain phenomena which impose restrictions on a language’s morphosyntax with respect to an argument’s animacy and specificity, such as e.g. the Definiteness Effect in English (and other Germanic languages). Therefore, he assigns a further feature [± SPECIFIC], but aims for a future, more principled and general account.
y and z are both in complement positions of the verb, but, for the successful unification with the Abstract Case feature [+LR], z needs to be specified as the final complement position [−HR] and is linked to the object function. The y argument is the non-final complement and thus [−HR, −LR], unifying with the empty Abstract Case feature, designating the indirect object function.

3.3.3 Quirky case and argument scrambling

Across languages, different types of case systems are employed. Kiparsky (1997) assumes that the cross-linguistic variation of case systems is the result of two factors: (i) languages differ with respect of their inventory of morphological cases, and (ii) languages allow for different instances of case feature mismatches. Case feature mismatches occur when the Abstract Case features of an argument fail to unify with the Morphosyntactic Case features assigned to it, either by morphology or by position. The Morphosyntactic Case features which cause a mismatch are referred to as being recessive by Kiparsky. However, certain Morphosyntactic Case features may ‘repair’ a mismatch induced by another feature in order to license the linking of an argument. These case features are referred to as dominant.

Quirky case is an example for a case feature mismatch due to morphology. Kiparsky (1997) exemplifies his approach to quirky case marking via an analysis of dative experiencer subjects in Icelandic. For example, the predicate líka ‘like’, whose argument structure and SF are represented in (46), has a dative experiencer subject in conjunction with a nominative object, see (45).

(45) Mér líka þessir bílar.
    Me-DAT like-PL these cars
    ‘I like these cars.’ (Kiparsky 1997, 22)

Kiparsky (1997) assumes that the dative case marking of experiencer/goal roles in Icelandic is an instance of ‘inherent’, i.e., quirky, case marking. In Kiparsky’s system, inherent case marking prespecifies certain thematic roles via the Morphosyntactic Case features [−LR] and/or [−HR] (similar to the analysis of idiosyncratic dative case in Zaenen et al. (1985)). In (46), the logical subject x, i.e., the experiencer argument, is pre-specified with the dative case features [−HR, −LR]. The y argument is marked as nominative by case and agreement, and receives the empty feature specification at the level of morphology. However, the morphological case pre-specification of the x argument with the [−HR] feature precludes unification with the abstract
case feature [+HR], and the experiencer argument can not be directly licensed as the subject. Kiparsky (1997) assumes that morphological case is recessive in Icelandic, while position is dominant and can override the morphological case features. With positional licensing being dominant, the experiencer argument \( x \) which is in SpecIP is allowed to be linked to the subject function in Icelandic.

\[
\text{líka: } \lambda y \lambda x \text{ [x LIKE y]} \quad \text{Thematic roles and SF}
\]

\[
\begin{array}{c|c|c}
 \text{Morphological case} & [+LR] & [+HR] \\
 \text{Abstract case} & [+LR] & [+HR] \\
 \text{Position} & [+LR] & [+HR]
\end{array}
\]

German is an example for the reverse scenario. In German, morphological case is dominant, whereas position is recessive. This is exemplified by the possibility to have argument scrambling in German, see the examples in (47), because positional mismatches are rescued by morphological case marking.

\[
\begin{align*}
(46) & \quad \text{lïka: } \lambda y \lambda x \text{ [x LIKE y]} \quad \text{Thematic roles and SF} \\
 & \quad | \quad | \\
 & \quad | \quad | \\
 & \quad [+LR] \quad [+HR] \\
 & \quad | \quad | \\
 & \quad [+LR] \quad [+HR] \quad \text{Morphological case} \\
 & \quad | \quad | \\
 & \quad [+LR] \quad [+HR] \quad \text{Abstract case} \\
 & \quad | \quad | \\
 & \quad [+LR] \quad [+HR] \quad \text{Position}
\end{align*}
\]

(47) a. Mir gefallen diese Autos.
   Me-DAT like-PL these cars
   ‘I like these cars.’

b. Diese Autos gefallen mir.
   (Kiparsky 1997, 23)

According to Kiparsky (1997), morphological case as well as position were recessive features of Old English. For example, the attested data for Old English (see, e.g., Allen 1995) shows that oblique experiencers only allowed for a subject analysis when they were realized as the first argument of the clause, despite Old English being a freely scrambling language. To license such configurations, Kiparsky (1997) assumes that position and case must have interacted in a complex, but yet systematic, fashion as partly joint, partly independent licensers. Over the course of time, English lost its morphological case system and positional licensing was established. Kiparsky connects the rise of positional licensing to the reorganization of the syntactic category I during Middle English, where morphology and case no longer functioned as an interacting system.
Kiparsky’s linking theory is able to account for the complex interrelation between case morphology and word order with respect to the licensing of grammatical relations from both a diachronic and synchronic perspective. However, his theory of linking assumes that dative experiencer subjects in Icelandic are an instance of idiosyncratic case marking, not accounting for the quite regular interrelation between datives and goal/experiencer roles. Moreover, as Svenonius (2002, 2006) has shown (see Section 2.7 from Chapter 2), dative case marking correlates with the structure of events in Icelandic. Per se, there exists no linking theory which explicitly factors in event semantic considerations. In this thesis, I draw on Ramchand’s (2008) first-phase syntax which presents a concise theory of event semantics at the syntax-semantics interface.

3.4 The first-phase syntax and event semantics

The first-phase syntax approach by Ramchand (2008) is based on the originally ‘generative-constructivist’ view that the syntactic projection of arguments is based on event semantics, following inter alia Hale and Keyser (1993) and Borer (2005). In Ramchand’s approach, syntax correlates directly with event semantics in that meaning is systematically built up in the syntax which is where event participants and event structure are represented. The first-phase syntax contains three components which correspond to the three subevents into which an event can be maximally decomposed: a causing or initiation subevent, a process-denoting subevent and a subevent denoting a result state. Each of these subevents has its own syntactic projection and is tied to a hierarchical embedding order as shown in (48) (taken from Ramchand 2008, 39).

The initiation projection (initP) as represented in the syntactic tree in (48) displays the outer causational projection which introduces the causation event and the external argument. The external argument is licensed in the specifier position of initP and is the ‘subject’ of cause or causer of the event which Ramchand (2008) refers to as INITIATOR. The tree’s central component is procP, the process projection which is in complement position of the init head. procP is present in every dynamic predicate representing change through time. The process projection moreover specifies the nature of the change or process described by the verb and licenses the event participant undergoing a change or process, i.e., the ‘subject’ of process or UNDERGOER (Ramchand 2008). In general, the runtime of the complete event is
homomorphic to the runtime of \textit{proc}P. The presence of a \textit{proc}P does not necessarily imply the existence of an \textit{init}P. An \textit{init}P is only needed when the verb expresses a state that causes or initiates the process. The result projection (\textit{res}P) represents the lowest projection in (48) as complement of \textit{proc} and expresses an event’s result state, licensing the participant that eventually holds the result state: the \textsc{resultee} or ‘subject’ of result. Again, a \textit{res}P is not obligatory with a \textit{proc}P. The \textit{res}P is only present when the lexical predicate explicitly expresses a result state. The ‘subject’ or ‘theme’ of each projection licensed in the respective specifier positions, i.e., the \textsc{initiator}, the \textsc{undergoer} and the \textsc{resultee}, form the set of core argument roles proposed by Ramchand (2008).

(48)

\begin{itemize}
  \item \textit{init}P (causing projection)
  \item \textit{proc}P (process projection)
  \item \textit{res}P (result proj)
\end{itemize}

The first-phase syntax is semantically interpreted in a ‘post-Davidsonian’ manner, meaning that verbal heads in the decomposition represent eventuality descriptions of which each one has a single open position for a predicational subject (Ramchand 2008). Furthermore, the semantics of event structure and event participants are interpreted directly off the structure by means of a regular and predictable semantic combinatoric process. The ‘leads-to’ relation (\(\rightarrow\)) is the basic primitive rule...
of event composition in Ramchand’s system (Ramchand 2008, 44):

\[(49) \quad \text{Event Composition Rule}\]
\[e = e_1 \rightarrow e_2: \text{e consists of two subevents, } e_1, e_2, \text{ such that } e_1 \text{ causally implicates } e_2\]
\[\text{(cf. Hale and Keyser 1993)}\]

Moreover, there are two general primitive predicates over events which represent the basic subevent types: states and processes, whereas a process is defined as “an eventuality that contains internal change” (Ramchand 2008, 44). Both the initiation event and the result event are states with their interpretation resulting from their hierarchical position in the syntactic structure given in (48). If a state is in \textit{init}, then it is interpreted as causing the process event; if in \textit{res}, the state is interpreted as being caused by the process. Based on these assumptions and the event composition rule, Ramchand (2008, 44) defines the following two derived predicates over events.

\[(50) \quad \text{IF } \exists e_1, e_2 \left[\text{State}(e_1) \& \text{Process}(e_2) \& e_1 \rightarrow e_2\right], \text{ then by definition }\]
\[\text{Initiation}(e_1)\]
\[(51) \quad \text{IF } \exists e_1, e_2 \left[\text{State}(e_1) \& \text{Process}(e_2) \& e_2 \rightarrow e_1\right], \text{ then by definition }\]
\[\text{Result}(e_1)\]

The semantic denotations of the event heads \textit{init}, \textit{proc} and \textit{res} can then be built up in terms of lambda notation, independently of any lexical-encyclopedic information. Lexical-encyclopedic content is projected onto the event heads by particular lexical items and is represented as \textit{init’}, \textit{proc’} and \textit{res’} in the lambda denotations. Accordingly, the \textit{res} head which is placed lowest in the first-phase syntax describes a state that has a specific state ‘holder’, i.e., the \textit{RESULTEE}, in its specifier position and is interpreted as follows (Ramchand 2008, 45):

\[(52) \quad [[[\text{res}]]] = \lambda P \lambda x \lambda e[P(e) \& \text{res’}(e) \& \text{State}(e) \& \text{Subject (x,e)}]\]

The syntactic projection of \textit{res}, the \textit{res}P, is the state description in the complement position of the \textit{proc} head. The \textit{proc} head in turn introduces a process and licenses the \textit{UNDERGOER} of the process in its specifier position (Ramchand 2008, 45):

\[(53) \quad [[[\text{proc}]]] = \lambda P \lambda x \lambda e_1 e_2[P(e_2) \& \text{proc’}(e_1) \& \text{Process}(e_1) \& e = (e_1 \rightarrow e_2) \& \text{Subject (x,e_1)}]\]
3.4. The first-phase syntax and event semantics

The verbal head \(init\) is the highest verbal head of the first-phase syntax and describes a state which has to combine with the lower structure in order to become an initiating event. The specifier position of \(init\) is filled by the INITIATOR. The interpretation of the \(init\) head is formalized in the following (Ramchand 2008, 45):

\[
[[init]] = \lambda P\lambda x\lambda e \exists e_1,e_2 [P(e_2) \& init'(e_1) \& State(e_1) \& e = e_1 \rightarrow e_2 \& Subject(x,e_1)]
\]

Apart from the three core arguments in specifier position (INITIATOR, UNDERGOER, RESULTEE), Ramchand (2008) assumes that rhematic material can by definition occur in complement position to an event head (cf. ‘Figures’ vs. ‘Grounds’ in the sense of Talmy 1978):

“RHEMES, and as an important subcase PATHS, do not describe elements that are referentially individuated and predicated over within an event topology, but those that actually construct the specific predicational property (static or dynamic) that the ‘subject’ is asserted to have.”

(Ramchand 2008, 46)

If a RHEME argument (a PP or DP) occurs in an eventive head’s complement position, e.g., in complement position of \(proc\) as in (55) (Ramchand 2008, 46), no separate event is introduced, but the event introduced by the eventive head is further described by the RHEME.

\[
(55)
\]

\[\begin{align*}
&\text{initP} \\
&\quad \text{init} \quad \text{procP} \\
&\quad \quad \text{proc} \quad \text{DP} \\
&\quad \quad \quad \text{RHEME}
\end{align*}\]
RHEMES in the complement position of proc are PATHS as they are homomorphically related to the position/state or motion/change of the UNDERGOER. In complement position of res, RHEMES are the GROUNDS of result or RESULT-RHEMES with an inherent nongradable property further describing the result state (Ramchand 2008).

The core arguments may furthermore occupy more than one structural position and form a composite role. For example, when one argument holds an initiational state and undergoes a change or process at the same time given that the changing property is homomorphic with the event trace of the proc event, then the argument involved is an UNDERGOER-INITIATOR. Another composite role based on the same principles, but with a result instead of an initiational state, is the RESULTEE-UNDERGOER.

3.4.1 Dynamic predicates

As an example for a dynamic transitive verb which has an INITIATOR subject as well as an UNDERGOER object, Ramchand (2008, 65) provides the following syntactic tree for the English predicate push illustrated by the sentence ‘Jon pushed the cart’:

(56)

```
initP
  /\  \\
John
     /\  \\
init      procP
  /\          /\  \\
push       the cart
     /\          /\  \\
proc       XP
     /\  \\
< push >
```

The verb phrase built from push in (56) contains a subject DP which initiates the dynamic event as well as a different DP undergoing a change of location which is the object of the phrase. The lexical-encyclopedic content of push is projected onto both of the eventive heads and push is thus listed as an [init, proc] verb in Ramchand (2008).
The dynamic transitive predicate *eat* as given in (57) is also an \([\text{init}, \text{proc}]\) verb, but the \textit{initiator} and the \textit{undergoer} are both occupied by the subject DP. This reflects the status of ‘Mary’ as being continuously experiencing and undergoing the process of eating from initiation until the end. The object DP ‘the mango’ is neither \textit{initiator}, \textit{undergoer}, nor \textit{resultee}, but serves to describe the process further in defining its own progress through its scalar structure and is thus a rhematic \textit{path} object.

\[(57)\]

\[
\begin{array}{c}
\text{Mary} \\
\text{initP} \\
\text{init} \\
\text{eat} \\
< \text{Mary} > \\
\text{procP} \\
\text{proc} \\
< \text{eat} > \\
\text{DP} \\
\text{the mango}
\end{array}
\]

The English ditransitive predicate *give* contains all of the three eventive heads as *give* refers to a dynamic event \((\text{procP})\) that follows an initiation event \((\text{initP})\) and gives rise to a telic and punctual interpretation with a definite result \((\text{resP})\), see the structure given in (58) for the sentence ‘Alex gave the ball to Ariel’ (Ramchand 2008, 103). Ramchand (2008) positions the directional preposition *to* under *res* in (58) because she assumes that *to* by itself contains a *res* feature in its lexical entry and may thus combine with *give*. Moreover, ‘the ball’ is the \textit{resultee-undergoer}, while ‘Alex’ is the \textit{initiator} argument and ‘to Ariel’ is rhematic material.

In the double object version (‘Alex gave Ariel the ball’) illustrated in (59) however, *give* identifies *res* directly and must take a stative complement itself. Ramchand (2008) formalizes this complement as a PP and makes reference to Harley’s (2002) analysis which assumes a null possessional \P head with a DP complement for the double object alternant of *give*. In contrast to the *to*-variant where the direct objects is both \textit{undergoer} and \textit{resultee}, the double object version licenses the first ob-
ject ‘Ariel’ as the RESULTEE, but the UNDERGOER position remains unfilled given that the semantics of ‘Ariel’ are not fully consistent with an UNDERGOER role, but the event described is still a dynamic one.

(58) Alex gave the ball to Ariel.
3.4. The first-phase syntax and event semantics

(59) Alex gave Ariel the ball.

Ramchand’s first-phase syntax does not only account for dynamic predicates, but also applies to stative predicates as is shown next.

3.4.2 Stative predicates

The main difference between dynamic and stative verbs lies in the absence of a proc projection for stative verbs. According to Ramchand (2008), stative verbs are the result of an init head selecting rhematic material instead of a procP complement. The selected rhematic material can either be a DP, an AP, or a PP, see for example the DP/NP ‘nightmares’ in (60) which exemplifies the structure of the stative predicate fear in ‘Katherine fears nightmares’ (Ramchand 2008, 106). The init head is employed for stative predicates because, similarly to dynamic predications, the stative eventuality described by the predicate is based on or caused by properties inherent to the DP argument in the specifier position. For example in (60) the state
of fearing nightmares arises on the grounds of Katherine’s personality. With stative predicates, the INITIATOR is interpreted as a state holder and not as an agent.

(60) **State:** \( x \text{ fears } y \)

In sum, Ramchand’s first-phase syntax allows for a generic approach to verbal meaning in that the semantics of event structure and the event participants are built up compositionally in the syntax instead of being explicitly stated in the lexical entries of verbs. Events denoted by dynamic predicates consist of maximally three projections representing subevents: an initiation event, a process, and a result state. The initiation event introduces the INITIATOR argument which is licensed in specifier position of \( \text{initP} \), the process event takes an UNDERGOER argument licensed in specifier position of \( \text{procP} \), and the result state licenses the RESULTEE in \( \text{resP} \). The INITIATOR is causing the initiation event, the UNDERGOER is undergoing a change or process, and the RESULTEE holds the result state. With dynamic predicates, \( \text{procP} \) and \( \text{resP} \) can additionally have a RHEME argument in complement position. A RHEME in complement position of \( \text{procP} \) is interpreted as PATH, whereas a RHEME in complement position of \( \text{resP} \) further describes the GROUNDS of result. Syntactic arguments may moreover be mapped onto more than one event participant, resulting in composite roles. In contrast to dynamic predicates, stative predicates do not have a \( \text{procP} \). Instead, Ramchand (2008) assumes that an init head selects rhematic material with stative predicates, resulting in the interpretation of the INITIATOR as a state holder, with the RHEME further describing the stative eventuality. Table 3.1 provides an overview of the event structure of the different types of dynamic verbs discussed by Ramchand (2008).\(^4\)

\(^4\)Core arguments that are occupied by the same entity are coindexed.
Table 3.1: Verb classes in the first-phase syntax (taken from Ramchand 2008, 108).

### 3.5 Summary and conclusion

This chapter has introduced and discussed various theories of linking arguments to grammatical relations and a theory at the syntax-semantics interface which situates syntactic arguments in the event structure. The presented linking theories differ with respect to the particularities employed for linking. The standard Lexical Mapping Theory introduced in Section 3.2 accounts for the mapping between thematic roles and grammatical functions via a binary feature classification system. Kibort (2014) extends the Lexical Mapping Theory by dissociating thematic roles from arguments to provide a clearer formalism for argument changing operations (Section 3.2.8). Zaenen (1993) moreover incorporates Proto-Role entailments into Lexical Mapping Theory, accounting for lexical semantic generalizations which could not be captured on the basis of thematic roles (Section 3.2.9). However, although implemented at early stages of the theory by Zaenen et al. (1985), case marking is not integrated directly into the standard Lexical Mapping Theory. Section 3.3 presented Kiparsky’s linking theory which accounts for the diachronic trade-off between inflectional morphology and word order with respect to the licensing of grammatical relations. The theory uses three linkers, i.e., case, agreement, and position, which may interact in a complex fashion to license grammatical relations (Kiparsky 1987, 1988, 1997).

<table>
<thead>
<tr>
<th>[init, proc]</th>
<th>init. proc</th>
<th>initiator, undergoer</th>
<th>drive, push, paint</th>
</tr>
</thead>
<tbody>
<tr>
<td>I Transitive</td>
<td>init. proc</td>
<td>initiator, path</td>
<td>eat, read, paint</td>
</tr>
<tr>
<td>II Intransitive</td>
<td>init. proc</td>
<td>initiator, undergoer</td>
<td>run</td>
</tr>
<tr>
<td>[init, proc, res]</td>
<td>init. proc, res</td>
<td>initiator, undergoer</td>
<td>throw, defuse</td>
</tr>
<tr>
<td>III Transitive</td>
<td>init. proc, res</td>
<td>initiator, undergoer</td>
<td>enter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>resultee</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>result-rheme</td>
<td></td>
</tr>
<tr>
<td>IV Intransitive</td>
<td>init. proc, res</td>
<td>initiator, undergoer</td>
<td>arrive, jump</td>
</tr>
<tr>
<td></td>
<td></td>
<td>resultee</td>
<td></td>
</tr>
<tr>
<td>V Ditransitive</td>
<td>init. proc, res</td>
<td>initiator, undergoer</td>
<td>give, throw</td>
</tr>
<tr>
<td></td>
<td></td>
<td>resultee</td>
<td></td>
</tr>
<tr>
<td>[proc]</td>
<td>proc</td>
<td>undergoer</td>
<td>melt, roll, freeze</td>
</tr>
<tr>
<td>[proc, res]</td>
<td>proc, res</td>
<td>undergoer, resultee</td>
<td>break, tear</td>
</tr>
<tr>
<td>VII Intransitive</td>
<td>proc, res</td>
<td>undergoer, resultee</td>
<td></td>
</tr>
</tbody>
</table>


Svenonius (2002) showed that dative case marking in Icelandic is the result of particular event structural configurations (cf. Chapter 2). The first-phase syntax by Ramchand (2008), which was detailed in Section 3.4, is an approach to event decomposition which assumes that syntax is directly correlated with event semantics, providing essential ideas for a theory of linking which incorporates event semantics as developed within the scope of this thesis.

The core insights gained from the presented linking theories which are relevant for the theory of linking established in this thesis are listed in the following:

- **Binary features** allow for the cross-classification of arguments, grammatical relations, thematic roles, position, case, and agreement.

- **Argument positions** account for changes at argument structure and dual linking possibilities between semantic participants and grammatical relations.

- **Proto-Role entailments** capture more general lexical semantic properties of arguments than discrete thematic roles on the basis of Proto-Agent and Proto-Patent properties.

- **The rise of positional licensing**: Case and position may interact in a complex fashion to license grammatical relations. Moreover, case marking may become recessive or even be lost over time, whereas position becomes dominant for licensing grammatical relations.

- **Event structure**: Dynamic predicates license three event participants: INITIATOR, UNDERGOER, and RESULTEE. In addition, dynamic predicates may take a RHEME argument which describes a PATH if in complement position of the process, or the GROUNDS of result if in complement position of the result state. Stative predicates have a state HOLDER and a RHEME, with the RHEME further describing the state.

While this chapter presented the theoretical foundations for the linking analysis presented in Chapter 6, the following chapters will investigate the diachronic interaction between dative subjects, voice, lexical semantics, event semantics, and word order in Icelandic, in turn providing the empirical foundations for the theoretical analysis.
Chapter 4

Dative subjects and event semantics

4.1 Introduction

This chapter investigates the diachronic interaction between dative subjects and the factors voice, thematic roles and event structure in the Icelandic Parsed Historical Corpus (IcePaHC, Wallenberg et al. 2011). As discussed at length in Chapter 2, lexical semantic factors generally condition the occurrence of dative case in Icelandic: Dative case marking correlates with thematic goal and experiencer roles (see, e.g., Zaenen et al. 1985, Maling 2001, 2002, Jónsson 2003) and dative themes moreover occur regularly with motion verbs (Maling 2002). However, the interrelation between dative case and particular thematic roles is not regular enough to account for dative case marking solely on the basis of thematic roles. For example, dative case marked goals and themes differ from one another with respect to voice. Under middle formation, only goal arguments retain their dative case marking, whereas the dative case marking is lost with themes. On these grounds, Svenonius (2002) proposes an alternative approach to case marking in Icelandic (see Section 2.7 in Chapter 2), suggesting that case marking can be explained on the basis of event semantics. Still, the exact particularities which license dative subjects in Icelandic have not yet been brought to light.

Moreover, the continuous existence of a monolithic dative subject construction with stable lexical semantics during the Icelandic diachrony has been taken as evidence for the Oblique Subject Hypothesis which assumes that dative subjects are
a common Proto-Indo-European inheritance (see, e.g., Barðdal et al. 2012 and Section 2.5 in Chapter 2). Yet, there is clear evidence for Indo-Aryan, a related, but substantially older, subbranch of Indo-European, which shows that lexical semantic change led to the emergence of dative subjects over time (cf. Deo 2003, Butt and Deo 2013). Furthermore, the Icelandic system is currently changing with respect to dative subjects and lexical semantics in that dative case is becoming increasingly associated with experiencer semantics, a process which has been dubbed as ‘Dative Substitution’ (see, e.g., Smith 1996, Jónsson 2003, Barðdal 2011).

The goal of the investigations presented in this chapter is two-fold: First, the factors which condition the occurrence of dative subjects in Icelandic have to be identified in order to provide a window of understanding of the complex interrelation between case and grammatical relations in the language. Therefore, the interaction between dative subjects and the conditioning factors as proposed by the previous literature, i.e., thematic roles, voice, and event semantics, is examined carefully with the data available in IcePaHC. Second, in order to shed more light on whether dative subjects have been inherited from an earlier language stage or are in fact a historical innovation, I investigate whether the distribution of dative subjects with respect to the conditioning factors remained stable over time in IcePaHC. Investigating the diachronic interaction between dative subjects and thematic roles, voice, and event semantics in IcePaHC is by no means a trivial task given that multifactorial interactions within a complex set of data have to be taken into account. On these grounds, I visualized the data using the glyph visualization as presented in Schätzle and Sacha (2016) which enables an interactive and exploratory access to multiple interactions at different levels of detail while providing an overview. This immensely facilitates the analysis process.

The corpus investigations show that dative case marking indeed correlates with particular event structure configurations and that the diachronic distribution of dative subjects has been changing in the history of Icelandic. In particular, the use of dative subjects increases in the corpus over time, with a significant increase of dative subjects together with stative experiencer predicates in the period post-1900. The striking increase of dative subjects with experiencer semantics moreover correlates with an increasing use of verbs carrying middle morphology. This increase can not be attributed to Dative Substitution, but is taken to be the result of lexical semantic change effectuated by middle formation.
4.2 Event structure and dative case

This chapter proceeds as follows: In Section 4.2, I briefly review Svenonius’ event structural analysis of dative case in Icelandic and the first-phase syntax approach to event decomposition by Ramchand (2008) which is used for the demarcation of verbs according to their event structure in my investigations. This is followed by the first part of the corpus study in Section 4.3 which investigates the diachronic distribution of dative subjects with respect to the overall distribution of case in IcePaHC and the interaction between dative subjects and voice, providing the empirical basis for the event structure investigation. Then, I provide an intermediate analysis of the data which is relevant for the second part of the corpus study in Section 4.4. In Section 4.5, the verb classification and annotation employed for the investigation of event structure is introduced. The second part of the corpus study is presented in Section 4.6 which provides the relevant findings with respect to the interaction between dative subjects, event structure, and voice which were obtained by means of the glyph visualization. Section 4.7 summarizes the relevant findings of the corpus investigations and concludes the chapter.

4.2 Event structure and dative case

The investigation of the interrelation between dative subjects and event structure presented in this chapter is motivated by Svenonius’ (2002) observation that dative case marking correlates with particular event semantics in Icelandic. Decomposing the verb phrase into two components, \( v \) and V, Svenonius (2002) postulates that objects are marked with dative case when the two subevents introduced by \( v \) and V do not have the same temporal extension. For example, verbs of accompanied motion, e.g., \textit{kasta} ‘throw’, which have a dative marked theme object, describe an event in which the causing subevent in \( v \) does not completely overlap with the subevent by V, describing the movement of the object, as the causer does not accompany the moving object until its endpoint. Moreover, motion verbs in which the movement of the object is independent of the action of an agent or causer, e.g., \textit{sökkva} ‘sink’, take a dative object. Svenonius’ theory furthermore applies to the dative goal arguments of ditransitive predicates such as \textit{gefa} ‘give’ which describes an event in which the giving event and the event in which the goal argument receives the entity given do not overlap completely. However, the event structural conditions licensing dative case with experiencer subjects are not discussed in detail by Svenonius.
The first-phase approach by Ramchand (2008) for event decomposition allows for a nuanced understanding of the relation between syntactic arguments and event structure and constitutes the theoretical basis for the event structure classification established in this chapter. In the first-phase syntax, meaning is systematically constructed as part of the syntax. Therefore, events can be decomposed into maximally three subevents in the first-phase syntax as presented in (1) where each subevent has its own syntactic projection.

(1)

The hallmark of dynamic predicates is the \textit{procP} which represents a process subevent. Dynamic events may moreover have a causing or initiating subevent (\textit{initP}) and a result state (\textit{resP}). The \textit{init} projection introduces the external argument, i.e., the causer of the event which is the \textsc{initiator}. The event participant which undergoes a change or process, i.e., the \textsc{undergoer}, is licensed in the specifier position of the \textit{proc} projection. The \textit{res} projection licenses the event participant which holds the result state, i.e., the \textsc{resultee}. Apart from the three core arguments in specifier position, \textsc{rhemes} may occur in the complement position of the event heads \textit{proc} or \textit{res} and further describe the respective subevent.

Stative predicates do not contain a \textit{proc} projection, but consist of an \textit{init} projection only (Ramchand 2008). As Ramchand (2008) points out, the \textsc{initiator} does
not cause a process with stative predicates, but internally causes the stative eventuality on the basis of its own properties. Thus, the INITIATOR is interpreted as a state holder with stative predicates. The init head of stative predicates may furthermore select a RHEME argument which further describes the stative eventuality, see the event structure given in (2).

\[
\text{init} P \\
\rightarrow \text{INITIATOR} \\
\rightarrow \text{init} \begin{array}{c} \text{RHEME} \end{array}
\]

The decomposition of verbs into events and event participants proposed by Ramchand (2008) sets the stage for the investigation of the interrelation between event semantics and dative subjects in IcePaHC. The next section provides the empirical foundation for this investigation by presenting the diachronic overall distribution of dative subjects in the corpus.

4.3 Corpus study I: The diachronic distribution of dative subjects

This section examines the diachronic distribution of dative subjects in IcePaHC with respect to the overall developments of case marking patterns and moreover investigates the interaction between dative subjects and voice. For this purpose, I extracted the relevant constructions from the IcePaHC dataset introduced in Chapter 2 via a Perl script and calculated the respective occurrence frequencies. The obtained frequencies were divided into time periods as suggested by the existing literature for historical Scandinavian (Haugen 1984), see, e.g., the first column in Table 4.1. Moreover, I conducted χ²-tests to examine whether the observed distributions differ from what could be expected given the overall distributions of the constructions in the whole corpus. Statistical significances are indicated via p-values which are provided along with the frequency calculations with p<0.05 *, p<0.01 **, and p<0.001 ***, see, e.g., the last column in Table 4.1.
4.3.1 Subject and object case

As was established in Chapter 2, subjects are mostly nominative in Modern Icelandic, but non-nominative subjects, and in particular dative subjects, are a regular part of the language. To a lesser extent, Icelandic has accusative and genitive subjects. Objects are usually accusative, but dative and genitive objects are also possible. Furthermore, nominative objects generally occur when the subject is marked with a non-nominative case. As a starting point for my investigations, I tested whether these generalities are found in the distributions of subject and object case in IcePaHC. Therefore, I calculated the occurrence frequencies of the different subject and object cases across the investigated time periods. Table 4.1 shows the distribution of subject case marking in all matrix declarative sentences in IcePaHC.

<table>
<thead>
<tr>
<th>Subject case</th>
<th>Period</th>
<th>NOM</th>
<th>DAT</th>
<th>ACC</th>
<th>GEN</th>
<th>Total</th>
<th>χ²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1150–1349</td>
<td>95.0%</td>
<td>3.9%</td>
<td>0.9%</td>
<td>0.2%</td>
<td>13718</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1350–1549</td>
<td>95.9%</td>
<td>3.2%</td>
<td>0.8%</td>
<td>0.1%</td>
<td>18349</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>1550–1749</td>
<td>95.2%</td>
<td>3.7%</td>
<td>0.9%</td>
<td>0.1%</td>
<td>11210</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1750–1899</td>
<td>95.3%</td>
<td>3.8%</td>
<td>0.7%</td>
<td>0.1%</td>
<td>11257</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1900–2008</td>
<td>93.1%</td>
<td>5.8%</td>
<td>1.1%</td>
<td>0.1%</td>
<td>10861</td>
<td>***</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>95.0%</td>
<td>4.0%</td>
<td>0.9%</td>
<td>0.1%</td>
<td>65394</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Diachronic distribution of subject case in IcePaHC.

The ‘Total’ column in Table 4.1 provides information about how the 65394 data points contained in the dataset, i.e., all matrix declarative clauses with a distinctly case marked subject from IcePaHC, are distributed across the time periods. The last row (‘Average’) gives the shares of subject cases across the total number of attested sentences. Those shares serve as a guideline for the evaluation of the χ²-measures as they represent the average distribution of the different subject cases in all matrix declarative clauses from IcePaHC.

The second time period (1350-1549) differs significantly (*** p<0.001) from the expected distribution. However, compared to the average distribution, the shares only deviate marginally. The statistically significant difference could be owed to the fact that the second time period contains more clauses than the others, see the ‘Total’ column. A previous study on V1 (verb-first) order in IcePaHC has moreover uncovered that the third time period (1550-1749) is affected by a genre effect inherent to the corpus (Butt et al. 2014). The third time stage is mainly represented by
4.3. Corpus study I: The diachronic distribution of dative subjects

Religious texts and biographies, while the narrative texts (Sagas and modern fiction) are dominant in the other time stages. The deviating genres caused the percentages in the third time period to diverge from the overall developments in Butt et al.’s (2014) study. Genre affects the distribution of subject case only marginally, but the genre effect in IcePaHC will become more evident throughout the studies presented in this thesis.

Nevertheless, there are statistically significant deviations in IcePaHC regarding subject case. Subjects are most often nominative throughout the corpus (95%), and to a largely lesser extent marked by datives (4%). Accusative subjects are rare (0.9%) and genitive subjects are as good as nonexistent (0.1%). This corresponds to what is known about subject case marking in Icelandic (see, e.g., Thráinsson 2007). Those frequencies remain stable across the periods before 1900. All the same, nominative subjects are decreasing significantly from 95.3% to 93.1% in the period post-1900, while dative subjects are on the increase from 3.8 to 5.8%. Accusative subjects increase only marginally (from 0.7 to 1.1%) and genitive subjects remain stable.

Next, I examined the distribution of object case in IcePaHC. Therefore, I extracted all declarative main clauses that have a case marked direct object in conjunction with subject case marking. In total, my search yielded 18632 clauses for which I give the frequency distributions of object case in Table 4.2.

<table>
<thead>
<tr>
<th>Period</th>
<th>NOM</th>
<th>DAT</th>
<th>ACC</th>
<th>GEN</th>
<th>Total</th>
<th>χ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150–1349</td>
<td>3.3%</td>
<td>14.7%</td>
<td>76.1%</td>
<td>6.0%</td>
<td>3963</td>
<td>***</td>
</tr>
<tr>
<td>1350–1549</td>
<td>3.4%</td>
<td>16.0%</td>
<td>77.3%</td>
<td>3.4%</td>
<td>5215</td>
<td>*</td>
</tr>
<tr>
<td>1550–1749</td>
<td>3.9%</td>
<td>15.6%</td>
<td>77.1%</td>
<td>3.3%</td>
<td>3498</td>
<td>*</td>
</tr>
<tr>
<td>1750–1899</td>
<td>3.0%</td>
<td>20.1%</td>
<td>72.9%</td>
<td>4.0%</td>
<td>3210</td>
<td>***</td>
</tr>
<tr>
<td>1900–2008</td>
<td>4.2%</td>
<td>20.2%</td>
<td>72.8%</td>
<td>2.7%</td>
<td>2746</td>
<td>***</td>
</tr>
<tr>
<td>Average</td>
<td>3.5%</td>
<td>17.0%</td>
<td>75.6%</td>
<td>3.9%</td>
<td>18632</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: Diachronic distribution of direct object case in IcePaHC.

All the observed distributions in the different periods differ significantly from the expected values with respect to object case. Although the distributions fluctuate along the time periods, overall developments can be observed. Nominative case on direct objects increases over the whole time span from 3.3% to 4.2%. This increase directly correlates with the increasing use of dative subjects because nominative is the default case for objects in transitive constructions with a non-nominative sub-
Moreover, dative objects are on the increase with a salient change of about 4% on the verge of the fourth time period. Direct objects are most often accusative (75.6%) which is a known fact about Icelandic (cf. Thráinsson 2007), but accusative objects still decrease over time. The reduced accusative object shares are probably caused by the increasing share of dative objects in the last two time stages as they chronologically coincide. Genitive objects as, e.g., shown in (3), where the predicate leita ‘search’ occurs together with a nominative subject and a genitive object, furthermore decrease significantly over the whole corpus.

(3) Nú leitar Ófeigur borgunarmanna
    Now search.PRS.3SG Ófeigur.NOM bailsmen.the.GEN
    ‘Now, Ófeigur searches for the bailsmen.’
    (IcePaHC, 1350.BANDAMENNM.NAR-SAG,.946)

I also examined the distribution of indirect objects in IcePaHC to gain further insights on Icelandic object case marking. I found 3396 matrix declarative clauses with a case marked subject and a case marked indirect object. The distribution of indirect object case is shown in Table 4.3.

<table>
<thead>
<tr>
<th>Period</th>
<th>NOM</th>
<th>DAT</th>
<th>ACC</th>
<th>GEN</th>
<th>Total</th>
<th>χ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>0.0%</td>
<td>87.9%</td>
<td>11.4%</td>
<td>0.7%</td>
<td>720</td>
<td></td>
</tr>
<tr>
<td>1350-1549</td>
<td>0.0%</td>
<td>86.4%</td>
<td>13.5%</td>
<td>0.2%</td>
<td>1055</td>
<td></td>
</tr>
<tr>
<td>1550-1749</td>
<td>0.0%</td>
<td>87.9%</td>
<td>11.7%</td>
<td>0.4%</td>
<td>684</td>
<td></td>
</tr>
<tr>
<td>1750-1899</td>
<td>0.0%</td>
<td>89.9%</td>
<td>9.0%</td>
<td>1.1%</td>
<td>533</td>
<td></td>
</tr>
<tr>
<td>1900-2008</td>
<td>0.0%</td>
<td>90.0%</td>
<td>9.7%</td>
<td>0.3%</td>
<td>391 *</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>0.0%</td>
<td>88.0%</td>
<td>11.5%</td>
<td>0.5%</td>
<td>3383</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: Diachronic distribution of indirect object case in IcePaHC.

In general, dative objects of ditransitive predicates, such as gefa ‘give’ in (4), are annotated as indirect objects in IcePaHC. Note that IcePaHC also assigns the indirect object function to a restricted set of single objects including, for example, the predicates þakka ‘thank’ and hjálpa ‘help’ which take a dative object, see, e.g., (5) for an example with hjálpa.
4.3. Corpus study I: The diachronic distribution of dative subjects

(4) Steikari gaf mér hálfs annars dals
cook.the.NOM give.PST.3SG I.DAT half.GEN another.GEN coin.GEN
value.ACC

‘The cook gave me money with the value of one and a half coins.’

(IcePaHC, 1661.INDIAFARI.BIO-TRA,70.1223)

(5) Þorgeir hafði hjálpað honum í hvívetna með
Þorgeir.NOM have.PST.3SG help.PST.PTCP he.DAT in every.respect.DAT with
ráði og dáð.
advise.DAT and deed.DAT

‘Þorgeir had helped him in every respect with words and deed.’

(IcePaHC, 1907.LEYSING.NAR-FIC,.1295)

Indirect objects are most often marked by dative case with a frequency of around 87.6% which corresponds to cross-linguistic findings (e.g., see Haspelmath 2001) and the Icelandic standard (Maling 2001, Thráinsson 2007). Accusative case marks indirect objects to a lesser extent, see (6) for an example in which the verb eggja ‘incite, encourage’ was annotated as taking an indirect accusative object. Genitive indirect objects are rare, see (7) for an example with the verb frýja ‘question’. Indirect nominative objects did not occur in the corpus. The distributions are rather stable over time, with minor deviations in the last time stage.

(6) Jökull eggjaði þá fóður sinn að gifta
Jökull incite.PST.3SG then father.ACC his.ACC to give.in.marriage.INF
Hákoni Ketilríði.
Hákonur.DAT Ketilríður.ACC

‘Then, Jökull incited his father to give Ketilríður to Hákonur in marriage.’

(IcePaHC, 1400.VIGLUNDUR.NAR-SAG,.753)

(7) Enginn frýði Brandi heitnum
Nobody.NOM question.PST.3SG Brandur.DAT name.PTCP.DAT
Kolbeinssyni vits eða góðvilja
Kolbeinsson.DAT intelligence.GEN or good.will.GEN
‘Nobody questioned Brandur’s, named Kolbeinsson, intelligence and good will.’

(IcePaHC, 2008.OFSI.NAR-SAG,.1077)

Subsequently, I examined the different types of object case marking occurring in clauses with a dative subject. In what follows, I present my findings on the basis of the proportion of the examined clauses which contain a main verb. First of all, I
found that verb type as annotated in IcePaHC, i.e., main verbs, modals, ‘have’, ‘do’, ‘be’, and ‘become’, has no effect on the diachrony of the factors investigated in this study. Out of the total number of 65,394 examined matrix declarative clauses, the majority of 51,663 clauses contained a main verb which still displays a representative proportion of IcePaHC. Moreover, the investigation in this chapter includes an analysis of the interaction between dative subjects and the factor voice. As middle formation is only possible with main verbs, a comparative analysis of dative subjects, different types of dative subject predicates and the interaction with voice has to be based on the proportion of sentences containing a main verb.

In total, I found 2,004 matrix declarative clauses containing a dative subject and a main verb. Constructions in which a dative subject predicate takes an object are rare in comparison to the dative subject constructions without an object in IcePaHC. Out of the 2,004 clauses with a dative subject, only 510 contained a direct or indirect object, see the last column in the bottom row of Table 4.4. Table 4.4 moreover shows the object case distribution of the 510 dative subject clauses containing an object. Dative subject predicates usually take a nominative object, see, e.g., (8) in which the verb *líka* ‘like’ has a nominative object in addition to a dative subject.

<table>
<thead>
<tr>
<th>Period</th>
<th>NOM</th>
<th>DAT</th>
<th>ACC</th>
<th>GEN</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>88.9%</td>
<td>3.7%</td>
<td>7.4%</td>
<td>0.0%</td>
<td>108</td>
</tr>
<tr>
<td>1350-1549</td>
<td>91.1%</td>
<td>1.5%</td>
<td>6.7%</td>
<td>0.7%</td>
<td>135</td>
</tr>
<tr>
<td>1550-1749</td>
<td>95.7%</td>
<td>0.0%</td>
<td>4.3%</td>
<td>0.0%</td>
<td>94</td>
</tr>
<tr>
<td>1750-1899</td>
<td>90.7%</td>
<td>2.7%</td>
<td>6.7%</td>
<td>0.0%</td>
<td>75</td>
</tr>
<tr>
<td>1900-2008</td>
<td>88.8%</td>
<td>4.1%</td>
<td>6.1%</td>
<td>1.0%</td>
<td>98</td>
</tr>
<tr>
<td>Average</td>
<td>91.0%</td>
<td>2.4%</td>
<td>6.3%</td>
<td>0.4%</td>
<td>510</td>
</tr>
</tbody>
</table>

Table 4.4: Object case distribution in dative subject clauses from IcePaHC.

(8) Vel líkúðu goðröði góð róði.
    well like.PST.3PL Goðröður.DAT good.NOM.PL oars.NOM
    ‘Goðröður liked good oars well.’
    (IcePaHC, 1150.FIRSTGRAMMAR.SCI-LIN,.70)

Besides nominative objects, I also found accusative, genitive and dative objects together with dative subjects, although they do not exist in Icelandic according to Thráinsson’s (2007) grammar (see also the classification of possible case patterns in
Section 2.2 in Chapter 2). However, Wood (2015) shows that DAT-ACC constructions exist in Icelandic, but marks them as archaic. Moreover, Barðdal (2008) lists DAT-GEN constructions for Icelandic, but notes that they are very low in type frequency and almost extinct. The DAT-DAT constructions which I found are mainly passives of ditransitive constructions with two dative objects.\(^1\)

This section examined the overall diachronic distributions of subject and object case in IcePaHC in order to set the stage for the corpus investigations of factors conditioning the occurrence of dative subjects in the history of Icelandic conducted in this chapter. One such factor is voice, as dative subjects emerge under passivization and middle formation in Icelandic (see, e.g., Zaenen et al. 1985, Sigurðsson 1989). Yet, dative subjects are heavily constrained by lexical semantic factors with respect to middle formation in that dative case is only retained with goal or benefactive arguments. Therefore, investigating the diachrony of dative subjects necessitates an investigation of the interaction of dative subjects with voice and lexical semantics in turn. To provide an empirical basis for the investigation of the interaction between dative subjects, lexical semantics, i.e., event structure and thematic roles, and voice, I investigate the diachronic interrelation between subject case and voice in IcePaHC.

### 4.3.2 Dative subjects and voice

Morphosyntactic operations regarding voice, i.e., passivization and middle formation, give rise to dative subjects in Icelandic (see Section 2.6.2 in Chapter 2 for details). I found that voice has an effect on the diachronic distribution of subject case in IcePaHC. Tables 4.5-4.7 show the subject case distribution across the three voices active, middle, and passive. Nominative case, being the most frequently used subject case overall, also occurs most often with all three voices. Moreover, the large majority of matrix declarative clauses with a main verb contains active constructions (43,772 out of 51,663). Table 4.5 provides a detailed view of subject case in active constructions which I discuss first in this section.

In contrast to all sentences (Table 4.1), the distribution of subject case in active constructions does not change much over the whole time span and only undergoes minor fluctuations. Thus, the increase of dative subjects that was observed before cannot be explained by developments pertaining solely to active clauses. However,

---

\(^1\)Some of the DAT-DAT constructions are in fact right dislocations of the subject NP which were tagged as objects. Being very small in number and of no significance for the present study, I did not correct them.
if we look at the passive constructions in Table 4.6, we see a change over time with respect to datives.

<table>
<thead>
<tr>
<th>Period</th>
<th>NOM</th>
<th>DAT</th>
<th>ACC</th>
<th>GEN</th>
<th>Total</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>95.7%</td>
<td>3.0%</td>
<td>1.2%</td>
<td>0.1%</td>
<td>9182</td>
<td>***</td>
</tr>
<tr>
<td>1350-1549</td>
<td>96.8%</td>
<td>2.2%</td>
<td>1.0%</td>
<td>0.0%</td>
<td>13388</td>
<td></td>
</tr>
<tr>
<td>1550-1749</td>
<td>97.1%</td>
<td>1.7%</td>
<td>1.2%</td>
<td>0.0%</td>
<td>7189</td>
<td>***</td>
</tr>
<tr>
<td>1750-1899</td>
<td>96.6%</td>
<td>2.4%</td>
<td>0.9%</td>
<td>0.1%</td>
<td>7438</td>
<td></td>
</tr>
<tr>
<td>1900-2008</td>
<td>95.7%</td>
<td>2.6%</td>
<td>1.6%</td>
<td>0.1%</td>
<td>6575</td>
<td>**</td>
</tr>
<tr>
<td>Average</td>
<td>96.4%</td>
<td>2.4%</td>
<td>1.2%</td>
<td>0.1%</td>
<td>43772</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5: Subject case distribution for active constructions in IcePaHC.

<table>
<thead>
<tr>
<th>Period</th>
<th>NOM</th>
<th>DAT</th>
<th>ACC</th>
<th>GEN</th>
<th>Total</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>88.3%</td>
<td>9.9%</td>
<td>0.0%</td>
<td>1.8%</td>
<td>715</td>
<td>***</td>
</tr>
<tr>
<td>1350-1549</td>
<td>81.2%</td>
<td>17.0%</td>
<td>0.0%</td>
<td>1.8%</td>
<td>548</td>
<td></td>
</tr>
<tr>
<td>1550-1749</td>
<td>84.5%</td>
<td>14.9%</td>
<td>0.0%</td>
<td>0.6%</td>
<td>713</td>
<td></td>
</tr>
<tr>
<td>1750-1899</td>
<td>81.4%</td>
<td>16.2%</td>
<td>0.0%</td>
<td>2.4%</td>
<td>413</td>
<td></td>
</tr>
<tr>
<td>1900-2008</td>
<td>78.7%</td>
<td>21.1%</td>
<td>0.0%</td>
<td>0.3%</td>
<td>342</td>
<td>**</td>
</tr>
<tr>
<td>Average</td>
<td>83.6%</td>
<td>15.0%</td>
<td>0.0%</td>
<td>1.4%</td>
<td>2731</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.6: Subject case distribution for passive constructions in IcePaHC.

Dative subjects are increasingly used in passive constructions with marked increases between the first and the last two time stages. Concomitantly, nominative subjects decrease. Overall, the shares of dative subjects are larger and the ones of nominatives smaller than in the active constructions. Moreover, genitive subjects are still rare, but also more frequent than in active clauses. Accusative subjects do not occur in passive constructions as accusative objects become nominative subjects under passivization.

The strongest increase of dative subjects in IcePaHC is attested for middle constructions, see Table 4.7. Dative subjects increase with middles from the third time stage on and show a significant increase in frequency of more than 13% as of 1900. Again, nominatives decrease accordingly. Middle constructions with accusative or genitive subjects were not found in the corpus. Interestingly, the subject case distribution of middles in the second time period differs significantly from the expected distribution. The same deviance was observed in the subject case distribution in all matrix declaratives (see Table 4.1), but is not present in the active and passive
distributions. I initially attributed this deviance to a larger number of sentences in the second time period, but given the absence of this deviance with actives and passives, attributing the deviance to the changes which subject case undergoes in middle constructions seems to be more plausible to me.

<table>
<thead>
<tr>
<th>Period</th>
<th>NOM</th>
<th>DAT</th>
<th>ACC</th>
<th>GEN</th>
<th>Total</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>94.0%</td>
<td>6.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>1073</td>
<td>***</td>
</tr>
<tr>
<td>1350-1549</td>
<td>94.0%</td>
<td>6.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>1337</td>
<td>***</td>
</tr>
<tr>
<td>1550-1749</td>
<td>92.7%</td>
<td>7.3%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>940</td>
<td>**</td>
</tr>
<tr>
<td>1750-1899</td>
<td>89.2%</td>
<td>10.8%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>812</td>
<td></td>
</tr>
<tr>
<td>1900-2008</td>
<td>76.0%</td>
<td>24.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>997</td>
<td>***</td>
</tr>
<tr>
<td>Average</td>
<td>89.5%</td>
<td>10.5%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>5159</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.7: Subject case distribution for middle constructions in IcePaHC.

As the increase in dative subjects is strongly correlated with a rise in middle forms, I took a closer look at the interaction between dative subjects and voice for a better comparison. Table 4.8 gives the shares of active, passive and middle constructions that occurred together with a dative subject throughout the different time stages.

<table>
<thead>
<tr>
<th>Period</th>
<th>active</th>
<th>passive</th>
<th>middle</th>
<th>Total</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>67.4%</td>
<td>17.1%</td>
<td>15.5%</td>
<td>414</td>
<td>***</td>
</tr>
<tr>
<td>1350-1549</td>
<td>63.3%</td>
<td>19.7%</td>
<td>16.9%</td>
<td>472</td>
<td>***</td>
</tr>
<tr>
<td>1550-1749</td>
<td>41.1%</td>
<td>35.7%</td>
<td>23.2%</td>
<td>297</td>
<td>***</td>
</tr>
<tr>
<td>1750-1899</td>
<td>54.0%</td>
<td>19.9%</td>
<td>26.1%</td>
<td>337</td>
<td>***</td>
</tr>
<tr>
<td>1900-2008</td>
<td>35.7%</td>
<td>14.9%</td>
<td>49.4%</td>
<td>484</td>
<td>***</td>
</tr>
<tr>
<td>Average</td>
<td>52.6%</td>
<td>20.4%</td>
<td>26.9%</td>
<td>2004</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.8: Diachronic distribution of dative subject predicates by voice in IcePaHC.

IcePaHC’s genre effect is once more revealed in Table 4.8 and affects the distribution of dative subjects according to voice in the third time stage which clearly deviates from the overall developments. Out of the 2004 constructions which contain a dative subject, around 53% are actives, 20% passives, and 27% are middles. As expected, the use of dative subjects with middles increases over the centuries, with a striking increase of 23% in the period post-1900, reaching a share of almost 50%. This increase mainly correlates with an increasing use of the predicate finnast ‘find, think, feel, seem’ which carries middle morphology, see (9).^2

^2Example (9) was used to illustrate the IcePaHC annotation in Section 2.8 of Chapter 2.
(9) Mörgum þeirra fanst þeir vera útlagar úr many.DAT they.GEN seem.PST.MID.3SG they.NOM be-INF exiles.NOM out.of landi sínu meðan þeir voru hér. their.own.DAT while they.NOM be.PST.3PL there

‘It seemed to many of them that they are exiles from their own land while they were here.’ (IcePaHC, 1907.LEYSING.NAR-FIC,.763)

Apart from finnast, the experiencer and raising predicate sýnast ‘seem, appear’, which was counted as middle form, occurs frequently in the last time stage, see (10). Another middle form which occurs frequently with a dative subject in the last time stage is the predicate takast ‘manage, succeed’ as shown in (11).

(10) Presti sýndist þjáningarsvipur fara um priest.DAT seem.PST.MID.3SG suffering.look.NOM move.INF over andlítið á Ragnhildi.

‘To the priest seemed a suffering look to move over Ragnhildur’s face.’

(IcePaHC, 1908.OFUREFLI.NAR-FIC,.735)

(11) Í landafraðíprófi tókst mér einu sinni að in geography.exam.DAT manage.PST.MID.3SG I.DAT one.DAT time.DAT to kíkja á prófblað ... take.a.look.INF at exam.paper.ACC

‘In the geography exam, I managed to look at the exam once ...’

(IcePaHC, 1985.SAGAN.NAR-FIC,.1620)

While middles are increasing, passives with dative subjects are only slightly reduced, but active dative subject constructions lose almost half of their initial share, with a large decrease of about 18% between the last two periods. Overall, dative subjects are becoming more firmly associated with verbs carrying middle morphology over the history of Icelandic. The final scenario starkly contrasts with the initial distribution in the first time stage in which active constructions were dominant and claimed a share of 67%, while passives and middles were distributed roughly equally with shares of 17% and 16% respectively.

In sum, the corpus studies presented in this section showed that the average frequencies for subject and object cases in IcePaHC correspond to the existing figures about case marking in Icelandic. Moreover, dative subjects are on the increase overall, with a salient increase as of 1900. Furthermore, the investigation of the interaction between subject case and voice indicates that the increasing usage of
dative subjects in the period post-1900 correlates with an increase of dative subjects together with middle voice. In the following section, I provide an intermediate analysis of the predicates occurring together with a dative subject which is relevant for the subsequent investigations in this chapter. I argue that (i) the increase of dative subjects observed in the period post-1900 cannot mainly be attributed to Dative Substitution as has been suggested in the previous literature (e.g., in Barðdal 2011), and that (ii) there are two distinct sources for dative subjects with middle verbs, i.e., productive middle formation with dative goal objects and lexicalization as experiencer predicates over time.

4.4 Dative Substitution, lexicalization and middle formation

The results for subject case marking presented in Section 4.3.1 seem to stand in contradiction to the results presented by Barðdal (2011) for subject case in Old and Modern Icelandic. In Barðdal’s study, nominative subjects increase from Old to Modern Icelandic from 76.3% to 85% and dative subjects decrease from 18.4% to 10.3% (see Table 2.2 in Chapter 2), while my data indicates the converse trajectory. However, in contrast to the statistics given in my study, Barðdal does not count the occurrences of different subject cases per se, but evaluates how many different types of verbs occurred together with the respective subject cases. Yet, counting type frequencies only shows if a specific verb occurred together with a particular subject case, but does not reflect actual patterns of usage of a syntactic construction. Despite the decreasing dative subject type frequency, Barðdal (2008, 2011) argues for the productivity of the dative subject construction. She sees the decreasing type frequencies as the consequence of the increasing semantic coherence of dative subject constructions which in turn is interpreted as a precursor of Dative Substitution.

Interestingly, the onset of Dative Substitution has been attributed to the latter part of the 19th century and could indeed be a factor behind the increasing usage of dative subjects post-1900. However, contra Barðdal (2011), I assume that an increase instead of a decrease of the type frequency could indicate that Dative Substitution is a factor behind the increasing use of dative subjects as of 1900. An increasing type frequency points to new or different verbs that are drawn into the system and begin to mark their subject with dative case. On these grounds, I also counted and looked at the different types of main verbs which occurred together with a dative subject in
IcePaHC to test whether the observed increase of dative subjects in IcePaHC can be attributed to Dative Substitution. Table 4.9 shows the type and token frequencies of dative subject predicates in the different periods investigated within IcePaHC.

<table>
<thead>
<tr>
<th>Period</th>
<th>Types</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150–1349</td>
<td>86</td>
<td>414</td>
</tr>
<tr>
<td>1350–1549</td>
<td>84</td>
<td>472</td>
</tr>
<tr>
<td>1550–1749</td>
<td>107</td>
<td>298</td>
</tr>
<tr>
<td>1750–1899</td>
<td>86</td>
<td>337</td>
</tr>
<tr>
<td>1900–2008</td>
<td>105</td>
<td>484</td>
</tr>
</tbody>
</table>

Table 4.9: Type and token frequencies of dative subject predicates in IcePaHC.

Whereas Barðdal (2011) found 72 types of dative subject predicates in Old Icelandic texts from around 1200 CE and 48 types in modern Icelandic texts from the 1980s (cf. Table 2.2), my findings from IcePaHC differ substantially. Table 4.9 shows that around 86 different types of verbs occur with a dative subject in the first time period. This frequency remains stable until the last time period in which the type frequency increases. Thus, not only the overall usage of dative subjects increases as of 1900 in IcePaHC, but also more different types of predicates are used together with a dative subject.

Although Dative Substitution is a factor behind the increase of dative subject verb types as of 1900, the verbs driving the increase of dative subjects in the last time period, i.e., the verbs with the largest token frequencies are not subject to Dative Substitution. The verbs which occur most often with a dative subject in the last time period are the experiencer and raising predicates þykja ‘think, seem’ and finnast ‘think, find, feel, seem’. The predicate þykja occurs together with a dative subject from the earliest stages in IcePaHC on, see (12), and is most often used in the periods pre-1900. Being used frequently with a dative subject throughout the whole diachrony, the dative subject of þykja cannot be the result of Dative Substitution.

(12) a. Pá þótti konungi gaman aðerta jarl
    then think.PST.3SG king.DAT fun.NOM to provoke.INF earl.ACC
    ‘Then the king enjoyed provoking the earl.’
    (Lit.: ‘Then the king thought it was fun to provoke the earl.’)
    (IcePaHC, 1275.MORKIN.NAR-HIS.461)

The deviating frequency in the third time stage can be attributed to the genre effect. By looking at the individual verbs taking a dative subject in the corpus, I found that the verb types used in the third time stage differ from the verbs in the other stages due to the difference in genre of the corresponding texts.
b. Alveg þykir mér þetta rými gersamlega entirely seem.PRS.3SG I.DAT this.NOM room.NOM completely óþolandi unbearable.NOM
‘This room entirely seems to me to be completely unbearable.’
(IcePaHC, 2008.MAMMA.NAR-FIC,.763)

As of 1900, þykja is the second most frequent predicate and occurs 63 times, whereas finnast is used most frequently after a continuous diachronic increase, see (13) (repeated from (9)) for an example with finnast from IcePaHC. As exemplified by (13), finnast is an experiencer verb which can be used as a raising predicate. Out of the 484 instances in which a main verb takes a dative subject in the last time period, finnast occurs 129 times, constituting a large part of the overall increase of dative subjects. However, the increasing use of finnast cannot be correlated with Dative Substitution.

(13) Mörgum þeirra fanst þeir vera útlagar úr many.DAT they.GEN seem.PST.MID.3SG they.NOM be.INF exiles.NOM out.of landi sínu meðan þeir voru hér. theirOWN.DAT while they.NOM be.PST.3PL here
‘It seemed to many of them that they are exiles from their own land while they were here.’
(IcePaHC, 1907.LEYSING.NAR-FIC,.763)

The stative experiencer predicate finnast ‘think, find, feel, seem’ is diachronically based on the verb finna ‘find’ which describes a dynamic event of finding and does not take a dative argument, see (14-a) for an example. The ‘original’ middle form of finna also describes a finding event, meaning ‘be found, meet’ as in (14-b), still occurring without a dative argument.

(14) a. Hann finmur í taðinu dauðan mannsbút he.NOM find.PRS.3SG in manure.DAT dead.ACC man.stump.ACC
‘He finds a dead man’s stump in the manure.’
(IcePaHC, 1675.ARMANN.NAR-FIC,97.174)

b. og fundust þar hvorki menn né skip and be.found.PST.MID.3SG there neither men.NOM nor ship.NOM
‘and neither men nor ship were found there’
(IcePaHC, 1450.VILHJALMUR.NAR-SAG,9.89)
In its usage as an experiencer and raising predicate, *finnast* ‘think, find, feel, seem’ takes a dative subject and does not represent this original middle reading anymore. Therefore, I assume that the middle form *finnast* ‘be found, meet’ was lexicalized as an experiencer predicate with a dative subject over time, resulting in the reading shown in (13). A detailed analysis of the diachronic lexicalization process of former middle verbs as stative experiencer predicates with a dative subject is presented in Chapter 6.

Interestingly, Barðdal (2011) finds that the experiencer predicates *þykja* ‘think, seem’ and *sýnast* ‘seem, appear’ are the most frequently occurring dative subject predicates in her Old Icelandic corpus, while *finnast* is dominant in the Modern Icelandic corpus. These findings are also borne out in my study, where, in addition to *þykja*, *sýnast* occurs highly frequently from the earliest time period onwards, see, e.g., (16) for an example from IcePaHC, with *þykja* being the most frequent predicate until the last time stage which is when *finnast* takes over. In contrast to the lexicalized experiencer predicate *finnast*, the dative subject with *sýnast* ‘seem, appear’ is the result of productive middle formation of the ditransitive predicate *sýna* ‘show’ which takes a dative goal object, compare, e.g., example (15-a), where *sýna* ‘show’ has a dative object, with example (15-b), in which the middle *sýnast* ‘seem, appear’ has a dative subject. Under middle formation, the goal argument becomes the subject, retaining its dative case marking. In (15-b), *sýnast* describes a physical appearance and perception process which correlates with the corresponding base meaning ‘show’ in the sense that something is shown to somebody, who perceives the thing shown. However, in (16), the goal argument is reanalyzed as an experiencer and *sýnast* is used as stative experiencer and raising predicate.

(15) a. Þorsteinn *sýnir* nú konungi hrossin

   *Þorsteinn* shows the horse to the king now.’

   (IcePaHC, 1275.MORKIN.NAR-HIS,.1834)

b. Þeim *sýndist* heilög María Magdalena í svefni

   ‘The holy Maria Magdalena appeared to them in their sleep.’

   (IcePaHC, 1350.MARTA.REL-SAG,.492)
4.4. Dative Substitution, lexicalization and middle formation

(16) og sýndist hyggnum mõnnum að tvõ vandræði
and seem.PST.MID.3SG prudent.DAT men.DAT to two.NOM difficulties.NOM
væri á
be.PST.SBJV.3PL at
‘and the prudent men seemed to have had two difficulties’

(IcePaHC, 1260.JOMSVIKINGAR.NAR-SAG,.686)

Productive middle formation of ditransitive predicates with a dative goal object and the subsequent lexicalization as experiencer predicates of verbs such as sýnast is also discussed in further detail in Chapter 6. It is important for the following investigations to differentiate between dative subjects which arose through regular middle formation and new dative subject experiencer predicates which entered the language via lexicalization, e.g., finnast ‘find, feel, think, seem’, which I will refer to as lexicalized experiencer predicates throughout the remainder of this thesis.

The middle forms which appear with a dative subject in IcePaHC are mostly experiencer predicates. In previous studies (Schätzle et al. 2015, Schätzle and Sacha 2016), I found that dative subjects generally occur most often with experiencer predicates throughout IcePaHC, with a significant increase of experiencer predicates as of 1900. However, the studies have also shown that it is difficult to find generalizations across the lexical semantic verb classes which have been suggested for dative subjects in Icelandic (see Section 2.6 in Chapter 2). The lexical semantic verb classes are usually too fine-grained to let significant patterns emerge and the classes are moreover difficult to delimit. In this thesis, I investigate the alternative approach to case marking in Icelandic suggested by Svenonius (2002) which factors in event semantics. On these grounds, I propose a novel classification for dative subject predicates according to their event structure which is based on Ramchand’s (2008) event decomposition framework in the next section. This classification is used for the investigation of the interrelation between dative subjects, thematic roles, event structure and voice in IcePaHC presented in Section 4.6, providing for a more generic approach to case marking and grammatical relations in Icelandic.
Chapter 4. Dative subjects and event semantics

4.5 Annotating event structure

This section sets out to investigate the interrelation between dative subjects and event semantics by introducing the different types of event structures which apply to predicates with a dative subject in IcePaHC. Moreover, this section describes the annotation process of the established event semantic verb classes which allows for the quantitative and visual analytic study of dative subjects and event structure presented in Section 4.6.

4.5.1 Event structure classification

In order to investigate the interaction between dative subjects and event semantics, I classified the main verbs which occurred together with a dative subject according to their event structure using Ramchand’s (2008) event decomposition of the first-phase syntax (see Section 3.4 in Chapter 3). For this purpose, the verbal constructions which occurred with a dative subject in IcePaHC were categorized with respect to the subevents that they entail. In Ramchand’s system, dynamic predicates always have a process event and can additionally imply an initiating and/or result event. Stative predicates represent a state only. Moreover, I recorded the types of event participants which were involved in the verbal constructions, i.e., (state) holder, initiator, undergoer, resultee, rheme. The verb classification was performed in the form of ordered lists which were used to identify recurring patterns.

In the event structure classification and annotation, I accounted for specific verbal construction types instead of basing the investigations solely on lexical verbal meanings of individual lemmas because the corresponding meanings may differ considerably. For example, the predicate koma can take a dative object which surfaces as a dative subject in the passive, see (17-a). In this type of construction, koma means ‘get, take’ and is a verb of transfer. Most often, however, koma occurs intransitively with a nominative subject in active constructions, meaning ‘come’ as in (17-b). The

---

4Ramchand (2008) assumes that stative predicates consist of an init projection. With stative predicates, the initiator is interpreted as a state holder and not as an agent. To demarcate stative from dynamic predicates and for the purpose of the formal analysis which follows in Chapter 6, I refer to the initiator of stative predicates as holder throughout the remainder of this thesis and moreover assume that stative predicates consist of a single state projection, conceptually equivalent to Ramchand’s init projection with stative predicates. This is moreover in line with Ramchand (2018), where Ramchand distinguishes between different types of state holders in the context of adjective alternations.

5Throughout the thesis, the following abbreviations are used for the event participants: init stands for INITIATOR, und for UNDERGOER, res for RESULTEE, and rh for RHEME.
4.5. Annotating event structure

verb *koma* can also take a dative subject in active sentences, but only if the verb occurs in specific constructions, e.g., together with the adverb *vel* ‘well’ as *koma vel* ‘benefit’ or the prepositional phrase *í hug* ‘to mind’ as *koma í hug* ‘get an idea’, see (17-c) and (17-d) respectively. These two constructions are both stative experiencer predicates, differing considerably from the passive *koma* in meaning. Therefore, I had to extract the individual sentences occurring with the different types of main verbs taking a dative subject, which was done via a Perl script, and identify the different constructions manually before they could be classified.

(17) a. **Borgaralegu þjóðskipulagi var kómið á með civil.DAT social.system.DAT be.PST.3SG take.PASS.PTCP on with blóðugum bylingum, áréttaði Andri. bloody.DAT.PL upheavals.DAT emphasize.PST.3SG Andri.NOM**

‘The civil social system was established with bloody upheavals, emphasized Andri.’

(IcePaHC, 1985.SAGAN.NAR-FIC,.64)

b. **Hann kom heim í Skálholt með litlum he.NOM come.PST.3SG home to Skálholt.ACC with little.DAT mætti. power.DAT**

‘He came home to Skálholt with little power.’

(IcePaHC, 1210.THORLAKUR.REL-SAG,.415)

c. **Eigi kom mér þessi fregn vel né hentuglega. not take.PST.3SG I.DAT this.NOM news.NOM well nor practically.**

‘I did not benefit from these news.’

(IcePaHC, 1791.JONSTEINGRIMS.BIO-AUT,159.1516)\(^6\)

d. **Kom honum síðan það í hug að...**

‘Then he got the idea that ... ’

(IcePaHC, 1210.JARTEIN.REL-SAG,.30)

Following this approach, four different verb classes emerged which classify the verbal constructions occurring with a dative subject in IcePaHC according to their event structure and the event participants involved. Table 4.10 provides an overview of the four classes, including the corresponding event structures and combinations of

\(^6\)IcePaHC annotates *þessi fregn* ‘these news’ as accusative plural. However *fregn* is a singular form, but translates into the plural ‘news’. Moreover, *fregn* is feminine and *þessi* as a feminine demonstrative pronoun can only be nominative singular as is *fregn*. 
event participants. RHEME arguments are only given in Table 4.10 if they can refer to full NP arguments. All stative predicates from IcePaHC with a dative subject can be subsumed in one class (Class I in Table 4.10). The remaining classes refer to dynamic predicates which are further divided into transfer verbs, transitions, and scalar changes.

<table>
<thead>
<tr>
<th>Class</th>
<th>Event type</th>
<th>Event structure</th>
<th>Event participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Stative predicates</td>
<td>[state]</td>
<td>HOLDER RH</td>
</tr>
<tr>
<td>II</td>
<td>Transfer</td>
<td>[init, proc, res]</td>
<td>INIT RES RH</td>
</tr>
<tr>
<td>III</td>
<td>Transitions</td>
<td>[init, proc]</td>
<td>INIT UND</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>UND-RES</td>
</tr>
<tr>
<td>IV</td>
<td>Scalar changes</td>
<td>[init, proc, res]</td>
<td>INIT UND-RES</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>UND-RES</td>
</tr>
</tbody>
</table>

Table 4.10: Event structure classification for dative subject predicates in IcePaHC.

The full set of verbal constructions for each class as identified in IcePaHC is provided in the Appendix and a detailed description of the individual classes follows next.\(^7\)

### 4.5.1.1 Class I – Stative predicates

The stative predicates in Class I are experiencer predicates such as verbs describing psychological states, including verbs of emotion and verbs of cognition, verbs describing bodily states, verbs describing characteristics and properties, verbs of position and stance, and verbs of obligation. A stative event has a state HOLDER as subject and may additionally have a RHEME. This is generally the case in Icelandic where rhematic material occurs in the form of a full NP argument with stative predicates, and RHEMES in the form of an AP or PP are moreover possible. For example, in (18), the predicate *líka* ‘like’ takes a dative experiencer subject in active clauses which designates the state HOLDER and has a RHEME in the form of a nominative object which further describes the state.

---

\(^7\)When I refer to verbs which take a dative subject in the following, the different kinds of verbal constructions with a dative subject identified in IcePaHC are included.
4.5. Annotating event structure

(18) En Þorgrími líkaði það illa. but Þorgrímur.DAT like.PST.3SG this.NOM badly ‘But Þorgrímur disapproved of this.’

(IcePaHC, 1400.GUNNAR.NAR-SAG,.241)

Moreover, stative experiencer predicates which carry middle morphology occurred together with a dative subject in the corpus. These are lexicalized experiencer predicates such as finnast ‘find, think, feel, seem’, see (19), and verbs which have a dative subject due to regular middle formation of a dynamic predicate, whereby the middle form has acquired a stative experiencer reading, e.g., sýnast ‘seem, appear’ as shown in (20), repeated from (10).

(19) Ekki finnast með Sóroásters trúárbrögð eins not seem.PRS.MID.3SG I.DAT Sóroaster.GEN religion.NOM as herleg og þau sem Óðinn hafði og glorious.NOM as that.NOM which Óðinn.NOM have.PST.3.SG and kenndi.

teach.PST.3SG

‘Sóroaster’s religion does not seem to me as glorious as the one which Óðinn had and taught.’ (IcePaHC, 1790.FIMMBRAEDRA.NAR-SAG,.34)

(20) Presti sýndist þjáningarsvipur fara um priest.DAT seem.PST.MID.3SG suffering.look.NOM move.INF over andlíði á Ragnhildi.

face.the.ACC of Ragnhildur.ACC

‘To the priest seemed a suffering look to move over Ragnhildur’s face.’

(IcePaHC, 1908.OFUREFLI.NAR-FIC,.735)

As stative experiencer predicates, the middle forms have the same event structure as the active forms in this class, taking a state HOLDER which is the dative subject and a RHHEME argument, e.g., the nominative object in (19). I found no passive constructions with verbs belonging into this class, as stative predicates can generally not passivize.

4.5.1.2 Class II – Transfer verbs

The second class mainly consists of ditransitive predicates with a dative goal argument which becomes a dative subject under passivization. These verbs, for example gefa ‘give’ in (21), describe an action which involves some kind of transfer of a
Chapter 4. Dative subjects and event semantics

(concrete or abstract) theme to a goal argument. Ditransitive transfer verbs generally have an INITIATOR as subject, the dative goal object is a RESULTEE and the accusative-marked theme object is a RHEME argument (see Ramchand 2008 on double object constructions and Section 3.4 in Chapter 3). Although transfer verbs are dynamic and a process event, i.e., a procP in terms of Ramchand, is involved, they do not take an UNDERGOER. Under passivization, the INITIATOR is demoted and the RESULTEE, i.e., the dative goal argument, is realized as the subject.

(21) Dufgussonum voru grið gefin fyrir Dufgussonur.DAT be.PST.3PL mercy.NOM.PL give.PASS.PTCP for flutning Ólafs Svarþssonar.

pleading.ACC Ólafur.GEN Svarþsson.GEN

‘Dufgussonur was given mercy for Ólafur Svarþsson’s pleading.’

(IcePaHC, 1250.STURLUNGA.NAR-SAG,422.1145)

There are also a few transitive transfer predicates which have a dative subject under passivization, e.g., koma ‘get, take’ as was shown in (17-a). These verbs have a similar structure as the ditransitives, taking an INITIATOR as subject and a dative goal RESULTEE as object. Moreover, rhematic material in complement to res is possible. This material is not a full NP argument as with the ditransitives, but consists of an AP or PP.

The middle verbs in this group are middle forms which are derived from a ditransitive with the structure exemplified above. For example, the middle gefast ‘get, receive’ as exemplified in (22) is the counterpart of the ditransitive predicate gefa ‘give’. With middles, the former agent, i.e., the INITIATOR, is no longer present, and the dative goal, i.e., the RESULTEE, becomes the subject instead.

(22) Nú gafst Geirmundi færi að tala það, now get.PST.MID.3SG Geirmundur.DAT opportunity.NOM to tell.INF that sem í brjóstí bjó.

what in brest.DAT reside.PST.3SG

‘Now Geirmundur got an opportunity to tell what he had on his mind.’

(IcePaHC, 1902.FOSSAR.NAR-FIC,.1490)

Furthermore, two predicates occurred with a dative subject in active constructions in the corpus which pattern with the event structure of the middle forms in this class. These predicates are gefa and tilfalla which both were found with meanings and constructions similar to gefast ‘get, receive’, see e.g., (23-a) for an example
with *gefa*. Moreover, when taking a dative subject, *gefa* generally occurs in fixed expressions such as *gefa byr* in (23-a) and *gefa vel* in (23-b) which both mean ‘get a fair wind’.

(23) a. og gefur honum byr norður á
and get.prs.3sg he.dat fair.wind.nom northern into
Þorgeisfjörður.
Þorgeisfjörður.acc
‘and he gets a fair north wind into Þorgeisfjörður’
(IcePaHC, 1350.BANDAMENNM.NAR-SAG,.1114)

b. Gefur þeim allvel.
get.prs.3sg they.dat rather.well
‘They rather get fair wind.’
(IcePaHC, 1350.BANDAMENNM.NAR-SAG,.1133)

4.5.1.3 Class III – Transitions

In Class III are verbs in which an entity is undergoing a transition, including verbs that describe simple transitions, manner of motion verbs and verbs describing mental processes. With these verbs, the dative marked arguments are themes. The predicates with a dative subject in this class are verbs such as *skola* ‘be washed ashore’ which takes a dative subject in active constructions, see (24), and *velta* ‘roll’ which has a dative subject in the passive, see (25), both describing a motion event.

(24) Úr brimöldu draumins skolar mér upp að
Out.of surge.dat dream.the.gen be.washed.ashore.prs.3sg I.dat up to
þessari hvítu strönd.
this.dat white.dat beach.dat
‘Out of a surge of the dream, I am washed ashore up to this white beach.’
(IcePaHC, 1985.SAGAN.NAR-FIC,.527)

(25) Með stunum og óhljódum og miklum erföðismunum
with moans.dat and shoutings.dat and great.dat effort.dat.pl
var þungum kössum velt upp á bryggjuna . . .
be.pst.3sg heavy.dat crate.dat roll.pass.ptcp up onto jetty.the.acc
‘With moans and shoutings and great effort, the heavy crate was rolled up onto the jetty.’
(IcePaHC, 1907.LEYSING.NAR-FIC,.491)

---

8The noun *byr* ‘fair wind’ is annotated as accusative in the corpus, but *byr* is NOM in this clause.
Unaccusative manner of motion verbs, e.g., *skola* ‘be washed ashore’, take an UNDERGOER argument as subject (cf. Ramchand 2008). The unaccusative verbs contrast with transitive manner of motion predicates such as *velta* ‘roll’ in this class in that the transitive predicates have a causing event in addition to a procP, thus licensing an INITIATOR and an UNDERGOER. With both the intransitive and the transitive predicates, the dative argument is licensed as the UNDERGOER. While the dative UNDERGOER is realized as the subject in active constructions with the unaccusative predicates, the UNDERGOER corresponds to a dative object with the transitive predicates which surfaces as a subject in the passive, where the INITIATOR is demoted. Ramchand (2008) assumes that causativization is a general structure building process in English which automatically builds a null *init* head on top of the corresponding intransitive [proc] structure, relating the event structures of the unaccusatives and the transitive predicates in this class to one another.

I found no middle verb with a dative subject in IcePaHC describing a simple transition as under middle formation, dative theme objects lose their case marking and are realized as nominative subjects instead, see, e.g., (26), which shows the middle form of *velta* ‘roll’.

(26) og pabbi og mamma veltast í undarlegum
    and dad.NOM and mom.NOM revolve.PRS.MID.3SG in strange.DAT
    slagsmálum.
    fight.DAT
    ‘and dad and mom revolve around each other in a strange fight.’
    (IcePaHC, 1985.SAGAN.NAR-FIC,.993)

Another group of verbs which describe a transition are verbs describing mental processes. These predicates have an INITIATOR-UNDERGOER, i.e., a volitional agent which is continuously involved in the process (cf. Ramchand 2008, 54f. on UNDERGOER-INITIATORS of dynamic psych verbs), as well as a RHHEME. An example for a mental process verb which takes a dative subject in the passive is *gleyma* ‘forget’ as shown in (27). In (27-a), *gleyma* occurs in an active clause, where the nominative experiencer subject is the INITIATOR-UNDERGOER and the dative object is a RHHEME. Under passivization, the INITIATOR-UNDERGOER, i.e., the former subject, is demoted and the RHHEME argument becomes the subject instead, compare the passive construction in (27-b) with the active clause in (27-a). Again, the dative argument which is realized as the subject in the passive refers to a theme.
4.5. Annotating event structure

(27) a. Guð hefur gleymt mínu vondu
    God.NOM have.PRS.3SG forget.PST.PTCP my.DAT evil.DAT
    athæfi. behavior.DAT
    ‘God has forgotten my evil behavior.’
    (IcePaHC, 1611.OKUR.REL-OTH.30.386)

b. Og honum skal ekki verða gleymt.
    and he.DAT shall.PRS.3SG not become.INF forget.PASS.PTCP
    ‘And he shall not become forgotten.’
    (IcePaHC, 1908.OFUREFLI.NAR-FIC..1306)

4.5.1.4 Class IV – Scalar changes

Class IV consists of verbs which describe scalar changes, i.e., change of state predicates, verbs describing changes of bodily states, and verbs describing a change of location, e.g., verbs of directed motion. These predicates generally have a dative marked theme argument. In contrast to the verbs in Class III, the verbs in Class IV all denote an event with a concrete endpoint and a definite result. Thus, they license a resP in addition to a procP. As was the case with manner of motion verbs, the change which an event participant is undergoing can follow an initiation state, but this is not the case for all of the verbs. The active predicates in Class IV are unaccusatives, e.g., linna ‘stop’ in (28), which do not have an initP and take a single UNDERGOER-RESULTEE argument. With these predicates, the UNDERGOER-RESULTEE argument is realized as a dative subject in active constructions.

(28) og linnti þeirri ferð ekki ...
    and stop.PST.3SG this.DAT journey.DAT not
    ‘And this journey ended not . . . ’
    (IcePaHC, 1525.GEORGIUS.NAR-REL,.709)

Moreover, the change of bodily state predicates which occur with a dative subject are often unaccusatives, taking an UNDERGOER-RESULTEE argument. However, in contrast to the other groups of verbs describing a scalar change, change of bodily state predicates generally have an experimenter subject, see, e.g., batna ‘recover’ in (29).

(29) Stúlkunni batnaði.
    girl.the.DAT recover.PST.3SG
    ‘The girl recovered.’ (IcePaHC, 1725.BISKUPASOGUR.NAR-REL,.346)
Transitive predicates describing a scalar change consist of a process event with an initiating and a result state, in which the *undergoer* also holds the result state. Under passivization, the *initiator* is demoted, while the *undergoer-resultee* is promoted to subject and retains its dative case marking, see, e.g., (30) in which the predicate *ljúka* ‘end, finish’ takes a dative subject in the passive.

(30) og þá var þessari hátiðlegu athöfn lokið.
    and then be.PST.3SG this.DAT festive.DAT ceremony.DAT finish.PASS.PTCP
    ‘and then this festive ceremony was finished.’

(IcePaHC, 1882.TORFHILDUR.NAR-FIC,.795)

Similar to the predicates in Class III, there are no middle verbs with a dative subject in Class IV as the dative theme argument of the transitive predicates in this group do not retain their case marking under middle formation.

4.5.2 Event structure annotation

Each of the 2004 main verbs occurring in a dative subject clause from IcePaHC was annotated with the corresponding event structure information in a separate csv-file which forms the basis of my investigations with respect to event structure. The csv-file was structured according to the format specifications required by the glyph visualization which will be presented in Section 4.6.\footnote{The glyph visualization requires an additional column after the sentence ID which was initially used to indicate the age of a text. However, given that the sentence ID encodes the age of the corresponding text in the form of the respective year date, the age column is redundant and is left blank here.} In the csv-file, the annotated information is mapped onto the unique sentence ID of each clause, see Figure 4.1 which shows a snapshot of the file. This moreover allows for the mapping between the present csv-file and the full set of extracted information stored in the IcePaHC dataset as well as the original sentences in the corpus.

The annotation was performed automatically using a Perl script, matching the main verb of each dative subject clause with information stored in the form of ordered lists for the different verb classes. If a main verb could in principle occur in a verbal construction type including more than just the verb itself as exemplified for *koma* in (17), further elements were searched in the respective matrix clause in the corpus, e.g., the adverb *vel* ‘well’ was additionally searched with *koma* to potentially identify *koma vel* ‘benefit’, in order to be able to annotate the relevant meaning and event structure.
4.5. Annotating event structure

Figure 4.1: Event class annotation of sentences containing a dative subject predicate in IcePaHC.

Each main verb of a dative subject clause was annotated with the verb class indicating the respective event structure using the labels shown in the first column of Table 4.11, and with more fine grained verb class information, see the last column in Table 4.11 which shows the subclasses including their labels in parentheses. Additionally, the csv-file contains information about voice for each clause, see Figure 4.1, for the investigation of the interaction of dative subjects with event structure and voice in turn which is presented in the following section.

<table>
<thead>
<tr>
<th>Label</th>
<th>Event class</th>
<th>Lexical semantic verb classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>stative</td>
<td>Stative predicates</td>
<td>Psychological states (psy)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bodily states (bod)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Characteristics/properties (char)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Position/stance (pos)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Obligation (obl)</td>
</tr>
<tr>
<td>transfer</td>
<td>Transfer</td>
<td>Verbs of transfer (transf)</td>
</tr>
<tr>
<td>transitions</td>
<td>Transitions</td>
<td>Simple transitions (transit)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Manner of motion verbs (motmanner)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mental processes (mentalproc)</td>
</tr>
<tr>
<td>scalar change</td>
<td>Scalar changes</td>
<td>Change of states (statechange)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of bodily states (bodchange)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of location (locchange)</td>
</tr>
</tbody>
</table>

Table 4.11: Annotation labels for event classes with respect to dative subject predicates in IcePaHC.
The stative predicates which were found together with a dative subject in IcePaHC consist of verbs describing psychological states, bodily states, characteristics and properties, position and stance, and verbs expressing an obligation. In terms of Barðdal et al.’s (2012) fine-grained lexical semantic verb classification of dative subject predicates (listed in Section 2.6 of Chapter 2), the stative predicates identified in this thesis belong to the classes of verbs denoting emotions, verbs denoting attitudes, verbs of bodily states, stative verbs of cognition, verbs of modality, verbs of ontological states, verbs of hindrance, and verbs of evidentiality which denote epistemic judgements, e.g., ‘seem’. This subsumes the majority of verb classes employed in Barðdal et al. (2012). Most of these classes are furthermore categorized under the umbrella verb class of experiencer predicates in Barðdal (2011) and Barðdal et al. (2012). Jónsson (2003) classifies verbs taking a dative subject as experiencer verbs, verbs of convenience, verbs of success and failure, verbs of acquisition, motion verbs, change of state verbs, and ‘other’ verbs. With respect to this classification, the stative predicates in question can be mainly categorized as experiencer verbs with a few verbs of convenience, which are verbs describing characteristics/properties in my categorization. Moreover, some of the stative verbs are found in Jónsson’s ‘other’ class.

Verbs of transfer were not further divided into subclasses as the class is homogeneous in itself in that all verbs describe some sort of transfer, but the corresponding detailed subclasses are too heterogenous to arrive at a feasible number of categories for a quantitative analysis. Most predicates in this class are dative object verbs which take a dative subject in the passive. Passive dative subjects are not included in the lexical semantic verb classification by Barðdal et al. (2012). Yet, some of the middle forms of transfer verbs can be found in their classification, e.g., takast ‘succeed, manage’ is classified as a verb of success, and verbs of gain which do not refer to a scalar change, e.g., bjóðast ‘be offered’, verbs of speaking, as well as verbs of evidentiality denoting a physical appearance, e.g., birtast ‘appear’, correspond to my class of transfer predicates. Similar verb classes from Jónsson’s (2003) classification belong to the class of transfer verbs. These are the class of verbs of success and failure and verbs of acquisition.

Transition verbs describe simple transitions, the manner of motion, and mental processes. Verbs undergoing a simple transition and motion verbs are not part of the lexical semantic verb classes in Barðdal (2011) and Barðdal et al. (2012). Still, as an example for verbs of hindrance, Barðdal et al. (2012) cite missa ‘slip, stagger’
which is also a manner of motion verb. Moreover, the dynamic verbs of perception and cognition in the classification of Barðdal et al. (2012) correspond to the verbs describing a mental process in the present class. Jónsson (2003) employs a distinct class for motion verbs which contains some of the verbs describing the manner of motion.

Verbs describing a scalar change consist of change of state predicates, verbs describing changes of bodily states and a change of location. With respect to the verb classification provided by Barðdal et al. (2012), verbs of gain as well as verbs of bodily states describing a change of state belong into the present class. However, as stated before, motion verbs were in general not included in their classification. Barðdal (2011) lists the motion verb snúa ‘get, turned’ under ‘Miscellaneous’, belonging to the higher class category of happenstance predicates. Moreover, the change of state predicate ljúka ‘come to an end’, which can occur with a dative subject in active clauses when used intransitively, is classified as ‘Miscellaneous’ in Barðdal (2011). Some of the change of location predicates are found in the motion verb class in Jónsson (2003) and the verbs in Jónsson’s change of state verb class are similar to the verbs identified as change of state predicates in this thesis.

### 4.6 Corpus study II: Dative subjects and event structure

This section presents a corpus study which investigates the diachronic interaction between dative subjects, event structure and voice in IcePaHC using the glyph visualization developed by Schätzle and Sacha (2016). The glyph visualization was originally designed by Butt et al. (2014) for the investigation of the interaction between verb-first word order and different types of verbs and subjects in IcePaHC. In Schätzle and Sacha (2016), we extended the glyph visualization to cope with a larger number of interacting factors in IcePaHC for the investigation of lexical semantic verb classes and their interaction with dative subjects and voice. This study makes use of the extended version to account for the complex interaction between event structure, thematic roles and voice with respect to the licensing of dative subjects in Icelandic.

#### 4.6.1 Glyph visualization

The glyph visualization tool follows Shneiderman’s (1996) Information Seeking Mantra ‘Overview first, zoom and filter, then details-on-demand’. The visualization system
was implemented in Java and uses the basic visualizations and built in interactions of the Java Piccolo 2D framework. For the investigation of dative subjects, event structure and voice, the csv-file presented in the previous section which was partly shown in Figure 4.1 is used as input for the glyph visualization. Moreover, the glyph visualization takes files containing the original sentences from each text in IcePaHC (provided by the corpus distribution) as input to provide access to the underlying data in the visualization.

In the visualization, each text from IcePaHC is displayed as one composed glyph representing the relevant data dimensions. Glyph representations of documents are known from the field of Information Retrieval and have, for example, been employed in the TileBars technique (Hearst 1995). Using glyphs for the visual representation of documents allows for the compact and iconic representation of a document’s content with respect to certain query terms. Regarding the investigation of dative subjects and event structure, these query terms are the different classes of event structure with a dative subject and the factor voice. Figure 4.2 shows the different glyph representations available for each text from IcePaHC.

Each text glyph consists of mainly three parts: (i) a black horizontal sentence bar on top, (ii) a time-line on the right, and (iii) a matrix containing colored items, see, e.g., Figure 4.2-top left. The black sentence bar on top of the glyph indicates the length of the text in comparison to the longest text of the corpus, which would cover the whole width of the glyph. The light gray stripes drawn into the sentence bar correspond to the occurrence of a dative subject clause in the narrative flow of the text. On the right side of the glyph is a horizontal time-line representing the time span covered by IcePaHC. An orange tick-mark on the horizontal time-line indicates the age of the text in relation to the whole time-span. The central component of the glyph is a matrix containing colored items which encode different sentence features aggregated on the document level. Sentence features depicted on the rows of the matrix are encoded via different shapes, see Figure 4.2-bottom left, while features mapped onto columns are represented by different colors for better visibility. The position of each feature is fixed in the matrix of each text glyph for better comparison. Empty matrix cells indicate that the given feature combination did not occur in the text.

The glyph shown on the top left of Figure 4.2 encodes the different verb classes co-occurring with dative subjects in IcePaHC according to the classification given

---

10 http://www.piccolo2d.org
Figure 4.2: Different glyph representations for texts from IcePaHC with respect to dative subjects, verb classes and voice.
Chapter 4. Dative subjects and event semantics

in Table 4.10. Stative predicates are colored magenta, transfer verbs are light blue, transitions are dark blue and verbs describing a scalar change correspond to the light brown cells. The glyph representation of the verb classes can be extended on demand for more details via keystroke or clicking on a text to show the interaction of verb classes with voice, see Figure 4.2-bottom left, or for the investigation of the more detailed verb classes given in Table 4.11, see Figure 4.2-top right, and their interaction with voice as shown in Figure 4.2-bottom right. The more detailed verb classes are encoded via the same color as the corresponding higher class categories, compare, e.g., Figure 4.2-top left with Figure 4.2-top right. When the glyph is expanded for the interaction of verb classes with voice, the different voices are depicted on the rows and represented via different shapes, i.e., active sentences are indicated by circles, passives by rectangles, and middles are depicted as rounded rectangles, see Figure 4.2-bottom left.

The colored cells moreover encode whether a feature combination occurred more or less often than expected in the analyzed text via the scale given in Figure 4.3. If a colored cell is filled from outside, then the corresponding feature or feature interaction occurred more often than expected. If a colored cell is filled from inside, then the feature occurrence is smaller than expected. Both the observed and expected frequencies were calculated on the basis of the text length and the average occurrence of the respective feature in the whole corpus.

To provide an overview, each of the 61 text glyphs is plotted onto a fixed position on the y-axis according to its age. The oldest text is represented at the top of the y-axis and the newest text is depicted at the bottom, see Figure 4.5. The position of the text glyphs on the x-axis moreover indicates the genre of the texts which can be read as columns for the investigation of genre-specific differences. In order to facilitate the temporal comparison of text features, the user can switch between the overview with the genre-shifted alignment and an overview where the documents
are placed among each other horizontally, as shown in Figure 4.6. On demand, the
tick-mark can be extended to show the relative occurrence frequency of a hovered
feature interaction in each text, allowing the user to track the distribution of the
feature interaction throughout the whole diachrony. In this case, the tick-marks of
all features are repositioned to show the relative frequency of the hovered feature
value instead of indicating the text age, see, e.g., Figure 4.6, where the tick-marks
indicate the values of the light brown feature cells.

Several interaction techniques have been implemented into the visualization which
offer the possibilities of drilling down into the data if desired: Zooming and panning
interactions were added to navigate within the visualization’s viewport. Moreover,
details on demand are accessible through tool-tip operations providing information
about meta data such as the year date of a text or the absolute and relative occur-
rence frequencies of a feature interaction. Furthermore, the visualization provides
access to the underlying data by connecting the statistical analysis with the actual
sentences involved in the calculations, see, e.g. Figure 4.4, where the sentences with
stative predicates and a dative subject are displayed.

![Figure 4.4: The sentences which are involved in the calculation of a feature interac-
tion can be accessed via mouse-over in the glyph visualization.](image)

In the following, I present my findings obtained by means of the visualization
with respect to the interaction of dative subjects with event structure and voice. The
findings are moreover complemented with frequency calculations in the form of tables
where the data has been divided into five time periods as was done in Section 4.3.
Again, in the tables, p-values calculated via $\chi^2$-tests indicate whether the observed
distributions differ from what could be expected given the overall distribution of the
constructions in the corpus ($p<0.05 \,*,$ $p<0.01 \,**,$ $p<0.001 \,***$).
Figure 4.5: Glyph representations for each text from IcePaHC are positioned among each other on the vertical axis based on their diachronic order and are aligned on the horizontal axis with respect to their genre. Genre labels are shown on top and can be read as columns (SCI scientific texts, NAR narratives, REL religious texts, LAW law texts, and BIO biographies).
4.6. Corpus study II: Dative subjects and event structure

Figure 4.6: Glyph visualization in the temporal comparison layout. The tick-mark indicates the relative occurrence frequency of the selected/hovered feature which enables the temporal comparison of the feature value. Moreover, the temporal comparison layout allows for the emergence of visual patterns showing diachronic differences, see, e.g., the patterns A and B.
4.6.2 Dative subjects and event structure

Previous studies have shown that dative subjects most often occur with experiencer predicates in Icelandic and that this tendency moreover increases over time (e.g., Barðdal 2008, 2011). Given that the stative predicates in my classification are in essence experiencer predicates, I expect that dative subjects occur most frequently together with stative predicates in IcePaHC and that this co-occurrence furthermore increases diachronically. In the temporal comparison mode, two salient patterns became visible at-a-glance in the glyph visualization, see patterns A and B in Figure 4.6. Stative predicates (magenta) are common with a dative subject throughout the whole diachrony. However, pattern A is characterized by the striking absence of stative predicates as they occur a lot less often than expected in the respective text glyphs. By switching between the genre-shifted and the temporal comparison layout, the deviations observed in pattern A could easily be attributed to the genre effect which has been identified before. The texts in pattern A are religious texts or biographies stemming from the 16th to 17th century, whereas the large majority of preceding and following texts are narrative in nature. This has a major effect on the distribution of dative subjects and event classes in the third time stage in Table 4.12, in which the frequency of stative predicates is considerably lower than in the other time stages, causing the increased frequencies of the other event classes in the third time stage. Moreover, in the genre-shifted alignment as shown in Figure 4.5, the religious texts generally contain less stative predicates with a dative subject than expected across the whole corpus, showing once more that genre has an impact on the distribution of dative subjects in IcePaHC.

Pattern B in Figure 4.6 differs with respect to stative predicates from pattern A. Stative predicates and transition verbs occur more often than expected in texts from pattern B. Moreover, transfer verbs occur frequently with a dative subject. The texts in pattern B are narratives from the 20th century and correspond to the frequency calculations obtained for the last time period in Table 4.12. Altogether, stative predicates are increasing with dative subjects over the centuries, indicating an increasing systematic association between dative subjects and experiencer semantics. As expected, dative subjects generally occur most often with stative predicates from the earliest IcePaHC stages on, see Table 4.12 and the visualizations in Figures 4.5 and 4.6. Moreover, the usage of stative predicates increases significantly with a dative subject from 68.0% to 71.9% in the last time period. By expanding the glyphs for
the more fine-grained verb classes, as shown in Figure 4.7, it became clearly visible that mainly psych verbs describing a psychological state, i.e., experiencer predicates, are occurring with a dative subject within the class of stative predicates.

<table>
<thead>
<tr>
<th>Period</th>
<th>stative</th>
<th>transfer</th>
<th>transitions</th>
<th>scalar changes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>71.0%</td>
<td>15.0%</td>
<td>3.9%</td>
<td>10.1%</td>
<td>414</td>
</tr>
<tr>
<td>1350-1549</td>
<td>69.7%</td>
<td>19.3%</td>
<td>3.8%</td>
<td>7.2%</td>
<td>472</td>
</tr>
<tr>
<td>1550-1749</td>
<td>50.2%</td>
<td>29.6%</td>
<td>5.1%</td>
<td>15.2%</td>
<td>297</td>
</tr>
<tr>
<td>1750-1899</td>
<td>68.0%</td>
<td>13.9%</td>
<td>4.5%</td>
<td>13.6%</td>
<td>337</td>
</tr>
<tr>
<td>1900-2008</td>
<td>71.9%</td>
<td>14.3%</td>
<td>7.2%</td>
<td>6.6%</td>
<td>484</td>
</tr>
<tr>
<td>Average</td>
<td>67.3%</td>
<td>17.8%</td>
<td>4.9%</td>
<td>9.9%</td>
<td>2004</td>
</tr>
</tbody>
</table>

Table 4.12: Diachronic distribution of event classes with respect to dative subject predicates in IcePaHC.

Furthermore, verbs of transfer (light blue) constitute a large part of the dative subject predicates in IcePaHC (17.9%) and occur frequently in texts across the whole corpus, see, e.g., Figure 4.6. Their occurrence frequencies only decrease slightly from 15.0% to 14.3% over the whole time span and they remain rather stable overall. This is not surprising given that transfer verbs mainly consist of passives with a dative goal argument. For one, datives occur regularly on goal arguments overall (see, e.g., Maling 2001). For another, passive constructions were found to be stable across the whole diachrony (see Table 4.8). In addition, productive middle formation is possible with dative goal arguments.

Transition verbs (dark blue) were found less regularly with a dative subject than stative predicates and verbs of transfer, and have the lowest average occurrence frequency (4.9%). This was to be expected as the class of transition verbs is the smallest verb class investigated in this thesis. However, although they are generally rare, they are increasingly used in the corpus from the 19th century onwards showing an increase from 4.5% to 7.2% between the last two time stages, see Table 4.12 and the bottom part of Figure 4.6, including pattern B. Given that transition verbs generally have dative theme arguments which cannot undergo middle formation, the increase cannot be attributed to the overall increase of dative subjects with middle verbs. The increase correlates with an increasing use of verbs describing simple transitions and mental processes which occur more frequently and regularly in the latter part of the corpus, whereas manner of motion verbs are already commonly used in earlier stages, see Figure 4.7.
Figure 4.7: Glyph visualization in the temporal comparison layout showing detailed verb classes.
Verbs describing a scalar change (light brown) decrease significantly between the last two time stages from 13.6% to 6.6% although they generally occur frequently in the corpus, see Figure 4.6, with an average frequency of 9.9% in Table 4.12. While verbs describing a change of a bodily state only occur sporadically across the corpus, verbs of directed motion (change of location) were found regularly, see Figure 4.7. Still, change of state verbs occur most often with a dative subject within the class of verbs describing a scalar change. The decrease of verbs describing scalar change can in turn be attributed to the lack of middle forms with a dative subject, which occur frequently in the last time stage overall.

Yet, in order to shed light on the impact which voice has on the diachronic distribution of dative subjects across the different event structure verb classes and to find an explanation for the observed developments, a more nuanced analysis of the interaction between dative subjects, event structure, and voice is necessary. Moreover, a more detailed analysis is in order to be able to differentiate between developments effectuated by the lexicalized experiencer predicates which take a dative subject and the emergence of dative subjects via productive middle formation. This is done in the next section, where I investigate the diachronic interrelation between dative subjects, event structure, and voice in IcePaHC by means of the glyph visualization.

4.6.3 Dative subjects, event structure and voice

By expanding the glyphs for voice, further interesting findings were obtained. Stative predicates were found with active and middle constructions, but not with passive constructions, see the magenta matrix cells in Figure 4.8, which shows the bottom part of the glyph visualization when expanded for voice, and the frequency calculations in Table 4.13. This was to be expected given that stative predicates cannot form passives. Overall, dative subjects occurred most often with stative predicates in active constructions, with an average occurrence frequency of 67.0% for active clauses. However, whereas stative predicates with a dative subject decrease in active constructions, they increase continuously with middle voice. The most striking change happens in the period post-1900, where stative predicates in active constructions decrease significantly from 65.5% to 43.7%, while middles increase strikingly from 34.5% to 56.3%, making up most of the constructions with a dative subject and a stative predicate in the last time period. This moreover became visible in the visualization where middle constructions occur a lot more often than expected with stative predicates in the text glyphs after 1900, while active constructions occur less
often, see Figure 4.8-left. The increasing use of stative predicates as of 1900 thus correlates with an increasing use of stative predicates with middle voice.

<table>
<thead>
<tr>
<th>Period</th>
<th>active</th>
<th>passive</th>
<th>middle</th>
<th>Total</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>82.0%</td>
<td>0.0%</td>
<td>18.0%</td>
<td>294</td>
<td>***</td>
</tr>
<tr>
<td>1350-1549</td>
<td>80.9%</td>
<td>0.0%</td>
<td>19.1%</td>
<td>329</td>
<td>***</td>
</tr>
<tr>
<td>1550-1749</td>
<td>63.8%</td>
<td>0.0%</td>
<td>36.2%</td>
<td>149</td>
<td></td>
</tr>
<tr>
<td>1750-1899</td>
<td>65.5%</td>
<td>0.0%</td>
<td>34.5%</td>
<td>229</td>
<td></td>
</tr>
<tr>
<td>1900-2008</td>
<td>43.7%</td>
<td>0.0%</td>
<td>56.3%</td>
<td>348</td>
<td>***</td>
</tr>
<tr>
<td>Average</td>
<td>67.0%</td>
<td>0.0%</td>
<td>33.0%</td>
<td>1349</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.13: Diachronic distribution of voice with stative predicates taking a dative subject in IcePaHC.

Figure 4.8: Glyph visualization in the temporal comparison layout expanded for voice showing event classes (left) and more detailed verb classes (right) in texts shortly before and after 1900.

In Section 4.6.2, the increase of stative predicates was attributed to an increase of dative subjects with experiencer/psych predicates. Expanding the detailed verb classes for voice furthermore shows that the increasing use of stative verbs with
middle voice is driven by an increase of psych verbs with middles, see Figure 4.8-right. By mousing over the corresponding feature cells, the highly frequent use of the lexicalized experiencer verb *finnast* ‘think, find, feel, seem’ could be identified as main factor behind the increase of middle voice with psych verbs in the period post-1900. The frequent occurrence of the experiencer predicate *sýnast* ‘seem, appear’, which is the result of lexicalization of a middle form derived from a transfer verb, adds up to the high frequency of middle voice with stative predicates in the last time period.

In contrast to stative experiencer predicates, the transfer verbs are mainly passives, with an average frequency of 70.6%, see Table 4.14. Yet, passives decrease strikingly with transfer verbs in the last two time stages from 78.7% to only 37.7% which corresponds to a decrease of over half of their share, and cannot be driving the slight overall increase of transfer verbs observed before.

<table>
<thead>
<tr>
<th>Period</th>
<th>active</th>
<th>passive</th>
<th>middle</th>
<th>Total</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>1.6%</td>
<td>80.6%</td>
<td>17.7%</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>1350-1549</td>
<td>5.5%</td>
<td>75.8%</td>
<td>18.7%</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>1550-1749</td>
<td>3.4%</td>
<td>79.5%</td>
<td>17.0%</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>1750-1899</td>
<td>2.1%</td>
<td>78.7%</td>
<td>19.1%</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>1900-2008</td>
<td>0.0%</td>
<td>37.7%</td>
<td>62.3%</td>
<td>69</td>
<td>***</td>
</tr>
</tbody>
</table>

Table 4.14: Diachronic distribution of voice with transfer verbs taking a dative subject in IcePaHC.

In the visualization shown on the left of Figure 4.8, passive transfer verbs generally occur less frequently than expected, see the light blue cells in the middle row. At the same time, transfer verbs increase significantly with middle voice from 19.1% to 62.3%, making up the majority of transfer verbs in the last time stage which is furthermore visible in Figure 4.8-left. This increase correlates with an increasing use of the middle form *takast* ‘manage, succeed’, see (31), which takes a dative subject, originally derived via productive middle formation from the goal object of the ditransitive transfer verb *taka* ‘get, obtain, take’, see (32). Overall, active constructions are rare and only consist of the fixed expressions *gefa byr* and *gefa til* meaning ‘to get a fair wind’ as was exemplified in examples (23-a) and (23-b).
While stative predicates and transfer verbs increase significantly with middle forms in the corpus, transition verbs do not occur with middle voice at all, see Table 4.15. This is expected as dative theme arguments are not retained under middle formation. Transition verbs with dative subjects usually occur in passive constructions (77.8%). With passives, the transition verbs moreover increase and in particular in the period post-1900, which is visible at-a-glance in the visualization, see the dark blue cells in Figure 4.8. However, the last time period does not differ significantly from the voice distribution across the whole corpus.

<table>
<thead>
<tr>
<th>Transition verbs</th>
<th>active</th>
<th>passive</th>
<th>middle</th>
<th>Total</th>
<th>χ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1150-1349</td>
<td>25.0%</td>
<td>75.0%</td>
<td>0.0%</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>1350-1549</td>
<td>27.8%</td>
<td>72.2%</td>
<td>0.0%</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>1550-1749</td>
<td>26.7%</td>
<td>73.3%</td>
<td>0.0%</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>1750-1899</td>
<td>26.7%</td>
<td>73.3%</td>
<td>0.0%</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>1900-2008</td>
<td>14.3%</td>
<td>85.7%</td>
<td>0.0%</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>22.2%</td>
<td>77.8%</td>
<td>0.0%</td>
<td>99</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.15: Diachronic distribution of voice with transition verbs taking a dative subject in IcePaHC.

The glyph visualization moreover shows that transition verbs describing a simple transition, e.g., halda ‘hold, maintain, keep’ in (33), and verbs describing a mental process, e.g., líta til ‘look towards’ in (34), exclusively occur with a dative subject in passive constructions, see, e.g., Figure 4.8-right. Overall, however, mental process verbs and simple transitions are rare with a dative subject in the corpus. Verbs describing the manner of motion on the other hand, which are also classified as transition verbs, occur frequently across the whole corpus either in the form of unaccusative predicates which take a dative theme subject in active constructions,
4.6. Corpus study II: Dative subjects and event structure

...e.g., *skóla* ‘be washed ashore’ in (35) (repeated from (24)), or taking a dative theme object which surfaces as a dative subject in the passive, e.g., *velta* ‘roll’ in (36) (repeated from (25)).

(33) þá hafði biskupnum verið haldið í at that time have.PST.3SG bishop.the.DAT be.PTCP.PST hold.PTCP.PASS in myrkrastofu í næstum heilt år dungeon.DAT for nearly whole.ACC year.ACC

‘At that time, the bishop has been held in the dungeon for nearly a whole year.’ (IcePaHC, 2008.OFSI.NAR-SAG,.689)

(34) Geirmundi varð litið til Gróa, sem Geirmundur.DAT become.PST.3SG look.PTCP.PASS towards Gróa.GEN who sat gagnvart honum og þó ekki mjög næri. sit.PST.3SG opposite.to he.DAT and still not very near

‘Geirmundur happened to look (lit. ‘became looked’) towards Gróa who sat opposite to him and still not very near.’ (IcePaHC, 1902.FOSSAR.NAR-FIC,.674)

(35) Úr brimöldu draumins skolar mér upp að Out.of surge.DAT dream.the.GEN be.washed.ashore.PRS.3SG 1.DAT up to þessari hvítu strönd. this.DAT white.DAT beach.DAT

‘Out of a surge of the dream, I am washed ashore up to this white beach.’ (IcePaHC, 1985.SAGAN.NAR-FIC,.527)

(36) Með stunum og óhjóðum og miklum erfiðismunum with moans.DAT and shoutings.DAT and great.DAT effort.DAT.PL var þungum kössum velt upp á bryggjuna ... be.PST.3SG heavy.DAT crate.DAT roll.PTCP.up onto jetty.the.ACC

‘With moans and shoutings and great effort, the heavy crate was rolled up onto the jetty.’ (IcePaHC, 1907.LEYSING.NAR-FIC,.491)

Verbs describing a scalar change did not occur with a dative subject in middle construction, see Table 4.16, given that verbs describing a scalar change take a dative theme argument. The absence of middle voice with scalar change predicates can moreover easily be identified by means of the extended glyph visualization, where verbs describing a scalar change (light brown) do not occur in the last row of the matrix showing the interaction with voice, see Figure 4.8. Verbs describing scalar changes can occur with a dative subject in active constructions together with unac-
Chapter 4. Dative subjects and event semantics

cusative predicates, which often have a transitive correlate, where the dative argument corresponds to the object function and becomes a subject under passivization. For example, the change of state predicate *ljúka* ‘end’ occurs with a dative subject in active constructions, see (37) and is moreover found as a dative subject in the passive, see (38) (repeated from (30)). This is manifested in the data in that the average shares of active and passive constructions are close to one another with verbs describing a scalar change (59.3% vs. 40.7%), see Table 4.16.

(37) og lýkur þar þinginu.
and end.PRS.3SG there meeting.DAT
‘and the meeting ends there.’

(IcePaHC, 1450.BANDAMENN.NAR-SAG,43.936)

(38) og þá var þessari hátiðlegu athöfn lokið.
and then be.PST.3SG this.DAT festive.DAT ceremony.DAT finish.PASS.PTCP
‘and then this festive ceremony was finished.’

(IcePaHC, 1882.TORFHILDUR.NAR-FIC,.795)

<table>
<thead>
<tr>
<th>Period</th>
<th>active</th>
<th>passive</th>
<th>middle</th>
<th>Total</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>78.6%</td>
<td>21.4%</td>
<td>0.0%</td>
<td>42</td>
<td>**</td>
</tr>
<tr>
<td>1350-1549</td>
<td>67.6%</td>
<td>32.4%</td>
<td>0.0%</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>1550-1749</td>
<td>44.4%</td>
<td>55.6%</td>
<td>0.0%</td>
<td>45</td>
<td>*</td>
</tr>
<tr>
<td>1750-1899</td>
<td>56.5%</td>
<td>43.5%</td>
<td>0.0%</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>1900-2008</td>
<td>50.0%</td>
<td>50.0%</td>
<td>0.0%</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>59.3%</td>
<td>40.7%</td>
<td>0.0%</td>
<td>199</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.16: Diachronic distribution of voice with verbs describing scalar changes which take a dative subject in IcePaHC.

Still, active constructions with a dative subject and a verb describing a scalar change decrease across the whole corpus, while passives increase. In the last time stage, the shares for both construction types are equal. The visualization moreover shows that mainly change of state predicates and verbs describing a change of location occur increasingly with passive constructions within the class of scalar changes, see Figure 4.8-right. Verbs describing changes of bodily states, e.g., *batna* ‘recover’ as was shown in (29) which is repeated below in (39), are rare, but mostly occur in active constructions with a dative subject across the corpus.
Verbs describing a bodily state such as \textit{batna} ‘recover’ differ from the other verbs describing a scalar change in that they take a dative experiencer argument instead of a dative theme. Thus, as both verbs describing a scalar change and transition verbs are increasingly used with passives involving a dative theme argument, I conclude that dative themes are increasingly confined to being realized as objects. I take this to be the result of the diachronically increasing systematic association between dative subjects and experiencer semantics, whereby the Icelandic system becomes gradually regularized, avoiding dative-case marked theme subjects with active constructions. This assumption is discussed further in Chapter 6.

4.7 Summary and conclusion

This chapter examined the diachronic interaction between dative subjects and the factors voice, thematic roles and event structure in a detailed corpus study using IcePaHC. By means of the glyph visualization, I was able to conduct a nuanced analysis of the interaction between dative subjects, event structure, and voice across all texts contained in the corpus. As expected, dative subjects occurred by far most often with stative experiencer predicates in IcePaHC. Moreover, transfer verbs, which take a dative goal argument, were found frequently with a dative subject in passive and middle constructions. Transition verbs and verbs describing scalar changes, which have a dative theme argument, were found less often in the corpus.

Altogether, the study showed that the use of dative subjects increases across the time stages covered by the corpus, with a salient increase as of 1900. The increase correlates with a significant increase of stative experiencer predicates with a dative subject, indicating an increasing systematic association between dative subjects and experiencer semantics. The increasing use of stative experiencer predicates is moreover associated with an increasing use of middle forms which have been lexicalized as experiencer and raising predicates with a dative subject over time. These predicates are historically derived from middle forms of dynamic predicates and the dative case marking of the subject cannot be attributed to Dative Substitution.

Moreover, the overall frequency of dynamic verbs of transfer remains stable in the corpus. But, although verbs of transfer occur mainly with a dative subject in passive
constructions, there is a striking increase of transfer verbs with a dative subject in middle constructions in the period post-1900. With these middle constructions, the dative subject is the result of productive middle formation of a dative goal object.

Transition verbs and verbs describing a scalar change generally take a dative theme argument. As theme objects do not retain their dative case marking under middle formation, I found no middles with a dative subject in these classes. The active constructions in both classes consist of unaccusative verbs with a dative theme subject, apart from verbs describing the change of a bodily state, a subcategory of verbs describing a scalar change which usually take a dative experiencer subject. The corpus study showed that dative subjects decrease with transition verbs and verbs describing a scalar change in active constructions, while their usage increases in the passive. These developments were attributed to the increasing systematic association between dative subjects and experiencer semantics in the history of Icelandic, which increasingly confines dative themes to the object function. Overall, the corpus study showed that the lexical semantic distribution of dative subjects has been changing in the history of Icelandic. This argues against the Oblique Subject Hypothesis, which assumes that a monolithic dative subject construction has been handed down over the centuries.

With respect to event participants, dative subjects occur most frequently in constructions in which the subject is a state holder, i.e., with stative experiencer predicates in active and (former) middle construction as holder and with verbs of transfer in passive and middle constructions as resultee. Moreover, while the dative case marking of the resultee argument is retained under middle formation with transfer verbs, dative case is lost with the undergoer(-resultee) argument in middles of transition verbs and verbs describing a scalar change.

In Chapter 6, I will provide a precise analysis and detailed account of the interrelation between dative case licensing, grammatical relations, voice, and event structure, explaining the observed diachronic developments as well as the deviant behavior of theme arguments with respect to dative case marking with middle voice. In order to be able to account for the complex interacting system which links arguments to grammatical relations and licenses case in Icelandic, the diachronic interaction between subject case and word order needs to be examined in more detail. Syntactic position has been identified as a main criterion for the identification of dative subjects in Icelandic (see, e.g., Barðdal and Eythórsson 2003). However, word order has been noted to have changed over the history of Icelandic (see, e.g., Röguvaldsson
4.7. Summary and conclusion

and Thráinsson 1990, Rögnvaldsson 1996) and word order changes have moreover been identified as a relevant factor behind the reanalysis of dative arguments as subjects in the history of Indo-European (see Allen 1995, Haspelmath 2001). On these grounds, the next chapter presents an investigation of the diachronic interaction between subject case marking and word order with respect to subject positions and verb placement in IcePaHC, trying to shed light on the role which word order plays in the complex Icelandic system.
Chapter 5

Dative subjects and the rise of positional licensing

5.1 Introduction

This chapter presents an investigation of the diachronic interaction between case marking, grammatical relations and word order in IcePaHC, focusing in particular on the interrelation between subject case marking, subject positions and verb placement in Icelandic. In general, languages tend to use either word order, case and/or agreement as signal for grammatical relations (Kiparsky 1987, 1988, 1997). Yet, over time, one strategy to mark grammatical relations may be replaced by another. For example, in the history of Germanic, the loss of inflectional morphology generally correlates with the development of a rigid word order (see, e.g., Kiparsky 1997). Kiparsky (1997) noted that this correlation is a unidirectional implication: lack of inflectional morphology implies a fixed order of direct nominal arguments, but the converse is not true. English is an example for a language which has developed a rigid word order over time, while at the same time inflectional morphology has been lost. English underwent several structural changes which exemplify the development of a fixed order of constituents, e.g., the change from OV (Object-Verb) to VO (Verb-Object) order in the verb phrase (see, e.g., Kiparsky 1996) and the fixing of the order of direct and indirect objects (Allen 1995). Furthermore, subjects became obligatory in the history of English and SpecIP became the subject licensing position in the language (Hulk and van Kemenade 1995). On the whole, Kiparsky (1997) traces these changes back to the introduction of functional categories in his-
Chapter 5. Dative subjects and the rise of positional licensing

torical English (see also Kiparsky 1995, 1996). In particular, Kiparsky (1997) sees the rise of the category Infl/I, which was introduced in Old English and has become obligatory by the time of the Middle English period, as the common structural cause for the word order changes, leading to the increasing fixation of word order. In Kiparsky’s analysis, the rise of I in English is taken to ultimately be a consequence of the impoverishment of morphological case marking on NPs in the history of the language.

Icelandic is an example for the unidirectionality of Kiparsky’s implication: While the language sports a fairly fixed word order, a rich case and verbal morphology has been retained over the centuries (see Section 2.2 in Chapter 2). Thus, a rigid word order does not necessarily imply the loss of inflectional morphology. Kiparsky (1997) takes Icelandic to be a language in which position licenses grammatical relations independently of case marking, allowing for the possibility to have dative subjects in the language. This moreover parallels the diachronic developments in English, where according to Kiparsky, dative experiencers could only be licensed as subjects when they were placed in a particular structural position, i.e., SpecIP. As SpecIP as subject licensing position is a historical innovation, the possibility to have dative subjects in English goes hand in hand with the observed word order changes in the language. Kiparsky (1997) hereby mainly draws on evidence brought forward by Allen (1986, 1995) showing that dative experiencers could only be analyzed as subjects when they were structurally parallel with nominative subjects.

Since word order changes are taken to be a main factor behind the possibility to have dative subjects in the history of English, examining the diachrony of dative subjects in Icelandic in conjunction with the Oblique Subject Hypothesis necessitates an analysis of the interaction between dative subjects and word order in the history of the language. Several changes have been observed with respect to word order in the history of Icelandic: For one, Icelandic follows the Germanic shift from OV to VO order in the VP (e.g., see Kiparsky 1996, Rögnvaldsson 1996, Hróarsdóttir 2000). Moreover, although V1 constructions are still common in present-day Icelandic, the usage of V1 has been decreasing over the centuries (Sigurðsson 1990, Butt et al. 2014). In general, word order used to be less rigid in Old Icelandic than it is in the modern language (Rögnvaldsson 1995). Previous studies looked at the increase of dative subjects and the observed word order changes separately, treating them as mostly independent developments (e.g., Barðdal 2011, Rögnvaldsson 1996). The in-
vestigation presented in this chapter however shows that the phenomena are more closely interlinked than has been previously acknowledged.¹

The corpus study in this chapter investigates the interaction between subject case and word order in IcePaHC using the HistoBankVis visualization system (Schätzle et al. 2017). HistoBankVis allows for the sophisticated investigation of multiple interacting features from different linguistic dimensions at different levels of details, supporting the researcher in the process of hypothesis testing and generation. The relevant insights gathered by means of HistoBankVis are moreover combined with the corpus data and the LFG account presented in Booth et al. (2017), where we looked at the interaction between subject positions, V1 constructions and dative subjects in IcePaHC by means of a standard corpus linguistic approach.²

The investigation in this chapter provides evidence for the gradual development of structure in the history of Icelandic, leading to the rise of positional licensing in the language (in line with Kiparsky 1995, 1997). In particular, the data shows that, over time, the clause-initial and prefinite SpecIP position, which was identified as topic position in the history of Icelandic, becomes the preferred subject position in the language. However, dative subjects consistently lag behind subjects overall in being realized in a particular structural position. This argues against the inheritance of a stable dative subject construction from Proto-Indo-European, given that the dative experiencers do not follow the positional licensing constraints established for nominative subjects in the history of Icelandic straightaway. This is moreover consistent with the historical English data, where dative experiencers could only be licensed as subjects when they followed the structural requirements established for nominative subjects in the language, i.e., placement in SpecIP (see Kiparsky 1997).

The chapter is structured as follows: Section 5.2 presents the empirical corpus findings with respect to the interaction between subject case and word order in

¹Faarlund (1990, 2001) generally objects to an analysis of dative arguments as subjects in Old Icelandic arguing on the basis of word order. However, he takes Old Icelandic to have been strictly non-configurational, which is against the general view of the Icelandic literature (see, e.g., Rögnvaldsson 1995). But, partially in line with the empirical findings obtained by means of the study presented in this chapter, Faarlund (1990, 2001) sees the development of SpecIP as a specialized subject position in the history of Icelandic as a factor behind the occurrence of dative subjects in the language.

²Hannah Booth from the University of Manchester conducted the research on the interrelation between V1 and expletives presented in Section 5.3, whereas I thoroughly examined the diachronic distribution of the different word order patterns, in particular with respect to subject case. The corpus data shown in this chapter is similar to the data presented in Booth et al. (2017), but has been revised for the purpose of this thesis.
Chapter 5. Dative subjects and the rise of positional licensing

IcePaHC and introduces the HistoBankVis system which was employed as method of investigation. In Section 5.3, I discuss the interaction between V1 structures and the clause-initial expletive það in Icelandic, which sheds light on the functional purpose of the clause-initial position as regards information structure and grammatical relations. Section 5.4 provides a formal LFG account of the diachronic word order developments as presented in Booth et al. (2017), building on Sells’ (2005) formal analysis of Icelandic phrase structure and Kiparsky’s (1995, 1997) ideas regarding the rise of positional licensing. The analysis of the diachronic interaction between dative subjects and word order is postponed to Chapter 6, where all factors involved in the licensing of dative subjects in Icelandic are accounted for. Section 5.5 summarizes and concludes the chapter.

5.2 Corpus study: The diachrony of subject case and word order

Identifying and understanding complex interactions between various linguistic structures over time is the core remit of historical linguistics. The HistoBankVis visualization system was specifically designed to facilitate the analysis of diachronic interactions between a flexible number of linguistic features and structures, fostering the emergence of visual patterns of historical linguistic change. This section presents a corpus study which investigates the interaction between word order changes and subject case in IcePaHC with the aid of the HistoBankVis tool. In particular, I investigated the interaction between subject position, verb placement and dative subjects in a series of detailed visual analyses of the data.

5.2.1 The HistoBankVis system

We developed the HistoBankVis system with the overall goal of providing a generically applicable system for the analysis of the type of high-dimensional and complex data typically underlying historical linguistic research. HistoBankVis allows a researcher to interact with the data directly and efficiently while exploring correlations between linguistic features and structures. The system in effect consigns to history the painstaking work of finding patterns of language change across various different tables of features, numbers and statistical significances by granting an interactive exploratory access to a complex data set via the combination of different layers of
Corpus study: The diachrony of subject case and word order

5.2. Corpus study: The diachrony of subject case and word order

data representation with a structured statistical analysis process. The core piece of HistoBankVis is the visualization of the dimension interactions which makes use of Parallel Sets (Bendix et al. 2005, Kosara et al. 2006), a visualization technique coming from the field of visual analytics, allowing for the flexible analysis and visual presentation of the correlations between a potentially high number of features from different data dimensions, i.e., linguistic factors. To provide a better understanding of the visual analytic approach used in my corpus study, I briefly explain the main components of HistoBankVis in the following.

5.2.1.1 System implementation

The HistoBankVis system is implemented as an online browser app and can be accessed with every modern browser (e.g., Mozilla Firefox or Google Chrome) via the corresponding website. Additionally, a standalone implementation of the system can be provided as download on demand which executes the server and the client on the local machine, enabling the off-line usage of the system. Both the tool and the source code are freely available and open source. The back-end of the system is implemented in Java and is in charge of the data processing and management, the computation of statistical tests as well as the preparation of the data for the different visualization layers. The data which contains the extracted linguistic dimensions and the corresponding texts are stored and managed in a relational database. As database, the system supports SQLite and PostgreSQL. HistoBankVis was developed with respect to the concrete research task of investigating syntactic change in IcePaHC. The IcePaHC dataset as described in Section 2.8 of Chapter 2 can be explored via HistoBankVis by default, but the system also makes provision for the upload of other datasets. Moreover, a file containing the original Penn Treebank data is stored in the system, allowing for the connection of the extracted data with the underlying sentence annotation via the unique sentence IDs from IcePaHC. Information about the necessary format specifications for the upload of datasets and the meta-data is provided on the website along with an example. A REST (REp-
resentational State Transfer) interface (Fielding 2000, Richardson and Ruby 2008) handles the communication between the server and the client-side. The client-side is implemented in JavaScript with the AngularJS\textsuperscript{7} framework. The different visualizations are designed in D3.js\textsuperscript{8} as Scalable Vector Graphics (SVG) and can moreover be downloaded by the user.

5.2.1.2 Iterative multilayer approach

The HistoBankVis system follows the iterative workflow displayed in Figure 5.1. The iterative workflow is created by the availability of different layers of data representation and visualization. The concept behind HistoBankVis and the available representation layers are detailed below.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Fig5_1.png}
\caption{Iterative workflow of HistoBankVis.}
\end{figure}

**Data processing** First, the corpus data needs to be processed by extracting the linguistic factors which are relevant for the research task at hand, e.g., as described in Section 2.8 of Chapter 2. The relevant linguistic factors are usually identified by consulting the respective theoretical literature. In HistoBankVis, linguistic factors are stored as *data dimensions*, with the corresponding values referred to as *features*. For example, the linguistic factor *voice* is a data dimension and has the three features *active*, *middle* and *passive*.

**Filter and dimension selection** In the filtering component of HistoBankVis, the analyst can filter for a subset of the data and select the data dimensions which are of interest for a given research question. Besides filtering for data within a specific temporal range, the system’s filtering component allows for the visual construction of SQL-like filters, based on logical AND/OR functions, for features in the underlying database. In addition to filtering for data with specific features, the user can moreover select the data dimensions for analysis. For example, Figure 5.2 shows

\textsuperscript{7}https://angularjs.org/
\textsuperscript{8}https://d3js.org/
the filtering component when filtering for sentences from IcePaHC between 1900 and 2008 CE with a dative subject together with verbs carrying middle morphology adhering to a specific set of word orders. Moreover, the result table shown in Figure 5.2 displays the data with respect to the selected data dimensions subject case, voice, word order, and the verb involved in each sentence. If no user-defined filter settings are entered, all available features occur for the selected data dimensions. On demand, the Penn Treebank annotation for each sentence in the result table can be displayed in conjunction with all available extracted features.

<table>
<thead>
<tr>
<th>Sentence Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>From year 1900 to 2008</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>sbj_case</td>
<td>sbj_DAT</td>
</tr>
<tr>
<td>word_order</td>
<td>O1SV, V01S, SVO1, S01V, VSO1, O1VS</td>
</tr>
<tr>
<td>voice</td>
<td>middle</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Result Table</th>
<th>53 records</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Export Records</td>
</tr>
<tr>
<td>ID</td>
<td>verb</td>
</tr>
<tr>
<td>1902.FOSSAR.NAR-FIC.,1554</td>
<td>virða</td>
</tr>
<tr>
<td>1902.FOSSAR.NAR-FIC.,1404</td>
<td>deprast</td>
</tr>
<tr>
<td>1902.FOSSAR.NAR-FIC.,1050</td>
<td>finna</td>
</tr>
<tr>
<td>1902.FOSSAR.NAR-FIC.,782</td>
<td>virða</td>
</tr>
</tbody>
</table>

Figure 5.2: Filtering component of HistoBankVis.

Define time periods  

Before visualizing the historical developments of the selected data dimensions with respect to the filtered dataset, time periods need to be defined by the researcher for the temporal comparison. HistoBankVis allows the user to enter fully customized time periods or to split the data into an arbitrary number of equally sized temporal ranges. For the IcePaHC data, the system automatically
supports two divisions into time periods. These are (i) the division of the data into Old and Modern Icelandic, i.e., 1150–1549 and 1550–2008 CE (cf. Thráinsson 1994) and (ii) the division of the data into more fine-grained periods as defined per Haugen (1984), i.e., 1150–1349, 1350–1549, 1550–1749, 1750–1899, and 1900–2008 CE.

**Visualization** Different visualization components are available in HistoBankVis, allowing the user to interactively explore significant patterns contained in the data. The researcher can compare the distributions of all selected features and dimensions of the filtered sentences across different time periods and moreover investigate potential interactions between data dimensions. All views are equipped with mouse interaction techniques, providing details-on-demand of each part of the visualizations. A detailed description of the visualization layers follows in Section 5.2.1.3.

**Hypothesis generation and feedback loop** A central issue of historical linguistics is that the factors underlying a change are often unknown or at least highly debated. HistoBankVis allows for an iterative process of hypothesis generation and validation by combining knowledge-based and data-driven modeling. The user can immediately react to the insights collected by means of HistoBankVis and feed the knowledge gained back into the system by interacting directly with the different components and switching back and forth between the different layers of analysis.

### 5.2.1.3 Visualizing change over time

The visualization component of the HistoBankVis system consists of three main visualization layers which enable a structured analysis process by providing different views of the data at different levels of detail. The first layer is a compact matrix visualization which provides an overview of differences between the data in the selected time periods. HistoBankVis moreover contains a difference histograms visualization which allows for a detailed investigation of changing features and data dimensions. As uncovering and understanding complex interactions between various linguistic dimensions and features is the central task of historical linguistic work, HistoBankVis implements a dimension interactions visualization which focuses on the visualization of multiple interacting features between a set of data dimensions.

**Compact matrix visualization** The compact matrix visualizes the differences between the selected data dimensions across the selected time periods. In the ma-
5.2. Corpus study: The diachrony of subject case and word order

trix, each row and each column corresponds to one time period, see Figure 5.3-left. The matrix design particularly facilitates the comparison of the first period to all others and every period with its predecessor along the diagonal of the matrix, leading to the emergence of visual patterns of change. The system supports two modes of comparison: statistical significance and distance measurement. The size of the difference between two periods is indicated via a colormap where red indicates a high and white a low significance or distance. As measurement of statistical significance, the system employs $\chi^2$-tests and maps the computed $p$-values onto the color map: Red corresponds to $p = 0$ and white to $p \geq 0.2$. Statistically significant differences (with $\alpha = 0.05$) between two periods are marked by •. X in the middle of a cell signals the absence of the necessary preconditions for a $\chi^2$-test. When the necessary preconditions for $\chi^2$ are not met, e.g., when dealing with problems of data sparsity, the Euclidean distance can be used as an alternative measurement. The compact matrix visualization component is equipped with a bar chart showing the record counts per time range which provides information on whether there are sparsely populated time periods, see, e.g., Figure 5.3-right. When using Euclidean distance in the compact matrix, red depicts a high Euclidean distance, reflecting a large difference between the compared distributions as is shown in Figure 5.3-middle where the period 1900–2008 differs overall. Besides providing an overview of the data, the compact matrix visualization serves as a measure of quality in that potential changes can be spotted easily in combinations of periods which differ significantly from each other.

Figure 5.3: Compact matrix visualization showing differences between ranges via $\chi^2$ (left) or Euclidean distance (middle) and a bar chart displaying record counts per range (right).
**Difference histograms visualization**  Whereas the compact matrix visualization provides a quick overview of changes across time periods, the difference histograms allow for a more nuanced investigation of the diachronic development of individual features. In the difference histograms visualization, each time period is visualized as one composed bar chart, i.e., a difference histogram, see Figure 5.4. The data dimensions are encoded via different colors in each time period and can be inspected in parallel. For example, in Figure 5.4, the features of the dimension subject case occur as blue bars in each time period, while the dimension word order is colored orange. The height of a bar corresponds to the percentage of sentences in which a feature occurs compared to all features from the corresponding data dimension. The percentages of all features from a data dimension sum up to 100% in each time period. Additional information, e.g., the exact absolute and relative frequencies and the relative size of the feature occurrence in comparison to the overall text size, can be accessed via mouseover.

The difference histograms visualization moreover allows for the comparison of bar heights across the different time periods which provides insights into whether features are changing over time. To facilitate the temporal comparison of features, the difference between two neighboring time periods is computed and visualized as a separate bar chart below each feature bar. A green bar indicates that a feature has been increasing over time, while a red bar signals a decrease, see, e.g., Figure 5.4 where SVO1 (Subject-Verb-Direct Object) increases, while VSO1 (Verb-Subject-Direct Object) decreases in the last time period compared to the previous time stage. Moreover, the size of the green/red bars represents the size of the change. Besides comparing each period with its previous time range, the system supports further comparison modes. On demand, each time period can be compared with the first or last period, with the average of all time periods, or with the average of all periods before the current one.

Although the difference histograms provide an insight into the diachrony of individual features from different dimensions, one can not directly correlate simultaneously occurring changes of data dimensions with each other. That is, whether a change of a feature from one dimension is linked to the development of a specific feature from another dimension can not yet be determined by means of the difference histograms. To this end, HistoBankVis contains a separate layer of visualization using the concept of dimension interactions.
5.2. Corpus study: The diachrony of subject case and word order

Figure 5.4: Difference histograms for the distribution of subject case and word order in transitive sentences from IcePaHC.
**Dimension interactions visualization** For the analysis of interrelations between multiple data dimensions and in particular between multiple features of different data dimensions, HistoBankVis provides a dimension interaction visualization which employs the Parallel Sets (Bendix et al. 2005, Kosara et al. 2006) technique. Parallel Sets apply the idea of Parallel Coordinates (Inselberg 1985, 2009) to a frequency-based representation of categorical data. Parallel Coordinates visualize relations between individual data points from a multidimensional dataset on a 2D plane. In a Parallel Coordinates visualization, each data dimension is represented as a vertical axis and the related features of the dimensions are connected by a line. With respect to the analysis of linguistic data, Culy et al. (2011b) developed Structured Parallel Coordinates (SPC), a specialized version of Parallel Coordinates. SPC have mainly been used for the analysis of co-occurrences of words, see e.g., Figure 5.5, where frequencies of word sequences, which are connected via lines, are plotted onto the vertical axes.\(^9\)

![Figure 5.5: Structured Parallel Coordinates visualization of co-occurrence frequencies developed by Culy et al. (2011b).](http://www.eurac.edu/en/research/autonomies/commul/projects/Pages/Linfovis/programs.aspx)

\(^9\)Figure 5.5 is taken from [http://www.eurac.edu/en/research/autonomies/commul/projects/Pages/Linfovis/programs.aspx](http://www.eurac.edu/en/research/autonomies/commul/projects/Pages/Linfovis/programs.aspx).
5.2. Corpus study: The diachrony of subject case and word order

The SPC technique has moreover been applied to the visualization and analysis of historical linguistic data by Lyding et al. (2012) who investigate the diachrony of modal verb meanings within historical academic discourse. Moreover, Theron and Fontanillo (2015) track the evolution of meanings as represented in historical dictionaries via diachronlex diagrams which are based on parallel coordinate plots.

Instead of drawing individual data points as polylines across the different dimensions, Parallel Sets visualize the frequency of each feature as proportions of equally spaced vertical axis which represent the data dimensions. In a Parallel Sets visualization, the data dimensions are connected by colored boxes. The size of a box represents the share which a feature holds of a feature from another dimension. For example, the dimension interaction in Figure 5.6 shows the interaction between the data dimension voice and word order. The shares of the different voices are mapped onto the share they hold of different word order possibilities from left to right.

![Figure 5.6: Dimension interaction for voice and word order in transitive sentences with a dative subject from 1150–1349.](image)

In Figure 5.6, all three voices occur most often with the VSO1 word order. The first dimension in the Parallel Sets visualization, i.e., the leftmost dimension in Figure 5.6, determines the color which is used for the corresponding series of interactions throughout the visualization. The version of Parallel Sets which was implemented as the dimension interactions visualization of HistoBankVis allows for the reordering of data dimensions and features via drag&drop which enables a flexible investigation
of different types of interactions. Additionally, the features on each dimension axis can be sorted alphabetically and according to size in an ascending or descending order which provides a better overview and furthers the emergence of significant patterns. Moreover, detailed information about the feature correspondence and occurrence counts corresponding to an interaction box can be displayed via mouse interaction techniques, see the mouseover in Figure 5.6.

The remaining sections present the empirical observations obtained by means of the corpus study which investigates case marking, grammatical relations and word order in IcePaHC using the HistoBankVis system. In particular the dimension interactions visualization and the flexibility of the system facilitated the investigation of potential interrelations between subject case and word order immensely, leading to novel and unexpected insights into the Icelandic diachrony.

5.2.2 Subject case and word order

To obtain an overview about the diachrony of possible word order patterns and their potential interrelation with subject case, I first looked at the diachronic development of word order in transitive sentences vis-à-vis subject case via the difference histograms visualization component of HistoBankVis. I used the filtering component to filter for sentences containing a subject (S), a finite verb (V) and a direct object (O1) and selected the data dimensions subject case and word order for visualization. I visualized the data with respect to the second set of predefined time periods, i.e., the more-fine grained division as per Haugen (1984). The compact matrix visualization which is displayed in Figure 5.3 showed at-a-glance that the distribution of subject case and word order changes significantly over time with the largest change occurring in the last time period. Figure 5.4 provides the difference histograms for subject case and word order visualizing differences between the time periods compared to the previous period.

The difference histograms show that SVO1 is continuously increasing, with the most striking increase between the last two periods, see the green bars under the SVO1 feature in Figure 5.4. Concomitantly, VSO1 is decreasing (red bars). Interestingly, dative subjects increase at the same time. Whether there is a correlation between the observed word order changes and the growing use of dative subjects can easily be investigated by means of the dimension interactions.

\[10\]We used the code available on https://www.jasondavies.com/parallel-sets/.
5.2. Corpus study: The diachrony of subject case and word order

The dimension interactions provide detailed insights into the diachronic correlation between individual features from different data dimensions. Figure 5.7 shows the dimension interactions for subject case and word order in the first and the last time period. Since nominatives represent the largest share of all subjects in IcePaHC, see the difference histograms in Figure 5.4, nominative subjects pattern like subjects overall with respect to word order. That is, in the initial time stage, see Figure 5.7-top, SVO1 is already the most prominent word order for nominative subjects, albeit the proportions of SVO1 and VSO1 are still rather equal. Nominative subjects are increasingly associated with SVO1 over time and the large majority of nominative subjects occurs together with the SVO1 order in the last time stage, see Figure 5.7-bottom.

Figure 5.7: Dimension interactions for subject case and word order from 1150–1349 (top) and 1900–2008 (bottom).
Dative subjects pattern differently with respect to word order. In contrast to nominatives, dative subjects preferably occur in the VSO1 word order in the first time stage as is shown in Figure 5.7-top. VSO1 remains the dominant word order for dative subjects until the last time period which is only when SVO1 finally prevails with dative subjects, see Figure 5.7-bottom. Thus, whereas the difference histograms indicate an overall move towards SVO1, with subjects favoring the preverbal position, the dimension interactions showed that dative subjects lag behind with respect to this development.

5.2.3 Dative subjects, word order and voice

Voice is well-known to be a conditioning factor for the occurrence of dative subjects in Icelandic (e.g. Zaenen et al. 1985, Sigurðsson 1989). However, whether there is a correlation between voice, subject case and word order has not yet been investigated. I have shown in Chapter 4 that the distribution of voice is changing over time for dative subjects, in particular with respect to an increasing use of dative subjects together with middle forms, which are mainly lexicalized experiencer predicates, post-1900. In order to see whether there could potentially be a correlation between the eventual realization of dative subjects in SVO1 and the increase of the lexicalized experiencer predicates, I filtered the data once more to only include sentences with a dative subject and middle voice from 1900–2008 in addition to being restricted to the selected set of word orders. In the result table, see Figure 5.2, I selected the data dimensions verb, word order, voice and subject case and noticed that SVO1 is particularly frequent in this subset of the data. Moreover, the experiencer predicate finnast ‘think, seem, feel’, which is leading the increase of middle voice as of 1900 and the increase of dative subjects overall, occurs mainly together with SVO1. It is thus likely that the changes not only coincide, but are moreover interrelated.

To test this hypothesis, I disabled the previous filter for middle voice and visualized the filtered dative subjects data with respect to the data dimensions voice and word order. The dimension interactions show that in the first time period, VSO1 is the preferred word order with all three voices in dative subject sentences, see Figure 5.6. As was shown before, the preference for dative subjects to occur with VSO1 is lost in the last time period. Instead, dative subjects mainly occur together with SVO1. Whereas SVO1 is the preferred option with middles and active constructions in the last time period which is shown in Figure 5.8-bottom, the voices pattern differently with respect to word order in the period from 1750–1899.
In the second to last period, dative subjects still occurred most frequently in active constructions which in turn were mainly found together with VSO1, see Figure 5.8-top. In contrast to active sentences, the SVO1 word order is already the dominant word order for middle constructions in the period before 1900. These findings show that there is indeed a correlation between the increasing realization of dative subjects in the SVO1 word order and an increasing use of dative subjects with verbs carrying middle morphology, i.e., the lexicalized experiencer predicates. There is thus a correlation between the factors conditioning dative subjects and word order.

The overall increasing tendency to occur in the SVO1 word order indicates an increasing preference for subjects to be realized in the prefinite position, with dative
subjects following the overall developments at a later stage. However, by filtering the data for a specific set of word orders in transitive sentences, the investigations so far only looked at a subset of the data. Thus, in what follows, I look at the general distribution of subject positions and their interaction with subject case in more detail to obtain a more holistic picture.

5.2.4 Case and subject positions

For the analysis of the positional distribution of subjects in relation to subject case, I visualized the full IcePaHC dataset without any filter specifications, choosing the data dimensions subject position and subject case for visualization. The compact matrix showed that the distributions are changing significantly over time with respect to the fine-grained division into time periods, again with a salient change towards the last period. The difference histograms given in Figure 5.9 confirm my previous findings with respect to an increasing use of dative subjects as of 1900 and moreover show that subjects are increasingly realized in the prefinite position, while subjects in the postfinite position are decreasing accordingly.

Again, the largest increase is situated in the last time period. In the first time period, the proportion of prefinite subjects is only slightly bigger than the proportion of postfinite subjects and the subject positions are thus almost equally distributed in the initial stage. The respective numbers are provided in Table 5.1 with the percentage of prefinite subjects given in the penultimate column and \( \chi^2 \)-tests showing whether the observed distributions differ from what could be expected (with \( p < 0.05^* \), \( p < 0.01^{**} \), \( p < 0.001^{***} \)).

<table>
<thead>
<tr>
<th>Period</th>
<th>prefinite</th>
<th>postfinite</th>
<th>Total</th>
<th>% prefinite</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>7045</td>
<td>6672</td>
<td>13717</td>
<td>51.4%</td>
<td>***</td>
</tr>
<tr>
<td>1350-1549</td>
<td>10091</td>
<td>8258</td>
<td>18349</td>
<td>55.0%</td>
<td>***</td>
</tr>
<tr>
<td>1550-1749</td>
<td>6076</td>
<td>5134</td>
<td>11210</td>
<td>54.2%</td>
<td>***</td>
</tr>
<tr>
<td>1750-1899</td>
<td>6490</td>
<td>4767</td>
<td>11257</td>
<td>57.6%</td>
<td></td>
</tr>
<tr>
<td>1900-2008</td>
<td>7924</td>
<td>2937</td>
<td>10861</td>
<td>73.0%</td>
<td>***</td>
</tr>
<tr>
<td>All</td>
<td>37626</td>
<td>27768</td>
<td>65394</td>
<td>57.5%</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: Diachronic distribution of subject positions in IcePaHC.
Figure 5.9: Difference histograms for subject case and position in IcePaHC.
By looking at the dimension interactions, see Figure 5.10, I found that nominative subjects conform to the overall developments of all subjects with respect to subject position as shown in the difference histograms and in Table 5.1. Yet, dative subjects pattern differently once more.

Figure 5.10: Dimension interaction for subject case and position from 1150–1349 (left) and 1900–2008 (right).

Albeit the tendency to occur in the prefinite position is increasing over time for both nominatives and datives, nominative subjects occur in the prefinite position from the first time stage on, while dative subjects occur mainly postfinally until the second to last time period, see Figure 5.10-top for the dimension interaction in the first time period and the frequencies in Table 5.2. The prefinite position

11 The overall count given for dative subject constructions in Table 5.2 (and also in Table 5.4) differs from the count given in the tables in Chapter 4. The count is different because all types
only becomes dominant with dative subjects in the last time period, see Figure 5.10-bottom.

<table>
<thead>
<tr>
<th>Period</th>
<th>prefinite (DAT)</th>
<th>postfinite (DAT)</th>
<th>Total</th>
<th>% prefinite (DAT)</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>131</td>
<td>404</td>
<td>535</td>
<td>24.5%</td>
<td>***</td>
</tr>
<tr>
<td>1350-1549</td>
<td>126</td>
<td>465</td>
<td>591</td>
<td>21.3%</td>
<td>***</td>
</tr>
<tr>
<td>1550-1749</td>
<td>119</td>
<td>298</td>
<td>417</td>
<td>28.5%</td>
<td>*</td>
</tr>
<tr>
<td>1750-1899</td>
<td>151</td>
<td>277</td>
<td>428</td>
<td>35.3%</td>
<td></td>
</tr>
<tr>
<td>1900-2008</td>
<td>353</td>
<td>273</td>
<td>626</td>
<td>56.4%</td>
<td>***</td>
</tr>
<tr>
<td>All</td>
<td>880</td>
<td>1717</td>
<td>2597</td>
<td>33.9%</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.2: Diachronic distribution of subject position for dative subjects in IcePaHC.

Given the striking increase of subjects in the prefinite position in the period post-1900, I wondered whether the data can be interpreted as showing the development of a designated subject position in the history of Icelandic. In order to test this hypothesis, I investigated the options for verb placement in IcePaHC and the respective interaction with subject case in more detail.

### 5.2.5 Subject case and V1

In an earlier study (Butt et al. 2014), we investigated the distribution of V1 matrix declaratives in IcePaHC. We showed that although V1 declaratives can be attested throughout all attested stages of Icelandic, V1 undergoes a marked decrease as of 1900. Replicating this study with the present dataset yielded the frequencies displayed in Table 5.3 which confirms the findings presented in Butt et al. (2014). Initially, V1 clauses represent 20.6% of all matrix declarative sentences. This initial share decreases significantly over time and in particular between the last two periods where the frequency drops strikingly from 18.4% to 2.7%.\(^{12}\)

Whether dative subjects differ with respect to V1 can be investigated with Histo-BankVis by means of just a few clicks. For an investigation of the interaction between V1 constructions and subject case, I simply had to choose the data dimensions subject case and V1 for visualization. The difference histograms show the decrease of

---

\(^{12}\)The third time stage in the V1 distribution is affected by a genre effect which has been uncovered in Butt et al. (2014) via the glyph visualization introduced in Chapter 4.
Chapter 5. Dative subjects and the rise of positional licensing

V1 constructions together with the concomitant increase of dative subjects in the last time period, see Figure 5.11.

<table>
<thead>
<tr>
<th>Period</th>
<th>V1</th>
<th>non V1</th>
<th>Total</th>
<th>% V1</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>2829</td>
<td>10888</td>
<td>13717</td>
<td>20.6%</td>
<td>***</td>
</tr>
<tr>
<td>1350-1549</td>
<td>3656</td>
<td>14693</td>
<td>18349</td>
<td>19.9%</td>
<td>***</td>
</tr>
<tr>
<td>1550-1749</td>
<td>1654</td>
<td>9556</td>
<td>11210</td>
<td>14.8%</td>
<td>***</td>
</tr>
<tr>
<td>1750-1899</td>
<td>2072</td>
<td>9185</td>
<td>11257</td>
<td>18.4%</td>
<td>***</td>
</tr>
<tr>
<td>1900-2008</td>
<td>292</td>
<td>10569</td>
<td>11257</td>
<td>2.7%</td>
<td>***</td>
</tr>
<tr>
<td>All</td>
<td>10503</td>
<td>54891</td>
<td>65394</td>
<td>16.1%</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3: Diachronic distribution of V1 matrix declaratives in IcePaHC.

Visualizing the interactions between the data dimensions offered further valuable insights. The visualization shows that in general, dative subjects have a stronger overall tendency to occur in V1 constructions than nominative subjects, i.e., the majority of subjects. Despite minor fluctuations, which are furthermore shown in Table 5.4, the frequency of dative subjects in V1 constructions is going down, as do V1 constructions in general. The drop of V1 is very striking for dative subjects between the last two time stages (from 29.4% to only 3.2%) which emerges visibly in the dimension interaction for the period post-1900, Figure 5.12-bottom, in comparison to the interactions in the period from 1750–1899 displayed on the top of Figure 5.12. In addition to the overall findings with respect to word order and subject position, the decrease of V1 could be a further indicator that subjects become increasingly associated with the prefinite position over time in Icelandic. Once more, the preference to occur clause-initially is lower for dative subjects than for subjects overall.

<table>
<thead>
<tr>
<th>Period</th>
<th>V1 (DAT)</th>
<th>non V1 (DAT)</th>
<th>Total</th>
<th>% V1 (DAT)</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>173</td>
<td>362</td>
<td>535</td>
<td>32.3%</td>
<td>**</td>
</tr>
<tr>
<td>1350-1549</td>
<td>254</td>
<td>337</td>
<td>591</td>
<td>43.0%</td>
<td>***</td>
</tr>
<tr>
<td>1550-1749</td>
<td>106</td>
<td>311</td>
<td>417</td>
<td>25.4%</td>
<td></td>
</tr>
<tr>
<td>1750-1899</td>
<td>126</td>
<td>302</td>
<td>428</td>
<td>29.4%</td>
<td></td>
</tr>
<tr>
<td>1900-2008</td>
<td>20</td>
<td>606</td>
<td>626</td>
<td>3.2%</td>
<td>***</td>
</tr>
<tr>
<td>All</td>
<td>679</td>
<td>1918</td>
<td>2597</td>
<td>26.1%</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4: Diachronic distribution of dative subjects in V1 declaratives in IcePaHC.
Figure 5.11: Difference histograms for subject case and V1 in IcePaHC.
In sum, the study shows that subjects are increasingly realized in the prefinite position, while the usage of V1 decreases, with the 19th century as a major key turning point. Overall, this points towards the development of structure and positional licensing in the history of Icelandic, in particular with respect to the emergence of a designated subject position. Dative subjects generally have a weaker tendency to be realized in a particular position and follow the overall developments only at a later point. Visualizing the data with HistoBankVis allowed for a flexible investigation of interactions between linguistic factors, while at the same time being able to keep an overview and inspect individual data points in detail. Moreover, the system enables the verification of a priori hypotheses within minutes, generating an iterative and productive analysis process.
5.3 V1 and expletives

In order to show whether the V1 decrease indeed indicates the development of a designated subject position in Icelandic, the factors which potentially led to the decrease of V1 in the language have to be further investigated. For historical German, Axel (2007) has observed that the diachronic decrease of V1 structures is connected to the rise of clause-initial expletives in the language. The decrease of V1 declaratives in Icelandic has previously been noted in the literature and the rise of clause-initial expletives has also been suggested as a factor behind the decrease (see, e.g., Sigurðsson 1990, Franco 2008). However, the exact interaction between V1 and expletives has not yet been studied in detail. In present-day Icelandic, the overt expletive það occurs clause-initially in various constructions, for example, in presentationals, see (1-a), placing the verb in second position, but is typically absent in older stages, where presentationals are usually realized as V1 sentences, see (1-b).

(1) a. Það var tölumerður snjór yfir öllu.
   ‘There was a considerable amount of snow over everything.’
   (IcePaHC, 2008.OFSI.NAR-SAG,.772)

   b. Var þá gleði mikil í kongs höll.
   ‘There was then great joy in the king’s hall.’
   (IcePaHC, 1480.JARLMANN.NAR-SAG,.48)

In Booth et al. (2017), we examined the interaction between V1 and expletives more closely by means of a corpus study using IcePaHC in order to find out whether the occurrence of the overt expletive það could be a factor behind the decrease of V1 in Icelandic. The empirical findings adduced by this study are detailed in the following.

5.3.1 The diachrony of expletive það

The data given in Table 5.5 shows the proportion of instances in which expletive það is present in the clause-initial prefinite position, rendering the sentence V2, and compares it to the V1 structures in which það is absent, i.e., the V1 structures for which an empty expletive has been annotated in IcePaHC.13

---

13 The expletive data presented in Table 5.5 was adduced by Hannah Booth (University of Manchester).
Table 5.5: Diachronic distribution of prefinite expletives in IcePaHC.

<table>
<thead>
<tr>
<th>Period</th>
<th>overt expl (V2)</th>
<th>no expl (V1)</th>
<th>Total</th>
<th>% overt expl (V2)</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>16</td>
<td>153</td>
<td>169</td>
<td>9.5%</td>
<td>***</td>
</tr>
<tr>
<td>1350-1549</td>
<td>26</td>
<td>205</td>
<td>231</td>
<td>11.3%</td>
<td>***</td>
</tr>
<tr>
<td>1550-1749</td>
<td>13</td>
<td>87</td>
<td>100</td>
<td>13.0%</td>
<td>***</td>
</tr>
<tr>
<td>1750-1899</td>
<td>59</td>
<td>92</td>
<td>151</td>
<td>39.1%</td>
<td>***</td>
</tr>
<tr>
<td>1900-2008</td>
<td>160</td>
<td>28</td>
<td>188</td>
<td>85.1%</td>
<td>***</td>
</tr>
<tr>
<td>All</td>
<td>274</td>
<td>565</td>
<td>839</td>
<td>32.7%</td>
<td></td>
</tr>
</tbody>
</table>

The frequencies in Table 5.5 show that the prefinite expletive það occurs rarely in the older texts from IcePaHC, but its usage increases over time, with a marked increase as of 1900. This coincides with a simultaneous decrease of V1 structures with an empty expletive and the overall V1 decrease. It seems thus plausible to assume that the expletive það is a factor behind the decrease of V1, given that the overt expletives in clause-initial position render older V1 structures as V2. However, the expletive story cannot fully explain the decrease of V1 clauses. Not all V1 declaratives in IcePaHC are constructions which lack an overt expletive. Consider for example (2) in which a prototypical referential subject (drottinn ‘the lord’) appears in the postfinite position and an overt expletive needs not to be postulated. Such examples occur frequently across the corpus, suggesting that the rise of expletive það cannot be the sole reason behind the decrease in V1.

(2) Sýndi drottinn mikla miskunn vin sínum show.PST.3SG lord.NOM.DEF great.ACC mercy.ACC friend.DAT his-own.DAT sankti Georgíum saint.DAT George.DAT

‘The Lord showed great mercy to his friend St. George.’

(IcePaHC, 1525.GEORGIUS.NAR-REL,.535)

Yet, the expletive story sheds more light on the function of the clause-initial, prefinite position in Icelandic. The diachronic findings for subject positions outlined in Table 5.1 led to the assumption that Icelandic developed a newly designated clause-initial and prefinite subject position as of 1900, and the decrease of V1 was interpreted as a corollary thereof. However, the expletive það does not straightforwardly qualify as a subject and therefore poses a problem for such an analysis.
5.3. V1 and expletives

5.3.2 V1, expletives and subject positions

In non-subject fronting contexts, subjects typically invert with the verb in Icelandic, retaining a V2 structure, compare, e.g., the subject initial clause in (3-a) with the inverted structure in (3-b), where the object is topicalized.

(3) a. Ég gleymdi þeim fljótt.
   I.NOM forget.PST.1SG they.DAT quickly
   ‘I quickly forgot them.’

b. Þeim gleymdi ég fljótt.
   they.DAT forget.PST.1SG I.NOM quickly
   ‘Them I quickly forgot.’

With respect to topicalization, expletive það does not behave like a prototypical subject. The expletive það is restricted to the clause-initial prefinite position in almost all construction types in Modern Icelandic (see, e.g., Sells 2005, Sigurðsson 2007, Thráinsson 2007). In contexts where another element occurs in initial position, það is generally absent, see the impersonal passive construction from IcePaHC in (4-a) in which the negation ekki is fronted. The corresponding construction without fronting has a clause-initial það and is exemplified in (4-b).

(4) a. Ekki var minnst á önnur dýr.
   not be.PST.3SG mention.PTCP on other.ACC animals.ACC
   ‘There was no mention of other animals.’

   (IcePaHC, 1985.SAGAN.NAR-FIC,.248)

b. það var ekki minnst á önnur dýr.
   EXPL be.PST.3SG not mention.PTCP on other.ACC animals.ACC
   ‘There was no mention of other animals.’

In contrast, the Swedish expletive det is overtly realized in both the pre- and the postfinite position in impersonal passives, see (5).

(5) 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.’
Data like (4) and (5) are standardly adduced to support the claim that the Swedish expletive *det* qualifies as a syntactic subject, but the Icelandic expletive *það* does not (see, inter alia, Platzack 1983 and Maling 1988, and also Faarlund 1990 on ‘expletive topics’).

Further evidence against the subject status of the Icelandic expletive *það* is provided on the basis of the grammaticality of so-called “Transitive Expletive Constructions”, see (6), in which the expletive co-occurs together with an overt thematic and logical subject (margir jólasveinar ‘many Christmas trolls’) in Icelandic. It is thus unlikely that *það* fulfills the role of a subject in (6). Furthermore, Transitive Expletive Constructions such as (6) are not permitted in Swedish.

(6) *það* hafa margir jólasveinar borðað
  EXPL have.PRES.3PL many.NOM Christmas-trolls.NOM eat.PST.PTCP
  þúðing. pudding.ACC
  ‘Many Christmas trolls have eaten pudding.’

  (Bobaljik and Jonas 1996, 209; gloss modified)

Given that the expletive *það* is increasingly realized in the prefinite position (cf. Table 5.5), but does not straightforwardly qualify as a subject, the initial hypothesis that the observed word order changes indicate the development of a newly designated subject position has to be revised. Instead, an alternative account for the structural developments observed is articulated in the next section which relies on the assumption of having an information structural motivation for V1 and expletives. The account is inspired by the core observation of Rögnvaldsson and Thráinsson (1990) that the function of expletive *það* is to license a clause in which there is no topic in Icelandic, not even the subject which otherwise tends to be the topic by default (see also Zaenen 1983).
5.4 The rise of positional licensing – an LFG analysis

An information structural motivation for V1 in Germanic has been previously articulated in the literature (e.g., Hinterhölzl and Petrova 2010) and taken into account by Sells (2005) who provides a formal analysis of Icelandic within LFG. This section presents an LFG analysis for the rise of positional licensing in the history of Icelandic based on the corpus evidence shown in Section 5.2. This analysis builds on Sells’ (2005) formal proposals. In line with Sells (and others, e.g., Rögnvaldsson and Thráinsson 1990), my analysis takes the function of það to be the licensing of topicless clauses, motivated by information structure. However, unlike Sells, I do not assume that það plays the role of a subj in Icelandic.

The analysis moreover draws on Kiparsky’s (1995, 1997) ideas regarding the rise of positional licensing and concomitant “growth” of syntactic structure. Kiparsky (1995) argues that the Germanic languages innovated structures and associated functional categories which were not present in Indo-European, the ancestor of the Germanic languages. On the basis of the Old English data brought forward by Hulk and van Kemenade (1995) and van Gelderen (1993), Kiparsky (1997) moreover takes the stand that the functional projection I was optional in Old English and only became obligatory at a later stage. The idea of the growth of structure in that a flat structure lacking functional properties may develop into a more articulated structure has been applied to Icelandic by Börjars et al. (2016) explaining functional changes in the Icelandic noun phrase. Moreover, Kinn et al. (2016) provide evidence for the overall rise of positional licensing in Icelandic by showing that the frequency of referential null subjects has decreased over the history of Icelandic.

For the analysis of the structural developments of Icelandic as evidenced in my corpus study, Kiparsky’s (1995, 1997) ideas on the development of structure and rise of positional licensing for grammatical relations are combined with an information structural account for word order. Kiparsky’s (1997) hypothesis for Icelandic is that while morphological marking has not been lost, it has become recessive in determining grammatical relations. Instead, the positional licensing of grammatical relations has become dominant, which in turn accounts for the increasing word order rigidity. My analysis moreover relies on the restrictive approach to functional categories in LFG (see Kroeger 1993, Börjars et al. 1999) which assumes a functional category when a functional feature is associated with a structural position.

---

14 The analysis presented here is similar to the analysis we provided in Booth et al. (2017).
Chapter 5. Dative subjects and the rise of positional licensing

5.4.1 V1 and information structure

In the earliest data, V1 structures are still prominent. With respect to V1 and information structure, I follow the account proposed by Hinterhölzl and Petrova (2010) for V1 in historical German, which assumes that the finite verb served as an information structural boundary, separating topic from comment. V1 thus marked topicless sentences, e.g., presentationals, in which the whole clause was placed in to the scope of assertion, i.e., the whole clause was focused. Along the lines of the analysis developed for Urdu by Butt and King (2004), I assume a flat sentence structure for the earliest Icelandic stages (pre-IcePaHC) in which case morphology was used to mark grammatical relations and word order was used for information structural purposes.

The information-structural role of the finite verb is, in combination with a number of other factors, what eventually led to the growth of structure. The finite verb becomes more fixed in second position due to its information structural function, i.e., separating topic and comment, leading to the rise of I. Accordingly, the initial position becomes increasingly associated with a specific discourse function, representing given or topical information. With respect to historical English, Kiparsky (1997) provides evidence for a scenario in which the functional category I was optional in Old English, before it became an obligatory category in the English phrase structure. Extending Kiparsky’s analysis, the following clause structure is emerging for Icelandic (with DF=Discourse Function):

\[
\text{IP} \quad \overleftarrow{\text{I}}' \quad \overrightarrow{\text{VP}}
\]

Thus, assuming the structure in (7), the prefinite position in SpecIP is associated with a specific discourse function, e.g., topics. Subjects tend to be topical because they generally encode the more agentive, sentient clausal participant which usually makes for a better topic than an inanimate patient. Given the cross-linguistic tendency of subjects to be topics, it has been argued that subjects are in fact grammaticalized topics (Givón 1990). Based on this assumption and the corpus findings, I
5.4. The rise of positional licensing – an LFG analysis

posit that the SpecIP topic position becomes increasingly associated with the subject function over the history of Icelandic as exemplified by the clause structure in (8).

(8)

\[
\text{IP} \rightarrow (↑\text{SUBJ})=↓ \text{I'} \rightarrow \text{XP} \rightarrow \text{I} \rightarrow \text{VP}
\]

5.4.2 Expletives and positional licensing

However, the tree in (8) is underspecified as Modern Icelandic still allows for constructions in which the subject is placed in the postfinite position. For example, in clauses like the Transitive Expletive Construction, repeated below in (9), the thematic subject occurs after the finite verb.

(9) Það hafa margir jólasseinar bórðað.

‘Many Christmas trolls have eaten pudding.’

(Bobaljik and Jonas 1996, 209; gloss modified)

Sells (2001, 2005) proposes the blueprint for clausal structure in Icelandic given in (10), which is able to account for the availability of postfinite subjects. The main idea of the structure is that the prefinite position in SpecIP is associated with a discourse function and a grammatical function (GF). Subjects can occur in the clause-initial position, but are not restricted to this position as the tree is equipped with further landing positions for grammatical functions under I’ and in the VP. In particular, the tree accounts for subjects occurring in the immediately postfinite position, i.e., after I, when the prefinite position is occupied by a fronted element, the expletive það, or in the context of V1. Sells assumes that the clause structure of Icelandic is governed by a set of linear constraints.

Yet, under the current assumption that the expletive það is not a subject, the possibility to have það in the prefinite position, which is reserved for grammatical functions, poses a problem for the clause structure in (10). Arguing for the subject
status of það, Sells treats Transitive Expletive Constructions as cases in which the information from the thematic subject of the clause, i.e. margir jólasveinar ‘many Christmas trolls’ in (9), is unified with the information coming from the expletive það via linear constraints. However, direct evidence for the subjecthood of það is not given by Sells. As stated before, in contrast to Sells, I take examples like the Transitive Expletive Construction in (9) as providing evidence against the idea of það being a canonical subj.

(10)

Thus, in my view, the structure in (10) has yet to be refined with respect to the elements hosted by the SpecIP position because það does not qualify as a subject. Instead, following Rögnvaldsson and Thráinsson (1990), það fulfills an information-structural role in that it marks sentences which lack a topic. Taking this into account, the possibilities for the SpecIP position in Icelandic are given in (11).

(11)
5.4. The rise of positional licensing – an LFG analysis

For one, SpecIP can be a topic position, typically hosting subjects, but also other topical grammatical functions. Alternatively, the expletive það can occur in SpecIP. When this is the case, then the sentence has no topic explicitly stated in (11). If the SpecIP position remains unfilled, the resulting structure is V1. According to the tree in (11), the resulting V1 structure would then lack a topic. This is the case with presentational V1 sentences such as in (1-b), repeated below in (12). However, not all V1 structures are structures which lack a topic, e.g., the narrative inversion-type of V1 has an overt topical subject directly after the finite verb (see Sigurðsson 1990), as shown in (13) (taken from Booth 2018, 102; gloss added). In example (13), the pronominal subject encodes discourse-old, i.e., given, information. Thus, the immediately postfinite position can in general also be analyzed as a topic position in Icelandic.

(12) Var þá gleði mikil í kóngs höll.
be.PST.3SG then joy.NOM great.NOM in king.GEN hall.DAT
‘There was then great joy in the king’s hall.’

(IcePaHC, 1480.JARLMANN.NAR-SAG,.48)

(13) Var hann þar þrjár nætur
be.PST.3SG he.NOM there three.ACC nights.ACC
‘He was there for three nights.’

(IcePaHC, 1310.GRETTIR.NAR-SAG,.201)

Yet, the corpus findings indicate that, over time, the information structural packaging of Icelandic is reorganized in favor of a designated prefinite topic position, since V1 decreases overall and the expletive það becomes obligatory in the clause-initial position as marker for topicless sentences. Moreover, the prefinite position becomes increasingly associated with subjects. In sum, the present findings and the analysis are in line with Kiparsky’s proposal for the rise of positional licensing in historical English, indicating that Icelandic developed structure over time.

15In Icelandic, the clause-initial position can moreover signal contrastive focus. Moreover, scene setting adverbials can be placed clause-initially (e.g., in clauses like ‘In the garden eats Mary the oranges’). In those instances, the verb does not function as separator between topic and comment. More research has to be conducted to separate out the precise discourse function of the clause-initial position and the immediately post-finite one.

16Moreover, constructions which have a stage topic in first position allow for the possibility to have a postfinite subject that is the topic of the clause. These constructions have not been taken into account in the present thesis, but will also be part of future work. Thanks go to Annie Zaenen for pointing this out to me.
A significant finding which has not as of yet been discussed in this chapter is that dative subjects consistently lag behind in being realized in a particular position and only follow the overall licensing constraints developed in the language in the period post-1900. The diachronic interaction between dative subjects and the rise of positional licensing is discussed in detail in Chapter 6, where I provide an analysis of the diachronic development of dative subjects using my own theory of linking, accounting for the complex interacting system which licenses case and grammatical relations in Icelandic. Moreover, in Chapter 6, the interaction between information structure and word order as observed in this corpus study is incorporated into my linking theory in terms of a reference frame which is based on the distinction between figure and ground by Talmy (1978). The precise interrelation which I assume to hold between the figure-ground division, information structure and word order is described in more detail in Section 6.2.2 of Chapter 6.

5.5 Summary and conclusion

The investigation of the interrelation between subject case marking and word order in IcePaHC presented in this chapter provides evidence for the development of structure and the rise of positional licensing in the history of Icelandic (in line with Kiparsky 1995, 1997). In particular, the corpus investigation showed that over time, subjects are increasingly realized in the prefinite position, whereas V1 constructions decrease concomitantly. In the formal analysis of the corpus evidence which is given in Section 5.4, I argue that the data indicates the development of a designated, prefinite topic position in SpecIP during the Icelandic diachrony, which begins to increasingly host subjects. Moreover, SpecIP is not yet exclusively a subject position in Modern Icelandic, given that V1 declaratives are still possible in the language and the prefinite expletive það, which is not a subject, can occur clause-initially, e.g., in conjunction with a postfinite subject in the Transitive Expletive Construction. A significant finding of the present investigation, which was uncovered by means of the HistoBankVis visualization system, is that dative subjects show a weaker tendency to be realized in a particular structural position than subjects overall. Only as of 1900, dative subjects follow suit with the majority of dative subjects occurring in the prefinite, clause-initial position. In conclusion, given that in older stages of the language, dative subjects differ significantly from subjects in general as regards the occurrence in a particular structural position, it is rather unlikely that dative sub-
jects are a stable part of the language, which argues in favor of the Object-to-Subject Hypothesis.

This chapter was concerned with the examination of the diachronic interaction between subject case and word order in Icelandic. The following chapter provides a detailed analysis of the empirical findings obtained via the investigations conducted in this thesis, discussing the complex and moreover changing interrelation between case, word order, lexical semantics and event structure in Icelandic. I introduce a novel linking theory in the next chapter, which serves to account for the complexity of the system which licenses grammatical relations in the language as well as the diachronic changes evidenced by the present thesis.
Chapter 6

Linking events, case and grammatical relations in Icelandic

6.1 Introduction

The corpus studies presented in Chapters 4 and 5 of this thesis yielded results which now need to further be understood. For one, the investigation of the interaction between event semantics, thematic roles, voice, and dative subjects in Chapter 4 showed that the usage of dative subjects increases strikingly in the history of Icelandic. This increase correlates with an increasing systematic association between dative subjects and stative experiencer verbs. In the previous literature, the increasing systematic association between experiencer semantics and dative subjects in Icelandic has been attributed to the process of Dative Substitution (see, e.g., Barðdal 2011 and Section 2.6 in Chapter 2). However, the IcePaHC data shows that Dative Substitution is not the main factor behind the increase of dative subjects. Instead, verbs which carry middle morphology that have been lexicalized as stative experiencer predicates with a dative subject over time are driving the increase. The study furthermore showed that the possibility of having dative subjects in Icelandic generally correlates with particular event structural configurations.

The study investigating the interaction between word order changes and subject case marking in the history of Icelandic presented in Chapter 5 showed that the increasing tendency for dative subjects to be realized in the prefinite, clause-initial position correlates with the increase of dative subjects with experiencer semantics. The overall picture that emerges is that case marking and grammatical relations
are part of a complex system in Icelandic in which position, event structure, lexical semantics, case and grammatical relations interact with one another in a systematic fashion. Moreover, the system has been changing over the history of Icelandic, leading to the rise of positional licensing and an increasingly systematic association between dative case and experiencer semantics. This in turn argues against the idea of dative subjects as a stable part of the Icelandic diachrony, questioning the inheritance of a common monolithic dative subject construction from Proto-Indo-European.

In order to account for the complex interacting system which licenses case and grammatical relations in Icelandic, I propose a novel theory for linking in this chapter. The theory incorporates the interrelation between positional licensing and information structure in terms of a reference frame based on Talmy’s (1978) figure-ground division which mediates between position, event structure and lexical semantics. With respect to event structure, crucial insights from Ramchand (2008) are integrated into the system. As per Svenonius (2002), event structure is taken to interact with case marking in my linking theory. In my system, a set of lexical semantic entailments follow from the event structure, the reference frame, and the sentience/animacy of the event participants. The lexical semantic entailments are implemented in the form of Dowty’s (1991) Proto-Role entailments which are employed for linking arguments to grammatical relations as per Zaenen’s (1993) linking theory. I apply my linking theory to the analysis of the empirical findings obtained by means of the corpus studies in this chapter and argue against the Oblique Subject Hypothesis by postulating different sources for the dative case marking of subjects in Icelandic. The linking analysis shows that the combination of event structure with the reference frame for linking is key to understanding the Icelandic system. In the analysis, specific combinations of event structure and the reference frame surface together with dative subjects and the emerging patterns allow for a general and concise account of dative subjects in Icelandic.

The chapter proceeds as follows: Section 6.2 introduces the new linking system. In Section 6.3, I provide a summary of the interrelation between event structure and case which I postulate throughout the analysis. Section 6.4 sets out to present my theory regarding the historical development of dative experiencer subjects with stative predicates, factoring in the rise of positional licensing in the history of Icelandic. Subsequently, I provide an analysis of the development of new lexicalizations instantiating stative experiencer verbs with a dative subject via middle formation.
and diachronic reanalysis in Section 6.5. In Section 6.6, I present an analysis of ditransitive predicates with a dative goal argument, arguing for the licensing of dative case marking based on the corresponding event structure configuration. Basing myself on the theory of case marking developed by Svenonius (2002) for Icelandic and developing his ideas further, I show in Section 6.7 that dative case is licensed with theme objects of dynamic predicates when there is a temporal mismatch between two subevents of a single and indivisible event. Moreover, with unaccusative predicates, the licensing conditions are similar to the ones that are postulated for verbs with a dative theme object. The analysis furthermore shows that dative themes and goals are licensed under different event structural preconditions which explains their contrasting behavior with respect to case marking under middle formation. Section 6.8 summarizes the theoretical implications of the analyses presented and concludes the chapter.

6.2 Linking system

The linking theory developed within the scope of this thesis accounts for the complexity of the system linking case and grammatical relations to syntactic arguments in Icelandic. In general, linking refers to the generalizations which can be found in the mapping of predicate-argument structures to a syntactic representation, explaining the relationship that holds between lexical semantics, argument structure, case marking, grammatical relations and syntactic structure in a language (see Butt 2006). Various linking theories exist and the theories differ with respect to the particularities employed for linking (see Chapter 3 and Butt 2006 for an overview). However, none of the existing theories includes all the features relevant for linking arguments to grammatical relations in Icelandic, i.e., case, word order, thematic roles, and event structure. Therefore, I propose a novel linking theory for Icelandic in this section, which combines the crucial components.

Building on the enhancements of the Lexical Mapping Theory, i.e., LFG’s standard linking theory, from Kibort (2014) and Zaenen (1993) with respect to argument positions and lexical semantics, my theory employs a reference frame which incorporates information structure in the form of FIGURE and GROUND (see Talmy 1978) as mediator between position, lexical semantics and event structure as sketched in the scheme in (1).
Event structure is incorporated into the present theory in the form of event participants borrowed from the event decompositional framework proposed by Ramchand (2008). In line with Svenonius (2002), I assume that case marking interacts with the structure of events in Icelandic. Grammatical relations (GR) are mapped to arguments via particular lexical semantics as per Zaenen’s (1993) feature classification based on Dowty’s (1991) Proto-Role entailments, which follow from the lexical semantic properties associated with the event participants and the reference frame. Moreover, grammatical relations are linked to structural positions via the reference frame. In the present proposal, event structure is taken to moreover interact with the reference frame, associating specific case markers with particular grammatical relations, structural positions and lexical semantics.

### 6.2.1 Event participants and argument positions

In accordance with Kibort (2014), I employ argument positions as an independent tier of representation, allowing for non-default mappings between semantic participants and grammatical relations. In Kibort’s system, argument positions are fixed at a-structure and represent the relative syntactic prominence of a predicate’s arguments as specified by its subcategorization frame (see Section 3.2.8 in Chapter 3). Following Kibort, I employ argument positions as argument slots at a-structure in order to separate out structural position and lexical semantics from event participants.

\[
\begin{array}{c|c|c|c|c}
\text{INIT} & \text{UND} & \text{RES} & \text{RH} \\
\hline
\end{array}
\]

(2) \quad \text{PRED} < \text{arg}_1 \quad \text{arg}_2 \quad \text{arg}_3 \quad \text{arg}_4 >

A predicate’s event structure is incorporated into linking via the event participants established by Ramchand (2008). The event participants selected by a predicate are mapped onto the argument positions from left to right as shown in (2), according to the hierarchy given in (3). In the present approach, four argument positions are
6.2. Linking system

assumed at a-structure to account for the mapping of each event participant to an individual argument.

(3) \text{INIT(ITOR)} \succ \text{UND(ERGOER)} \succ \text{RES(ULTEE)} \succ \text{RH(EME)}

The hierarchy of event participants directly follows from the hierarchical order of event projections assumed in the first-phase syntax (see Ramchand 2008 and Section 3.4 in Chapter 3). Ramchand (2008) takes the process subevent, in which the undergoer is licensed, to be the hallmark of dynamic predicates. The \text{INITIATOR} is part of the init projection in Ramchand’s decompositional framework which denotes the initiating or causing event. Furthermore, the \text{INITIATOR} is highest in the hierarchy given in (3) because it is the external argument, generally realized as a subject (cf. Ramchand 2008). Moreover, \text{RHEME} arguments can occur in complement position of either the proc head, which is part of the process projection, or the res head, which is part of the result projection, always denoting the argument lowest in the structure.

Stative predicates consist of an initiation event only and have an \text{INITIATOR} as state holder that internally causes the stative eventuality on the basis of its own properties (see Ramchand 2008). To demarcate stative from dynamic predicates, and in analogy to the annotation scheme employed in the corpus study in Chapter 4, I refer to the \text{INITIATOR} of stative predicates as (state) \text{HOLDER} in this thesis.\footnote{In a more recent paper on adjective alternations, Ramchand (2018) distinguishes between two different types of state \text{HOLDERS}. This is mostly in line with the approach taken in this thesis.} In addition to a state \text{HOLDER}, stative predicates can license a \text{RHEME} which further describes the stative eventuality. An example for the linking of event participants to argument positions with stative predicates is given in (4), showing the available pattern for the predicate \text{líka} ‘like’ as used in example (18) from Chapter 4, which is repeated in (5).

\begin{align*}
\text{HOLDER} & \quad \text{RH} \\
\text{líka} \ ‘\text{like}’ & < \quad \text{arg}_1 & \text{arg}_2 >
\end{align*}

(5) En Þorgrími líkaði það illa.
\begin{center}
but Þorgrímur.DAT like.PST.3SG this.NOM badly
\end{center}
\begin{center}‘But Þorgrímur disapproved of this.’\end{center}

(IcePaHC, 1400.GUNNAR.NAR-SAG,.241)
In general, more than one event participant, apart from the rheme, can be linked onto the same argument position, resulting in a composite role. For instance, the first argument of the predicate sjá `see' as exemplified in (7) is a volitional agent, but also continuously involved in the seeing process. The corresponding argument is thus mapped to two participants, the INITIATOR and the UNDERGOER (see Ramchand 2008, 54f. on UNDERGOER-INITIATORS of dynamic psych verbs), in the corresponding linking scheme in (6).\footnote{I refer to UNDERGOER-INITIATORS as INITIATOR-UNDERGOERS in the following.} The entity seen, i.e., ‘Illugi’s cabin’ in (7), which corresponds to the second argument position, is a rheme as it further describes and specifies the seeing event.

\begin{center}
\begin{tabular}{cccc}
INIT & UND & RES & RH \\
\hline
sjá `see' & $\arg_1$ & $\arg_2$ \\
\end{tabular}
\end{center}

(7) Hann sér skála Illuga undir fellinu.
he.NOM see.PRS.3SG cabin.ACC Illugi.GEN under hill.the.DAT
‘He sees Illugi’s cabin at the bottom of the hill.’

(IcePaHC, 1650.ILLUGI.NAR-SAG,.968)

The empirical findings obtained in Chapter 5 indicated that, in accordance with Kiparsky (1997), Icelandic developed positional licensing constraints for grammatical relations over the centuries. Moreover, information structure, i.e., the development of a designated topic position, has been identified as a determining factor behind the rise of positional licensing. The interaction between information structure and position in Icelandic is implemented in the form of a reference frame, which furthermore mediates the linking of event participants to grammatical relations. The reference frame is based on the classical division of objects into FIGURE and GROUND which was borrowed from Gestalt psychology and established in linguistics by Talmy (1978).

\subsection*{6.2.2 Reference frame}

Talmy’s (1978) distinction between FIGURE and GROUND sets a clause’s reference frame by specifying the reference relation between two distinct entities and their relation to the whole event. The relation which holds between the FIGURE and GROUND of a clause is best exemplified on the basis of events which refer to the
6.2. Linking system

movement or location of a physical object with respect to another object, as shown in the following examples (taken from Talmy 1978, 419):

(8) a. the pen lay on the table
    b. the pen fell off the table

In both clauses from (8), ‘the pen’ is the FIGURE argument, i.e., the argument which refers to a (conceptually) moving point. Moreover, the FIGURE argument represents the path or site which is conceived as the salient issue. The GROUND argument of both clauses is ‘the table’ which serves as stationary reference-point with respect to which the FIGURE is characterized (Talmy 1978).

The division into FIGURE and GROUND moreover correlates with the information structural distinction between TOPIC and COMMENT. A topic refers to the entity in a clause whose referent the sentence is about and the comment provides information about what is being said about the topic (see, e.g., Reinhart 1981, Krifka 2008). Moreover, a topic usually contributes given information to the discourse. Thus, like the FIGURE argument, the topic represents the more salient issue in comparison to a reference-point in the discourse. As the study from Chapter 5 has shown, the clause-initial position is the preferred topic position in Icelandic. Thus, I assume that FIGURE-GROUND does interact with word order in that the FIGURE, which is a prototypical topic, usually precedes the GROUND argument.3

For example, in sentence (7), the INITIATOR-UNDERGOER argument hann ‘he’ is the topic of the clause and designates the FIGURE argument, i.e., the salient issue. The RHÈME argument skála Illuga ‘Illugi’s cabin’ in (7) serves as stationary reference-point with respect to the FIGURE argument and refers to the GROUND. The reference frame is incorporated into the linking scheme for the predicate sjá as shown in (9).

---

3I generally assume that focused constituents, i.e., constituents that contribute new or contrastive information to the discourse, can also be realized as FIGURE arguments. However, the distinction between discourse topics and focused elements is complex and theories differ with respect to the notions employed (see, e.g., Krifka 2008 for an overview). An interesting approach is taken by Dalrymple and Nikolaeva (2011), who argue for the possibility of having more than one sentence topic. Subjects are usually primary topics and refer to the pragmatically more prominent element. Objects tend to be secondary topics, which gives rise to Differential Object Marking in some languages, differentiating between subject and non-subject topics. Overall, more research on information structure in Icelandic is necessary to be able to fully account for the interrelation between FIGURE-GROUND and information structure.
The correlation between position and reference frame is moreover exemplified by English, where the reference frame switches depending on the structural positions of the arguments. Consider, for example, the clauses in (10) (see Talmy 1978, 421 for a similar example). Whereas in (10-a), ‘John’ is used as the reference-point (GROUND) which characterizes the FIGURE’s (‘Anna’) appearance, ‘Anna’ is the reference-point and ‘John’ is the FIGURE argument in (10-b), where the arguments occur in the reversed order. As English uses word order to signal grammatical relations, FIGURES are usually subjects and GROUNDS denote objects. In general, FIGURES, i.e., topics, are most often subjects (see, e.g., Givón 1990, Dowty 1991, Dalrymple and Nikolaeva 2011).

(10) a. Anna resembles John.
    b. John resembles Anna.

Furthermore, locative inversion in Chichewa is correlated with a particular discourse function which implies a change of the FIGURE-GROUND relation. Locative inversion refers to a structure in which a location is fronted to signal presentational focus (see, e.g., Bresnan and Kanerva 1989 and Section 3.2.7 of Chapter 3). Under locative inversion, the referent of the former subject is (re)introduced to the scene referred to by the preposed, and focused, locative, which is no longer the reference-point and GROUND object, but has become the FIGURE and subject instead.

6.2.3 Sentience and animacy

Although I have argued in Chapter 5 that the rise of positional licensing in Icelandic correlates with the development of a designated topic position, FIGURE and GROUND are not directly mapped to grammatical relations in my new system. Instead, I assume that the semantic interpretation of the event participants and the reference frame in combination with information about the animacy/sentience of the arguments lead to a set of lexical semantic entailments, which regulate the mapping to grammatical relations via Zaenen’s (1993) binary feature classification (see Sec-
Cross-linguistically, the more sentient/animate argument of a transitive clause is generally realized as the subject, whereas the argument lower on the animacy hierarchy given in (11) is usually the object (see, e.g., Silverstein 1976, Hopper and Thompson 1980, Comrie 1989, Dowty 1991, Aissen 2003).

(11) 1st person > 2nd person > 3rd person > proper names > humans > animates (non-humans) > inanimates

Sentience is implemented into my linking system via the [+sentient] feature, which marks a sentient (and animate) event participant. Applied to the linking of sjá ‘see’ with respect to the sentence given in (7), this results in the figure argument hann ‘he’, which is a third person pronoun and high on the hierarchy in (11), receiving the [+sentient] feature, see (12), where hann ‘he’ corresponds to arg₁.

(12) sjá ‘see’ < arg₁ arg₂ >

The interrelation between the event participants, the reference frame, sentience and the lexical semantic entailments employed for the mapping of arguments to grammatical relations is described in the following.

6.2.4 Lexical semantic entailments

The notions of figure and ground are also related to Dowty’s (1991) Proto-Role entailments. Agentive and sentient entities, i.e., Proto-Agents, usually make for better topics than inanimate patients (see, e.g., Givón 1990). Moreover, a figure is generally a (conceptually) moving object causing its own movement and is therefore agent-like, see Dowty’s Proto-Agent entailments in (13) (repeated from (31) in Chapter 3). Ground objects have a stationary setting within a reference frame with respect to the figure object which corresponds to Dowty’s Proto-Patient entailments in (14) (repeated from (32) in Chapter 3).

---

\(^4\)Initiator arguments are underspecified with respect to animacy/sentience in Ramchand (2008).
Proto-Agent entailments

a. Volitional involvement in the event or state
   Example: John in John is ignoring Mary.

b. Sentence/perception
   Example: John in John sees/fears Mary.

c. Causing an event or change of state in another participant
   Example: unemployment in Unemployment causes delinquency.

d. Movement (relative to the position of another participant)
   Example: tumbleweed in The rolling tumbleweed passed the rock.

e. Independent existence
   Example: John in John needs a new car.

Proto-Patient entailments

a. Change of state
   Examples: mistake in John made a mistake., error in John erased the error.

b. Incremental theme
   Example: apple in John ate the apple.

c. Causally affected by another participant
   Example: Mary in John kicked Mary.

d. Stationary relative to another participant
   Example: rock in The rolling tumbleweed passed the rock.

e. Existence not independent of event
   Example: house in John built a house.

Dowty’s Proto-Role entailments are incorporated into the standard Lexical Mapping Theory, i.e., LFG’s linking theory, by Zaenen (1993) in order to be able to account for lexical semantic generalizations without having to assume a thematic role hierarchy (see Section 3.2.9 in Chapter 3). Zaenen (1993) implements Dowty’s Proto-Role entailments by using the binary feature classification of the Lexical Mapping Theory which classifies arguments as [±r(estricted)] and [±o(bjective)] (see Section 3.2.5 in Chapter 3). Zaenen’s feature classification is given in (15) (repeated from (34) in Chapter 3).
6.2. Linking system

(15) **Feature Classification**
1. If a participant has more patient properties than agent properties, it is marked $-r$.
2. If a participant has more agent properties than patient properties, it is marked $-o$.
3. If a participant has an equal number of properties, it is marked $-r$.
4. If a participant has neither agent nor patient properties, it is marked $-o$.

For the mapping between grammatical relations and the feature specifications, Zaenen furthermore assumes a set of association principles, see (16) (repeated from (35) in Chapter 3):

(16) **Association Principles**
order the participants as follows according to their intrinsic markings:
$-o < -r < +o < +r$
order of the GR [grammatical relations] as follows:
$\text{subj} < \text{obj} < \text{obl}_\theta < \text{(obl)}$
Starting from the left, associate the leftmost participant with the leftmost GR it is compatible with.

In my linking system, the lexical semantic entailments in the form of Proto-Agent and Proto-Patient properties follow from the lexical semantics associated with the event participants, the FIGURE-GROUND assignment, and the [+sentient] feature. Zaenen’s feature classification and association principles are then used to map grammatical relations to the corresponding arguments.

To illustrate this, the linking of arguments to grammatical relations for sjá ‘see’ with reference to sentence (7) is detailed in the following. In (7), sjá has two arguments, i.e., hann ‘he’ which corresponds to arg$_1$ and skála Illuga ‘Illugi’s cabin’ which is arg$_2$, see (17). arg$_1$ is a sentient INITIATOR-UNDERGOER denoting the FIGURE of the clause. INITIATOR-UNDERGOERS refer to volitional event causers and are therefore prototypical agents according to Dowty’s Proto-Role entailments, see (13). FIGURES are Proto-Agents as they represent the salient issue and the conceptually moving entity. Moreover, sentient entities are Proto-Agents. Thus, arg$_1$ overall has three Proto-Agent (P-A) properties.
The rheme argument, which is mapped to arg\textsubscript{2} in (17), is neither a Proto-Agent nor a Proto-Patient, as the purpose of a rheme is to further describe the event. Yet, arg\textsubscript{2} has one Proto-Patient (P-P) property as it is the ground argument. Grounds are generally Proto-Patients as they denote the, often stationary, reference-point in relation to another participant. In (17), the number of stars (*) indicates the respective number of Proto-Role properties.

\[
\begin{aligned}
\text{(17)} & \quad \text{INIT} \quad \text{UND} \quad \text{RES} \quad \text{RH} \\
\text{síá } \text{‘see’} & < \quad \text{arg}_1 \quad \text{arg}_2 > \\
\text{FIGURE} & \quad \text{GROUND} \\
\text{[+sentient]} & \\
\text{P-A:***} & \quad \text{P-P:}* \\
\text{[−o]} & \quad \text{[−r]} \\
\text{SUBJ} & \quad \text{OBJ}
\end{aligned}
\]

As per Zaenen’s feature classification in (15), arg\textsubscript{1} receives the [−o] feature specification because it has more agent than patient properties. arg\textsubscript{2} has more patient than agent properties and is therefore classified as [−r], see (17). Following Zaenen’s association principles as given in (16), the argument with the [−o] specification, i.e., arg\textsubscript{1}, is linked to subj and arg\textsubscript{2} is realized as an obj given its [−r] feature specification.

### 6.2.4.1 The Asymmetrical Object Parameter

Although Zaenen’s linking principles fit in nicely with my linking system, the feature classification system has to be modified in order to be able to account for the linking of ditransitive predicates. As will be shown in Section 6.6, applying Zaenen’s feature classification to the analysis of ditransitive transfer predicates results in two [−r] marked arguments in addition to the subj argument, which is marked [−o]. In keeping with the wellformedness conditions of the standard Lexical Mapping Theory (see Section 3.2.6 in Chapter 3), each argument has to be linked to a unique function. Moreover, by the Asymmetrical Object Parameter which is postulated by Bresnan and Moshi (1990), see (18), only one argument can be classified as [−r] and be linked to the primary object function.
6.2. Linking system

(18) **The Asymmetrical Object Parameter (AOP)**

* \(\theta \ldots \theta\)  
| \[−r\] | \[−r\]*

(Bresnan and Moshi 1990, 172)

In Bresnan and Moshi (1990), linking with two \([−r]\) objects is rescued by adding the \([+o]\) feature specification to one of the arguments, which therefore is linked to OBJ\(\theta\). I implement the Asymmetrical Object Parameter into Zaenen’s feature classification by extending the first principle:

(19) **Feature Classification (extended)**

1. If a participant has more patient properties than agent properties, it is marked \(−r\). If there are two \(−r\) marked participant, mark one of the participants as \(+o\).
2. If a participant has more agent properties than patient properties, it is marked \(−o\).
3. If a participant has an equal number of properties, it is marked \(−r\).
4. If a participant has neither agent nor patient properties, it is marked \(−o\).

The linking of ditransitive predicates and the application of the extended first principle is detailed in Section 6.6.

6.2.4.2 Oblique arguments

Zaenen’s linking system does also not properly account for the linking of oblique arguments, which are neither Proto-Agents nor Proto-Patients in the sense of Dowty. In Section 6.4, I argue for the historical development of dative experiencer subjects from locative predications. In locative predications, see, e.g., (20), a state **HOLDER**, which can be sentient, is the **FIGURE** argument and the location is a **RHEME** and **GROUND** argument. The **FIGURE** argument in (20) is a Proto-Agent and linked to **SUBJ** via the \([−o]\) feature, see (21). As **GROUND** argument, the **RHEME** is a Proto-Patient. Proto-Patients are marked with the \([−r]\) feature in Zaenen (1993), which precludes linking to an **OBL**. Instead, as was shown for the dynamic predicate *sjá ‘see’* in (17), RHEME arguments which refer to the **GROUND** are linked to the **OBJ** function.
via Zaenen’s feature classification. However, when the ground refers to a location in a stative setting, it has neither Proto-Agent, nor Proto-Patient properties, and should be $[-o]$. Moreover, locations are semantically restricted, i.e., $[+r]$, as they are generally realized as obliques (cf. Bresnan and Kanerva 1989). As $[-o,+r]$ arguments, the ground location can be linked to OBL. Therefore, I postulate a $[+\text{location}]$ feature for ground locations, to demarcate them from rhemes and grounds which are objects. The linking with locative predications is discussed in more detail in Section 6.4.

(20) Hann var í stofunni.
    he.NOM be.PST.3SG in living.room.the.DAT
   ‘He was in the living room.’

(21) \[
\begin{array}{c|c|c}
\text{HOLDER} & \text{RH} \\
\hline
\text{arg}_1 & \text{arg}_2 & > \\
\text{FIGURE} & \text{GROUND} & \\
[+\text{sentient}] & [+\text{location}] & \\
\text{P-A:}** & \text{P-A:}, \ 	ext{P-P:} & \\
[-o] & [-o,+r] & \\
\text{SUBJ} & \text{OBL} & \\
\end{array}
\]

This section showed that the lexical semantics associated with the event participants and the reference frame are crucial for the linking of arguments to grammatical relations in Icelandic. The interrelation between the reference frame and lexical semantics with respect to licensing grammatical relations furthermore plays a crucial role for passivization and middle formation. In my analysis, passivization and middle formation are taken to be operations which change a clause’s reference frame. This in turn changes the lexical semantic entailments so that the linking possibilities of a predicate change. This is in line with the general spirit of passivization in Lexical Mapping Theory (see, e.g., Bresnan and Kanerva 1989, Bresnan and Zaenen 1990, Kibort 2014), but provides a new, more semantically realistic basis for linking.
6.2.5 Passivization

Passivization changes the reference frame of a clause in that the former object, which usually denotes the GROUND argument with transitive predicates, is promoted to subject and becomes the FIGURE. Consider, for example, the active clause containing the transitive predicate *opna* ‘open’ in (22).

(22) Hann opnaði kirkjudyranna.
    he,NOM open,PST.3SG church.doors.the,ACC
    ‘He opened the church doors.’

In (22), the FIGURE argument is hann ‘he’ and the reference-point and GROUND argument is ‘the church doors’, being stationary in relation to the FIGURE. Moreover, the FIGURE argument causes the process described by the predicate and is an INITIATOR, see the linking scheme for *opna* in (23). The GROUND argument undergoes the ‘opening’ process and holds a result state. Therefore, the GROUND argument is linked to the UNDERGOER as well as the RESULTEE in (23). As a volitional causer (INITIATOR) and sentient FIGURE argument, arg₁ in (23) is a Proto-Agent, receiving the [−o] feature, and is linked to subj. The GROUND argument, i.e. arg₂ in (23), has two patient properties as it undergoes a change of state as an UNDERGOER-RESULTEE in addition to referring to the GROUND. It therefore receives the [−r] feature and is linked to obj.

(23) \[ opna \text{ ‘open’} < \quad \text{INIT} \quad \underbrace{\text{UND}} \quad \text{RES} \]

\[
\begin{array}{l}
\text{FIGURE} \\
[+\text{sentient}] \\
\text{P-A:**} \\
[−o] \\
\text{SUBJ}
\end{array}
\quad \begin{array}{l}
\text{GROUND} \\
\text{P-P:**} \\
[−r] \\
\text{OBJ}
\end{array}
\]

Example (24) shows the passive version of the active sentence given in (22). In the passive, the former GROUND argument is the salient argument and becomes the FIGURE.
Chapter 6. Linking events, case and grammatical relations in Icelandic

(24) Kirkjudyrnar voru opnaðar (af honum).
    church.doors.the.NOM be.PST.3PL open.PTCP.PASS.NOM (by he.DAT)
    ‘The church doors were opened (by him).’

As figure, arg₂ no longer has two patient properties, but an equal number of Proto-Agent and Proto-Patient properties. Arguments with an equal number of properties are marked [−r] in Zaenen (1993). The [−r] feature specification generally allows for the mapping of the corresponding argument to the subject function as, for example, unaccusative predicates have a single Proto-Patient argument linked to subj (see, e.g., Zaenen 1993). Therefore, as figure, the UNDERGOER-RESULTEE of opna in (24) is promoted to the subject function, see the linking scheme in (25). With passives, the INITIATOR argument is moreover demoted and suppressed for linking. Therefore, it can only be realized as an adjunct, see (24). Yet, as both the INITIATOR and the UNDERGOER-RESULTEE are still available for linking, passivization is based on the same event structure and argument structure as the active version.

(25) \[
\text{opna passive} \quad < \quad \text{arg}_1 \quad \text{arg}_2 \quad > \\
\text{FIGURE}
\]

\[
\begin{array}{c}
\text{P-A:*, P-P:*} \\
\text{[−r]} \\
\text{SUBJ}
\end{array}
\]

6.2.6 Middle formation

Similar to passivization, middle formation changes a clause’s reference frame in that a former GROUND object is promoted to subject and becomes the FIGURE argument. The crucial difference between passivization and middle formation is that whereas an INITIATOR argument is in principle still available for linking in the passive, INITIATORS cannot be linked with middles. As per Kaufmann’s (2007) analysis of the middle, agent arguments, i.e., INITIATORS, are structurally absent in middle constructions, although they are still available on a pragmatic or inferential level. This is in line with Wood (2015), who showed that agentive modifiers, purpose clauses and by-phrases are incompatible with Icelandic middles (see Section 2.6.2 of Chapter 2).
6.2. Linking system

For instance, *opna* ‘open’ has the anticausative middle form *opnast* as shown in (26) with the linking pattern given in (27). In (27), the INITIATOR argument of *opna* is not linked. Instead, the UNDERGOER-RESULTEE is linked to the subject. As FIGURE argument, the UNDERGOER-RESULTEE is not only a Proto-Patient, but also a Proto-Agent. Given that the UNDERGOER-RESULTEE has an equal number of Proto-Agent and Proto-Patient properties, the corresponding argument receives the \([-r]\) feature, which allows for linking to SUBJ.

(26) Kirkjudyrnar opnast
church.doors.the.NOM open.PST.MID.3PL
‘The church doors opened.’ (IcePaHC, 1985.MARGSAGA.NAR-FIC,.418)

(27)

\[
\begin{array}{c}
\text{INIT} \\
\text{UND} \\
\text{RES}
\end{array}
\]

\[\text{opna middle} < \arg_2 > \]

\text{FIGURE}

\text{P-A:*, P-P: *}

\[\text{[−r]} \]

\[\text{SUBJ}\]

Cross-linguistically, middles are also known to have a reflexive or reciprocal reading (see, e.g., Kemmer 1993), which implies that FIGURE and GROUND are mapped onto the same entity. In Icelandic, for example, the predicate *klæða* ‘dress’ has the reflexive middle form *klæðast* as shown in (28) (see also Sigurðsson 1989).

(28) a. Napóleon klæddi sig.
Napoleon.NOM dress.PST.3SG REFL.ACC
‘Napoleon dressed himself.’

b. en Napóleon klæddist
but Napoleon.NOM dress.PST.MID.3SG
‘but Napoleon got dressed.’ (IcePaHC, 1861.ORRUSTA.NAR-FIC,.213)

The linking pattern for *klæða* ‘dress’ as used in example (28-a) is given in (29). With *klæða*, \(\arg_1\), i.e., *Napóleon* ‘Napoleon’ in (28-a), is an INITIATOR argument and \(\arg_2\) which corresponds to the reflexive pronoun *sig* ‘himself’ is an UNDERGOER-
RESULTEE. As sentient FIGURE, the INITIATOR is a Proto-Agent linked to SUBJ. The UNDERGOER-RESULTEE is the GROUND argument and a Proto-Patient which is linked to OBJ via the \([-r]\) feature specification.

(29) \[ \text{klaða ‘dress’} < \begin{array}{ccc} \text{INIT} & \text{UND} & \text{RES} \\ \text{arg}_1 & \text{arg}_2 & > \\ \text{FIGURE} & \text{GROUND} & \end{array} \] \[ [+\text{sentient}] \] 
\[ \begin{array}{cc} \text{P-A:***} & \text{P-P:**} \\ \text{[−o]} & \text{[−r]} \end{array} \] 
\[ \begin{array}{cc} \text{SUBJ} & \text{OBJ} \end{array} \]

The scheme given in (30) shows linking for the corresponding middle klaðast as exemplified in (28-b). With the reflexive middle klaðast, FIGURE and GROUND collapse onto the same argument, i.e., the UNDERGOER-RESULTEE, because the INITIATOR is no longer available for linking in the middle. In (30), the UNDERGOER-RESULTEE has two Proto-Agent properties given that the participant is a sentient FIGURE argument. As UNDERGOER-RESULTEE and GROUND, the corresponding argument additionally has two Proto-Patient properties. With an equal number of properties, the argument is specified as \([-r]\) which allows for linking to SUBJ.

(30) \[ \text{klaða middle} < \begin{array}{ccc} \text{INIT} & \text{UND} & \text{RES} \\ \text{arg}_2 & > \\ \text{FIGURE} & \text{GROUND} & \end{array} \] \[ [+\text{sentient}] \] 
\[ \begin{array}{cc} \text{P-A:**} & \text{P-P:**} \\ \text{[−r]} \end{array} \] 
\[ \text{SUBJ} \]

In sum, voice operations change a clause’s reference frame in that a former GROUND becomes the FIGURE argument. Middle formation has the effect that the INITIATOR is no longer available as an argument for linking to grammatical rela-
tions although the event structure of the predicate remains intact overall. With
passives on the other hand, the INITIATOR argument is suppressed for linking, but
the argument is still present at the level of a-structure.

6.2.7 Summary

This section concludes the presentation and discussion of the basic linking appa-
ratus developed in this thesis. My linking theory builds on Kibort’s (2014) and
Zaenen’s (1993) enhancements of the Lexical Mapping Theory with respect to ar-
gument positions and lexical semantics, but combines them in a novel way via a
reference frame which functions as interface between position, event structure and
lexical semantics. Event structure is incorporated via the event participants postu-
lated in Ramchand’s (2008) first-phase syntax. The event participants are mapped
to argument positions which are associated with discourse functional properties in
the form of the reference frame. The reference frame distinguishes between the FIG-
URE and the GROUND argument of a clause (see Talmy 1978). Based on the event
participants, the reference frame, and information about the sentience/animacy of
the arguments, a set of lexical semantic entailments can be deduced, which are in
turn used to link grammatical relations to arguments via Zaenen’s (1993) incorpora-
tion of Proto-Role entailments into LFG’s linking theory. Moreover, in my system,
voice operations are taken to change a clause’s reference frame which interacts with
the linking possibilities of arguments to grammatical functions.

In the following sections, I provide an analysis of the data presented in Chapters 4
and 5, accounting for the diachronic interaction between case, word order, lexical
semantics and grammatical relations in Icelandic. Following Svenonius (2002), the
structure of events is taken to interact with case marking in this analysis. The
interrelation between event structure and case marking which I posit for Icelandic is
detailed next.
6.3 Event structure and case

The study presented in Chapter 4 showed that in line with Svenonius’ (2002) theory of case, dative case marking correlates with particular event structural configurations in Icelandic. In the following sections, I will analyze these configurations further, accounting for the linking of case and grammatical relations in Icelandic and the diachronic changes observed in the data. As Svenonius (2002) primarily focuses on the licensing of dative case on objects, I present my own ideas regarding the interrelation between dative case marking on subjects and event structure which are based on the IcePaHC data. Moreover, my analysis of dative goal objects differs from Svenonius’ approach. In order to provide an overview of the interrelation between case marking and event structure that I postulate throughout my analyses, I briefly summarize the mapping between case and event participants, which in turn affects the reference frame.

In Icelandic, the default case marking for subjects is nominative (see, e.g., Zaenen et al. 1985, Thráinsson 1994). This is confirmed by IcePaHC, where 95% of all subjects are marked nominative, see Table 4.1 in Chapter 4. Thus, nominative is the default case for figure arguments, which generally refer to the initiator with dynamic and the holder with stative predicates, designating the highest semantic participant available for the respective types of predicates (i.e., [+HR] in Kiparsky 1997). Direct objects are most often marked accusative, see Table 4.2 in Chapter 4, and therefore, accusative is the default case for ground arguments. This entails that the undergoer and resultee arguments of transitive predicates, and moreover rheme arguments of ditransitives and stative predicates are generally marked via accusative case, representing the lowest semantic participant available (i.e., Kiparsky’s [+LR] feature specification).

In accordance with Svenonius (2002), I assume that with dynamic predicates, dative case is licensed when two subevents of a single, indivisible event do not overlap completely with respect to their temporal runtime. I will argue in the following that a temporal mismatch between the initiating event and the process event results in the dative case marking of an undergoer argument. Moreover, the resultee argument of a ditransitive construction, i.e., the goal/recipient, is marked with dative case in Icelandic because it refers to a sentient state holder. In IcePaHC, dative subjects were most often found together with stative experiencer predicates, designating
6.3. Event structure and case

a state holder. With non-nominative subjects, the object, i.e., the ground argument, is usually marked nominative and refers to a rheme. This applies to stative predicates with dative subjects (state holder) as well as passives and middles of ditransitive constructions, where a state holder, i.e., the resultee, which is marked with dative case, becomes the subject and a rheme argument receives nominative case marking. This is moreover in line with Kiparsky’s empty feature specification of the nominative as it is compatible with both the highest and the lowest argument.

Moreover, my theory assumes that with composite roles, the higher event participant from the hierarchy given in (3) determines the possibilities for case marking of the corresponding argument (in line with evidence from case marking with complex predications in Urdu, see, e.g., Butt 1995, 2013). The event structural configurations which correlate with case marking in Icelandic are further summarized in (31), where $\neg$ $(\text{init} \circ \text{proc})$ indicates a temporal mismatch between the initiating and the process event.

(31) Principles for Case Assignment

1. Nominative
   a. Default figure (initiator or holder)
   b. If a non-nominative resultee/holder is the figure, then the ground (rheme) is marked nominative.

2. Accusative
   a. Default ground (undergoer, resultee or rheme)
   b. holder

3. Dative
   a. If $\neg$ $(\text{init} \circ \text{proc})$, then the undergoer is marked dative.
   b. resultee (not ground)
   c. holder

5 There are moreover stative experiencer predicates which take an accusative subject in Icelandic, e.g., langa ‘want, long for’, and thus also have a state holder. I will however not discuss these predicates in any detail in the present thesis, focusing on the diachronic developments of dative subjects as evidenced by my corpus studies. A possible analysis of those predicates would be that the accusative case is the result of default object case marking with the accusative object being reanalyzed as subject over time. The accusative case marking of the subject then has to be analyzed as an instance of quirky, i.e., idiosyncratic, case marking. This is quite plausible, given that the accusative subjects of stative experiencer verbs are systematically replaced by datives in the context of Dative Substitution.

6 Genitive case marking is not analyzed in this thesis, given that genitive subjects and objects occurred only rarely in the corpus (see Tables 4.1-4.3).
6.4 Dative subjects and the rise of positional licensing

Most dative subjects occurred together with a stative experiencer predicate in the corpus. Stative experiencer predicates which take a dative subject, e.g., líka 'like' in (5), have a dative marked state HOLDER as subject and can take a nominative RHEME as object as shown in the linking scheme given in (32). The interpretation of arg₁ as an experiencer is the result of the interplay between the different levels of linguistic information involved in linking: event structure, information structure (i.e., the reference frame), and lexical semantics. With the stative predicates in question, arg₁ is the state HOLDER which is furthermore a sentient and animate FIGURE. Thus, arg₁ is a Proto-Agent in the sense of Dowty (1991), rather than being a Proto-Patient. As a state HOLDER and Proto-Agent, arg₁ has the semantic interpretation of an experiencer. Moreover, stative experiencer predicates usually involve a theme/stimulus argument corresponding to the RHEME argument which is neither a Proto-Agent nor a Proto-Patient, but refers to the GROUND. As GROUND, it has one Proto-Patient property. As Proto-Agent, the state HOLDER and FIGURE, i.e., the experiencer, is linked to SUBJ, while the RHEME argument is linked to OBJ.

(32)

\[
\text{stative experiencer predicate} < \begin{array}{c|c|c}
\text{HOLDER} & \text{RH} \\
\hline
\text{arg₁} & \text{arg₂} \\
\end{array} \begin{array}{c|c|c}
\text{FIGURE} & \text{GROUND} \\
\hline
[+sentient] & \\
\end{array}
\]

\[
\text{P-A.**} & \text{P-P.}^* \\
\text{[-o]} & \text{[-r]} \\
\text{SUBJ} & \text{OBJ} \\
\text{DAT} & \text{NOM}
\]

6.4.1 Experiencer subjects and locative inversion

Historically, stative experiencer predicates are often derived from locative predications. For example, in Sanskrit (Old Indo-Aryan), the experiencer is encoded as a location at which a theme or stimulus is placed, see example (33) which has a genitive experiencer, taken from Butt and Deo (2013).
6.4. Dative subjects and the rise of positional licensing

(33) ma-yi ced asti te pr̃īt-i
I-LOC.SG be-PRES.3.SG you.REFL.SG affection-NOM.SG
‘If you have love for me...’ (Lit. ‘If your affection is at me...’)
(Mahābhārata 1.161.14c)

Similar constructions can also be found for Icelandic, whereby the location/experiencer is marked dative, see (34) and (35).

(34) ...og þó er mér grunur
and still be.PRES.3.SG I.DAT suspicion.NOM
‘...and I am still suspicious’ (Lit. ‘suspicion is to me’)
(IcePaHC, 1275.MORKIN.NAR-HIS,.828)

(35) Mikil ógn stóð mér af þessum mann.
much.NOM fear.NOM be.situated.PST.3.SG I.DAT of this.DAT man.DAT
‘I was very afraid of this man.’ (Lit. ‘a lot of fear was situated at me’)
(IcePaHC 1300.ALEXANDER.NAR-SAG,.272)

Bresnan and Kanerva (1989) show that variable linking possibilities of thematic roles to arguments exist with locative predication, providing the classic LFG analysis for locative inversion in Chichewa (see Section 3.2.7 in Chapter 3). Under locative inversion in Chichewa, a locative phrase is preposed and becomes the subject, compare the inverted structure in (36-b) with the uninverted clause in (36-a) (repeated from example (18) in Chapter 3).

7-well 7SB-be 17-3-village
‘The well is in the village.’

17-3-village 17SB-be 7-well
‘In the village is a well.’

(Bresnan and Kanerva 1989, 2)

For the uninverted structure, Bresnan and Kanerva (1989) propose the linking given in (37) (repeated from example (19) in Chapter 3). The predicate khāla ‘remain’ takes two arguments, i.e., a theme and a location. As per the intrinsic feature classification of the Lexical Mapping Theory (see Section 3.2.5 in Chapter 3), the theme argument is classified as $[-r]$ and the location is $[-o]$. Moreover, Bresnan and Kanerva (1989) take the location to be $[+r]$ by default, based on the observation
that locatives are usually realized as oblique arguments. Thus, the theme argument is linked to SUBJ, while the location is realized as an OBL in the uninverted form, see the linking in (37).

(37) \[ \text{khála} < \text{theme} \quad \text{location} > \]

\[
\begin{array}{c|c|c}
\text{intrinsic:} & [-r] & [-o] \\
\text{defaults:} & [-r] & [+] \\
\hline
\text{SUBJ} & \text{OBL}_{doc} \\
\end{array}
\]

In the inverted structure, the location is linked to SUBJ instead and the theme argument becomes the OBJ, see (38) (repeated from (20) in Chapter 3). Bresnan and Kanerva (1989) argue that the subject linking of the location is possible because of the discourse functional properties of the preposed location, signaling presentational focus. The special focus context assigns the $[-r]$ feature to the location argument, which allows for the linking to SUBJ, while the theme is linked to OBJ.

(38) \[ \text{khála} < \text{theme} \quad \text{location} > \]

\[
\begin{array}{c|c|c}
\text{intrinsic:} & [-r] & [-o] \\
\text{defaults:} & [-r] & [+] \\
\hline
\text{OBJ} & \text{SUBJ} \\
\end{array}
\]

The basic insights from Bresnan and Kanerva (1989) on locative inversion are incorporated into the analysis which I posit for the historical development of stative experiencer predicates with a dative subject in Icelandic. I assume that, in line with Butt and Deo (2013), the stative experiencer predicates are historically derived from locative predications. With respect to datives in Icelandic, Maling (2002, 36) reports on the observation made by Nygaard (1906/1966) that dative case in Old Icelandic is a melting pot of the Indo-European instrumental, ablative and locative cases together with the dative marked recipients and experiencers. It is thus very plausible that the Icelandic dative experiencers originate from former locative predications and that dative experiencers are generally inherited from a language stage older than the attested record of Icelandic.

In the historically older locative predications, the dative case marked location is the GROUND argument, whereas the stimulus or theme argument is the FIGURE being located with respect to the GROUND, see the linking scheme in (39). Thus,
the stimulus/theme is the state **HOLDER** and the location is the **RHEME** argument. As **FIGURE** argument, the state **HOLDER** is a Proto-Agent, and thus \([-o]\), while the **GROUND** argument, which is a location as is indicated via the \([+\text{location}]\) feature, has neither Proto-Agent nor Proto-Patient properties. Therefore, the **GROUND** argument is also \([-o]\). As per the classic LFG analysis of locative predications (Bresnan and Kanerva 1989), I assume that locations are generally realized as oblique arguments in locative predications and are \([-r]\) by default in such configurations. Based on the feature specifications, the stimulus argument is linked to **SUBJ**, whereas the location is realized as an **OBL**.

\[
\begin{array}{c|c}
\text{HOLDER} & \text{RH} \\
\hline
\end{array}
\]

\[
\text{locative predication} < \begin{array}{c}
\text{arg}_1 \\
\text{arg}_2 \\
\end{array} > \\
\begin{array}{c|c}
\text{FIGURE} & \text{GROUND} \\
\hline
\end{array}
\]

\[
\begin{array}{c}
\text{P-A:*} \\
\text{P-A:} \\
\text{P-P:*} \\
\text{P-P:} \\
\hline
\text{[−o]} \\
\text{[−r]} \\
\hline
\text{SUBJ} \\
\text{OBL} \\
\text{OBJ} \\
\text{NOM} \\
\text{DAT} \\
\hline
\end{array}
\]

(39) (theme) (location)

\[
\begin{array}{c|c}
\text{HOLDER} & \text{RH} \\
\hline
\end{array}
\]

\[
\text{locative inversion} < \begin{array}{c}
\text{arg}_1 \\
\text{arg}_2 \\
\end{array} > \\
\begin{array}{c|c}
\text{FIGURE} & \text{GROUND} \\
\hline
\end{array}
\]

\[
\begin{array}{c}
\text{P-A:*} \\
\text{P-P:*} \\
\hline
\text{[−o]} \\
\text{[−r]} \\
\hline
\text{SUBJ} \\
\text{OBJ} \\
\text{NOM} \\
\text{DAT} \\
\hline
\end{array}
\]

(40) (location) (theme)

In the inverted structure, see (40), the location has a specific discourse function and is thus the more salient issue, i.e., the **FIGURE** argument, which is the state
HOLDER. Moreover, the theme argument denotes the GROUND argument with respect to which the \textit{figure} is located. As \textit{figure} and state \textit{holder}, the location is a Proto-Agent with the $[-o]$ feature and linked to \textit{subj}. When GROUND arguments are not marked with the $[+\text{location}]$ feature, they are Proto-Patients in the sense of Dowty. As Proto-Patient, the theme argument is linked to \textit{obj} via the $[-r]$ feature.

Experiencers tend to be sentient and animate arguments, see, e.g., the dative experiencer in (34). Sentient entities usually make for better topics than inanimate stimuli (cf. Givón 1990). As topics usually denote the \textit{figure} argument, sentient experiencers tend to occur as \textit{figures} and state \textit{holders}, while the theme/stimulus argument is the GROUND with respect to which the \textit{figure} is characterized. This corresponds to the linking given for locative inversions in (40), but the \textit{figure} additionally carries the $[+\text{sentient}]$ feature with experiencers, see the linking for stative experiencer predicates given in (32), which is repeated below in (41).

(41)

\begin{align*}
\text{stative experiencer predicate} & < \quad \text{HOLDER} \quad \text{RH} \\
& \quad \text{FIGURE} \quad \text{GROUND} \\
& \quad [+\text{sentient}] \\
& \quad \text{P-A}^{**} \quad \text{P-P}^{*} \\
& \quad [-o] \quad [-r] \\
& \quad \text{SUBJ} \quad \text{OBJ} \\
& \quad \text{DAT} \quad \text{NOM}
\end{align*}

As sentient \textit{figure}, the state \textit{HOLDER}, i.e., the dative experiencer, is a Proto-Agent which is linked to the subject per the $[-o]$ feature. The theme/stimulus is, as \textit{rhememe} and \textit{ground}, a Proto-Patient receiving the $[-r]$ feature and linked to \textit{obj}. I assume that over time, the linking in (32) becomes the preferred option for the linking of dative experiencers with stative predicates over time, with experiencers becoming more firmly linked to subjects than to obliques, leading to the increasing use of dative subjects together with stative experiencer predicates. This fits in nicely with the analysis posited for the rise of positional licensing in Chapter 5 as detailed in the following.
6.4. Dative subjects and the rise of positional licensing

6.4.2 Dative experiencers and positional licensing

Although dative experiencers can be linked to SUBJ with stative predicates in Modern Icelandic, it is rather unlikely that they were straightaway licensed as subjects in the history of the language. In Chapter 5, I have argued for the gradual development of structure and positional licensing in Icelandic, which I have taken to be motivated by the increasing association of topics with the clause-initial SpecIP position. The corpus study in Chapter 5 showed that subjects, which are prototypical topics (cf. Givón 1990, Dalrymple and Nikolaeva 2011), are increasingly realized in the prefinite position in IcePaHC, with 1900 as major key turning point, see Table 6.1, which is repeated from Table 5.1.

<table>
<thead>
<tr>
<th>Period</th>
<th>prefinite</th>
<th>postfinite</th>
<th>Total</th>
<th>% prefinite</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>7045</td>
<td>6672</td>
<td>13717</td>
<td>51.4%</td>
<td>***</td>
</tr>
<tr>
<td>1350-1549</td>
<td>10091</td>
<td>8258</td>
<td>18349</td>
<td>55.0%</td>
<td>***</td>
</tr>
<tr>
<td>1550-1749</td>
<td>6076</td>
<td>5134</td>
<td>11210</td>
<td>54.2%</td>
<td>***</td>
</tr>
<tr>
<td>1750-1899</td>
<td>6490</td>
<td>4767</td>
<td>11257</td>
<td>57.6%</td>
<td></td>
</tr>
<tr>
<td>1900-2008</td>
<td>7924</td>
<td>2937</td>
<td>10861</td>
<td>73.0%</td>
<td>***</td>
</tr>
<tr>
<td>All</td>
<td>37626</td>
<td>27768</td>
<td>65394</td>
<td>57.5%</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.1: Diachronic distribution of subject positions in IcePaHC.

Once the prefinite position has been firmly established as hosting subjects in Icelandic, with over 70% of subjects appearing prefinitely in the data as of 1900, see Table 6.1, dative experiencers also increasingly go there, conforming to the overall structural change of the language. However, the dative experiencers do not lead the change, but rather follow suit. That is, whereas nominative subjects preferably occur in the prefinite position from the earliest attested stages of Icelandic onwards, dative experiencer subjects only began to show a preference for the prefinite position in the period post-1900 in IcePaHC, see Table 6.2, repeated from Table 5.2.

I provide the following explanation for the comparably weaker tendency of dative subjects to occur prefinitely in IcePaHC: As dative experiencers are usually sentient FIGURE arguments, i.e., topics, they tend to occur in the clause-initial, prefinite SpecIP position. However, as they can also be linked to OBL, as was show in (39), their overall tendency to occur in SpecIP is weaker than for canonical nominative subjects. Over time, SpecIP increasingly hosts subjects. Once subjects have become more firmly associated with SpecIP, the dative experiencers become more firmly associated with the subject function, with the linking in (41) becoming dominant.
Table 6.2: Diachronic distribution of subject position for dative subjects in IcePaHC.

<table>
<thead>
<tr>
<th>Period</th>
<th>prefinite (DAT)</th>
<th>postfinite (DAT)</th>
<th>Total</th>
<th>% prefinite (DAT)</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150-1349</td>
<td>131</td>
<td>404</td>
<td>535</td>
<td>24.5%</td>
<td>***</td>
</tr>
<tr>
<td>1350-1549</td>
<td>126</td>
<td>465</td>
<td>591</td>
<td>21.3%</td>
<td>***</td>
</tr>
<tr>
<td>1550-1749</td>
<td>119</td>
<td>298</td>
<td>417</td>
<td>28.5%</td>
<td>*</td>
</tr>
<tr>
<td>1750-1899</td>
<td>151</td>
<td>277</td>
<td>428</td>
<td>35.3%</td>
<td></td>
</tr>
<tr>
<td>1900-2008</td>
<td>353</td>
<td>273</td>
<td>626</td>
<td>56.4%</td>
<td>***</td>
</tr>
<tr>
<td>All</td>
<td>880</td>
<td>1717</td>
<td>2597</td>
<td>33.9%</td>
<td></td>
</tr>
</tbody>
</table>

6.4.3 Case versus position

This is moreover in line with Allen’s (1995) findings for oblique experiencers in the history of English. Allen provides data showing that oblique experiencers already share some properties with canonical subjects in Old English. When nominative on subjects becomes obligatory in the 14th century, experiencer subjects follow suit and are eventually realized as nominatives. The analysis of dative subjects and positional licensing given here is furthermore mostly in line with Kiparsky (1997), who thoroughly discusses experiencer subjects. In Kiparsky’s linking system, features for position, thematic role, case and agreement morphology are combined for linking (see Section 3.3 in Chapter 3). As regards Icelandic experiencer arguments, the features assigned by the dative case marking generally preclude a linking to subject in his analysis. However, the features associated with the position of dative experiencers, i.e., SpecIP, are able to override the inherent features of the dative argument and regulate the subject linking. That is, because positional nominative case can be assigned in SpecIP, the features associated with positional nominative case are valid in this position. For Icelandic, Kiparsky takes morphological case to be recessive as it does not determine the grammatical relation referred to by an argument. Instead, position is dominant in Icelandic due to the rise of positional licensing in Germanic.

This is implemented into my linking analysis by means of the reference frame. Dative experiencers can only be linked to SUBJ when they correspond to the FIGURE argument, which correlates with the occurrence of the dative experiencer in the clause-initial SpecIP position. Being the FIGURE of a clause adds to the number of Proto-Agent properties, which regulates the linking to SUBJ in my system, see (41). In the alternative linking in (39), which I postulate to be historically older, the dative marks a location which is the GROUND argument. The lexical semantics associated
with dative locations preclude linking to \textit{subj} (and \textit{obj}), which cannot be ‘rescued’ by the structural properties associated with the \textit{ground}.

The next section relates the present theory for the diachronic development of dative subjects with stative experiencer predicates to the lexicalized middles, which were found to mainly drive the increasing use of dative subjects with stative experiencer predicates in the period post-1900.

### 6.5 Lexicalized experiencer and raising predicates

The corpus study presented in Chapter 4 showed that the increase of dative subjects correlates with an increasing use of middle forms which have been lexicalized as stative experiencer predicates over time. The increase is primarily driven by the frequent use of the experiencer and raising predicate \textit{finnast} ‘find, feel, think, seem’ with a dative subject in the period post-1900, see example (42) (repeated from (13) in Chapter 4).

\begin{equation}
\text{Mörgum þeirra fanst þeir vera útlagar úr many.DAT they.GEN seem.PST.MID.3SG they.NOM be.INF exiles.NOM out.of landi sínu meðan þeir voru hér. land.DAT theirOWN.DAT while they.NOM be.PST.3PL here ‘It seemed to many of them that they are exiles from their own land while they were here.’}
\end{equation}

\text{(IcePaHC, 1907.LEYSING.NAR-FIC,.763)}

The non-middle counterpart of \textit{finnast} is the predicate \textit{finna} ‘find’, which describes a dynamic event of finding, see, e.g., (43). As such, \textit{finna} ‘find’ has a nominative subject which is an \textit{INITIATOR-UNDERGOER}, see the linking scheme given in (44). The predicate moreover takes an accusative object which is the \textit{RESULTEE} of the event.

\begin{equation}
\text{Gunnar fann seint hrossin um daginn Gunnar.NOM find.PST.3SG late horse.the.ACC during day.the.ACC ‘Gunnar found the horse late during the day.’}
\end{equation}

\text{(IcePaHC, 1400.GUNNAR.NAR-SAG,.281)}

\footnote{Kiparsky’s linking theory for Icelandic implicates that dative experiencers can only be analyzed as subjects in the history of the language once V2 has been established. However, whether this is true is difficult to determine on the basis of the corpus data as it is not easy to assess the subject status of the dative experiencers in the older texts from IcePaHC.}
Linking in (44) proceeds as follows: The INITIATOR-UNDERGOER, who is causing the event, is sentient and the FIGURE of the clause. Thus, the corresponding argument (arg1) has more Proto-Agent than Proto-Patient properties and is classified as [−o]. The RESULTEE, i.e., arg2, undergoes a change of state and is the GROUND. From this follows that arg2 is a Proto-Patient receiving the [−r] feature. Via the feature specification, the INITIATOR-UNDERGOER is linked to SUBJ, while the RESULTEE is realized as an OBJ. By default, the FIGURE argument is marked by nominative, whereas the GROUND has accusative case marking.

### 6.5.1 Reanalysis via middle formation

With reciprocal middles, FIGURE and GROUND are mapped onto the same argument and the argument referring to the INITIATOR is no longer available for linking. The original middle version of finna can have the meaning ‘meet’ and exemplifies these features. For example, in sentence (45), the event described by the middle predicate finnast ‘meet’ has no INITIATOR-UNDERGOER, i.e., no volitional causer, and the subject þeir feðgar ‘they, father and son’ is both the stationary reference-point and the salient argument. Thus, the RESULTEE denotes FIGURE and GROUND in the middle. As such, the corresponding argument has one agent property and two patient properties, still receiving the [−r] feature, see (46). The [−r] feature generally allows for linking to SUBJ (see Zaenen 1993). Therefore, with finnast ‘be found, meet’, the RESULTEE argument is linked to SUBJ. As FIGURE argument, the RHEME is assigned nominative case by default.
6.5. Lexicalized experiencer and raising predicates

(45) Nú finnast þeir feðgar.
      now meet.PRS.MID.3PL they.NOM father.and.son.NOM

      ‘Now they, father and son, meet.’

(IcePaHC, 1350.BANDAMENNM.NAR-SAG,.1100)

(46)

\[
\begin{array}{ccc}
\text{INIT} & \text{UND} & \text{RES} \\
\text{finnast ‘meet’} & < & \text{arg}_2 \\
\text{FIGURE} & > & \text{GROUND} \\
P-A:*, P-P:** & | & [-r] \\
\text{SUBJ} & | & \text{NOM}
\end{array}
\]

The middle form finnast also has a non-reciprocal meaning in the sense of ‘be found’. Without an INITIATOR or UNDERGOER, events described by finnast ‘be found’ have a stative interpretation in the sense of ‘be situated/located’. This is exemplified by (47), where finnast ‘be found’ occurs together with a dative case marked locative phrase, which reinforces the stative interpretation.

(47) En annars dags eftir fannst skipið
      and one.GEN day.GEN after be.found.PST.MID.3SG ship.the.NOM
      óspillt öðrum megin árinn
      undamaged.NOM one.DAT side.DAT lake.the.GEN

      ‘and one day later, the ship was found undamaged on one side of the lake’

(IcePaHC, 1210.JARTEIN.REL-SAG,.187)

Overall, finnast ‘be found’ with a nominative subject was often found together with a locative phrase in the corpus. Linking for finnast ‘be found’ as exemplified by (47) is given in (48). In addition to the RESULTEE which is linked to SUBJ, finnast ‘be found’ takes a dative case marked RHEME location in (47). With the locative, the RHEME is the stationary reference-point, i.e., the GROUND, characterizing the RESULTEE argument, which is the FIGURE. As was shown in Section 6.4, RHEME locations are linked to OBL via the [+location] feature. The FIGURE argument has equally many P-A and P-P properties and is linked to SUBJ.
Chapter 6. Linking events, case and grammatical relations in Icelandic

Like the structure postulated for locative predications in (39), the linking scheme in (48) for finnast has a state holder, i.e., the RESULTEE, as highest argument and a RHHEME is moreover the GROUND. Given these structural similarities, I postulate that finnast ‘be found’ was reanalyzed as a stative predicate over time, whereby the RESULTEE was reinterpreted as a HOLDER, resulting in the linking given in (49).

(49) $\text{FINNAST} {\text{‘be found, meet’}} < \begin{array}{c|c}
\text{ARG}_2 & \text{ARG}_3 \\
\text{FIGURE} & \text{GROUND} \\
\{+\text{location}\} & \\
\text{P-A:}* & \text{P-P:}* \\
\{[-r] & \{[-o, +r] \\
\text{SUBJ} & \text{OBL} \\
\text{NOM} & \text{DAT} \\
\end{array}$

In analogy to the trajectory exemplified in Section 6.4, I assume that the stative experiencer predicate finnast ‘find, feel, think, seem’ is historically derived from the stative version of finnast ‘be found’ with the structure shown in (49), whereby a sentient experiencer/location, which is marked dative, becomes reanalyzed as state HOLDER and FIGURE. The former nominative subject becomes the GROUND and RHHEME and is linked to OBJ instead. The linking for finnast as stative experiencer predicate with a dative subject meaning ‘feel, find’ is shown in (51) and an example is provided in (50).
6.5. Lexicalized experiencer and raising predicates

(50) og fannst honum nótt.
and feel.PST.MID.3SG he.DAT night.NOM
‘and he felt the night.’ (IcePaHC, 1861.ORRUSTA.NAR-FIC,.1670)

(51) finnast ‘find, feel’<
       HOLDER       RH
       |       |
       arg1       arg2
       FIGURE     GROUND

 [+sentient]

 P-A:**  P-P:*  
 [−o]      [−r]
       SUBJ       OBJ
       DAT        NOM

6.5.2 Secondary predication and raising

The stative experiencer predicate finnast ‘find, feel, think, seem’ can furthermore be used as a raising predicate denoting epistemic judgements. A diachronic relationship between epistemic raising predicates and verbs of perception or feeling can be observed cross-linguistically (Barron 2001). In order to become an epistemic raising predicate, further semantic bleaching is necessary which can be achieved through secondary predication (cf. Barron 2001 for the development of Old French sembler ‘look, have an appearance’ to Modern French sembler ‘seem’). Secondary predication is also given for finnast ‘find, feel’ in (52) via the adjectival complements.

(52) Finnst hönum þetta gott og dýrmætt öl.
find.PRS.MID.3SG he.DAT this.NOM good.NOM and valuable.NOM beer.NOM
‘He finds this to be good and valuable beer.’

(IcePaHC, 1675.ARMANN.NAR-FIC,98.215)

Within LFG, secondary predication is generally treated at the level of f-structure, projecting information stored in the lexical entry (see Section 3.2 in Chapter 3 for an introduction to f-structures). For example, Simpson (1983) provides an LFG account for secondary predication, showing that an XCOMP, i.e., an open complement function, in the form of an adjective, a nominal, a transitive prepositional phrase,
or an intransitive preposition can be added to a predicate’s a-structure via a lexical rule, with further information that the XCOMP is predicated of the object which is provided by the lexical entry. The f-structure in (53) corresponds to the sentence from example (52). The adjectival complements are represented as an XCOMP in (53). The control link between the subject of the XCOMP and the f-structure of the OBJ indicates that the syntactic OBJ of the main clause acts semantically as the subject of the XCOMP.

(53)  

\[
\begin{align*}
\text{PRED} & : \text{‘find (SUBJ, OBJ, XCOMP)’} \\
\text{TENSE} & : \text{past} \\
\text{OBJ} & : \begin{cases} \text{PRED} & : \text{‘beer’} \\
\text{CASE} & : \text{NOM} \\
\text{SPEC} & : \begin{cases} \text{PRED} & : \text{‘this’} \\
\text{CASE} & : \text{NOM} \\
\text{SUBJ} & : \end{cases} \end{cases}
\end{align*}
\]

In the context of secondary predication, the requirement to have a concrete and physically perceivable stimulus gradually disappears as the basis for the perceiver’s judgement can be indirect or abstract (Barron 2001), relating to the knowledge or cognition of the perceiver. Therefore, I assume that once secondary predication is possible, finnast also increasingly denotes epistemic judgments. Moreover, as a predicate denoting epistemic judgements, finnast begins to be increasingly used as a raising predicate, which is the most common usage of the stative experiencer predicate finnast ‘find, feel, think, seem’ in the period post-1900, as shown in the examples in (54).

\[\text{8}A\text{ similar LFG analysis for secondary predication at f-structure is presented in Ahmed et al. (2012), who account for resultatives via a PREDLINK which relates the predicator, i.e., the object, with the predication, e.g., an adjective. See also Müller (2002) for an LFG approach to secondary predication and resultatives.} \]
6.6 Dative goals and middle formation

(54) a. Nú fanst henni hann horfa á sig.
now seem.PST.MID.3SG she.DAT he.NOM look.INF at himself.ACC
‘Now he seemed to her to look at himself.’
(IcePaHC, 1908.OFUREFLI.NAR-FIC,.468)

b. Mér fanst það vera friður náttúrunar.
I.DAT seem.PST.MID.3SG that.NOM be.INF peace.NOM nature.the.Gen
‘That seemed to me to be the peace of nature.’
(IcePaHC, 1920.AGIN.REL-SER,.639)

Another stative experiencer and raising predicate which occurred frequently with a dative subject in the corpus, and in particular in the periods pre-1900, is sýnast ‘seem, appear’. As sýnast is the morphological middle form of the ditransitive transfer predicate sýna ‘show’ which takes a dative goal object, a different trajectory has to be postulated. An analysis of middle formation with transfer predicates is provided in the next section, and the lexicalization of sýnast as an experiencer and raising predicate is discussed there.

6.6 Dative goals and middle formation

Dative subjects occurred by far most often together with stative experiencer predicates in the corpus. Yet, dative subjects were also frequently found together with transfer verbs in IcePaHC, see Table 4.12 in Chapter 4. Transfer verbs usually are ditransitive predicates which describe a transfer of an entity to a goal object, whereby the goal is marked dative. With transfer verbs, the dative goal surfaces as a subject in the passive and under middle formation (see Table 4.14). A prototypical example for a ditransitive transfer predicate is gefa ‘give’ as shown in (55).

(55) og konungur gefur honum vopn í hönd sér.
and king,NOM give.PRS.3SG he.DAT weapon.ACC in hand.DAT oneself
‘and the king gives him the weapon with his own hands.’
(IcePaHC, 1275.MORKIN.NAR-HIS,.121)

Transfer verbs generally have an INITIATOR which is the FIGURE argument that causes the event. As sentient and volitional causer, the INITIATOR is a Proto-Agent linked to the SUBJ, see the linking scheme for gefa in (56). Various different analyses exist for ditransitive predicates (see, e.g., Butt 2006 for an overview). The analysis of ditransitive constructions employed here follows Ramchand’s (2008) analysis of
double object constructions which licenses the recipient/goal argument as the resultee and the entity given as the rheme of the event (see Section 3.4 in Chapter 3). Moreover, as per Ramchand’s analysis, the event does not license an undergoer, although the event entails a process.

In (55), the goal argument is a sentient resultee with equally many Proto-Agent as Proto-Patient properties, see arg\textsubscript{2} in (56). The thing given, i.e., \textit{vopn} ‘weapon’ in (55) designates the ground argument and is a Proto-Patient. Thus, both the resultee and the rheme are associated with the \([-r]\) feature. This exemplifies the dual linking possibilities of objects in ditransitive constructions in Icelandic. Zaenen et al. (1985) show that with \textit{gefa} ‘give’, both arguments can be linked to either \textit{obj} or \textit{obj}\(_{\theta}\). Yet, as Zaenen et al. (1985) moreover show, only NPs associated with the grammatical function \textit{obj} may passivize. Thus, in order for the dative goal argument to be realized as a passive dative subject, it has to be linked to \textit{obj}, whereas the accusative object has to be realized as \textit{obj}\(_{\theta}\). This is regulated via the AOP, which I implemented into linking as an extension to Zaenen’s feature classification, see the first principle given in (19). As per this principle, one of the \([-r]\) arguments is in addition \[+o\] to allow for linking of one of the objects to \textit{obj}\(_{\theta}\). Applying this to the linking for \textit{gefa} ‘give’ in (56) results in the rheme argument being additionally classified as \[+o\]. In this way, the resultee can be linked to \textit{obj}, while the rheme is \textit{obj}\(_{\theta}\).

\[\begin{array}{ccccc}
\text{transfer verb} & \text{INIT} & \text{UND} & \text{RES} & \text{RH} \\
< & \text{arg}_1 & \text{arg}_2 & \text{arg}_3 & > \\
\text{FIGURE} & +\text{sentient} & +\text{sentient} & & \\
\text{GROUND} & \text{P-A:***} & \text{P-A:*, P-P:*} & \text{P-P:*} & \\
& \text{[−o]} & \text{[−r]} & \text{[−r, +o]} & \\
& \text{SUBJ} & \text{OBJ} & \text{OBJ}\(_{\theta}\) & \\
& \text{NOM} & \text{DAT} & \text{ACC} & \\
\end{array}\]

\footnote{Note that this contrasts with the IcePaHC annotation which generally annotates the dative goal argument of ditransitives as an indirect object. This results in the high frequencies of dative case on indirect objects in the corpus as shown in Table 4.3 of Chapter 4.}
6.6. Dative goals and middle formation

The INITIATOR and FIGURE argument is marked nominative by default. As lowest available semantic participant, the RHEME is marked accusative with ditransitives. I assume that the RESULTEE receives dative case marking because it is a sentient state holder which has both Proto-Agent and Proto-Patient properties. Svenonius (2002) assumes that the dative case marking on goal objects of ditransitive predicates arises because of a mismatch between the temporal runtimes of two of the three subevents instantiated by the predicate, see Section 2.7 in Chapter 2. However, he remains rather vague about the nature of this mismatch. Assuming that RESULTEES are prototypically dative case marked because they are sentient state holders allows for a more concise account for case marking in Icelandic. This will become more evident throughout the remainder of this chapter.

6.6.1 Passivization with transfer verbs

In IcePaHC, dative subjects were frequently found together with transfer verbs in the passive across the whole corpus. Under passivization, the dative goal becomes the FIGURE and is linked to SUBJ as shown in (57) and exemplified by (58).

(57) transfer verb<sub>passive</sub> < \text{arg}_2 \quad \text{arg}_3 > \\
\text{INIT} \quad \text{UND} \quad \text{RES} \quad \text{RH} \\
\text{FIGURE} \quad \text{GROUND} \\
[+sentient] \\
\text{P-A:*}, \text{P-P:*}, \text{P-P:*} \\\n\text{P-P:*} \\
L \quad L \\
\text{SUBJ} \quad \text{OBJ} \\
\text{DAT} \quad \text{NOM} \\

(58) og sìðan var okkur matur gefinn. \\
and then be.pst.3sg we.dat food.nom give.ptcp.pass \\
‘and then food was given to us.’

(IcePaHC, 1628.OLAFUREGILS.BIO-TRA.,241)

---

10This is in line with Primus (1999, 2002), who argues that dative case marked recipient/goal objects are neither prototypical agents nor patients in the sense of Dowty. Therefore, she reformulates Dowty’s Proto-Role entailments and introduces a third, ‘Proto-Recipient’, role.
In the passive, the resultee receives the \([−o]\) specification given that as sentient figure, it has more agent than patient properties. The rheme is a Proto-Patient and linked to obj. The initiator is demoted, but can in principle still be linked and can be realized as an adjunct. When the initiator is demoted, the resultee is, as figure, the highest available semantic argument. The figure argument is usually marked nominative. However, I assume that the resultee preserves its dative case marking because it still refers to a sentient state holder with both Proto-Agent and Proto-Patient properties. With a sentient state holder as figure and an inanimate rheme as ground, the linking pattern of gefa ‘give’ in the passive is similar to the linking of stative experiencer predicates as shown in (41). In analogy to this, the former accusative marked obj\(\theta\), which is the ground argument, receives nominative case marking in the passive. This moreover fits in with the Icelandic Association Principles postulated by Zaenen et al. (1985) in that the highest available grammatical function is generally assigned nominative, given that the passive subject is already marked dative and not available.

### 6.6.2 Middle formation with transfer verbs

Overall, transfer verbs were found most frequently with a dative subject in passive constructions. Yet, middle forms of transfer verbs show a striking increase with dative subject in the period post-1900, making up over 60% of the transfer verbs with a dative subject in the last time period (see Table 4.14). Middle formation with transfer verbs is exemplified by means of the transfer predicate gefa ‘give’. With the middle form of gefa, i.e., gefast ‘get, receive’, the initiator argument is no longer available for linking, see example (22) from Chapter 4, repeated in (59), and the linking pattern given in (60). As two separate arguments remain after middle formation with ditransitives, figure and ground are mapped onto two entities.

\[
\text{(59) Nú gafst Geirmundi færi að tala það, now get.PST.MID.3SG Geirmundur.DAT opportunity.NOM to tell.INF that sem í brjósti bjó. what in brest.DAT reside.PST.3SG ‘Now Geirmundur got an opportunity to tell what he had on his mind.’} (\text{IcePaHC, 1902.FOSSAR.NAR-FIC,.1490})
\]
I assume that, in analogy to the passive, the dative case marking of the RESULTEE is retained as it remains a sentient state holder with both Proto-Agent and Proto-Patient properties. As highest available semantic participant for linking and FIGURE argument, it is linked to SUBJ. As with passives, the RHEME and GROUND argument is linked to OBJ and receives nominative case marking.

In this respect, middles of transfer verbs are very similar to stative experiencer predicates as the arguments available for linking are a sentient state holder which is the FIGURE and a RHEME which is the GROUND. In Chapter 4, I showed by means of the glyph visualization that the diachronic increase of dative subjects together with middles of transfer verbs correlates with the frequent use of the middle takast ‘manage, succeed’ which is derived from the transfer verb taka ‘get, obtain, take’, see example (61). With takast ‘manage, succeed’, the goal argument has been reanalyzed as an experiencer, see (62).

(61) En því næst var þeim tekið öl og matur.
    but next be.PST.3SG they.DAT take.PST.PTCP beer.NOM and food.NOM
    ‘But next beer and food was taken to them.’
    (IcePaHC, 1480.JARLMANN.NAR-SAG,.211)

(62) mér tókst það loksins.
    I.DAT manage.PST.MID.3SG it.NOM finally
    ‘I finally managed it.’
    (IcePaHC, 2008.OFSI.NAR-SAG,.284)

Despite that takast still denotes a dynamic event, the reading exemplified in (62) is very stative-like, given that the RESULTEE as a state holder is the highest semantic participant. It is therefore quite plausible that middles of transfer verbs can in general be reanalyzed as stative experiencer predicates over time. I suppose that
this has happened with *sýnast* ‘seem, appear’, the middle form of the transfer predicate *sýna* ‘show’ which can be used as a stative experiencer and raising predicate, according to the trajectory described in the following section.

### 6.6.3 Transfer verbs and stative experiencer predicates

As a ditransitive verb describing a transfer, *sýna* ‘show’ has a dative goal and an accusative theme object as shown in (63), corresponding to the linking scheme postulated in (56). The predicate *sýna* ‘show’ entails a perception event caused by the initiator, which makes the rHEME object visible to the resultee. In the middle, the former initiator is not available for linking, resulting in the reading of *sýnast* as physical appearance of the rheme to the resultee, see (64).

(63) Þeir sýndu Snorra bréfin
they.NOM show.PST.3PL Snorri.DAT letter.the.ACC
‘They showed the letter to Snorri.’
(IcePaHC, 1250.STURLUNGA.NAR-SAG,430.1399)

(64) En á hinni sömu nótt eftir sýndist henni
but in that.DAT same.DAT night.DAT later appear.PST.MID.3SG she.DAT
maður í svörtum klæðum
man.NOM in black.DAT clothes.DAT
‘But later during the same night, a man in black clothes appeared to her.’
(IcePaHC, 1210.JARTEIN.REL-SAG,.359)

Without an initiator, the resultee no longer refers to the goal argument of the whole event with *sýnast* ‘appear’ as used in (64), but denotes a perceiver instead. Therefore, the resultee can be reanalyzed as experiencer. This is in line with Ramchand (2008), who notes that psych resultees are in general experientially affected. Without neither initiator nor undergoer argument, the event is moreover likely to be reanalyzed as stative, given that a state holder and figure is the subject, while a rHEME, which denotes the ground, is the object. This is exemplified by the usage of *sýnast* ‘appear’ as stative experiencer predicate in (65).

As per Barron (2001), secondary predication is a precondition for stative experiencer predicates to become raising predicates denoting epistemic judgements (see Section 6.5). Secondary predication is given for *sýnast* in (65), and the usage of *sýnast* ‘seem, appear’ as a raising predicate denoting epistemic judgements is shown in (66) (repeated from example (20) in Chapter 4). The corresponding linking is
given in (67) and assumed to be analog to the linking of the stative experiencer and raising predicate *finnast* ‘find, think, feel, seem’, which was shown in Section 6.5.

(65) Geirmundi sýndust þau djúp og hrein eins og himinbláminn á heðu vetrarkvöldi ... ský.blu.e.the.NOM on clouless.DAT winter.evening.DAT 'To Geirmundur appeared they deep and pure like the blue sky on a cloudless winter evening ...'

(66) Presti sýndist þjáningarsvipur fara um andlitið á Ragnhildi. físta.the.ACC of Ragnhildur.ACC 'To the priest seemed a suffering look to move over Ragnhildur’s face.'

(67) *

\[
\begin{array}{l}
\text{HOLDER} & \text{RH} \\
\mid & \mid \\
\text{arg}_1 & \text{arg}_2 \\
\text{FIGURE} & \text{GROUND} \\
\text{P-A:} & \text{P-P:} \\
\text{[+sentient]} & \\
\text{[−o]} & \text{[−r]} \\
\mid & \\
\text{SUBJ} & \text{OBJ} \\
\text{DAT} & \text{NOM}
\end{array}
\]

6.6.4 **Interim summary**

The overall picture which emerges at this point is that dative subjects became productive in the history of Icelandic with constructions which have a state holder, i.e., either a HOLDER or RESULTEE, as subject and a RHEME argument as object. These constructions are in particular lexicalized stative experiencer predicates and middle forms of ditransitive predicates with a dative goal argument.

Apart from the possibility to have dative goal objects in Icelandic, theme objects may also be marked dative and surface as dative subjects in the passive. Dative
themes as passive subjects were found together with two classes of verbs throughout the whole corpus: transition verbs and verbs describing scalar changes. The dative theme object verbs from both classes have similar event structures in that the subject usually is an INITIATOR and the object an UNDERGOER.\footnote{Mental process verbs, which take a nominative experiencer subject that is an INITIATOR-UNDERGOER, occurred only rarely in the corpus. Therefore, linking with these predicates is not discussed in this thesis.} Yet, in contrast to transitions, i.e., simple processes, scalar changes have a concrete endpoint and thus license a RESULTEE argument. With both verb classes, dative subjects also occurred with unaccusative predicates in the corpus. With the unaccusative constructions, no INITIATOR argument is licensed. The next sections detail the linking of the event participants to grammatical relations of transition verbs and verbs describing scalar changes exemplified on the basis of motion verbs which are prototypical representatives of both groups. I provide an analysis of dative themes and middle formation, showing that event structural changes effectuate the loss of the dative case marking with the respective middles.

6.7 Dative themes and event structure

Overall, transition verbs and verbs describing a scalar change were found less often in the corpus than stative experiencer predicates and transfer verbs (see Table 4.12 in Chapter 4). Yet, changes could be observed. With transition verbs and verbs describing a scalar change, dative subjects were increasingly found with a dative subject in the passive (see Tables 4.15 and 4.16), while the usage of unaccusative constructions with a dative subject decreased accordingly.

6.7.1 Caused and accompanied motion

Verbs of caused and accompanied motion occur regularly with transition verbs and verbs describing a scalar change in the corpus. The motion verbs included in the class of verbs describing a transition are manner of motion predicates. Manner of motion predicates occur regularly with a dative subject in the passive in IcePaHC. An example for a manner of motion predicate is \textit{velta} ‘roll’ as in (68), which describes a process caused by an INITIATOR in which the UNDERGOER takes part, see the linking pattern for \textit{velta} in (69). The INITIATOR is a volitional and sentient causer and therefore a Proto-Agent. As highest argument, the INITIATOR is realized as the
nominative subject. The **UNDERGOER** is a Proto-Patient as it designates the object moved which is causally affected and refers to the **GROUND**. As a Proto-Patient with the \([-r]\) specification, the **UNDERGOER** is linked to the object.

\[(68)\] 
\[
\text{Veltir hann honum síðan dauðum út í sjóinn.}
\]
\[
\text{roll.PRS.3SG he.NOM he.DAT then dead.DAT out to sea.the.ACC}
\]
\[
\text{‘He then rolls him, who is dead, out to the sea.’}
\]

(IcePaHC, 1400.GUNNAR2.NAR-SAG., 221)

\[(69)\]
\[
\text{velta ‘roll’}
\]
\[
\begin{array}{c|c}
\text{INIT} & \text{UND} \\
\hline
\text{ARG1} & \text{ARG2} \\
\hline
\text{FIGURE} & \text{GROUND} \\
\hline
\text{+[sentient]} & \\
\hline
\text{P-A:***} & \text{P-P:**} \\
\hline
\text{[-o]} & \text{[-r]} \\
\hline
\text{SUBJ} & \text{OBJ} \\
\hline
\text{NOM} & \text{DAT}
\end{array}
\]

Per Svenonius’ (2002) analysis of case, verbs in which the movement of the object is not accompanied by an agent or causer throughout the whole runtime of the event have a dative object in Icelandic. With **velta** ‘roll’, the **INITIATOR** does not accompany the movement of the **UNDERGOER** throughout the whole ‘rolling’ event. Therefore, in accordance with Svenonius, dative case is licensed on the object. Given that the **UNDERGOER**, but not the **INITIATOR**, is present throughout the whole process, I assume that the temporal mismatch is elicited by the **INITIATOR** which is thus marked via || in (69).

Events described by verbs such as **velta** contrast with caused motion verbs in which the causer accompanies the moving object throughout the whole event. According to Svenonius, these predicates have an accusative object as the two subevents involved are temporally indistinguishable. Consider, for example, the predicate **flytja** ‘move, transport, carry’ as used in (70) which has an **INITIATOR** that accompanies the **UNDERGOER** throughout the whole event, and the object is marked with accusative case by default, see the corresponding linking in (71).
(70) Þorfinnur flutti alla menn heim af skipinu til 
Þorfinnur.NOM transport.PST.3SG all.ACC men.ACC home by ship.DAT to 
sín. 
his.GEN

‘Þorfinnur transported all men home with his ship.

(IcePaHC, 1310.GRETTIR.NAR-SAG,.462)

(71)

<table>
<thead>
<tr>
<th>INIT</th>
<th>UND</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

flytja ‘transport’ < arg₁ arg₂ >

figure ground

[+sentient]

P-A:***  P-P:**

[−o] [−r]

SUBJ OBJ

NOM ACC

The motion verbs in the class of verbs describing a scalar change entail a change of location and have an UNDERGOER-RESULTEE instead of a pure UNDERGOER. Like the manner of motion verbs, these verbs describe a caused motion when taking a dative object and occur regularly in the corpus. An example is kasta ‘throw’ as given in (72) with the linking scheme from (73). With kasta ‘throw’, the motion of the object is not accompanied by the causer throughout the whole event (see also Svenonius 2002). In line with my analysis of manner of motion verbs, I assume that the temporal mismatch which licenses dative case marking with kasta is due to the nature of the INITIATOR, which is causing the movement of the object, but does not accompany the moving object throughout the full extent of the motion event.

(72) Hann kastaði þá öxinni 
he.NOM throw.PST.3SG then axe.the.DAT

‘He then threw the axe.’   (IcePaHC,1400.GUNNAR.NAR-SAG,.533)
6.7.2 Dative themes and passivization

Under passivization, the dative theme surfaces as a dative subject with the motion verbs, see, e.g., (25) in Chapter 4 for an example with velta ‘roll’, which is repeated in (75), and the linking for passive constructions with a motion verb in (74). In the passive, the UNDERGOER(-RESULTEE) is the FIGURE argument. Therefore, the corresponding argument has one agent and one patient property. Thus, the UNDERGOER(-RESULTEE) is [−r] and realized as the subject. Since both the INITIATOR and the UNDERGOER(-RESULTEE) argument are in general available for linking in the passive, retaining the temporal mismatch, dative case marking is still licensed on the UNDERGOER(-RESULTEE).

(74) \[ \text{motion verb}_{passive} < \arg_1 \arg_2 > \]
(75) Með stunnun og óhjólóðum og miklum erfiðismunum with moans.DAT and shoutings.DAT and great.DAT effort.DAT.PL
var þungum kössum velt upp á bryggjuna ... be.PST.3SG heavy.DAT crate.DAT roll.PTCP up onto jetty.the.ACC
‘With moans and shoutings and great effort, the heavy crate was rolled up onto the jetty ... ’ (IcePaHC, 1907.LEYSING.NAR-FIC.,491)

6.7.3 Dative themes and middle formation

While passive constructions with motion verbs and a dative subject were found quite frequently in the corpus, middle forms of the respective verbs did not occur with a dative subject in the corpus. In general, dative theme objects do not retain their dative case marking under middle formation in Icelandic (see, e.g., Wood 2015 and Section 2.6.2 in Chapter 2). I assume that dative case cannot be licensed with the corresponding middles because the INITIATOR role, which was formerly in charge of the temporal mismatch, can no longer be linked. This contrasts with the middles of goal arguments which are licensed as RESULTEES and are marked with dative case because they are sentient state holders.

With the middle forms of the motion verbs in question, the FIGURE argument is an UNDERGOER or an UNDERGOER-RESULTEE which is linked to SUBJ, see the linking given in (76). Without INITIATOR, the temporal mismatch which licensed dative case is no longer retained. Thus, the UNDERGOER(-RESULTEE) receives nominative case marking by default in the middle construction, see, for example, the usage of the middle of velta in (26), repeated in (77). In (26), the UNDERGOER is also the GROUND argument which is exemplified by the linking scheme in (76).

(76) motion verb $middle$ <

| INIT || UND | RES |
|---|---|---|
| $arg_2$ > |
| FIGURE |
| GROUND |
P-A:* P-P:**

<table>
<thead>
<tr>
<th>[-T]</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUBJ</td>
</tr>
<tr>
<td>NOM</td>
</tr>
</tbody>
</table>
6.7. Dative themes and event structure

(77) og pabbi og mamma veltast í undarlegum slagsmálum.
and dad.NOM and mom.NOM revolve.PRS.MID.3SG in strange.DAT fight.DAT
‘and dad and mom revolve around each other in a strange fight.’
(IcePaHC, 1985.SAGAN.NAR-FIC,.993)

6.7.4 Unaccusative predicates

In the corpus, transition verbs and verbs describing a scalar change moreover occurred in the form of unaccusative verbs taking a dative subject in active constructions. With transition verbs, unaccusative verbs have a dative case marked undergoer argument which is the subject, see, e.g., the manner of motion verb *skola* ‘be washed ashore’ in example (24) from Chapter 4, repeated in (78), and the corresponding linking in (79).

(78) Úr brimöldu draumins skolar mér upp að þessari hvítu strönd.
Out.of surge.DAT dream.the.GEN be.washed.ashore.PRS.3SG I.DAT up to this.DAT white.DAT beach.DAT
‘Out of a surge of the dream, I am washed ashore up to this white beach.’
(IcePaHC, 1985.SAGAN.NAR-FIC,.527)

(79) skola ‘be washed ashore’ < arg₁ >

In Ramchand (2008), unaccusative predicates are verbs which lack an INITIATOR argument (see Section 4.3 in Ramchand 2008). This is exemplified by *skola* ‘be washed ashore’ as used in (78), which has a single UNDERGOER argument only. The UNDERGOER designates the object moved in (78) and therefore has a Proto-Patient
property. As figure, the undergoer moreover has a Proto-Agent property. Thus, it has equally many properties and is assigned [−\(r\)], which allows for linking to subj with unaccusatives (cf. Zaenen 1993). Svenonius (2002) postulates that when the moving object has its own independent trajectory, i.e., the movement is independent of the action of an agent or causer, dative case is licensed on the moving object. Following Svenonius, I assume that the subject is marked dative with skola, given that there is no event causer, i.e., no INITIATOR, and the undergoer has its own independent trajectory.

Similar constructions were found with unaccusative predicates in the class of verbs describing a scalar change, consider, e.g., the intransitive use of ljúka ‘end’ in (80) (repeated from (37) in Chapter 4), where the undergoer-resultee is linked to the subj as a Proto-Patient, given that it is undergoing a change of state. The undergoer-resultee furthermore designates the figure argument and has a Proto-Agent property, see the linking in (81). In analogy to what was shown for the undergoer argument with skola, the undergoer-resultee with ljúka has equally many Proto-Agent and Proto-Patient properties, therefore receiving the [−\(r\)] specification. In this way, the undergoer-resultee is linked to subj.

\[(80)\] og lýkur þar þinginu.
and end.prs.3sg there meeting.dat
‘and the meeting ends there.’

(IcePaHC, 1450.BANDAMENN.NAR-SAG,43.936)

\[(81)\]

\[
ljúka \text{ ‘end’ } < \text{ fig } \]
\[
\text{P.-A.}* , \text{ P.-P.}* 
| \quad [−\(r\)] 
| \quad \text{subj} 
| \quad \text{dat}
\]

Although ljúka ‘end’ does not describe a motion event, it describes a change of state which has its own independent trajectory given that there is no INITIATOR. In (37), the ‘meeting’ ends without an explicit causing force being present. Along
the lines of Svenonius’ theory for motion verbs, I take the dative case marking of the subject to be resulting from the independence of the event from an agent or causer.

6.7.5 Diachronic perspective

Unaccusative transition and scalar change predicates with a dative subject can be attested throughout the Icelandic diachrony. However, they are generally rare and their usage decreases over time. This is not surprising, given that they do not correspond to the prototypical event structure configuration identified for dative subjects in this thesis, i.e., event structures which have a state holder as subject and a RHEME as object. The occurrence of transition and scalar changes remains relatively stable over time with a dative subject in the passive. I explain this on the basis of the increasing systematic association between dative subjects and experiencer semantics by which the system becomes regularized to preferably mark state holders in HOLDER/RESULTEE-RHEME configurations with dative case. As a consequence, dative themes, i.e., UNDERGOERS, become increasingly confined to the object function.

Moreover, the unaccusative predicates which take a dative theme subject can in general be causativized, see, e.g., (82) for a causative example with the scalar change predicate ljúka ‘end’. In the causative version, an INITIATOR is licensed in addition to the UNDERGOER-RESULTEE and linked to the subject, which corresponds to the linking pattern of kasta ‘throw’ given in (73). This moreover contrasts the unaccusative predicates with the middle forms, where no INITIATOR argument can be linked. With the transitive version of ljúka ‘end’, dative case is licensed on the object, i.e., the UNDERGOER-RESULTEE, as the INITIATOR, which is causing the ‘ending’ event, does not accompany the UNDERGOER-RESULTEE throughout the whole event, given that, e.g., in (82), the causing event partially precedes the ending process of ‘his speech’.

(82) Lúka þeir nú tali sínu.
end.PRS.3PL they.NOM now speech.DAT his.DAT
‘They now bring his speech to an end.’
(IcePaHC, 1450.BANDAMENN.NAR-SAG,28.151)

I assume that the transitive versions of the dative theme predicates, in which the dative theme argument is linked to an object, are historically older than the corresponding unaccusatives, given that they are more stable throughout the corpus, representing the typical licensing conditions for dative themes. Moreover, with the
Chapter 6. Linking events, case and grammatical relations in Icelandic

unaccusatives, the subjects have prototypical object characteristics because they are UNDERGOER-RESULTEES with Proto-Patient properties. Accordingly, I take dative themes to have been originally licensed as objects. Once FIGURE arguments became more firmly associated with the subject role and dative subjects became possible in the language (see Section 6.4), dative themes could also be linked to subjects with unaccusatives.

6.7.6 Expericner subjects and scalar changes

An exception to the class of verbs describing scalar changes are verbs describing a bodily change which take a dative case marked expericner subject in active constructions. The event structure of those predicates is similar to the prototypical dative subject event structure of stative predicates. For example, the predicate batna ‘recover’ in (83) has an UNDERGOER-RESULTEE as subject and a RHEME as object. The UNDERGOER-RESULTEE is the sentient FIGURE argument and is therefore linked to the subject. As RESULTEE, the UNDERGOER-RESULTEE refers to a sentient state holder. The RHEME refers to the GROUND and is linked to OBJ, see the linking in (84).

(83) en Vilborgu batnaði svefnleysið.
    and Vilborg.DAT recover.PST.3SG insomnia.the.NOM
    ‘and Vilborg recovered from the insomnia.’

(IcePaHC, 1882.TORFHILDUR.NAR-FIC,.1157)

(84)

\[
\begin{array}{cc}
\text{batna ‘recover’} & \text{UND} & \text{RES} & \text{RH} \\
\text{arg}_1 & \text{arg}_2 & > \\
\text{FIGURE} & \text{GROUND} & \\
\text{[+sentient]} & \\
\text{P-A:**, P-P:*} & \text{P-P:*} \\
\text{[-o]} & \text{[-r]} & \\
\text{SUBJ} & \text{OBJ} & \\
\text{DAT} & \text{NOM} & \\
\end{array}
\]
As such, the linking with batna ‘recover’ conforms to the typical dative subject pattern with a sentient state holder as subject and a RHEME as object. Thus, I assume that with verbs such as batna ‘recover’, the subject has acquired an experiencer reading over time, which makes way for the possibility to have dative subjects with verbs denoting a change of bodily state.

6.8 Summary and conclusion

In this chapter, I introduced a novel linking theory as an extension of LFG’s Lexical Mapping Theory, accounting for the complex interrelation between case marking, event structure, lexical semantics, word order and grammatical relations observed in Icelandic. Building on Zaenen’s (1993) and Kibort’s (2014) enhancements of the Lexical Mapping Theory, I propose a linking system which implements a reference frame in the form of Talmy’s (1978) FIGURE-GROUND division, functioning as mediator between lexical semantics, word order, and event structure in the licensing of grammatical relations. The reference frame encodes information structural properties which interact with word order and lexical semantics, linking grammatical relations to particular structural positions. Motivated by Svenonius’ (2002) event structural account for case marking in Icelandic, I moreover incorporated the event participants proposed in the first-phase syntax approach for event decomposition by Ramchand (2008) into my linking system. In the present approach, dative case is taken to be licensed under particular event structural conditions which are mapped to grammatical relations via the reference frame. The reference frame in turn interacts with lexical semantic entailments in the form of Proto-Agent and Proto-Patient properties (cf. Dowty 1991, Zaenen 1993) which are linked to grammatical relations via a binary feature classification system.

The linking theory presented in this chapter serves to explain the diachronic developments and synchronic patterns evidenced in the corpus studies from Chapters 4 and 5. I have shown in Section 6.4 that the increasing systematic association between dative subjects and experiencer semantics results from the overall tendency of sentient participants to be realized as the FIGURE argument with stative predications, which more firmly links dative experiencers to state HOLDERS and subjects in turn. Furthermore, experiencers are in principle open to various linking possibilities and are non-canonical subjects, showing a weaker overall tendency to follow the positional licensing constraints developed in the language. In line with this argumentation, I have provided an account for the preponderance of dative subjects together
with lexicalized stative experiencer and raising predicates carrying middle morphology in the period post-1900 in Section 6.5. I have argued that these predicates were instantiated as dative subject predicates only after they have been reanalyzed from describing a dynamic event to being a stative predication via middle formation.

In Sections 6.6 and 6.7, I furthermore provided detailed analyses of the interaction between event structure and case marking with dynamic predicates. I have shown that the event structural conditions which license dative case marking on goal and theme arguments differ from one another in a way that affects the licensing of case marking under middle formation. While dative case is licensed with theme arguments via a temporal mismatch between the initiating and the process event of dynamic predicates, goal arguments are generally licensed as resultees and receive dative case marking because they are sentient state holders. Under middle formation, the sentient resultee, i.e., the dative goal argument, refers to the figure argument and is linked to subj. With a sentient state holder as subject and a rHEME argument as object, the event structure of middles of ditransitives is similar to the structure postulated for stative experiencer predicates taking a dative subject. This moreover explains the diachronic rise of experiencer predicates via middle formation. Dative goals contrast with dative themes which are licensed due to a temporal mismatch between the causing and the process event caused by the initiator. Being unavailable for linking in the middle, the initiator can no longer cause the mismatch and dative case is no longer licensed on the corresponding theme arguments. Furthermore, dative themes are generally most often realized as objects, given that they do not correspond to the prototypical dative subject pattern which has a state holder as subject and a rHEME as object.

In conclusion, the linking analysis presented in this chapter has provided a theoretical account for the diachronic development of dative subjects in Icelandic, showing that case marking is part of a complex system with many interacting parts which may moreover change over time. Overall, lexical semantic changes led to changes at argument structure, effecting changes with respect to event participants and case marking in turn. Furthermore, the rise of positional licensing induced an increasing systematic association between figure arguments and the prefinite position, tying dative experiencers and subjects more firmly together. In sum, these findings postulate that the licensing conditions for case and grammatical relations have been changing in the history of Icelandic, arguing against the inheritance of a monolithic and common Proto-Indo-European dative subject construction in the Icelandic diachrony.
Chapter 7

Summary and conclusion

This thesis investigated the diachronic interrelation between dative subjects, lexical semantics, event structure, voice and word order in the Icelandic Parsed Historical Corpus (IcePaHC, Wallenberg et al. 2011) in order to provide an understanding of the complex interacting system which licenses case and grammatical relations in Icelandic. Moreover, the investigations presented in the thesis shed more light on the diachronic development of dative subjects in Icelandic, contributing to the on-going discussion on whether dative subjects are a common Proto-Indo-European inheritance. In order to account for the complexity of the Icelandic linking system and the diachronic developments as evidenced by the corpus investigations, I developed a novel linking theory which factors in the relevant features for licensing case and grammatical relations in Icelandic, i.e., event structure, lexical semantics, position, and information structure.

Motivated by Svenonius’ (2002) event structural approach to case marking in Icelandic, the corpus study presented in Chapter 4 investigated the interaction between dative subjects, voice, thematic roles and event structure, using the glyph visualization developed by Schätzle and Sacha (2016). For the analysis of event structure, I decomposed the dative subject predicates from IcePaHC into subevents and event participants as per Ramchand’s (2008) event decompositional framework of the first-phase syntax which allowed for a more nuanced approach to the interaction between case and event semantics than provided by Svenonius. In Ramchand’s first-phase syntax, dynamic events can be decomposed into maximally three subevents, i.e., initiation, process and result, licensing three event participants: INITIATOR, UNDERGOER, and RESULTEE. In addition, the process and the result subevent can license a
RHEME argument which further describes the corresponding subevent. Stative events cannot be decomposed in the same way and consist of a state HOLDER and a RHEME argument, which further describes the stative eventuality in question.

In my investigation, four different event structure verb classes emerged with a dative subject: (i) stative predicates, (ii) transfer verbs, (iii) transition verbs, and (iv) verbs describing a scalar change. Stative predicates have a dative case marked state holder as subject and can take a nominative RHEME object. The state holder moreover refers to an experiencer argument. Transfer verbs are prototypically ditransitives with a dative goal object, which becomes a subject under passivization and middle formation. With transfer verbs, the subject is an INITIATOR marked with nominative case, the dative goal object is licensed as a RESULTEE, and the secondary object is an accusative case marked RHEME. Transition verbs can be further divided into simple transitions, manner of motion verbs and mental process verbs, generally having a dative marked theme argument. With simple transitions and manner of motion verbs, the dative argument is licensed as an UNDERGOER, either as a subject when the predicate is an unaccusative verb, or as an object, which is realized as a dative subject in passive constructions. When the dative UNDERGOER is the object, a nominative INITIATOR is the subject which in turn becomes demoted in the passive. Mental process verbs have a nominative subject which is both INITIATOR and UNDERGOER, and a dative case marked RHEME as object which is realized as a dative subject in the passive. Verbs describing a scalar change are similar to transition verbs, but license a dative marked UNDERGOER which is at the same time a RESULTEE. The class of verbs describing a scalar change consists of change of state predicates, verbs describing a change of bodily state or a change of location, i.e., verbs of directed motion. As with transition verbs, verbs describing a scalar change can have a dative subject with unaccusative predicates or take a dative subject in passive constructions, when an INITIATOR is demoted. With verbs describing a bodily change, the dative UNDERGOER-RESULTEE can be an experiencer, while it is generally a theme argument with other verbs describing a scalar change.

By means of the glyph visualization, I was able to conduct a detailed investigation of the complex interaction between dative subject predicates, event structure, thematic roles and voice in IcePaHC. The investigation showed that dative subjects occurred most often together with stative predicates as experiencer arguments in the corpus. Moreover, the association between dative subjects and experiencer semantics increases systematically over time. This increase correlates with an increasing
use of stative experiencer predicates carrying middle morphology. These predicates are lexicalized experiencer and raising predicates such as finnast ‘find, think, feel, seem’, which is the predicate that occurs most often with a dative subject in the period post-1900. The lexicalized experiencer predicates carrying middle morphology are historically derived from middle forms of dynamic verbs and are not subject to Dative Substitution, a process by which accusative experiencer subjects are systematically replaced with datives (see Chapter 2). The corpus study furthermore showed that dative subjects are increasingly used with transfer verbs carrying middle morphology, whereby the dative goal argument is often reanalyzed as an experiencer. Moreover, there is an interrelation between middles of transfer verbs and stative experiencer predicates, as, e.g., the stative experiencer predicate and middle form síñast ‘seem, appear’ corresponds to the middle of the ditransitive transfer verb sína ‘show’ which has a dative goal object. Transition verbs and verbs describing a scalar change occurred less frequently with a dative subject in the corpus than stative predicates and transfer verbs, and become increasingly confined to passive constructions. Furthermore, both verb classes were generally not found with a dative subject in middle constructions, as dative case is not retained with theme arguments under middle formation.

In order to be able to account for the diachronic interrelation between word order and case marking with respect to licensing grammatical relations in Icelandic, I conducted a second corpus study which investigated the interaction between subject case and word order in IcePaHC using the HistoBankVis visualization system (Schätzle et al. 2017), focusing on the diachronic interplay between dative subjects, subject positions, and verb placement. The investigation of subject case and word order was presented in Chapter 5 and showed that Icelandic is changing over time with respect to word order. In particular, I found that subjects are increasingly placed in the prefinite, clause-initial position in IcePaHC and V1 declarative structures are decreasing concomitantly. I therefore postulated that the corpus data provided evidence for the development of structure and the rise of positional licensing in Icelandic, which I accounted for in a formal syntactic analysis using LFG. However, as V1 declaratives are still part of the modern language and the expletive það, which is not a subject, is restricted to the clause-initial position, I argued that the clause-initial, prefinite SpecIP position does not license subjects straightaway, but first becomes identified as topic position in the history of Icelandic. Therefore, I opted for an information structural motivation behind the observed developments. As subjects are
usually topics, they are also increasingly placed in the SpecIP position, which then becomes the preferred structural option for subjects in Icelandic. The investigation of the interaction between subject case and word order conducted with HistBankVis moreover uncovered that dative subjects differ with respect to the overall structural developments. Dative subjects show a weaker tendency to be realized in a particular position than subjects overall. While subjects in general occur preferably in the prefinite position from the earliest time stages attested in IcePaHC on, dative subjects only start to mainly occur prefinally in the period post-1900, which coincides temporally with the increasingly systematic association between dative subjects and experiencer semantics in Icelandic.

The overall picture which emerges from the investigations presented in Chapters 4 and 5 is that the distribution of dative subjects has been changing over time. This speaks against the inheritance of a monolithic and stable dative subject construction in Icelandic which has been passed down from earlier language stages as is assumed by the Oblique Subject Hypothesis (see, e.g., Barðdal and Eythórsson 2003, 2009, Barðdal et al. 2012). Moreover, in conducting the corpus investigations, using visual analytics for historical linguistics (HistLingVis) for data analysis has proved to be an extremely valuable methodology, which facilitated the identification of changing features as well as the detailed examination of interactions between multiple different data dimensions, i.e., linguistic structures, across time.

In order to provide a theoretical analysis of the diachronic corpus findings and to account for the complex interaction between case, grammatical relations, word order, lexical semantics, and event structure in Icelandic, I developed and presented a novel linking theory as an extension to LFG’s Lexical Mapping Theory in Chapter 6. My linking theory combines the relevant features for linking case and grammatical relations in Icelandic and builds on the enhancements of Lexical Mapping Theory with respect to lexical semantics and argument positions brought forward by Zaenen (1993) and Kibort (2014). As central component, the linking system introduced in this thesis implements a reference frame in the form of Talmy’s (1978) figure-ground division which represents the interaction between information structure and word order in Icelandic. The lexical semantic properties of the reference frame and the event structure are moreover taken to yield a set of lexical semantic entailments in the form of Dowty’s (1991) Proto-Roles, which regulate the licensing of grammatical relations via the binary feature classification proposed by Zaenen (1993). This in turn links grammatical relations to particular structural positions. Case marking is
linked to arguments via the event participants posited in Ramchand’s (2008) first-phase syntax approach, which were used for the investigation of the interaction between event structure and dative subjects in Chapter 4. Moreover, event structure interacts with the reference frame which functions as mediator between case, lexical semantics, position, and grammatical relations.

In Chapter 6, I provided a theoretical analysis of the corpus findings using my linking theory. By means of the theory, I posited that the increasing systematic association between experiencer semantics and dative subjects with stative predicates is effectuated by the tendency of sentient event participants, i.e., Proto-Agents in the sense of Dowty (1991), to be realized as state holder and figure arguments, which in turn are prototypical topics. Once topics have been firmly associated with the clause-initial SpecIP position, dative experiencers are increasingly placed there and are eventually linked to subjects. This is mostly in line with Kiparsky’s (1997) linking analysis of experiencer subjects in Icelandic, who argues that the features associated with the SpecIP position in Icelandic are able to override the inherent dative case features, regulating the linking of dative arguments to the subject function. I moreover provided an analysis for the diachronic development of dative subjects together with lexicalized experiencer predicates carrying middle morphology which mainly drive the increase of dative subjects in the period post-1900. My analysis postulated that these predicates were instantiated as dative subject predicates only after they have been reanalyzed as stative predicates, with the reanalysis process being effectuated by middle formation.

My linking theory is moreover able to account for the differences which arise between theme and goal arguments with respect to dative case marking under middle formation. As per Svenonius (2002), I assume that, with dynamic predicates, dative case is licensed when two subevents of a single, indivisible event do not have the same temporal extension. Using Ramchand’s event decomposition, I show that the event structural configurations which license dative case on theme arguments differ from the structural constraints applying to dative goal arguments. With transition verbs and verbs describing a scalar change, dative case is licensed on the undergoer(-resultee), i.e., the theme argument, on the grounds of a temporal mismatch between the initiation and the process subevent caused by the initiator which is only partially involved in the process described by the verb. I assume that, in line with Kaufmann’s (2007) analysis of the middle, although the initiator is in principle still available at the level of event structure, no syntactic argument can
be linked to the INITIATOR in the middle. When no INITIATOR can be linked, the mismatch between the initiating and the process event is lost with transition verbs and verbs describing a change of state. Thus, in the corresponding middle, dative case cannot be licensed. In contrast to this, transfer verbs, which take a dative goal argument, license dative case because the goal argument is a sentient RESULTEE, i.e., a state holder. In the middle, the dative argument still refers to a sentient state holder, which is, similar as with stative experiencer predicates, the FIGURE argument. Therefore, dative case is retained in the middle with goal arguments. Overall, the IcePaHC data indicates that the Icelandic system becomes regularized over time in that dative subjects become increasingly associated with event structural configurations which have a sentient state holder as FIGURE argument and a RHEME as GROUND. These are mostly stative predicates with an experiencer subject, but regular middle forms of transfer verbs add to the overall increase of dative subjects in the history of Icelandic. As a consequence of this, dative themes, i.e., UNDERGOER and UNDERGOER-RESULTEE arguments, become increasingly confined to the object function.

The thesis at hand shed light on the diachrony of the complex interacting system licensing grammatical relations and case in Icelandic, offering new perspectives for future research. For one, much more needs to be done to be able to understand how information structure was expressed across the attested stages of Icelandic in order to fully capture the role which information structure plays in the complex system. Moreover, the knowledge gained about the interrelation between case marking, event structure, word order and grammatical relation could be encoded in a computational resource to build a (morpho-)syntactic parser or various other NLP tools for Icelandic. Lastly, as HistLingVis is a newly emerging field of research, the full potential that visual analytics has to offer for historical linguistic research has not yet been tapped, leaving an abundance of open research possibilities.
Appendix

Event structure classification of dative subject predicates established in Chapter 4:

1. Class I – Stative predicates

- Psychological states


- **Bodily states:**

  - **active**: bleða ‘bleed’, koma ekki dår á auga ‘not sleep a wink’, koma varla blundur á auga ‘not sleep a wink’, slá niður ‘have a relapse’, svíða ‘be hurt, feel pain’

- **Characteristics and properties:**

  - **active**: ánægja ‘suffice’, duga ‘be enough’, fara vel/illa ‘suit well/badly’, gagna ‘be of use’, hæfa ‘suit, fit’, hætta til/við ‘have a bad tendency to’, hlýða ‘befit, be fitting’, nægja ‘suffice’, sama ‘suit, become/be proper for sby’, varða ‘be liable to, finable, punishable’

  - **middle**: gagnast ‘be of use, be helpful’
• Postion/stance:
  – active: ægja saman ‘be mixed up’, hátta ‘be arranged, constructed’, seinka ‘be delayed’
  – middle: dveljast ‘stay longer than intended, get a hindrance’, hankast ‘be caught’

• Obligation:
  – active: bera ‘have an obligation’, byrja ‘have an obligation’

2. Class II – Transfer verbs

• Transfer verbs:
  – active: gefa vel ‘get a fair wind’, gefa byr ‘get a fair wind’, tilfalla ‘get, receive’
### 3. Class III – Transitions

#### Mental processes:


#### Simple transitions:


#### Manner of motion:

- **active**: fylgja ‘accompany, be accompanied with’, muna ‘move’, reika ‘happen to wander’, skoła ‘be washed ashore’, slá ‘move quickly’
4. Class IV – Scalar changes

- Change of bodily state:
  - active: batna ‘get better, recover’, draga til dauða ‘be led to death by sth’, elna ‘become worse (sickness)’, fórla ‘grow weaker’, gróa ‘be healed (of wounds)’, hitna ‘get warmer’, sortna fjyrir augum ‘lose consciousness’, svala ‘get colder’, versna ‘get worse, deteriorate’
  - passive: mishyrma ‘mistreat, torture, inflict bodily harm’, slátra ‘slaughter’

- Change of state:

- Change of location (directed motion):
  - active: ljósta ‘blow/shoot up of a sudden’, snúa ‘get turned’, vökja ‘turn’
Bibliography


Bibliography


Bibliography


Mayer, Thomas, Wälchli, Bernhard, Hund, Michael and Rohrdantz, Christian. 2014. *From the extraction of continuous features in parallel texts to visual analytics of heterogeneous areal-typological datasets*. Amsterdam/Philadelphia: John Benjamins.


Ramchand, Gillian. 2006. Verb meaning and the lexicon, Manuscript, University of Tromsø.


Bibliography


Walkden, George. 2015. HeliPaD: the Heliand Parsed Database, version 0.9.

Wallenberg, Joel C., Ingason, Anton Karl, Sigurðsson, Einar Freyr and Rögnvaldsson, Eiríkur. 2011. Icelandic Parsed Historical Corpus (IcePaHC), version 0.9.


