Aspect, tense and modality: theory, typology, acquisition
Boland, J.H.G.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
This study searches for similarities in adult and child language systems in the domains of tense, modality and aspect (TMA). It takes its starting point in Functional Grammar, in which grammatical TMA expressions are classified according to their scope. Wider scope correlates with an increase in cognitive complexity and a decrease in communicative motivation. This leads to the expectation that TMA expressions with wider scope behave differently from expressions with narrower scope both in adult and in child language.

Several studies are conducted. A detailed analysis of TMA expressions in English serves to find out if their scope plays a role in synchrony and diachrony. A typological study is carried out to answer the question how scope differences are reflected in diachronic developments, synchronic configurations, frequency, and morphosyntactic properties of TMA expressions. The acquisition order of TMA morphemes is examined in detail for English and more generally for 24 other languages. Here the acquisition order of TMA expressions with different scopes, and the crosslinguistic similarities in the acquisition of TMA expressions are studied.

The study provides new theoretical insights in TMA. It reveals remarkable similarities in the development of TMA expressions in diachrony, synchrony and first language acquisition that provide a strong argument for more interdisciplinary research on language and language use. The book is of interest to typologists, psycholinguists and those interested in functional models of language.
Aspect, tense and modality:
Theory, typology, acquisition

Volume I
Aspect, tense and modality:

Theory, typology, acquisition

Volume I
Promotiecommissie
Promotores: Prof. dr. A.E. Baker
Prof. dr. P.C. Hengeveld

Overige leden: Prof. dr. H. Behrens
Prof. dr. O.C.M. Fischer
Prof. dr. W. Klein
Prof. dr. J.L. Mackenzie
Prof. dr. F.P. Weerman

Faculteit der Geesteswetenschappen
Universiteit van Amsterdam
CONTENTS

Volume I

ACKNOWLEDGEMENTS ............................................................................................................... XV

Chapter 1  Introduction .............................................................................................................. 1
  1.1 Introduction .................................................................................................................... 1
  1.2 Linguistic universals ..................................................................................................... 2
  1.3 Universals in first language acquisition .................................................................... 9
  1.4 Relations between universals in adult and child language .................................. 15
  1.5 Functional linguistic theory ....................................................................................... 16
  1.6 Structure of this thesis ............................................................................................... 21

PART I  THEORY ......................................................................................................................... 23
  Introduction to Part I ........................................................................................................... 23

Chapter 2  An Outline of Functional Grammar ...................................................................... 25
  2.1 Introduction .................................................................................................................. 25
  2.2 Levels of analysis ........................................................................................................ 25
  2.3 The representational level .......................................................................................... 26
    2.3.1 The predicate and arguments ............................................................................. 26
    2.3.2 The predication ................................................................................................. 29
    2.3.3 The proposition ................................................................................................. 29
  2.4 Modifiers ...................................................................................................................... 30
  2.5 The Scope Hierarchy ................................................................................................... 33

Chapter 3  Tense, Aspect and Quantification ......................................................................... 35
  3.1 Introduction .................................................................................................................. 35
  3.2 Situation types ............................................................................................................ 36
  3.3 Aspect .......................................................................................................................... 40
    3.3.1 Introduction ...................................................................................................... 40
    3.3.2 Defining aspect ................................................................................................. 42
    3.3.3 Categories in detail ......................................................................................... 44
    3.3.4 Summary .......................................................................................................... 51
  3.4 Tense ............................................................................................................................ 52
    3.4.1 Introduction ...................................................................................................... 52
3.4.2 Topic time .......................................................... 54
3.4.3 Defining tense ....................................................... 55
3.4.4 Scope of tense ....................................................... 57
3.4.5 Tense and aspect ................................................... 58
3.4.6 Summary ............................................................. 60

3.5 Quantification ............................................................ 61
3.5.1 Defining quantification ............................................ 61
3.5.2 Property quantification ........................................... 62
3.5.3 Event quantification ............................................... 63
3.5.4 Summary ............................................................. 64

3.6 Conclusion ............................................................... 65

Chapter 4 Modality, Evidentiality and Irrealis .................. 67
4.1 Introduction ............................................................... 67
4.2 Modality ................................................................. 67
4.2.1 Sense: potentiality, disposition, weak necessity and necessity 68
4.2.2 Source: internal, external and epistemic ....................... 70
4.2.3 Scope: participant-, event- and proposition-oriented ........ 74
4.2.3.1 External, non-deontic modality ............................... 77
4.2.3.2 External, deontic modality .................................... 77
4.2.3.3 Epistemic modality ............................................ 79
4.2.4 Classification of modality ......................................... 82
4.2.5 Summary ............................................................. 85

4.3 Related domains ........................................................ 85
4.3.1 Future tense and prospective aspect .......................... 85
4.3.2 Irrealis ................................................................. 87
4.3.3 Evidentiality .......................................................... 88

4.4 Conclusion ............................................................... 89

Chapter 5 Implications of the Scope Hierarchy ............. 91
5.1 Introduction ............................................................... 91
5.2 Grammaticalization ................................................... 92
5.3 Cognitive mechanisms and communicative needs ............ 96
5.3.1 Pragmatic inference .............................................. 96
5.3.1.1 Metaphor and metonymy .................................... 97
5.3.1.2 Polysemy and ambiguity .................................... 99
5.3.2 Automaticity ....................................................... 100
5.3.3 Communicative needs ............................................ 101

5.4 Questions and hypotheses .......................................... 103
5.4.1 Diachronic development ......................................... 104
5.4.2 Frequency and variation ....................................... 106
CONTENTS

5.4.3 Synchronic configurations ......................................................... 109
5.4.4 Expression form .......................................................................... 112
5.4.5 Expression order .......................................................................... 113
5.4.6 Language acquisition ................................................................... 114
5.5 Conclusion ...................................................................................... 115

PART II TYPOLOGY .................................................................... 117
Introduction to Part II ......................................................................... 117

Chapter 6 TMA in English............................................................... 119
6.1 Introduction ................................................................................... 119
6.2 Inventory of TMA expressions ...................................................... 120
6.2.1 Criteria for grammaticality .......................................................... 120
6.2.2 Aspect ........................................................................................ 126
6.2.3 Tense .......................................................................................... 130
6.2.3.1 Past ......................................................................................... 130
6.2.3.2 Present .................................................................................... 130
6.2.3.3 Future ..................................................................................... 135
6.2.4 Event quantification ................................................................. 135
6.2.4.1 Habitual .................................................................................. 135
6.2.4.2 Frequentative ........................................................................ 138
6.2.5 Modality ...................................................................................... 139
6.2.5.1 Potentiality ............................................................................. 140
6.2.5.1.1 Participant-oriented potentiality ........................................... 140
6.2.5.1.2 Event-oriented potentiality .................................................. 141
6.2.5.1.3 Proposition-oriented potentiality ...................................... 142
6.2.5.2 Disposition .............................................................................. 145
6.2.5.2.1 Participant-oriented disposition ....................................... 145
6.2.5.2.2 Event-oriented disposition ............................................... 146
6.2.5.2.3 Proposition-oriented disposition ..................................... 146
6.2.5.3 Weak necessity ......................................................................... 147
6.2.5.3.1 Participant-oriented weak necessity .................................. 148
6.2.5.3.2 Event-oriented weak necessity ......................................... 149
6.2.5.3.3 Proposition-oriented weak necessity ............................... 150
6.2.5.4 Necessity ................................................................................ 151
6.2.5.4.1 Participant-oriented necessity .......................................... 151
6.2.5.4.2 Event-oriented necessity .................................................. 153
6.2.5.4.3 Proposition-oriented necessity ...................................... 154
6.2.6 Irrealis ......................................................................................... 155
6.2.7 Summary ..................................................................................... 156
6.3 Diachrony ...................................................................................... 159
CONTENTS

6.3.1  Tense inflection ................................................................. 159
6.3.2  Aspectual constructions .................................................. 159
6.3.3  Central modals ............................................................... 161
6.3.4  Other modals ................................................................. 166
6.3.5  Habitual ........................................................................... 169
6.3.6  Irrealis ............................................................................. 170
6.3.7  Discussion and conclusion ............................................. 171

6.4  Frequency ........................................................................... 172
6.4.1  Token-frequency .............................................................. 172
   6.4.1.1  Data-selection ............................................................. 172
   6.4.1.2  Coding of utterances .................................................. 173
   6.4.1.3  Results ........................................................................ 174
6.4.2  Size of classes .................................................................. 175

6.5  Synchronic configurations .................................................... 176

6.6  Expression form .................................................................... 178

6.7  Expression order .................................................................... 183

6.8  Conclusion ............................................................................ 185

Chapter 7  TMA in Languages of the World .............................. 187

7.1  Introduction ................................................................. 187

7.2  Diachrony ........................................................................... 187
7.2.1  Introduction ................................................................. 187
7.2.2  Method ........................................................................... 188
7.2.3  Results ........................................................................... 189
7.2.4  Conclusion ..................................................................... 195

7.3  Methodology for H2-H5 .................................................... 196
7.3.1  Selection of forms .......................................................... 198
7.3.2  General coding .............................................................. 199
7.3.3  Details of the analysis .................................................... 203

7.4  Frequency ........................................................................... 211
7.4.1  Crosslinguistic frequency ............................................... 211
7.4.2  Size of operator classes .................................................. 212
   7.4.2.1  Method ........................................................................ 212
   7.4.2.2  Results ........................................................................ 214
   7.4.2.3  Summary ..................................................................... 215
7.4.3  Reconsidering H2b and H2c .......................................... 216
7.4.4  Conclusion ..................................................................... 218

7.5  Synchronic configurations .................................................... 219
7.5.1  Implicational presence .................................................... 219
   7.5.1.1  Creole languages ......................................................... 221
   7.5.1.2  Conclusion ................................................................. 222
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.5.4</td>
<td>Discussion and conclusion</td>
<td>309</td>
</tr>
<tr>
<td>8.6</td>
<td>Qualitative analysis: Variation in use</td>
<td>310</td>
</tr>
<tr>
<td>8.6.1</td>
<td>Introduction</td>
<td>310</td>
</tr>
<tr>
<td>8.6.2</td>
<td>General method and analysis</td>
<td>311</td>
</tr>
<tr>
<td>8.6.3</td>
<td>Acquisition of inflection</td>
<td>312</td>
</tr>
<tr>
<td>8.6.4</td>
<td>Acquisition of auxiliaries</td>
<td>316</td>
</tr>
<tr>
<td>8.6.5</td>
<td>Acquisition of modal domains only</td>
<td>320</td>
</tr>
<tr>
<td>8.6.6</td>
<td>Discussion and conclusion</td>
<td>325</td>
</tr>
<tr>
<td>8.7</td>
<td>Qualitative analysis: Construction of paradigms</td>
<td>326</td>
</tr>
<tr>
<td>8.7.1</td>
<td>Introduction</td>
<td>326</td>
</tr>
<tr>
<td>8.7.2</td>
<td>Method and analysis</td>
<td>326</td>
</tr>
<tr>
<td>8.7.3</td>
<td>Results</td>
<td>329</td>
</tr>
<tr>
<td>8.7.4</td>
<td>Discussion and conclusion</td>
<td>331</td>
</tr>
<tr>
<td>8.8</td>
<td>Polysemous markers</td>
<td>332</td>
</tr>
<tr>
<td>8.9</td>
<td>Discussion and conclusion</td>
<td>333</td>
</tr>
<tr>
<td>9</td>
<td>Functions of TMA in Child English</td>
<td>335</td>
</tr>
<tr>
<td>9.1</td>
<td>Introduction</td>
<td>335</td>
</tr>
<tr>
<td>9.2</td>
<td>Methodology</td>
<td>337</td>
</tr>
<tr>
<td>9.2.1</td>
<td>Data selection</td>
<td>337</td>
</tr>
<tr>
<td>9.2.2</td>
<td>Coding</td>
<td>337</td>
</tr>
<tr>
<td>9.3</td>
<td>Results</td>
<td>340</td>
</tr>
<tr>
<td>9.3.1</td>
<td>Distribution in adult-adult language</td>
<td>340</td>
</tr>
<tr>
<td>9.3.2</td>
<td>Distribution in child language</td>
<td>341</td>
</tr>
<tr>
<td>9.3.2.1</td>
<td>Progressive</td>
<td>342</td>
</tr>
<tr>
<td>9.3.2.2</td>
<td>Simple past</td>
<td>344</td>
</tr>
<tr>
<td>9.3.2.3</td>
<td>Simple present</td>
<td>346</td>
</tr>
<tr>
<td>9.3.2.4</td>
<td>Conclusion</td>
<td>348</td>
</tr>
<tr>
<td>9.3.3</td>
<td>Distribution in the input</td>
<td>348</td>
</tr>
<tr>
<td>9.3.3.1</td>
<td>Progressive</td>
<td>348</td>
</tr>
<tr>
<td>9.3.3.2</td>
<td>Simple past</td>
<td>350</td>
</tr>
<tr>
<td>9.3.3.3</td>
<td>Simple present</td>
<td>352</td>
</tr>
<tr>
<td>9.3.3.4</td>
<td>Conclusion</td>
<td>353</td>
</tr>
<tr>
<td>9.4</td>
<td>Explanations</td>
<td>354</td>
</tr>
<tr>
<td>9.4.1</td>
<td>Aspect Before Tense Hypothesis</td>
<td>354</td>
</tr>
<tr>
<td>9.4.2</td>
<td>Prototype Account</td>
<td>356</td>
</tr>
<tr>
<td>9.4.3</td>
<td>Discourse Topic Hypothesis</td>
<td>357</td>
</tr>
<tr>
<td>9.4.3.1</td>
<td>Introduction</td>
<td>357</td>
</tr>
<tr>
<td>9.4.3.2</td>
<td>Relevant combinations</td>
<td>359</td>
</tr>
<tr>
<td>9.4.3.3</td>
<td>Children's development</td>
<td>361</td>
</tr>
<tr>
<td>9.5</td>
<td>Discourse topics</td>
<td>361</td>
</tr>
</tbody>
</table>
9.5.1 Use of progressive .................................................................362
  9.5.1.1 Progressive for talking about the here-and-now ..........362
  9.5.1.2 Progressive for describing past events .......................364
  9.5.1.3 Summary ........................................................................366
9.5.2 Use of simple past ...............................................................367
  9.5.2.1 Simple past for talking about the here-and-now ..........367
  9.5.2.2 Simple past for describing past events .......................369
  9.5.2.3 Summary ........................................................................372
9.5.3 Use of simple present ..........................................................373
  9.5.3.1 Simple present for talking about the here-and-now .......373
  9.5.3.2 Simple present for commenting on the world ...............374
  9.5.3.3 Summary ........................................................................375
9.5.4 Conclusion ............................................................................376
9.6 Discussion and conclusion .....................................................376

Chapter 10 Acquisition of TMA across Languages ...... 379
10.1 Introduction ............................................................................379
10.2 Methodology ..........................................................................381
  10.2.1 Data collection ..................................................................381
  10.2.2 Classification of TMA .......................................................383
10.3 The Data ................................................................................384
  10.3.1 Altaic: Turkish ..................................................................385
  10.3.2 Australian: Warlpiri .........................................................389
  10.3.3 Bantu: Sesotho .................................................................391
  10.3.4 Eskimo-Aleut: West Greenlandic .....................................392
  10.3.5 Finn-Oegric: Finnish .........................................................396
  10.3.6 Indo-European .................................................................498
    10.3.6.1 Baltic: Lithuanian .........................................................498
    10.3.6.2 Germanic: Dutch, German, Swedish .........................499
    10.3.6.3 Greek ..........................................................................415
    10.3.6.4 Romance: French, Italian and Spanish .......................420
    10.3.6.5 Slavic: Polish and Russian ...........................................430
  10.3.7 Japanese ............................................................................438
  10.3.8 Kartvelian: Georgian .........................................................443
  10.3.9 Korean ...............................................................................446
  10.3.10 Mayan: Quiché and Yucatan ..........................................452
  10.3.11 Semitic: Modern Hebrew .................................................457
  10.3.12 Sino-Tibetan: Mandarin and Cantonese .........................458
  10.3.13 Trans-New Guinea: Kaluli ..............................................467
10.4 Intermediate conclusion .........................................................469
10.5 Discussion ..............................................................................472
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.5.1</td>
<td>Introduction</td>
<td>472</td>
</tr>
<tr>
<td>10.5.2</td>
<td>Acquisition of tense and aspect</td>
<td>473</td>
</tr>
<tr>
<td>10.5.2.1</td>
<td>Functions of early TA forms</td>
<td>473</td>
</tr>
<tr>
<td>10.5.2.2</td>
<td>Universal acquisition order of tense and aspect functions</td>
<td>481</td>
</tr>
<tr>
<td>10.5.2.3</td>
<td>Conclusion</td>
<td>491</td>
</tr>
<tr>
<td>10.5.3</td>
<td>Acquisition of modality</td>
<td>492</td>
</tr>
<tr>
<td>10.5.3.1</td>
<td>Reanalysis of Korean</td>
<td>492</td>
</tr>
<tr>
<td>10.5.3.2</td>
<td>Universals in the acquisition of modality</td>
<td>497</td>
</tr>
<tr>
<td>10.5.3.3</td>
<td>Explanations of acquisition order</td>
<td>500</td>
</tr>
<tr>
<td>10.6</td>
<td>Conclusion</td>
<td>501</td>
</tr>
</tbody>
</table>

**PART IV**

**CONCLUSIONS**

**Chapter 11**

Conclusions and Implications

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.1</td>
<td>Introduction</td>
<td>509</td>
</tr>
<tr>
<td>11.2</td>
<td>Summary and conclusions</td>
<td>509</td>
</tr>
<tr>
<td>11.3</td>
<td>Implications for future research</td>
<td>515</td>
</tr>
</tbody>
</table>

**ABBREVIATIONS**

**APPENDICES**

**LIST OF TABLES AND FIGURES**

**REFERENCES**

**SAMENVATTING (SUMMARY IN DUTCH)**

**CURRICULUM VITAE**
Acknowledgements

During the six years of investigating tense, aspect and modality, many people and institutions have supported me. Without their help, this project would not have succeeded. The Universiteit van Amsterdam, the Landelijke Onderzoeksschool Taalwetenschap (LOT) and the Amsterdam Center for Language and Communication (ACLC) provided the necessary means to do my work. I am particularly grateful to Machtelt Bolkestein (†), Folkert Kuiken and Anne Baker for all their efforts.

I am indebted to many teachers for my intellectual growth during the years. Several courses in the LOT schools were of great help to me. Special thanks go to Jürgen Bohnemeyer, who introduced me in all the ins and outs of tense and aspect, to Michael Tomasello, who convinced me of the idea that children construct a language, and to Melissa Bowerman, who inspired me to think more about the relation between conceptual and linguistic development. I have also very much enjoyed the classes of Nick Enfield on language contact, of Bill Croft on typology and universals and of Marian Klamer on linguistic fieldwork. A very special occasion was the Linguistic Institute in Santa Barbara, 2001. The discovery of the richness of functional linguistics has been an important source of inspiration in my work. Particularly the classes of Roger Andersen, Peter Austin, Elizabeth Bates, Joan Bybee and Adele Goldberg were very stimulating.

I want to express my deep gratitude to my supervisors Kees Hengeveld and Anne Baker. They have been an almost perfect supervising couple. Anne provided the necessary structure, both in the planning of the project as a whole and at the text level. Kees made sure that there was always enough enthusiasm and positive energy to continue. My discussions with them have been of major help during the years, to clarify my ideas and to improve my work. They were both of great support when I suffered from physical problems. Anne has spent much time and probably patience on correcting my English. She helped enormously in unravelling the chaotic child data. Kees has spent hours on helping to analyze the different languages and on discussing my almost endless adjustments of the theoretical issues. Thank you.

I am very grateful to Heike Behrens, Olga Fisher, Wolfgang Klein, Lachlan Mackenzie and Fred Weerman for accepting to be a member of the committee. I want to thank Niels Smit and Elma Blom who read large parts of the manuscript and provided many useful comments. Thanks to them, I could improve the book in many respects. I am also very grateful to Angela van Leeuwen who designed Figure 10-3.
At several occasions, I was delighted by the friendliness with which colleagues from around the world have approached me. A lovely event in my first year as PhD student was the Hungarian and Dutch Language Acquisition (HuDuLa) project at NIAS in Wassenaar: the collaboration with, among others, Jeroen Aarssen, Steven Gillis, Casper de Groot, Evelien Krikhaar and Szilvia Papp was very pleasant. At several conferences and symposia I enjoyed the time I spent with Helena Taelman, Karen Schauwers and Judith Rispens. I want to thank Hella Olbertz for the great weeks in Santa Barbara. I also want to thank the participants of the FG group for the stimulating discussions and all the researchers who have so kindly responded to the questions I sent to them.

I have had many roommates during the years, in order of appearance: Margriet Heim, Česke Niewold, Gerry Wanders, Margot Rozendaal, Martin Salzmann, Norval Smith, Miriam van Staden, Rafael Fischer, Jorge Gomez Rendon, Eva van Lier, Marina Dyakonova, and Niels Smit. They made my working days very agreeable. A few of them need a special word of thanks. Margot was of endless help in discussing the coding scheme and the actual coding of the English files. It was great to have a mate with whom I could discuss the beautiful and difficult child utterances. With Česke, I had many talks about our lives as PhD students and it always cheered me up to find out that our strugglings were very comparable. Česke’s irony is beyond compare. Thanks to Gerry I did not have to worry about pencils or printing paper. She was very sensitive to when I could use some support in the form of kind words or cookies. And, as if this was not enough, she also read through different versions of the manuscript and prevented me from many inconsistencies.

The contacts between the PhD-students from the ACLC were very pleasant. At first we were a small group including, to mention a few, Wim, Barbertje, Nel, Victoria, Margot, Hedde, and Suzanne, later growing with, among others, Maren, Diana, Marian, and Daniela. The sharing of fascinations and frustrations made me like my work much more. I thank everyone from het oude ATW. I have enjoyed the atmosphere of friendship and familiarity.

Many thanks go to my friends, my family and my family in law for their support and their interest in my well being during the years. In the last year, many people took care of Jozef so that I could continue writing. Marieke Schoenmakers, Lydia Kes, Marjolijn van Luit, Marie Anne Remmelink, Hans and Truus Boland, Hilly Kes, and above all, Klaas Kes, have been of endless help. If it had not been for their help, this dissertation would still be far from finished. Finally, I thank my love, Jan. He supported me mentally, by listening to my doubts and worries and by having confidence in me, as well as practically, by taking care of Jozef and the household. He was my anchor during the years. It is to him that I dedicate this work.
Chapter 1

Introduction

1.1 Introduction

The majority of human beings take one of their most typical characteristics for granted: their ability to use language. However, this ability is one of the most extraordinary and sophisticated properties of human beings. Seemingly simple aspects of specific languages are practically impossible to describe correctly, yet human beings use them without any problems. Although it takes a few years, children learn to speak like their parents, including the most complex constructions, without any explicit instruction. There is large variation in language systems all over the world and all these different systems are adequate systems for human communication. What, then, is language exactly? What is it used for and how does it work? And what are the cognitive mechanisms that make language use possible?

Unfortunately, the above questions are in general studied independently. Theoretical linguists are concerned with the question what language systems look like and, possibly, how they function in communication. Psycholinguists on the other hand are concerned with the question how our brains work and how language is processed and acquired. The two groups of scholars do not often turn their attention to each other’s field of research. However, there are probably strong relations between the two areas. On the one hand, cognitive processes may be of strong influence on possible language systems. What is easier to process and to acquire may be crosslinguistically more frequent. Language systems cannot be fully understood if it is unclear how the system develops in individual language users. On the other hand, the knowledge on variation in language systems may help in understanding the mental processes that play a part in language use and in language acquisition. Recurring linguistic properties may reveal important insights in cognitive processes. Although there is large variation in the acquisition of forms in different languages, from a more abstract viewpoint, there may be universals in the acquisition of the communicative and semantic functions that are expressed by language. Here, the study on linguistic variation might provide helpful insights. It is thus necessary to combine the knowledge of the two disciplines in order to establish
a deeper understanding of language and language users, their communicative needs and cognitive capacities.

In this thesis an attempt is made to examine the relation between universals in typological variation and in first language acquisition. This relation will be examined in the semantic domains of aspect, tense, and modality and the related domains of irrealis, evidentiality and quantification. These domains have been subject to a considerable number of studies on the theoretical side, on typological issues and acquisitional phenomena. This study focuses on the question to what extent the variation in grammatical systems of aspect, tense and modality is identical in languages of the world and in stages of language acquisition. This chapter discusses in more detail the significance of universals in adult language (1.2) and in child language (1.3) and the relation there might be between the two (1.4). Section 1.5 discusses how a functional grammatical theory might help to investigate and explain linguistic phenomena. Section 1.6 discusses the main research question and the structure of this thesis.

1.2 LINGUISTIC UNIVERSALS

At present there are about 6,000 different languages known, including extinct languages. These languages exhibit a rich diversity in every possible aspect of their linguistic systems: sounds, word and sentence structure, lexicon, social rules for appropriate language use, et cetera. Despite the fact that the variation is immense, there are limits on variation: only a subset of all logically possible combinations of linguistic properties is actually attested. Variation in language systems is not arbitrary. There are systematic similarities, presumably determined by essential universal human characteristics. One of the ultimate aims of linguistics is to discover uniformities between languages and the patterns in rule systems. If it is known what are possible and impossible languages, this will help to understand how the human mind works.

The quest for language similarities or linguistic universals is based on crosslinguistic comparisons (Comrie 1989; Croft 2003; Greenberg 1963). Linguistic universals are of different types. Firstly, there is a distinction between on the one hand unconditional or unrestricted universals and on the other hand implicational universals. Unconditional universals state a constraint on possible language types on one specific parameter, independent of other linguistic properties, such as ‘Indefinite articles never have more than two syllables’ or ‘No language has only VC or CVC syllables’. More often however the presence or absence of a linguistic property is related to the presence or

---

absence of another property, so that the universal is implicational. An example of an implicational universal is the presence of reflexives: If there is a pronominal reflexive for first person, then there is also one for second person and third person (Faltz 1985). In other words, no language has first person reflexives without also having second and third person reflexives, whereas the opposite configuration does occur. Implicational universals reveal important information about languages as they connect linguistic properties that are logically independent. The general format of an implicational universal is: If Q, then also P, which implies that there are languages with both properties, without both properties or with only property P, but there are no languages with only property Q. This is presented in (1):

\[(1) \quad \text{IF } Q, \text{ THEN } P\]

<table>
<thead>
<tr>
<th>P</th>
<th>Q</th>
<th>Attested</th>
<th>Unattested</th>
</tr>
</thead>
<tbody>
<tr>
<td>−</td>
<td>−</td>
<td>Attested</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>−</td>
<td>Attested</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>Attested</td>
<td></td>
</tr>
<tr>
<td>−</td>
<td>+</td>
<td>Unattested</td>
<td></td>
</tr>
</tbody>
</table>

A second distinction in types of universals is between absolute and statistical universals: Absolute universals are by definition exceptionless, as far as we know. Statistical universals, however, have counterexamples, but they do reveal interesting information about relations between different linguistic properties. Suppose there are two properties, R and S, then theoretically there are four possible linguistic systems: with both properties, without both properties or with one of the properties. If there is no relation between the properties, it may be expected that the distribution between all four types would be randomly distributed, for example as in (2):

\[(2) \quad + R \quad - R\]

<table>
<thead>
<tr>
<th>S</th>
<th>+</th>
<th>−</th>
<th>26%</th>
<th>22%</th>
</tr>
</thead>
<tbody>
<tr>
<td>−</td>
<td>23%</td>
<td>29%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If, however, the distribution is not random and this deviation is statistically significant, it may be assumed that R and S are somehow related, even though this relation is not absolute. An example of such a relation is: If the dominant word order is VSO, the language is probably prepositional (Dryer 1992). A
possible non-random distribution of the two properties R and S is presented in (3):

<table>
<thead>
<tr>
<th></th>
<th>+ R</th>
<th>− R</th>
</tr>
</thead>
<tbody>
<tr>
<td>− S</td>
<td>42 %</td>
<td>24 %</td>
</tr>
<tr>
<td>+ S</td>
<td>27 %</td>
<td>7 %</td>
</tr>
</tbody>
</table>

In (3), the presence of S seems to be dependent on the presence of R. There are hardly any languages (only 7%) in which S is present, but R is not. Such a strong tendency is remarkable and needs an explanation.

In short, if both distinctions are combined, there are four different types of universals:

1. Absolute unconditional universals (all languages have P).
2. Absolute implicational universals (if a language has Q, then it also has P).
3. Unconditional tendencies (nearly all languages have P).
4. Implicational tendencies (if a language has Q, then it will probably have P).

It appears that most linguistic universals are of the fourth type; implicational tendencies. In many cases, there may be dependencies between more than two properties and they can be ordered as a sequence of implications, such as presented in (4):

\[ M \subset N \subset O \]

in which ‘\( \subset \)’ can be interpreted as: ‘is implied by’. The presence of M is implied by the presence of N, and the presence of N is implied by the presence of O. These sequences are called implicative hierarchies. An example of such a (statistical) hierarchy in the domain of phonology is presented in (5) (Jakobson 1941/1968):

\[ /p/ \subset /t/ \subset /k/ \]

The presence of the phoneme /p/ is implied by the presence of /t/ and the presence of /t/ is implied by the presence of /k/. In other words, a language with the phoneme /k/, also has the phonemes /t/ and /p/ and a language with the phoneme /t/, also has the phoneme /p/. The relation does not hold the other way around. An example within the domain of the lexicon is presented in (6) from Berlin & Kay (1969):
This hierarchy predicts that if a language has a color term more to the right, it also has the preceding color terms on the left.

There has been a recent debate about the usefulness and relevance of implicational hierarchies. Cysouw (2003) has warned against the practice to claim an implicational relationship on the basis of distributional frequency of different configurations of linguistic properties. He argues rightly that the distribution of configurational patterns is only relevant if it is significantly different from what can be expected on the basis of frequency of the individual linguistic properties alone. He concludes that there are only mutual statistically significant correlations between linguistic properties, but no true implications or one-way dependencies. In reaction to Cysouw, Maslova (2003), Dryer (2003) and Plank (2003) have objected to this latter stance and have convincingly shown that different types of unidirectional implicational relationships do exist and can be objectively established by statistical tests. They all, however, stress the importance of finding meaningful interpretations of relationships between linguistic properties after having established a statistically significant distribution of configurational patterns.

What is the information that implicational hierarchies provide? Firstly, they describe the possible or most likely configurations of linguistic properties in languages of the world within a specific domain. The hierarchy in (5) predicts that languages are likely to occur that have only /p/, or /p/ and /t/, or /p/, /t/ and /k/ are likely to occur, whereas languages with only /t/, only /k/ or with /t/ and /k/ but no /p/ are impossible or unlikely.

Secondly, implicational hierarchies describe the frequency of occurrence of certain items, both crosslinguistically and within a single language. The more to the left of the hierarchy an item is positioned, the more frequently the item will be found within languages of the world, but probably also within a single language. With respect to the phonological hierarchy this means that /t/ is more frequent than /k/ crosslinguistically, but also within one language.

Thirdly, implicational hierarchies provide ways to systematically describe differences between languages. From the hierarchy, the ‘cut-off point’ of a language may be determined: languages may have properties of the hierarchy up to a certain point. For example, if a language has property M and N of the hierarchy in (4) but not property O, then the cut-off point for that language lies between N and O. Since the hierarchy exhaustively defines the permitted configurations, no linguistic system will exist that does not conform to the
hierarchy. Consequently, it is only at the cut-off point that a language will change: it either loses property N or it acquires property O. This relation between typological and diachronic data was first hypothesized in Greenberg (1978) and later confirmed in many case studies. It is also at the cut-off point that there is regional variation and that language users are insecure about the grammaticality of the involved characteristics.

It is an ongoing debate how to explain linguistic universals in general and implicational hierarchies in particular. Within formal approaches to language, linguistic variation is presumably constrained by an innate universal grammar (UG) that contains possible values for different linguistic parameters. Within functional approaches on the other hand, the paradigm in which this thesis is embedded, universals are explained directly by general cognitive capacities of human beings, and properties of human communication (e.g. Comrie 1989: 24-29; Dik 1997a: 7). Implicational hierarchies are said to reflect scales of universal preferences (Haspelmath 2004; Vennemann 1983), also called scales of typological markedness (Croft 2003) and they are also referred to as markedness or preferential scales.

How do cognitive capacities and properties of communication influence linguistic systems? Human beings are dependent on their cognitive capacities in order to produce and interpret language: it may be assumed that linguistic structures that are easy to process are the most efficient and functional in communication and have the greatest chance to emerge and to survive in language use. Structures that cognitively tax our brains, on the other hand, disturb the fluency of communication. Such user-unfriendly structures have little chance to arise and will quickly change to structures that may be processed more efficiently (cf. Haspelmath 2004: 565). The way our brains process information will determine and restrict the possibilities for efficient linguistic rule systems. It is known for example that the processing of subordinate clauses is easier—all other things being equal—when they are at the beginning or end of the sentence (left- or right peripheral), than when they are in the middle of the main clause. It appears that there is a strong tendency for the order of noun and relative clause to correlate with the basic word in such a way that the preferred order arises, i.e., in SOV languages relative clauses tend to occur on the left, and in VSO languages on the right (Comrie 1989: 27).

Although it is not exactly known how the mind works, it is certain that different factors play a part in efficient processing, such as perceptual saliency, frequency, iconicity and economy. Perceptual saliency helps the hearer identify an element; frequent use of an element or structure facilitates storage and retrieval. The principle of iconicity is that ‘the structure of language reflects in some way the structure of experience’ (Croft 2003: 102). It is, for example, easier to process a sentence that iconically reflects the temporal ordering of
events—he locked the door before he left—than a sentence that does not—before he left, he locked the door. A one-to-one relation between form and function may also be considered a reflection of iconicity. The principle of iconicity competes with the principle of economy: ‘the expression should be minimized where possible’ (2003: 102). It is more economic, but less iconic, to express different meanings in one morpheme (more functions in one form), such as in homonymy or when person and number are expressed in one inflectional morpheme.

Another way in which human cognition plays a part in linguistic systems is through the influence of conceptual structures. A recent approach to linguistic universals is the semantic map model: these are universal structures of conceptual knowledge, represented by the semantic or conceptual space, that describe the universal relations between conceptual values (Croft 2001: 105). Languages differ in how they map linguistic constructions onto the conceptual space, and in what meanings are covered by a single linguistic element (e.g. Anderson 1982, 1986; Croft 2001, 2003; Van der Auwera & Plungian 1998). A linguistic element always maps onto one or more concepts that are related within a conceptual space (Croft 2001: 96; Van der Auwera & Plungian 1998: 112). An example of such a conceptual space or conceptual continuum is presented in Bowerman & Choi (2001) for static spatial relationships, consider Figure 1-1:

![Figure 1-1. Simplified conceptual space for static spatial relationships and crosslinguistic variation in semantic mapping (adapted from Bowerman & Choi 2001: 485)](image)

2 Bowerman & Choi (2001) do not present a form for the expression of the four relations in the middle for Japanese.
It appears that languages do not categorize spatial situations in arbitrarily different ways. All of them appeared to be constrained by an underlying gradient—an implicational hierarchy—that orders spatial situations in the way shown in [Figure 1-1] (…) there was variation in how many spatial terms a language used to cover the situations, and in where the territory of one term left off and that of the next began, but if a term was used for more than one segment of the gradient, it covered adjacent segments. (Bowerman & Choi 2001: 485-86)

Besides cognitive skills of the language users, the function of language itself will be decisive in the development of linguistic systems. Language is an instrument for expressing communicative needs of human beings, such as asking questions; referring to entities, situations or ideas; making other people do, know or believe something; maintaining social relationships; describing events; expressing ones needs and wishes, etcetera. As the functions of language are assumed to be universal, the crosslinguistic variation in linguistic systems will be restricted: every language is an instrument for the same communicative functions. Although there are many different instruments, not every logically possible linguistic system could be an adequate instrument of human communication. For example, probably all languages have means to distinguish between interrogatives, declaratives and imperatives. A common expression form for this distinction is intonation and/or word order, whereas there is no language that expresses tense or number by a change in word order or intonation. It may be assumed that the former functions are very urgent or basic to communication and therefore need to be substantially different from each other: word order and intonation are perceptually very salient distinctions. A rather straightforward example of an assumed relation between form and communicative function is that a warning, for example for an approaching elephant, buffalo, or truck driver, is not efficient or functional if the part of the linguistic structure that signals the warning is at the end of a long construction. Finally, the fact that different linguistic rules have to collaborate and operate as a well-oiled machine probably blocks certain configurations of properties, as they would lead to ineffective systems, unsuitable for communication. Some universal linguistic phenomena thus ‘serve to make language more functional, either as a communication system in general, or more particularly relative to the communicative needs of humans’ (Comrie 1989: 26-27).

No sharp boundary can be drawn between the influence of cognition and of communicative needs, since they are interrelated. It may be assumed that the way human beings conceptualize the world influences what they want to talk about and how they structure their linguistic information. Concepts that are relevant to human beings (entities, activities or properties people naturally pay attention to) will in general also be relevant or basic aspects in communication. Thus, universal
cognitive bias and communicative needs are closely related to each other, and are expected to have influence on language systems.

In sum, universal cognitive capacities and properties of communication determine the boundaries for linguistic variation; they are the restraining framework in which languages are shaped and continuously reshaped. Since language users all around the world have comparable cognitive capacities for processing information and constructing concepts, linguistic rule systems will look alike. And since language users all around the world try to reach similar communicative goals, the instruments they use, i.e. their languages, will show similarities. This influence could be schematically represented as in Figure 1-2.

![Figure 1-2. Constraints on typological variation](image)

There are different explanations for different universals. Every linguistic universal needs an explanation in cognitive or communicative terms, rather than in terms of an innate universal grammar (UG) (cf. Slobin 1985). In this view implicational hierarchies reflect extra-linguistic sequences going from least to most abstract in meaning, from most to least relevant to communication, most to least efficient in processing or perceptually most to least salient. None of these sequences is by itself a sufficient explanation for all possible restrictions: they complement each other. Future research may reveal further general social-cognitive characteristics in human beings that constrain possible language types.

### 1.3  UNIVERSALS IN FIRST LANGUAGE ACQUISITION

Linguistic universals are not only of importance to the description of linguistic systems of adults. In the study of first language acquisition there is an ongoing debate about what is universal in the process of language acquisition and what is language-specific (see Berman 1986; Bowerman & Levinson 2001b; Lieven 1997; Slobin 1985, 1997). Children start at the same, universal point, but end up
speaking different languages. On the one hand there is evidence that children follow similar developmental paths crosslinguistically, for example, active constructions are acquired before passive constructions and main clauses before subordinate clauses, but on the other hand there is evidence that children immediately pay attention to language-specific characteristics, for example, the acquisition of spatial expressions in English, Mayan languages and Korean follows language-specific patterns from the start (Bowerman & Choi 2001; P. Brown 2001; de León 2001; Levinson 2001).

What is universal in child language and what is language-specific? Similar to what was assumed for adult human beings, cognitive capacities of children are in principle universal (although there is individual variation in age and speed of development) and the same holds for their communicative needs. It thus may be assumed that cognitive capacities and communicative needs form a universal constraining framework in which language is acquired. The fact that children learn to speak different languages is obviously the consequence of a difference in language input.

One hotly debated issue is what the cognitive capacities are that children bring to the job. The debate has largely concentrated on the question of how children acquire the meaning of words. This may at first sight seem a rather uncomplicated aspect of language acquisition, but when one considers all the logically possible referents of a word, the problem of assigning meaning to words turns out to be virtually unsolvable. This can be illustrated by the famous example of Quine (1960: 29), who described this ‘problem of referential indeterminacy’ in the context of a native speaker who says *gavagai* to a field linguist at the moment a rabbit runs by. The linguist immediately assumes that *gavagai* means ‘rabbit’, but how can he be sure that the word does not refer to this specific rabbit only, to animals in general, to a specific part or property of the rabbit such as his head, furiness or color, to the activity of scurrying, etcetera? Besides, as Bloom (2000: 4) noticed, *gavagai* could equally well mean ‘Look!’ or ‘I’m bored’, or it could be the case that the chain of sounds consists of two or more words.

Children continuously need to assign meaning to words. But how do they succeed if there are simply too many possible referents? Apparently, children do not come up with all the logical options: there must be some restrictions on what they consider potential referents. One way to think of these restrictions is to suppose that there are very specific constraints on word learning (see Landau, Smith, & Jones 1988; L. B. Smith 2001; Waxman 1990, 1994). Children could for example assume that words only refer to whole objects or entities (the entire rabbit) and not to parts or properties of it (the ears, the softness) (see Golinkoff, Mervis, & Hirsh-Pasek 1994; Macnamara 1982; Markman 1989, 1994), or that words only refer to a whole category of objects (rabbits), rather than to a
subgroup of it (running rabbits, eating rabbits) (Markman & Hutchinson 1984). There is disagreement whether constraints or word learning principles are innate or learned, generally cognitive or specifically linguistic, universal or language-specific and at what age they play a role in children’s word learning.

A different way to face the problem of word learning is the recent social-pragmatic, usage-based account of first language acquisition, in which the idea of a priori constraints on word learning is weakened or even rejected (see P. Bloom 2000, 2001; Clark 2003; Slobin 2001; Tomasello 1999, 2001, 2003). On the contrary, this approach assumes that general social-cognitive abilities guide the child in language acquisition. In this view, Quine’s description of the problem of referential indeterminacy is mistaken, in that it ignores the influence of a shared context in communication. Children acquire language in social interaction and they make use of the shared context between parent and child for establishing possible referents for a word. Bloom (2000) states that:

words are learned through abilities that exist for other purposes. These include an ability to infer the intentions of others, an ability to acquire concepts, an appreciation of syntactic structure, and certain general learning and memory abilities. These are both necessary and sufficient for word learning. (p.10)

Tomasello (1999; 2001; 2003) poses that understanding of intentional and mental states of other persons is the crucial property that enables human beings to use language and to communicate. He argues that from around nine months, the ability emerges in human infants to understand others as intentional agents like themselves (Tomasello 2003: 3, 21-28). This means that children understand goal-directed behavior and attention of others as manifest in their behavior. This ability becomes manifest by some important changes in the behavior of children around this age, which do not take place in non-human primates. Firstly, children become able to check the attention of others by following gaze: they begin to share attention with others to an object or event of mutual interest (joint attentional frame). Such situations of joint attention create a domain ‘of current relevance’ for the child, a common contextual frame for interpreting the communicative intentions of the other. Secondly, children start monitoring emotional reactions of adults: they begin to understand that the communicative intention of the other is to change their own intentional states, for example, the other intends for the child to change his or her attention towards an object. Tomasello (2001: 133) states that ‘the understanding of intentions—specifically, the understanding that other persons have intentions towards my intentional states—is the very foundation on which

\[^3\] It is only around age four that children understand the other as a mental agent: a person with plans, beliefs, and desires that are not necessarily visible in behavior (cf. Tomasello 1999: 179).
language acquisition is built’. The third change that Tomasello mentions is that, because children now understand the other as an intentional agent, they become able of imitative learning, also called social or cultural learning. The crucial point about imitative learning is that the child not just mimics the adult’s action: the child identifies with the adult, and understands the underlying goal or intention of the adult’s action. The child thus intentionally reproduces the adult’s intentional act and imitatively learns the use of tools and instruments. In this approach, the use of language as an instrument to communicate intentions is one of the domains in which social learning takes place. Children start to imitatively use language. The imitation of this intentional linguistic behavior leads to the acquisition of linguistic symbols, as a by-product of social interaction.

Tomasello further states that the process of assigning meaning to sounds takes place in contexts of joint attention and that children’s interpretations of adult speech are based on the assumption that this speech will be relevant to the ongoing social interaction. The ‘social contexts serve to “constrain” the interpretive possibilities’ (Tomasello 2001: 135): children simply do not come up with all the possible hypotheses for linguistic symbols, since they would not make sense in the context. In addition, Tomasello (2001: 135-36) assumes that, when children begin acquiring language, they have ‘an adult-like understanding of at least some aspects of the social activities in which they participate’. They have a growing ability to conceptualize the world similarly to adults, which also restricts them from certain hypotheses about word meanings. These conceptualizations, however, are not a priori tied to language in specific ways.

The process of word learning depends fundamentally on the child being biased to conceptualize the world in certain ways (similar to adults’ conceptualizations), it is just that the connection of conceptualizations to language must be learned in communicative interactions with others. (Tomasello 2001: 153)

In a similar vein, Slobin (2001) argues that the acquisition of grammatical elements is restricted by the communicative context in which these elements are used because this context simply restrains children from postulating certain notions as meanings for these forms. As opposed to his earlier work (1985), he now claims that there are no prelinguistic (grammatically) notions in the individual child’s mind that operate as constraints on interpreting linguistic elements. Language variation is limited not because of a universal restricted set of semantic entities that can be expressed grammatically, but because of ‘conditions on the processing, social use and learning of form-function relations’ (Slobin 2001: 438). It is the social-historical development of language that creates forms to express what is relevant, important or salient to human experience and communication.

Finally, Clark (1987; 1993; 2003) proposes that from an early age children are sensitive to the pragmatic principles of conventionality and contrast, which
encompass the assumptions that there is a conventional form for expressing a certain meaning, and that a difference in form signals a difference in meaning. Now if a speaker does not use the conventional form in an appropriate context, then the addressee infers that he must mean something different from what the conventional form would have signified. This pragmatic understanding also helps in assigning meaning to words.

Contrary to Bloom (2001: 178), both Slobin (2001: 439) and Tomasello (2001: 155) argue that the same learning mechanisms apply across the lexicon, including both content words, purely lexical items, and functors, purely grammatical items. I subscribe to the position that the acquisition of grammatical elements as well as whole linguistic constructions are embedded in a social-pragmatic context and is guided by their communicative functions.

What is universal in language acquisition is the capacity of children to interpret communicative intentions and to imitate these intentions in addition to general learning abilities, such as various kinds of pattern finding or categorization. A further domain that probably is of influence to language acquisition is the general conceptual development of children. The conceptualizations that children make are to a large extent universal and independent of their native language, although recent attention to the Sapir-Whorf hypothesis (Whorf 1956), stating that language-specific characteristics cause their speakers to have different conceptualizations of the world, suggests that ‘some very moderate form of “Whorfianism” may be unavoidable’ (Bowerman & Levinson 2001a: 13). Nevertheless, children learning different languages have to discover which concepts are denoted by which linguistic elements in their native language. These relations are not pre-determined or constrained. In the words of Levinson (2001):

Many linguistic categories are simply not natural in any straightforward sense at all: they have to be learnt from instances of usage. Sure, they may be built out of underlying “natural” concepts, and moreover the range of variation may be limited. But the point is that languages construct concepts that otherwise might not have been. And that is precisely the added cognitive value of language: it provides “un-natural concepts,” complex conceptual wholes which connect across natural capacities (…), and which can be processed as units in working memory, thus vastly increasing the power of our mental computations. (…) This picture is radically opposed to the standard line in child languages research, which assumes that language rests directly on the fundaments of preexisting categories. (p.584)

The linguistic input to children in social interaction is the primary source for children that reveals form-function relations and that helps children to construct linguistic categories. The input to children is not universal: it varies in many respects, both structurally and in social-interactional ways, both crosslinguistically

---

4 Cf. Tomasello (2003: 4) for a summary.
and cross-individually (see e.g. Lieven 1994; Pine 1994). However, the possible variation in input languages, as argued in section 1.2, is limited: the communicative intentions for which language is used are similar in all different input languages and the possible structural and semantic properties are also restricted. Furthermore, the communicative needs of children themselves are in all probability universal. Children all over the world develop similar communicative needs: they want to describe what they see, express desires and intentions, ask questions and give orders, etcetera.

To conclude, there are several important universal factors involved in the process of first language acquisition: general processing and learning capacities, conceptualizations, functional and structural universals in language input, and communicative intentions of the children. These factors are expected to restrict the possible linguistic systems that may arise. During every stage of language acquisition, the child’s linguistic system, however primitive, is evidently controlled by the cognitive capacities of the child and it functions as an instrument for expressing communicative intentions of the child. As was argued in the previous section for adult languages, this will restrict the possible variation in child language.

Language acquisition is thus dependent on cognitive as well as communicative factors, just as similarities between adult languages of the world are dependent on cognitive as well as communicative factors. See Figure 1-3 for a schematic representation.

![Diagram](image-url)

**Figure 1-3.** Constraints on language variation
The limits on variation may be described using words of Tomasello:

There are (...) universals in the way symbols are created, learned, and used across languages, of course, reflecting universals both in the way human beings experience the world and in the ways they interact and communicate with one another socially. (Tomasello 2001: 135)

1.4 RELATIONS BETWEEN UNIVERSALS IN ADULT AND CHILD LANGUAGE

If human cognition and communicative intentions are the restraining factors in linguistic variation, both in typology and in acquisition, it may be assumed that there is a connection between typological phenomena and language acquisition: universals will identically be reflected in the typological domain and in stages of language acquisition. What is cognitively more efficient, conceptually more salient or communicatively more basic or relevant will be more frequent within languages of the world as well as within different stages of language acquisition.

In section 1.1 it was explained that variation in adult languages is often described by implicational hierarchies, which reflect scales of cognitive and communicative factors, such as perceptual saliency, communicative relevance, processing efficiency, etcetera. Consider once more the general format of a hierarchy:

\[(7) \quad P \subset Q \subset R\]

in which ‘\(\subset\)’ should be interpreted as: ‘is implied by’. The presence of property P is implied by the presence of property Q. Another approach is to call P a less marked property than Q. Here, I use ‘less marked’ to refer to conceptually less difficult and communicatively more motivated. Less marked properties have higher frequency of use and occur more often crosslinguistically. It has often been demonstrated that implicational hierarchies are reflected in languages of the world. For example, Bybee (1985) has shown that the ordering of morphemes is motivated by the iconic principle of Behaghel (1932: 4) ‘what belongs together mentally is placed close together syntactically’: morphemes are ordered in such a way that they reflect their relevance to the meaning of the stem they modify.

If the same types of language-external constraints hold for child and adult language, then the assumption is justified that implicational hierarchies or

5 See Haspelmath (t.a.) for a critical discussion of the use of the term markedness.
6 Contrary to Haspelmath (t.a.), I assume that conceptual difficulty is primarily the cause for low frequency of use in adult languages, and not the effect of it.
markedness scales also describe the variation within child language systems. Analogous to the prediction for adult linguistic systems, it may be predicted that within the process of language acquisition every stage will consist of a linguistic system that is in accordance with the implicational hierarchy. The hierarchy in (7) predicts that there should be no stage in which element Q is present, but P is not, or in which element R is present, but P and Q are not. If variation in child language is actually restricted by similar factors as variation in adult language, then the hierarchies exactly predict in what order elements are acquired. Only if children acquire first the least marked element P, then the more marked element Q and finally the most marked element R, their linguistic system is at every stage in accordance with the hierarchy, i.e. with cognitive or communicative constraints that hold for every human being. Another possibility is that children acquire the different properties simultaneously: for example, children could start out with property P and Q simultaneously and only later acquire property R. The acquisition order will thus follow the direction of the hierarchy: what is more to the left, will be acquired earlier than what is more to the right. If this is not the case, then a system arises in which for example the more marked property Q is present but the less marked property P is. This system would cross the limits of variation.

Implicational hierarchies or markedness scales established for describing adult language variation are thus hypothesized to be reliable predictors of universals in language acquisition. The relation between possible language variation and first language acquisition may even be so strong that the explanation for (some) implicational hierarchies is to be found in the learning capacities of children: the absence or infrequency of certain structures in adult languages might not be caused by the fact that adults would not be capable of using them, but by the fact that children have much difficulty learning them.

Whether the variation in adult and child languages is limited in identical ways will be the main subject of this thesis. The predictive power of implicational hierarchies for phenomena in languages of the world and for acquisition order will be investigated in the domain of aspect, tense, and modality.

1.5 FUNCTIONAL LINGUISTIC THEORY

Before turning to the actual study, the role of linguistic theory will be discussed. This thesis is embedded in the framework of Functional Grammar (Dik 1997a), a grammatical model of language, based on semantics and pragmatics. Within a functional framework, language is considered an instrument for social interaction, meant to establish communicative relations between human beings. This instrument, however, is structured by rules: a rule system of semantic, syntactic, morphological and phonological rules governs the constitution of
linguistic expressions while communicative rules or principles account for the patterns of verbal interaction, in which the linguistic expressions are used. Within this linguistic rule system, phonology is viewed as instrumental with respect to morphosyntax; morphosyntax is viewed as instrumental with respect to semantics and semantics is viewed as instrumental with respect to communicative functions.

The aim of Functional Grammar is to describe the rule systems of languages, from the starting point of communicative intentions or functions and semantics. The model should be general enough to describe every possible language, and specific enough not to describe impossible language systems. Furthermore, it should not only account for sentences but also for connected discourse. These requirements are defined as the descriptive adequacy of a theory (cf. Chomsky 1965). As different grammars could be constructed with the same descriptive adequacy, an even more important criterion for a linguistic model is its ‘explanatory adequacy’ (Dik 1997a: 13). In order to arrive at explanatory adequacy, the grammatical model may not clash with the restrictions on linguistic variation. Therefore the model should be compatible both with cognitive and communicative factors.

Following up on Dik (1997a: 12-15), Boland (1999) defined the criteria to which the model of Functional Grammar should comply as two standards of explanatory adequacy, namely the standard of cognitive adequacy and the standard of communicative adequacy, and as two domains of application, namely the domain of typological variation and the domain of acquisitional variation. The standards of explanatory adequacy serve as the constraining framework within which a theory can be developed. The standard of cognitive adequacy captures general cognitive features of human beings that might play a part in language function and form. A linguistic model should be compatible with cognitive models of linguistic behavior, processing capacities and conceptualizations of human beings. Givón (1988) provides a good example of relating cognition to language use. He explains the fact that in most languages the first position of the clause has a special function by a psychological feature of humans: it attracts more attention from the addressee. Givón states that ‘the string-initial position invites the hearer to pay more attention, and thus to store and retrieve the information more efficiently’ (1988: 276). This kind of relation is precisely what should be understood by the standard of cognitive adequacy. Human cognitive capacities have an impact on language and this relation should be made clear in a grammatical model.

The standard of communicative adequacy requires that a grammatical model should fit in a broader pragmatic theory of verbal interaction. Within a functional approach verbal interaction is described as follows: a speaker wants to communicate a certain intention to an addressee for which he uses a verbal
expression. In the production as well as the interpretation of verbal expressions speaker and addressee make use of pragmatic information: this is the knowledge of the preceding context, the current speech situation and general knowledge of the world. The speaker anticipates the interpretation of the addressee of his utterance by estimating the pragmatic information of the addressee, and shapes his intention in such a way that the addressee will be capable of reconstructing the intention of the speaker. In interpreting the speaker’s utterance, the addressee uses his own pragmatic information but he also makes an estimation of the pragmatic information of the speaker. The utterance in itself does not establish but rather mediates the relation between the intention of the speaker and the interpretation of the addressee. A linguistic model ultimately describes what the communicative intentions and functions of language are, and provides explanations on how different linguistic elements may function as a clue for encoding and decoding communicative intentions.

The two standards thus set the boundaries within which a linguistic model has to fit: a functional grammatical model may not be in conflict with what is known about communicative and cognitive factors, and should preferably reflect universal cognitive and communicative principles. The goal of the grammatical model is then to provide adequate descriptions of languages and systematic explanations for differences and similarities between natural languages. There are two domains of application of a theory: the theory should account for possible linguistic variation in languages of the world and in stages of language acquisition.

The first domain of application is typological variation. In Dik’s words:

A typologically adequate theory reveals the most fundamental recurrent properties of natural languages, properties which have sedimented into the systems of languages through centuries of intensive use in verbal interaction. It is a reasonable working hypothesis, then, that those principles which are most generally characteristic of natural languages are at the same time the principles which have the most fundamental psychological and pragmatic significance. (Dik 1997: 15)

I propose to define the domain of typological variation as including both crosslinguistic and diachronic variation. As Dik remarks, it is not only the case that a grammatical model should account for descriptions and predictions for specific languages and typological variation, but crosslinguistic comparison can also point to underlying cognitive or communicative factors. Universals may produce ‘hypotheses of cognitive structure that can be tested and confirmed or rejected by cognitive psychological research’ (Croft 2003: 203).

---

7 Bakker (1998) proposed a separate standard of diachronic adequacy. In my view, however, diachronic variation may be covered by the notion typological variation.
But the typological domain is not the only descriptive challenge. The assumption is justified that language acquisition is dependent on cognitive as well as communicative factors, just as similarities between adult languages are dependent on cognitive as well as communicative factors. The cognitive and communicative universals that restrict typological variation will also restrict the possible linguistic systems in every stage of first language acquisition. If the grammatical model is in accordance with the standards of communication and cognition, it should not only supply correct descriptions of languages of the world but also of the different stages within the process of language acquisition. In particular, the domain of acquisition includes normal first language acquisition. I assume that in the process of language acquisition of disordered children and in language attrition, there may arise atypical linguistic structures as there are atypical factors involved, such as possible deficits in the cognitive capacities of the language users. The process of second language acquisition will also have specific peculiarities: there may for example be interference of the first language.

From a functional perspective, it is fruitless to describe linguistic rule systems in isolation from the question why language is organized the way it is. The organization of verbal interaction is influenced by several human qualities and skills, such as biological properties, social capacities, cognitive development, information processing mechanisms, knowledge of the world, etcetera. One of the most intriguing questions is how all these different mechanisms co-operate in such a way that newborn human beings eventually become capable of verbal communication and speak like adults. In order to answer this question, it is crucial to examine how language is learned and what language is. The ultimate theory of language should answer both questions. A language theory that correctly describes the communicative purposes and linguistic rules of adult language, but cannot account for the developmental aspects of language is inadequate. A grammar of adult speech should be rejected if it is incompatible with the developmental stages and psychological processes of language acquisition. On the other hand, a theory of language acquisition is not complete if it does not explain why children eventually use the same rule system as adults to encode communicative intentions and decode linguistic expressions.

Why do we need a linguistic theory in the quest for universals? According to Dryer (t.a.) a theory-neutral “metalanguage” would do. Haspelmath (2004: 569) also argues that a phenomenological description of linguistic systems is enough for discovering linguistic universals: a grammatical model does not need to reflect the real underlying mental patterns of language users. His main objection to functional theoretical models is that they often claim to be cognitively real, whereas there are many subjective decisions in the construction of such models. He does, however, recognize that descriptive models may be more or less
adequate and that external evidence can determine which of two observationally adequate descriptions should be chosen (2004: 574).

Theoretical models that comply with explanatory adequacy may provide helpful information in understanding the limits on possible linguistic systems. This is exactly what the theory of Functional Grammar aims at: describing relations between different semantic and pragmatic functions of language in such a way that they are in accordance with general cognitive characteristics and communicative principles. Whether Functional Grammar represents real mental structures is not the most relevant question: to date, it is unknown what the real mental structures look like and we are far from understanding how neurological processes lead to conceptualizations, attention, memory, etcetera. Only provisional models of the human brain are available. The function of a linguistic model is to relate these mental models to linguistic structure and to account for the relation between communicative needs and their semantic and formal structuring. Descriptions of these relations and structures are helpful or even necessary for crosslinguistic research since the definition of a category for crosslinguistic comparison must be based on semantics, pragmatics, discourse function etcetera, and not on morphosyntactic criteria (see for example Croft 2003: 13-14). A further goal of functional linguistic models is to account for the relations between different aspects of the language system. A phenomenological description of grammar lacks an explanation for the relations between different linguistic aspects: establishing an ever-growing list of linguistic universals without being able to unify them and understand relations between them does not seem to be a fruitful enterprise. A functional linguistic model does not only try to account for variation within a restricted domain, but also for relations between linguistic phenomena as they occur within an entire linguistic system. If the relation between cognitive and communicative functions and language is modeled, it becomes possible to investigate if these functions are reflected in languages of the world. And if linguistic universals in languages of the world can be interpreted in terms of cognition and communication, new hypotheses can be formulated about human cognition and communication.

To recapitulate, a functional model of language serves as an interface between the standards of adequacy and the domains of application. On the one hand a theoretical modal has to fit in the framework of communicative and cognitive properties and on the other hand it has to provide correct descriptions of actual adult and child language data. Whereas the cognitive and communicative standards operate as restrictions on the possible grammatical models, the typological and acquisitional domains operate as tests of the theory. On the one hand, typological research and research of language acquisition may point to cognitive and communicative factors through the intermediate of a
linguistic theory: on the other hand, the modeling of cognitive and communicative factors is a helpful tool in describing and accounting for phenomena in typology and acquisition. This is represented in Figure 1-4. In this thesis I will show how the semantic representations in Functional Grammar are necessary tools for the crosslinguistic and cross-stage comparison of categories.

**Figure 1-4.** Demands on a functional model of language

### 1.6 Structure of this Thesis

The main research question in this thesis is:

Are the limits on variation across adult languages of the world identical to the limits on variation across stages of language acquisition?

The underlying assumption is that universal cognitive and communicative aspects constrain possible language systems, both in adult language and in child language. Since implicational hierarchies or markedness scales can often describe these restrictions within a linguistic domain, I predict that such hierarchies or scales are not only reflected in the typological domain but that they are also predictive with respect to the order of language acquisition. As mentioned before, this hypothesis will be tested in the semantic domains of aspect, tense, and modality, and the closely related domains of quantification,
irrealis and evidentiality. Functional Grammar provides an analysis of grammatical expressions of these notions that relates cognitive constraints and communicative needs to linguistic structure. On the basis of this analysis it is possible to predict limits on variation in adult and child language.

This thesis is divided in four parts. Part I is dedicated to theoretical approaches to aspect, tense and modality. In Chapter 2 a broad outline is given of the theory of Functional Grammar and in particular the analysis of grammatical expressions of aspect, tense and modality is explained. This analysis leads to the formulation of an implicational hierarchy. Chapter 3 defines the semantics of the domains of aspect and tense and the related domains of quantification and situation types. Chapter 4 presents a classification of different types of modality and the related domains of irrealis, future tense and evidentiality. Finally, in Chapter 5, the specific research questions for this study are formulated and motivated, using the implicational hierarchy of Chapter 2. Part II investigates the limits on variation for languages of the world. Chapter 6 discusses quantitative and qualitative properties of morphemes that express notions of aspect, tense and modality in English. Chapter 7 investigates universals in the systems of aspect, tense and modality in a sample of 76 languages. Part III is dedicated to first language acquisition. In Chapter 8, a detailed examination is presented of the acquisition of aspect, tense and modality in English. Chapter 9 examines the contexts of use of expressions of aspect and tense in English child language. Chapter 10 investigates the acquisition of aspect, tense and modality in a crosslinguistic perspective, using a survey of existing studies. Finally, Part IV consists of a discussion of the findings and the implications for further research.
Part I  Theory

INTRODUCTION TO PART I

Part I discusses theoretical issues concerning the expression of tense, modality and aspect by grammatical means. The semantic domains of tense, modality and aspect (hereafter, TMA) have been studied extensively in theoretical linguistics, from many different angles. Notwithstanding this enormous effort, the domains remain notoriously complex and hard to grasp. I will not discuss all different approaches to TMA. In this thesis, the description of TMA is embedded in the framework of Functional Grammar (Dik 1997a). This theoretical model of language provides an analysis of TMA that generalizes across different semantic functions within the domains, which leads to clear predictions about possible universals within the domain of TMA. The approach to TMA in Functional Grammar will be the subject of Chapter 2.

Although TMA is described rather precisely in the theory of Functional Grammar, the current state of the art within this model does not suffice in all respects for this thesis. Therefore, Chapters 3 and 4 are dedicated to a more precise description of the domain of TMA. Insights from Functional Grammar and other approaches are used to (re)define the functions of the broad TMA domains and of specific semantic functions within these domains and to classify TMA categories according to their scope. The proposed definitions and classifications of TMA are assumed to have universal value.

Finally, in Chapter 5, the relation between the domain of TMA and grammaticalization processes will be discussed. This discussion leads to the formulation of specific hypotheses concerning the limits on variation that may be expected to exist in TMA systems in languages of the world and in stages of first language acquisition.
Chapter 2

An Outline of Functional Grammar

2.1 INTRODUCTION

In Chapter 1 it was hypothesized that universals in general and implicational hierarchies in particular are reflected in languages of the world as well as in stages of first language acquisition. Is this hypothesis tenable within the domain of TMA? The semantics and pragmatics of these domains have to be established before TMA systems can be compared across languages in the world and across stages of language acquisition. In this thesis, the semantic description of TMA is embedded in the Theory of Functional Grammar, adopting the concept of scope. A TMA expression has scope over the maximal part of the utterance that is modified by that expression. Before the discussion of the domains of TMA in Chapters 3 and 4, the basic principles of Functional Grammar will be discussed in order to understand the concept of scope.

2.2 LEVELS OF ANALYSIS

The Theory of Functional Grammar (from now on FG) has as its aim to describe and explain the grammatical organization of natural languages. It is assumed that the grammatical organization of language is determined by its communicative functions. Accordingly, the basis for the model is constituted by the semantic and pragmatic functions that are universally relevant to languages, even though not every language makes use of the same functions. The model strives to account for the relations between pragmatics, semantics and morphosyntax and therefore tries to describe and define interpersonal communicative functions, semantic representations and expression forms.

In the research on FG, the exposition of the model by Simon C. Dik (1997a; 1997b) has long been the standard. However, in the course of the last decade, many adaptations to the model have been proposed culminating in the introduction of Functional Discourse Grammar (FDG) (Hengeveld 2004a). In FDG the relation between cognitive and communicative factors and linguistic structure is more fully developed. The FDG model starts from communicative intentions and describes linguistic structure at the level of discourse acts. Here,
those aspects of FG/FDG relevant for the present study will be discussed. For the sake of convenience, the term ‘FG’ will be used and not ‘FG/FDG’.

One of the basic principles of FG is that a linguistic expression conveys what the speaker talks about and why the speaker talks about it. An utterance is analyzed at four different levels: an interpersonal, a representational, a morphosyntactic and a phonological level. The interpersonal level accounts for the communicative intentions of the speaker, such as the distinctions between orders and requests, or the communicative acts of reference and ascription. The representational level accounts for the semantics of the communicated content of the utterance that is transferred from speaker to addressee. The structure of this level will be explained in more detail below. Expression rules relate the interpersonal and the representational layer to the morphosyntactic level that contains the language-specific formal elements and templates used to express the semantics and pragmatics. Finally, the phonological level is the interface between the formal structure and the actual articulation.

For this thesis, the representational or semantic level is the essential part of the FG model. The semantic structure of an utterance is subdivided into three hierarchically ordered layers that all serve different communicative functions. Firstly, by producing an utterance, the speaker describes a certain property or relation pertaining to one or more individuals, in other words, an event; the first layer of the semantic structure serves to describe a set of possible events. Secondly, the speaker relates the description of the set of possible events to the specific event the speaker has in mind; the second layer of the semantic structure serves to situate the event in a real or hypothesized world. Thirdly, the speaker transfers propositional content to the addressee; the third layer of the utterance serves to present the content of the speech act (Hengeveld 1989: 130). Expressions of TMA are described as operators that modify the different layers. Different TMA categories apply to different layers, they have different scopes, and consequently, they contribute to different communicative functions. In Chapters 3 and 4 specific TMA categories will be discussed. Here, the different layers at the representational level will be described in more detail.

2.3 THE REPRESENTATIONAL LEVEL

2.3.1 The predicate and arguments

The most basic communicative function that the semantic structure of an utterance fulfils is properly describing a state of affairs (also: event, situation).

---

1 See Anstey & Mackenzie (2005) for a thorough background in the development of FG, and Hengeveld & Mackenzie (in prep.) for a full account of the present state of the art.
This is done by the first layer of an utterance that primarily contains two semantic units: predicates and arguments. The predicate designates a property or relation and arguments designate entities involved in the property or relation expressed by the predicate. The description of an event is thus compositional.

Although languages universally have predicate-argument structures, it is language-specific how the predicate and argument slots may be filled to describe a certain real world event. All linguistic elements or lexemes are listed in the lexicon: the language user not only stores the form of the lexeme in question, but also information about the meaning or use and the communicative function for which it may be used. For each lexeme, it is specified how many arguments are involved, i.e., the entities that obligatorily participate in the property or relation designated by the lexeme. This is the quantitative valency of a lexeme. Each argument is labeled with a semantic function, specifying the part that the participant plays in the state of affairs. This is the qualitative valency of the lexeme. The representations of the lexemes read, man and old can illustrate this. See (1):

\[
\begin{align*}
\text{(1)} & \quad \text{a. (read) } (x_1)_{\text{Agent}} (x_2)_{\text{Patient}} \\
& \quad \text{b. (man) } (x_1)_{\text{Ø}} \\
& \quad \text{c. (old) } (x_1)_{\text{Ø}}
\end{align*}
\]

The verbal lexeme read describes the relation between two arguments; one argument \((x_1)\) has the semantic function Agent—the participant that reads—and the other \((x_2)\) has the semantic function Patient—the participant that is read. Both the nominal lexeme man and the adjectival lexeme old describe properties of only one argument \((x_1)\). This argument has the semantic function Zero, which means that the participant is primarily involved in a state. In 1b the participant has the property ‘man’ and in 1c it has the property ‘old’.

Second, the categorical status, verb, noun, adjective or adverb, is stored for each lexeme, which serves as information about the possible communicative functions for which the lexeme may be used. For example, different lexemes can be used predicatively. In the sentence The man reads a book, the verbal lexeme read is used as the predicate; in he is president the nominal lexeme president is used as the predicate; in the man is old the adjectival lexeme old is used as the predicate; and in she is here the adverbial lexeme here is used as the predicate. In addition to the predicative use, different lexemes or parts of speech have specific, distinguishing communicative functions. Without any modification, the function of a verbal lexeme is only predicative: it always designates a property

\[\text{\footnote{There is some variability in the valency of lexemes: for example, write in English may be used as a one-place predicate I'm writing or a two place predicate I'm writing a letter and there may also be a third participant, I'm writing a letter for John.}}\]
or relation that is ascribed to one or more arguments. Besides the predicative function, only a noun can be used with a referential function, to refer to an individual or a more abstract entity, such as the man is old. Adjectives and adverbs may be used for modifying, in addition to their potential predicative function: an adjective can be used to modify a nominal head and an adverb to modify a non-nominal head (Hengeveld 1992).

Depending on the event he tries to describe, the speaker chooses lexemes from the language-specific lexicon to serve as the predicate, designating the central relation or property in the event that is ascribed to a specific number and type of arguments. The speaker also chooses lexemes (or more complex constructions) that designate the entities that participate in the relation or property. In FG, the description of the event is formalized as in (2):

\[ (f_1: \text{Lexeme}_0) (x_1: \text{Lexeme}_0) \]

in which \( f_1 \) is the variable used as a symbol for the main predicate; the colon can be read as ‘such that’; \( \beta \) stands for the categorical status of the lexeme, verb, noun, adjective or adverb; \( (x_1)^n \) represents the required number and type of arguments. Arguments can be concrete individuals, such as John, the dog, or the book, but also abstract entities, such as the meeting, the idea or the question. In fact, the first layer of the semantic structure does not describe a specific event, but it describes a set of possible events. The description of the event at the first layer is not related to the real world; it is not specified which concrete event in the real world the speaker has in mind.

Speakers thus use different semantic units to construe a proper description of a set of possible events. In addition to the predicate \( (f) \) and the arguments \( (x) \), there may be units designating non-obligatory participants (adjuncts), for example with the role beneficiary, company or instrument, or properties of the event, such as the direction or path of an event. See (3):

(3) a. Diana, buying a present, for Margaret, \text{BENEFICIARY} 
   b. I went to Paris, with John, \text{COMPANY} 
   c. I want to write a letter, with a feather, \text{INSTRUMENT} 
   d. Peter and Angie, will cycle to France, \text{DIRECTION} 
   e. He drove through the tunnel, \text{PATH} 

Verbs need to be made a noun (nominalization) before they can be used to refer, such as in English by adding -ing fishing or eating as the name of the activities.
2.3.2 The predication

The speaker not only describes a set of possible events in an utterance. He also relates the description of this set of events to the concrete event he has in mind. The second communicative function of an utterance is thus situating the event. The semantic unit in an utterance that accounts for the function of situating the event is the second layer, that contains the **predication**. The predication designates an event and is symbolized by the variable ‘e’. This event e can be located in space, time or actuality. An illustration of the underlying structure of the predication is presented in (4) for ‘The old man reads a book’:

\[(4) e; [(f;read) \ (x_1;man_{N}\ :\ old_{A})_{Agent} \ (x_2;book_{N})_{Patient}]\]

‘The old man reads a book’

The predicate *read* provides two argument slots. The argument slot \(x_1\) \_Agent is filled with *the old man*, that itself is built up by the nominal lexeme *man* and the adjectival lexeme *old*. The argument slot \(x_2\) \_Patient is filled with *a book* that itself is construed on the basis of the nominal lexeme *book*. The predication, the semantic unit at the second layer, is thus construed by the semantic units of the first layer, the predicate, arguments and, potentially, adjuncts.

2.3.3 The proposition

Besides describing and situating, a speaker also transmits content to the addressee. In FG, the third layer of the underlying representation of an utterance serves to express the transferred content of the utterance. The semantic unit at this layer is the **proposition** that designates a potential fact or propositional content. A propositional content exists only in the mind of the speaker contrary to an event, which is part of the external world and exists independently of the speech act and speaker. A propositional content can be evaluated in terms of truth and the speaker may denote his personal attitude towards the propositional content. The proposition is represented by the variable ‘p’, which is restricted by a predication. Consider (5):

\[(5) p; [e; [(f; Preda) \ (x_1;p)]]\]

Each layer of the underlying representation is provided with a specific variable that symbolizes the entity designated at that layer: at layer 1 the variable \(f\) is used for the description of a property or relation and \(x\) for arguments, at layer 2 the variable \(e\) is used for the event and at layer 3 the variable \(p\) is used for the propositional content. The variables thus represent semantic units designating a certain type of entity. A simplified representation of the layers is presented in
Table 2-1. It is important to note that semantic units at lower layers form part of the semantic units at higher layers: the predicate and arguments form part of the predication and the predication forms part of the proposition. As a result the designated entities are more abstract and the semantic units are more complex at each next higher level.

<table>
<thead>
<tr>
<th>Linguistic unit</th>
<th>Structure</th>
<th>Function</th>
<th>Designation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicate + arguments</td>
<td>((f;\text{Pred}_0) (x_1)^n)</td>
<td>Describing</td>
<td>Relation / property + participants</td>
</tr>
<tr>
<td>Predication</td>
<td>([e; (f;\text{Pred}_0) (x_1)^n])</td>
<td>Situating</td>
<td>Event</td>
</tr>
<tr>
<td>Proposition</td>
<td>([p; [e; [(f;\text{Pred}_0) (x_1)^n]]])</td>
<td>Presenting content</td>
<td>Propositional content</td>
</tr>
</tbody>
</table>

The underlying semantic representation of an utterance in FG is thus based on the communicative functions that the utterance fulfils and its semantic designations. The semantic units are universal, that is, the predicate, the predication and the proposition can be expressed in any language, but each language has its own set of lexical and grammatical elements that may be used to build up the units.

### 2.4 Modifiers

The representational level of an utterance thus consist of layers that serve the functions of describing a set of possible events, situating the event and presenting the content of the utterance. Semantic units that designate different entities fulfil these functions: at the first layer, the predicate \((f)\) designates a property or relation and arguments \((x)\) designate the individuals that participate in the property or relations; at the second layer, the predication \((e)\) designates a state of affairs; at the third layer, the proposition \((p)\) designates the content of the utterance. Layers should be considered as semantic rather than syntactic units.

The designated entities at each layer of the semantic structure can be further modified by lexical or grammatical elements such as adverbial constructions, periphrases, inflection, particles or auxiliaries. The distinction between grammatical and lexical elements will be discussed in more detail in 5.2 and 6.2.1. For now, it suffices to say that the boundary between lexical and grammatical elements is not absolute, but rather a continuum ranging from purely lexical to purely grammatical elements. Grammatical modifiers have a general, often abstract, meaning and lexical modifiers a specific, often concrete,
meaning. For example, the grammatical expression of past tense in English, –ed, indicates that the event is temporally located before the moment of speech, whereas lexical expressions of past time, such as yesterday, two days ago or in 1907, indicate much more specifically when the event actually took place. Grammatical modifiers are called operators in FG and are represented by the symbol π in the semantic representation; lexical modifiers are called satellites and represented by the symbol σ. Grammatical expressions of TMA, the subject of this thesis, are thus grammatical modifiers, formalized as operators.

Modifiers contribute to the communicative functions of description, situating and presenting content, by modifying the semantic units and therefore specifying the designated entity of a semantic unit. The classification of modifiers in FG is primarily based on the scope of their modification, with which is meant the maximal semantic unit that is modified by the operator or satellite. The target of modifiers can be the property or relation designated by the predicate, the event designated by the predication or the propositional content designated by the proposition. Arguments can also be modified, but that is not relevant to this thesis and will not be discussed further. Modifiers are labeled with the number of the layer to which they apply: π1 and σ1 for grammatical and lexical modifiers of the predicate at the first layer, π2 and σ2 for modifiers of the predication at the second layer and π3 and σ3 for modifiers of the proposition at the third layer. A complete semantic representation of an utterance with its modifiers is presented in Figure 2-1.

![Semantic representation of the clause in FG](image)

Each language has its own set of expression rules that translate the underlying semantic representation into an actual linguistic expression. At the
point of translation morphosyntax comes into play. From a functional perspective morphosyntax is primarily a means to express semantics and pragmatics. Every detail of the syntactic organization of a language should thus have an underlying semantic or pragmatic explanation in the model of the utterance that accounts for these syntactic features or there should be general cognitive factors that explain the syntax.

This thesis is restricted to grammatical modification by \( \pi_1 \)-, \( \pi_2 \)-, and \( \pi_3 \)-operators. They will be referred to as predicate or \( \pi_1 \)-operators, predication or \( \pi_2 \)-operators and proposition or \( \pi_3 \)-operators. Notice in Figure 2-1 that (modified) semantic units of lower layers fall within the scope of semantic units at higher layers. Consequently, operators at lower layers fall within the scope of operators at higher layers. Note furthermore that several modifications may apply at the same time to the same semantic unit, so that operators at the same layer may have scope over each other. Scope relations between operators are thus not an a priori argument for locating those operators in different layers. The location of specific categories at a certain layer is entirely determined by its semantic contribution to the utterance.

How is the function of predicate operators defined? In Hengeveld’s definition (1989: 133) predicate operators (\( \pi_1 \)) contribute to ‘building up a proper description of the situation the speaker wishes to refer to.’ This is correct, but it is only a specific part of the event description that is influenced by predicate operators, that is, the designated property or relation. Predicate operators do not influence the designated participants or additional participants. The description of the participants can be modified but this does not influence the description of the predicate. The proper description of a state of affairs is thus established compositionally by the elements at the first layer: a (modified) predicate, (modified) arguments, and possibly, (modified) additional participants. I adhere to Hengeveld’s approach (1989) that this description of the event cannot be further modified as a whole, contrary to the position of Dik (1997a) and Cuvalay-Haak (1997). A predicate operator (\( \pi_1 \)) is thus a grammatical expression that changes the description of the property or relation that is ascribed to the argument(s), without modifying the description of the arguments itself. A \( \pi_1 \)-operator interacts with the semantics of the lexeme that designates the property or relation. The resulting complex property of relation is then applied to the arguments. Chapters 3 and 4 will show that the categories aspect, property quantification, and certain modal categories, such as ability and volition, are to be classified as \( \pi_1 \)-operators.

Predication operators (\( \pi_2 \)) contribute to the situating function of the predication, they ‘relate the description of a [state of affairs] to the occurrence of that [state of affairs] in a real or imaginary world.’ (Hengeveld 1989: 134). In Dik’s formulation (1997a: 218) \( \pi_2 \)-operators ‘leave the internal constitution of
the [state of affairs] intact, but either quantify it or locate it with respect to spatial, temporal and “objective” cognitive dimensions.’ In the next two chapters the semantic categories tense, irrealis, event quantification and certain modal categories are shown to be $\pi_2$-operators.

Proposition operators ($\pi_3$) contribute to the presentation of content, by evaluating this content. They specify how much responsibility the speaker takes for the propositional content or how reliable the speaker estimates that the proposition is, by indicating his personal attitude towards the proposition, his commitment to the truth of the proposition or the source of evidence he has for the proposition. In Chapter 4, it is shown that the categories of evidentiality and certain modal categories function as $\pi_3$-operators.

This thesis is restricted to operators at the representational level in FG. There are also modifiers that operate at the interpersonal level (2.2), and that express, among other things, the basic illocution of an utterance (declarative, interrogative), politeness strategies or discourse structure (finally, in short). These modifiers fall outside the scope of this thesis.

2.5 The Scope Hierarchy

As discussed above, FG analyzes an utterance as serving different communicative functions at the same time. The first layer fulfills the function of describing a set of possible events. The semantic units at this layer are the predicate and the arguments and possibly adjuncts. The predicate designates a property or relation and the arguments designate entities that participate in the property or relation. Predicate operators ($\pi_1$) have scope over the predicate only. They contribute to the description of a state of affairs by specifying additional features of the property or relation. The second layer fulfills the function of situating the event. It relates the description of the set of possible events to a concrete event in a real or imaginary world. The semantic unit at this layer is the predication that designates the event. Predication operators ($\pi_2$) have scope over the predication. They specify the setting and occurrence of the event, by situating the event in time or actuality or expressing the frequency of occurrence of the event. The third layer fulfills the function of presenting the propositional content. The semantic unit at this layer is the proposition that designates a propositional content. Proposition operators ($\pi_3$) have scope over the proposition. They are concerned with expressing the speaker’s personal attitude or commitment towards the propositional content (Hengeveld 1992: 130-32).

According to the model of FG, lower layers in the semantic representation form part of higher layers, which makes the relation between the layers hierarchical. Each higher layer contains a more complex semantic unit that
designates a more abstract entity. Operators like TMA expressions apply to
different layers and also stand in a hierarchical relation to each other, since
operators at higher layers have scope over all lower layers, including their
operators. As a consequence, operators at each higher layer modify a more
complex semantic unit ranging from a predicate to a predication to a
proposition, that specify increasingly abstract entities, from properties or
relations ($\pi_1$) to events ($\pi_2$) to propositional contents ($\pi_3$). The function of
operators at each higher layer thus seems to be cognitively more complex.

Furthermore, the function of operators at each higher layer seems to be
communicatively less motivated, that is, they present more redundant
information. The function of $\pi_1$-operators is most basic or motivated. In most
utterances, the property or relation ascribed to the arguments is important new
information and the modification of this property or relation by $\pi_1$-operators is
crucial to an adequate description of the event. $\pi_1$-Operators contain
information that is so specific that it cannot be predicted or inferred from
context and it needs to be expressed linguistically. The function of $\pi_2$
operators, modifying the relation of the event to the real or imagined world, is
also a basic communicative function, but here, the context often helps the
addressee to infer the temporal or spatial location, the actuality, or the
frequency. This means that $\pi_2$-operators more often than $\pi_1$-operators present
redundant information. Finally, the function of $\pi_3$-operators, modification of
the propositional content, seems to be communicatively least motivated: $\pi_3$
operators provide the speaker’s evaluation of the content, but in many
utterances, the expression of the speaker’s evaluation is not relevant. It is a
conversational maxim that the speaker is committed to the propositional
content: only when this is not the case or when the commitment of the speaker
is somehow special, the speaker’s evaluation has to be encoded linguistically.

It is thus assumed that if their scope is wider, operators are cognitively more
complex and communicatively less needed. In other words, operators with
wider scope have more marked functions than operators with narrower scope,
according to the Scope Hierarchy in (6):

\begin{equation}
\pi_1\text{-operator} \subset \pi_2\text{-operator} \subset \pi_3\text{-operator},
\end{equation}

in which ‘$\subset$’ stands for: ‘is less marked than’.

It is expected that the Scope Hierarchy is reflected in TMA systems of the
languages of the world and in stages of first language acquisition. In Chapters 3
and 4 it will be investigated how specific TMA domains should be understood
in terms of scope. In Chapter 5 hypotheses will be formulated on how the
Scope Hierarchy will be reflected in the variation of TMA domains in adult
languages and child languages.
Chapter 3

Tense, Aspect and Quantification

3.1 Introduction

In FG, TMA expressions are defined with respect to their scope, the part of the utterance that they modify. As the scope of an operator gets wider, the operator modifies a more complex semantic unit that designates a more abstract entity and it contributes to a less basic communicative function. See Table 3-1:

Table 3-1. Functions of operators

<table>
<thead>
<tr>
<th>Operator</th>
<th>Scope over</th>
<th>Modifies</th>
<th>Contributes to</th>
</tr>
</thead>
<tbody>
<tr>
<td>π1</td>
<td>Predicate</td>
<td>Property / relation</td>
<td>Describing</td>
</tr>
<tr>
<td>π2</td>
<td>Predication</td>
<td>Event</td>
<td>Situating</td>
</tr>
<tr>
<td>π3</td>
<td>Proposition</td>
<td>Propositional content</td>
<td>Presenting content</td>
</tr>
</tbody>
</table>

Operators with wider scope are more marked than operators with narrower scope, in that they are cognitively more complex and communicatively less relevant. The hierarchical relation between operators can be conceived of as a markedness scale or implicational hierarchy.

To determine the scope of specific TMA expressions, i.e. to which class of operators they belong, their semantics have to be described adequately: it has to be clear what the expression adds to the semantic structure in order to know as what type of operator it functions. As described in Chapter 2, predicate operators (π1) change the description of the property or relation that is ascribed to the argument(s) without modifying the description of the arguments or of additional participants. They specify additional features of the property or relation in such a way that it is applicable to the argument(s). Predication operators (π2) contribute to the situating function of the predication, they ‘relate the description of a [state of affairs] to the occurrence of that [state of affairs] in a real or imaginary world.’ (Hengeveld 1989: 134). Finally, proposition operators (π3) contribute to the function of presenting the content.

1 An earlier version of this chapter appeared in Boland (2005).
They specify how much responsibility the speaker takes for the propositional content or how reliable the speaker estimates the proposition to be.

In this chapter the semantics of the related domains tense, aspect and quantification are discussed together with the scope of specific categories in these domains. Section 3.3 is dedicated to the domain of aspect; the domains of tense and quantification will be discussed in sections 3.4 and 3.5, respectively. Quantification covers notions such as habitual, frequentative, repetitive, distributive and iterative, sometimes labeled as quantificational aspect. Although the discussion will be organized according to general TMA domains, note that I agree with the statement in Dahl (2000):

The basic units of description are not “the category of tense” and “the category of aspect” but rather what we call grams, i.e., things like Progressive in English, the Passé simple in French etc. Notions like tense, aspect, and mood are seen as ways of characterizing the semantic content of grams, or domains from which their meanings are chosen, but do not, in the typical case, represent structurally significant entities in grammatical systems. (p.7)

Before the semantics of TMA can be discussed, however, it is necessary to take a closer look at situation types or inherent temporal constituencies of events because aspect, tense and quantification are often interrelated with this component.

3.2 SITUATION TYPES

In the previous chapter it was explained that one of the major functions of an utterance is to describe an event, but it was not yet discussed what events actually are. In FG an event is defined as ‘the conception of something that can be the case in some world’ (Dik 1997a: 51). Events can be classified with respect to their inherent temporal structure. Vendler (1967), elaborating on Aristotle, was one of the first who recognized that verbs have different ‘time schemata’ and he proposed a classification of verbs. It is now, however, widely accepted that it is not the inherent temporal structures of a verb that should be taken into account but rather the inherent temporal structure of an event. This event is compositionally constructed by the (modified) predicate (often, but not necessarily, expressed by a verb), the (modified) arguments, and possibly (modified) additional participants. In the literature, different terms are used to refer to different temporal structures: lexical aspect, situation type or event structure.\(^2\) \(^3\)

---

\(^2\) Cf. Verkuyl (1972) for a systematic analysis of the compositional nature of event structure.

\(^3\) The term Aktionsart is also used to refer to event structure, but in my view it covers a related, though slightly different phenomenon, namely systematic marking of the type of relation or property designated by the verb, such as German *brennen* ‘burn’, *an-brennen* ‘start burning’ (of food), *ver-brennen* ‘burn up’.
In this thesis I will use the terms situation type and event structure interchangeably. I will not use the term lexical aspect in order to avoid confusion with (grammatical) aspect.

Different situation types are generally described by the universal features dynamicity, telicity and durativity (e.g., Chung & Timberlake 1985; Comrie 1976; Mourelatos 1978; C. S. Smith 1991). The first broad distinction is made between situations that are \([–\text{dynamic}]\) and situations that are \([+\text{dynamic}]\). Consider (1)-(3):

(1) Mary is searching for an envelope.
(2) She also needs a stamp.
(3) She has written a letter.

The difference between (2) on the one hand and (1) and (3) on the other is a matter of dynamicity. The situation type described in (2), `need a stamp`, is non-dynamic or static. It differs from the dynamic situation types `search for an envelope` and `write a letter` in that `need something` can continue forever, without any input of energy, `unless something happens to change that state` whereas `search for an envelope` and `write a letter` are actions started deliberately whereby the action `will only continue if it is continually subject to a new input of energy` (Comrie 1976: 49). Non-dynamic or static events are called states. Entities involved in a state are presented `as being or remaining the same at all points of the time interval during which the [state of affairs] obtains` (Dik 1997a: 107), whereas there is always some change involved in dynamic events.

A test that helps to distinguish dynamic and static events is to add an adverb of manner that expresses a sense of control, such as `deliberately` or `carefully`. Dynamic events (with animate agents) allow this combination, as shown in (4) and (6), whereas the combination with a state leads to a semantic anomaly, as shown in (5):

(4) Mary is searching for an envelope on purpose. / Mary is carefully searching for an envelope.
(5) *She needs a stamp on purpose. / *She carefully needs a stamp.
(6) She has written a letter on purpose. / She has carefully written a letter

Other examples of states are `know the Russian alphabet, want a new car, be a linguist, have three sisters, be blond, love John, see the eclipse, etcetera.`

Dynamic events can be further divided with respect to telicity. The above examples (1) and (3) differ in this respect. The event `search for an envelope` could in principle last forever: it is atelic. The event `write a letter` on the contrary has an inherent endpoint, the moment at which the letter is finished: it is telic.
Atelic events are internally homogeneous; any subpart of them has the same properties as the event as a whole. In contrast, telic events involve change through time: any subpart of the event ideally has properties different from those of any other subpart and different from those of the event as a whole. (Bohnemeyer 1998: 59-60)

In other words, a telic event involves a certain change of state, from a ‘source state’ to a ‘target state’ (Klein 1994). It describes ‘a process that leads up to a well-defined terminal point, beyond which the process cannot continue’ (Comrie 1976: 45), whereas atelic events have an arbitrary endpoint. Atelic events are referred to as activities or processes and telic events as events or state changes. I will use the term event or state of affairs as a neutral term and refer to static, dynamic, atelic and telic events or states of affairs. More examples of atelic events are dance, listen to the radio, hunt sharks, buy clothes. Some examples of telic events are climb the mountain, cross the river, kill the dragon and free the lady.

There are several criteria to determine the telicity of an event (Dik 1997a: 109). A first test is to combine it with a specification of duration. If an event is telic, it is possible to add a specification of the duration until the reaching of the endpoint, such as (within) an hour, whereas this is not possible with atelic events. Compare (7) and (8):

(7) Mary wrote the letter in an hour.
(8) *She searched for an envelope in an hour.

A second related test is illustrated in (9) and (10). A telic event can be combined with an expression like ‘it took X three hours to ...’, whereas this is impossible with an atelic event.

(9) It took her three hours to write a letter.
(10) *It took her three hours to search for an envelope.

A third criterion is the almost-test. If one adds almost to an atelic event, it implies that the event has never started, whereas with a telic event, it implies either that the event never started or that it was started but not finished. Compare (11) and (12):

(11) Mary almost wrote a letter (but she didn’t start/but she didn’t finish it).
(12) Mary almost searched for an envelope (but she didn’t start/*but she didn’t finish it).

Smith (1996) and Bohnemeyer (1998: 66) have noted that the above tests do not yield the same results in every language, as for example in Navajo and in
Yukatek Maya. The exact semantics of constructions like almost or within an hour are not identical in every language, which affect possible combinations. Bohnemeyer (1998: 60-61) suggests that the strongest test crosslinguistically is to use the question: If X is interrupted in the course of VERB-ing, has she then VERB-ed? With telic events, the answer is no, whereas with atelic events the answer is yes. Consider (13) and (14):

(13) If Mary is interrupted in the course of writing a letter, has she then written a letter? Answer: No.
(14) If Mary is interrupted in the course of searching for an envelope, has she then searched for an envelope? Answer: Yes.

Whether the event structure is telic or atelic depends for a large set of verbs on the presence of an object, whether the object refers to a specified or an unspecified quantity, or on the presence and type of an additional participant. See (15) and (16):

(15) a. *Bob painted within an hour. > atelic
  b. ?Bob painted landscapes within an hour. > atelic
  c. Bob painted five landscapes within an hour. > telic

(16) a. *Bill walked in Santiago within three hours. > atelic
  b. Bill walked to Santiago within three months. > telic

These examples illustrate that the event structure is not a property of the verb but rather the resultant semantics of the predicate and its arguments.

A third feature that distinguishes event structures is punctuality, i.e. whether events have duration in time or not. This feature is mainly relevant to telic events.

[punctual events] are conceived as having no duration: their beginning coincides with their terminal point: they occupy only one point in time. [non-punctual events] on the other hand, occupy a certain stretch of time, and have a distinct beginning and terminal point. (Dik 1997a: 111)

Telic punctual events are, for example, reach the summit, break the egg, sit down, etcetera. Punctual events do not combine very well with specifications of duration, such as for an hour. Telic non-punctual events are for example draw a still life, run a mile, bake a pie, etcetera. Their combination with duration is unproblematic.

By using the three parameters of dynamicity, telicity and punctuality, the different situation types can be described schematically as in Figure 3-1:
Smith (1991) demonstrated that there is a fifth situation type, which she called semelfactives. Examples are *jump*, *drip*, *knock*, *bounce* or *sneeze*. These verbs are often used to describe a repeated (iterative) event: for example, *jumping* in *I was jumping for an hour* involves a series of jumps that could be conceived of as an atelic event. Semelfactives describe one punctual event or a series of punctual events, but they do not involve a change of state: they are therefore atelic. They can be regarded as a special type of activity [+dynamic, -telic].

The above distinctions in event structures are commonly recognized and they seem to be universally applicable. However, in what way a specific real world event is semantically structured is a language-specific characteristic. For example, a specific real world event may in one language be described as a state (*know*), while another language describes it as a change of state (*come to know*, *realize*; one language may describe a specific event as a state (*be seated*) whereas another describes it as an activity (*sit*).

### 3.3 Aspect

#### 3.3.1 Introduction

To determine the scope of aspectual expressions, their semantics have to be defined. What do speakers do when they mark aspect? What do they add to the meaning of the utterance? The examples in (17)-(22) from Bohnemeyer (1998:

---

4 They are called ‘points’ in Moens (1987).
73) illustrate aspectual distinctions with respect to the description of an event of Mary writing a letter, which started at 6.30 and ended at 7 o’clock:

(17) Mary was going to write a letter.
(18) At 6.30, Mary started writing a letter.
(19) At 6.45, Mary was writing a letter.
(20) At 7, Mary finished writing a letter.
(21) At 7.15, Mary had written a letter / At 7.15, a letter was written.
(22) Mary wrote a letter from 6.30 to 7.

A common definition of aspect is that it marks the viewpoint or the perspective from which a speaker looks at an event. In the seminal work by Comrie, aspect is defined as ‘different ways of viewing the internal temporal constituency of a situation’ (1976: 3). One of the major aspectual distinctions is between perfective and imperfective aspect. When speakers use perfective aspect, as in the above example (22), they look ‘at a situation from outside, without necessarily distinguishing any of the internal structure of the situation’, ‘the whole of the situation is presented as a single unanalysable whole, with beginning, middle, and end rolled into one’. When speakers use imperfective aspect, as in (19), they look ‘at the situation from inside’ and make explicit reference to the internal temporal constituency of the situation, to ‘the various individual phases that make up the action’ (1976: 3-4).

The standard approach to aspect in FG is to distinguish different areas of aspectuality (Dik 1997a: 221-22): the first type of aspectuality concerns the distinction between perfective and imperfective aspect, similar to Comrie’s view. The second area is phasal aspectuality, which specifies the phase of development of the event in terms of beginning, continuation or end of an event (see (18)-(20)). The third area concerns perspectival aspectuality that relates the occurrence of the event to an outside temporal reference point. This includes the categories prospective (17) and perfect (21). Finally, Dik speaks of quantificational aspect that gives information about the frequency of an event. This last category is not considered aspect in this thesis, but it is treated under the notion of quantification in section 3.5.

In my view, the definition of aspect in FG is not entirely adequate. This results in considerable disagreement on the scope of different aspectual areas (Anstey 2002: 3). In order to arrive at a more decisive analysis, it is important to reconsider the semantics of different aspectual categories. Therefore, in the next section, other approaches to aspect will be discussed.

---

3.3.2 Defining aspect

In the early nineties, Smith (1991) proposed an analysis of aspect that is compatible with Comrie’s view, but nonetheless crucially different in formulation:

The aspectual viewpoint of a sentence functions as an independent lens on the situation talked about. Viewpoint makes visible all or part of a situation, without obscuring the conceptual properties of the situation types. (C. S. Smith 1991: 171).

Smith considers aspect as selecting a part of the event or the whole event, whereas Comrie uses the metaphor of looking at the internal temporal structure in a different way. In a similar vein as Smith, though logically better underpinned, Bohnemeyer (1998) proposes to treat different aspectual categories as one unified category of ‘boundary operators’ that select a specific part of the temporal structure of an event that is relevant to the conversation.6 What this means, is illustrated in Figure 3-2. The lower line in Figure 3-2 represents a time line with times related to examples (17)-(22). The upper figure represents the maximal referential projection of an event. This includes all parts of the temporal structure that can potentially be referred to by a linguistic expression, including a potential pre- and post-state. The dotted ovals indicate which part of the temporal structure is selected, representing the examples in (17)-(22). Bohnemeyer distinguishes six notional boundary operators: language-specific categories do not necessarily match these selections, but may select larger or smaller parts.

In each sentence, the speaker concentrates on a different part of the temporal structure. In (17), the speaker focuses on the ‘pre-state’ of the process of writing a letter: this is prospective aspect. In (18), the starting point of the process of letter-writing is focused on, while in (20) the endpoint is highlighted. These are ingressive and egressive aspect, respectively. In (19) the speaker selects a time interval in the middle of the temporal structure, without taking the initial or terminal boundary into consideration. This is both imperfective and progressive aspect. In (21) the speaker concentrates on the post-state of the

---

6 This idea is inspired among others by Chung & Timberlake (1985), Klein (1994), Moens (1987) and Smith (1991).
letter-writing, which is the counterpart of (17). This is perfect (had written) or resultative (was written) aspect. Finally, in (22) the speaker describes the complete event, including the beginning and endpoint, which is perfective aspect.

What Bohnemeyer means by pre- and post-states is different from pre-and post-times:

Pre-and post-times are time intervals that precede and follow the event, respectively. By contrast, pre- and post-states are related to the event in some non-temporal way in addition to temporal sequentiality. (...) Post-states reflect some kind of trace that the occurrence of the target event leaves behind in the world. (...) the target event may be the consequence or purpose of the pre-state, or an agent may be said to be in an intentional pre-state with respect to the target event. (...) I assume that pre- and post-states of an event E are events which do not form part of E, but encompass E in a natural 'causal chain'. (1998: 76-77)

Causality is thus the essential distinction between prospective and perfect aspect on the one hand, and future and past tense on the other. Dik (1997a: 239) explains the difference of a prospective compared to a future tense as ‘a prediction about what is going to happen in the future on the basis of what information the speaker has now, whereas a future is a simple statement about what will happen in the future.’

Now what exactly are the semantic distinctions between (17)-(22)? All utterances concern the same relation, designated by the predicate write, with the same participants, designated by the arguments Mary and a letter. What is
variable is that different parts of the temporal structure of ‘write’ are ascribed to Mary. For example, in (17) Mary is not writing, she is going to write, in (20) she is not going to write, she stops writing. In other words, the speaker selects exactly the part of the temporal structure of the property or relation that is relevant, and it is only this specific part that is predicated of the arguments.

Bohnemeyer’s analysis unites different types of aspect that are distinguished in FG: in his account there is no distinction between the function of perspectival aspect (prospective, perfect), phasal aspect (progressive, egressive, ingressive) or (im)perfective aspect: in principle, all aspectual expressions serve the same function, which is selecting the part of the temporal structure of the property or relation that is ascribed to the arguments. In this thesis, Bohnemeyer’s proposal is adopted and aspect is defined as follows:

D1. **Aspect** markers select the relevant parts of the temporal structure of a property or relation, including the pre- or post-state. It is only this selected part of the temporal structure that is the predicated property or relation of the argument(s).

Aspect markers modify the property of relation that is ascribed to the argument(s) and contribute to a proper description of the event; at the same time, they leave the description of the argument(s) intact. This implies that aspect markers have scope over the predicate only and that they should be classified as \( \pi 1 \)-operators in FG.

### 3.3.3 Categories in detail

Now that the semantics of aspect are broadly defined, a closer look is taken at the specific semantic functions within the domain of aspect, at their relation with the temporal structure of properties and at relations and possible combinations of aspect markers. The specific functions will be described starting from the maximal temporal structure of a property or relation. Notice that the extension of a property or relation in time does not mean that it also has location in time. As shown above, it is not always the maximal temporal structure that is predicated of the argument(s); it is often only part of the temporal structure that the speaker wants to assert. The selection of part of the temporal structure neither modifies the meaning of the arguments nor of the additional participants; it only modifies what is predicated of the argument(s). As a consequence, this has repercussions on the description of the set of possible events, since only the selected part of the temporal structure of the property or relation, only the modified predicate helps to build up the event description.
Although different properties or relations behave differently a general temporal structure is presented in Figure 3-3:

![Diagram](image)

**Figure 3-3.** Representation of maximal temporal structure of an unmodified relation or property: PRED (x)

The time span represented by the horizontal line represents a homogeneous or a heterogeneous interval, with or without a state change within the boundaries. The length of the time span is not exactly specified in the temporal structure of the unmodified property or relation. The initial and terminal boundaries, represented by the left and right vertical lines can represent an arbitrary beginning and endpoint but they can also represent the point of beginning and completion in case the predicate is used in the description of a telic event. However, whether the temporal boundaries are points of ending or of completion is determined by the combination of the predicate, the arguments and the additional participants. In case of states, there are no boundaries if the state is permanent, but there are boundaries if the state is temporary. When aspectual expressions are used, the temporal structure is modified: part of the temporal structure is selected or part of the structure is added and then selected as in the case of pre- and post-states. Only this selected part of the temporal structure is ascribed to the argument(s), and nothing more.

Common aspectual distinctions will be illustrated below, in Figure 3-4 to Figure 3-10: in these representations, the maximal temporal structure is represented by the dotted line. The dotted oval indicates the part of the temporal structure that is selected and ascribed to the argument(s). In addition to the selected part, some other parts of the temporal structure may be implied. Contrary to what I claimed in Boland (2005), the implied part is the interval earlier in the temporal structure. This can be inferred by general knowledge of temporal structures. For example, if an event is going on, it must have begun. However, it is not necessarily the case that anything earlier in time is ascribed to the participants: the initial boundary is not made explicit by the progressive (Figure 3-6), whereas it is made explicit by the continuative (Figure 3-7). The

---

7 The representations are inspired by Figures 4 and 34-38 in Bohnemeyer (1998).
complete part of the temporal structure that is implied is represented through the solid line, which is thus not necessarily equal to the part of the structure that is selected and ascribed to the argument(s), indicated by the dotted oval.

I will discuss aspectual categories in two sets: those categories that do not and those categories that do imply the complete temporal structure. Categories that do not imply the complete temporal structure are the prospective, the ingressive (inchoative), the progressive, the continuative and the imperfective. First, the prospective will be discussed. If a speaker uses a prospective, he adds a pre-state to the temporal structure. Only this pre-state of the event is predicated of the argument(s). So, the argument is not in the state of, for example, reading, but only in the state of going to read. See Figure 3-4:

![Figure 3-4. Representation of prospective: ‘going to PRED’](image)

When a participant is in the pre-state of an event, it is probable, but not necessary that the event will in fact also take place. A prospective may thus lead to the inference of the occurrence of the event, but this inference may be denied: I was going to read, but I couldn’t find my book and I ended up watching television.

The ingressive focuses on the initial boundary. In English, it can only be expressed lexically by start or begin, but in other languages there may be specific markers for ingressive. The ingressive is represented in Figure 3-5:

![Figure 3-5. Representation of ingressive: ‘start PRED’](image)

When an ingressive is combined with a state, it means that the state has begun and it may imply that the state still continues. In several languages, such as

---

8 The idea that the speaker adds a pre-state or a post-state to the temporal structure is based on Klein (subm.) even though Klein speaks of pre- and post-times. Bohnemeyer (1998) on the contrary supposes that the pre- and post-state are always part of the maximal temporal structure of a property or relation.
Russian, Ancient Greek, Mandarin Chinese and Spanish (Comrie 1976: 19), a perfective may be combined with a stative state of affairs, which yields the interpretation that the state begins, or, as an inference, that the state holds. When the initial boundary is selected, it implies in general that the event takes place for an unspecified duration. \textit{Jim started to read} implies that Jim read and if the starting point is just before speech time, it probably also implies that Jim is now in the process of reading. Note that when an ingressive is combined with a telic event, it does not imply that the event reaches its point of completion: \textit{Jim started to draw a circle} does not imply that Jim drew a circle.

By using a progressive the speaker selects a subinterval in between the boundaries to ascribe to the argument(s), without taking the initial or terminal boundary into account. Consider Figure 3-6:

![Figure 3-6. Representation of progressive: \textquote{PRED-ing}](image)

The starting point of the event is implied, but not ascribed to the participants. A progressive is most frequently used with dynamic and non-punctual events. However, depending on the particular language, there may be exceptions that raise specific interpretations. First, a progressive may be used with predicates that by default describe a permanent state, \textit{naughty} or \textit{a teacher}. In those cases, the state is described as if it were a temporary, deliberate activity, with boundaries in the temporal structure, such as \textit{being naughty} or \textit{being a teacher} (behaving naughtily or behaving like a teacher, rather than being a teacher by profession). Furthermore, a progressive may be combined with a predicate that in general describes properties or relations of very short duration, raising the interpretation that the property or relation has a longer duration, either because it is repeated (\textit{jumping, firing a gun}) or because the stage leading up to the state change is stretched (as in \textit{reaching the top}).

Closely related to the progressive is the continuative. It was argued that a progressive implies, but does not ascribe the initial boundary to the participants. When the initial boundary does form part of what is ascribed to the participants, we speak of continuative aspect. In English, this is expressed by the construction \textquote{keep-ing}. Consider Figure 3-7:
A continuative indicates that the argument(s) have started and still continue their activity.

Finally, imperfective aspect may cover the same meanings as progressive and continuative markers (see Figure 3-6 and 3-7). Imperfective aspect thus functions in the same way as phasal aspect, but makes less fine-grained selections. Speakers have to infer more about the intended focus.

The aspectual categories that target the complete temporal structure are the egressive, completive, perfect and perfective. The egressive and completive focus on the terminal boundary. They both select the terminal boundary, but a completive can only combine with telic events. It not only indicates that the event stopped, but also that it is completed. In English, egressive and completive aspect can only be expressed lexically by *stop, end, or finish*. Egressive and completive aspect are represented in Figure 3-8:

If a speaker uses a perfect, then a post-state is added to the temporal structure. It is exactly this post-state that is predicated of the argument(s). This view is compatible with Smith (1991: 148) who also claims that the post-state, selected by the perfect, should be considered the modified property ascribed to the argument:

Present perfect sentences *ascribe to their subjects a property* that results from their participation in the prior situation. If at some time Henry has laughed, danced, built a sandcastle, *the property of having done these things is ascribed of Henry.* (italics mine).

The perfect is the counterpart of the prospective and can be depicted as in Figure 3-9:
The post-state of an event starts immediately after the terminal boundary of an event, but it may stretch in time for an indeterminate period. The post-state may also be selected by a resultative ('was PRED-ed') in which the agentive argument has little or no relevance. Note that a perfect makes the complete temporal structure explicit. It is therefore possible to get the so-called continuative perfect reading as in *He has lived here for ten years* or *I have already run five miles*. Although the preceding event is implied, the focus is on the post-state, since the ‘having lived somewhere for ten years’ is only true in the post-state of this event (see Klein 1994: 113 for a similar analysis). In Bohnemeyer’s words:

> Pre-and post-state operators select semantic components of the causal chain in which the event is embedded, in the same manner in which the other boundary operators select parts of the event itself. (1998: 77)

Although there is a temporal relation between the pre- or post-state and the event (the event took place before the post-state or the event will take place after the pre-state), this temporal relation is only an implication and not part of the meaning. Only when this implication becomes an increasingly essential part of the meaning, as has happened to different extents in German and Dutch (cf. Boogaart 1999: 156), then it may grammaticalize into a tense marker.

Like imperfective, perfective aspect makes a less fine-grained selection than phasal aspect. A perfective expression selects the entire temporal structure including the initial and terminal boundary. This is depicted in Figure 3-10:

---

9 There is no real perfective marker in English, but the simple past is often interpreted as perfective past.
A perfective often covers the post-state too, but not necessarily. Moreover, it is important to note that perfective is not the same as completed, an assumption often incorrectly made, as argued by Comrie:

The perfective does indeed denote a complete situation, with beginning, middle, and end. The use of 'completed', however, puts too much emphasis on the termination of the situation, whereas the use of the perfective puts no more emphasis, necessarily, on the end of a situation than on any other part of the situation, rather all parts of the situation are presented as a single whole (Comrie 1976: 18).

Some of the aspectual categories may be combined. This is again a language-specific quality. In English, a perfect or prospective can have scope over a progressive. This is however no argument for positioning them at a different layer. The combination of a perfect with a progressive leads to a more refined selection of what is predicated of the argument, that is the argument is in a post-state of being engaged in an event, without the implication that this activity has ended. Consider Figure 3-11:

![Figure 3-11. Representation of perfect progressive: 'have been PRED-ing']()

Observe that a perfect progressive does not necessarily imply the terminal boundary. In contrast to John has written a letter, which implies that the letter is finished, in John has been writing a letter there is no implication that John has finished writing the letter. Therefore, the terminal boundary is not bold in Figure 3-11.

The opposite of the perfect progressive can also be expressed: John is going to be writing a letter. In this case, the pre-state of the activity of writing is ascribed to the argument(s), without making the boundaries of the temporal structure explicit. Consider Figure 3-12:

![Figure 3-12. Representation of prospective progressive: ‘going to be PRED-ing’](

As with a 'simple' prospective, the actual occurrence of the property or relation is not implied, only expected. The distinction between both categories is that
the expectation of the prospective progressive is confined to the interval in between the boundaries, whereas a normal prospective concerns the complete temporal structure. Finally, it is possible to combine a perfect or prospective with an ingressive or an egressive. Consider Figure 3-13 for the representation of prospective ingressive:

![Figure 3-13. Representation of prospective ingressive: 'going to start PRED']()

Perfect and prospective aspect can have other aspectual expressions in their scope, but this is not considered an argument for regarding these expressions as different types of operators. The communicative function of the complex aspectual expressions is equal to 'simple' aspectual expressions, that is, they select the part of the temporal structure that is ascribed to the argument(s). The combination of aspectual expressions results in a more complex selection of the temporal structure, with different parts of the structure taken into account. However, like with simple aspect, complex aspect does not change anything about the description of the arguments nor does it situate the event in the real world; the combination of expressions functions as a π1-operator.

### 3.3.4 Summary

Aspect was defined as a grammatical means to select a specific part of the temporal structure of a property or relation to be ascribed to the argument(s). This definition differs from the standard approach to aspect in FG in that it assumes a similar function for all possible aspect markers. Selecting the pre-state or post-state of a property or relation in principle does not differ from selecting one of the boundaries, a subinterval or the complete temporal structure. All categories of aspect have the same communicative function: they select part of the temporal structure of the property or relation that is designated by the predicate. Therefore, they all function as π1-operators. An aspect marker modifies what is predicated of the participants. It restricts or defines the part of the temporal extension that is ascribed to the argument(s). The fact that certain aspectual expressions may have scope over each other is no reason for locating them at different layers of the semantic representation. The combinations of aspect markers still function as a selection of the temporal structure, albeit a more complex selection.
It must be stressed that it is language-specific which aspectual categories are marked by grammatical means. In English for instance lexical verbs such as start or finish are needed to describe an initial or terminal boundary of an event; in fact, in English these verbs describe a state of affairs by themselves, the state of affairs of starting or finishing something. Note that simple tenses may imply, but do not encode a perfective or imperfective viewpoint. I will elaborate on this matter in the next section.

3.4 TENSE

3.4.1 Introduction
The second semantic domain to be discussed is the domain of tense. The most common temporal distinction is between past and non-past, where non-past may be split up into present and future. A less common distinction is between future and non-future. For the discussion of tense, consider the situation in which Mary is writing. Depending on when this situation takes place, the speaker could describe the situation in different ways. If the speaker does not make a clear selection of the event, the situation could be described as in (23)-(25). If the speaker wishes to focus for example on the post-state of the event, the situation could be described as in (26)-(28):

(23) Mary wrote a letter.
(24) Mary is writing a letter.
(25) Mary will write a letter.

(26) Mary had written a letter.
(27) Mary has written a letter.
(28) Mary will have written a letter.

In the classical analysis of tense by Reichenbach (1947), tense denotes a three-way relation between the time of speech (ST)\(^{10}\), the time of the event (ET) and the time of a reference point (RT). The time of speech is the time at which the utterance is spoken; the time of the event is the temporal location of the event and the point of reference is indicated either by an adverb or contextually. In Reichenbach’s analysis, past, present and future tense (23)-(25) indicate that ET is at RT and RT is located relative to ST, before, at or after ST, respectively. Thus, for the past tense the relation would be ET = RT and RT < ST; for the

\(^{10}\) Comrie and Reichenbach use S, R and E, but here ST, RT and ET are used.
present tense ST = RT = ET; and for the future tense ST < RT and RT = ET. In the case of a perfect, ET is located before RT. In the pluperfect RT precedes ST (26), that is; ET < RT and RT < ST. In the present perfect (27), RT overlaps ST, that is; ET < RT and RT = ST. In the future perfect (28), ST < RT and ET < RT, but there are three possible resultant interpretations: ET < ST < RT; ST < ET < RT, or; ST = ET and ET < RT.

Reichenbach’s approach was later modified by Comrie (1981; 1985). He criticized the idea that every tense marker involves a reference point by making a strict distinction between absolute and relative tense marking. Comrie argued that absolute tense, such as the simple past, present and future in English, marks the temporal relation between ET and ST directly; there is no RT involved. The past, present and future are thus analyzed as ET before ST, ET simultaneous to ST, and ET after ST, respectively. Relative tense, on the contrary, marks the relation between ET and a reference point RT: ET can be before, simultaneous to or after RT. The reference point (RT) is somehow given in the (linguistic) context, ‘the range of potential reference points being in principle all those compatible with the given context’ (Comrie 1985: 58). The pluperfect and the future perfect combine absolute and relative tense marking. In contrast to Reichenbach’s analysis, Comrie assumes that in those cases, there are two binary relations, between ET and RT and between RT and ST, rather than one three-way relation between ET, RT and ST. The future perfect for example is composed of ET before RT and of RT after ST. The relation between ST and ET is not established and therefore, different readings are allowed: ET may take place before, at or after ST. Comrie makes a distinction between the basic meaning of a tense category and its implicatures. The basic meaning of the past is ‘location in time prior to the present moment’ (1985: 41). There is often a conversational implicature that the situation does not continue up to or beyond the present, but that is not part of the basic meaning.

A further difference between Comrie and Reichenbach is that the former considers present perfect as essentially an aspectual category rather than a (relative) tense category, expressing ‘current relevance’ (Comrie 1976, 1981), as opposed to the pluperfect and future perfect, that are combinations of absolute and relative tense. It is a language-specific characteristic which temporal distinctions a language marks grammatically. The most commonly marked opposition is between past and non-past. Some languages divide the non-past into present and future.

The standard account of tense in the FG model closely resembles Comrie’s point of view: ‘Tense operators serve to locate the SoA [state of affairs] on the time axis in relation to some t,’ (Dik 1997a: 237). The moment t can represent the moment of speech (absolute tense) or another moment in time (relative
tense). This view needs some modification, however, as it is not completely satisfactory. This will be discussed in the next section.

3.4.2 Topic time

Although Comrie’s view on tense is widely used, Klein (1994: 22-23) has shown that it does not account for some common phenomena. The first problem with Comrie’s view is that a past tense marker is used for a state that still holds at the moment of speech; in *It was a beautiful book, but it was in Japanese*, there are two states marked for past tense, whereas the real world states (being a beautiful book and being in Japanese) are still true for the present time. Why then is it correct to use a past tense in these contexts? The opposite phenomenon also occurs: a future marker is used for a state that already holds at the moment of speech. It is perfectly fine to ask *will you be here at eight?* and to answer *yes, I will be here*, whereas in both cases, the state of being here already holds in the present. It can, therefore, not be maintained that tense markers locate the event in time, since the events in the above examples also hold in the present, whereas they are not encoded for present tense. Klein solves this problem by the notion of **topic time**, that is ‘the time span to which the speaker’s claim on this occasion is confined’ (1994: 4). In his view, tense does not locate the **event** on the time line, but rather the topic time. The topic time interval is located on the temporal axis by tense markers. In Klein’s view ‘tense does not directly specify the “time of the situation”; rather, it imposes a temporal constraint on the time for which the assertion is made.’ (1994: xiii). This implies that the speaker does not mark the temporal location of the complete situation or event (ET in Comrie’s approach); he merely asserts the time location for a selected part of the situation, the part he has focused on. In the earlier examples (26)-(28), only the post-state of the event of Mary’s writing a letter is located in time. Klein calls the time of the interval for which the assertion is made the topic time (from now on, TT), and the time of the real event the **Situation Time** (TSit). In his view absolute tense marks the relation between TT and the time of the utterance (TU) or speech time. The past tense indicates that TT < TU, which means that the speaker asserts something for a period of time (TT), which is located prior to TU. Accordingly, the present can be represented by TT overlapping TU; and the future by TT after TU.

Topic time also explains why aspectual expressions are used to select part of the temporal structure. In Klein’s view aspect serves to link the time of the event to the topic time interval (1994: 99). This is compatible with the definition of aspect presented in D1. The relevant part of the temporal structure that aspect selects is exactly the part of the structure that holds at topic time. An overview of Klein’s approach to relevant categories of tense and aspect is presented in Table 3-2.
Table 3-2. Temporal and aspectual relations according to Klein (1994)

<table>
<thead>
<tr>
<th>Tense categories</th>
<th>Aspect categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>past</td>
<td>TT before TU</td>
</tr>
<tr>
<td>present</td>
<td>TT overlaps TU</td>
</tr>
<tr>
<td>future</td>
<td>TT after TU</td>
</tr>
<tr>
<td></td>
<td>prospective</td>
</tr>
</tbody>
</table>

Notes. TT = topic time ; TU = Time of utterance ; TSit = Time of situation or event

Klein argues that relative tenses should be analyzed as prospective or perfect aspect in combination with tense. The pluperfect, then, is the combination of the selection of the post-state of a situation (TT), which is situated prior to TU, while the present perfect situates the post-state of a situation (TT) as overlapping TU.

3.4.3 Defining tense

I agree with Klein that tense does not necessarily locate the entire state of affairs in time. I also agree with Klein that there is a topic time, a time span about which the speaker is asserting something. I do, however, not agree with his view that tense locates the topic time interval on the temporal axis. In my view, topic time is not systematically marked linguistically. However, there are several cues that the language users can rely on to infer the topic time, such as tense markers, temporal adverbial phrases, world knowledge about the sequence of events, etcetera. This stance resembles the statement that Partee (1984) makes about reference times:

Reference times are not directly denoted by any part of the sentence; they are more like a part of the necessary context for interpreting tensed sentences (…), akin to the kind of locative frame of reference needed to interpret left and right and other locative expressions. And like the locative case, they are not bound to the actual context of the utterance but can be ‘constructed’ and shifted in the course of interpretation. (pp. 264-65)

The relation between the topic time interval and the temporal structure of a property or relation is crucial. Tense does locate the event in time, but only the part that is relevant to topic time. Speaker and addressee mutually understand what is the relevant time span about which the speaker asserts something. It is

11 This view is compatible with Dik’s remark (1997a: 50) that ‘the qualified [state of affairs] expressed in the core predication can be located in space and time by predication operators’ (emphasis mine).
because of the Conversational maxim of Quantity (Grice 1975)\footnote{12} that tense marking only concerns the part of the event that is relevant to topic time and not the complete event.

If a speaker selects a part of the temporal structure by aspectual operators, then this indicates that only that specific part of the temporal structure is relevant at topic time and ascribed to the argument(s). If no selection is made of the temporal structure, then the temporal structure of the property or relation determines what is the relevant part to topic time. In cases of a temporal structure with boundaries, it can be assumed that the complete state of affairs is relevant at topic time; in such cases the speaker wants to ascribe the entire temporal structure to the argument(s). In cases of a temporal structure without boundaries (properties conceived of as permanent states) only the part of the state that overlaps topic time is located in time. No explicit selection of the temporal structure is made but it is pragmatic inference that only a partial interval of the permanent state is located in time: locating the entire state of affairs is virtually impossible (or at least pragmatically very inconvenient) since it holds anterior to, simultaneous to and posterior to the topic time. Language users know that once a permanent state holds for a certain topic time, it will also hold for other topic times. So, only the part of the state that is relevant to the topic time needs to be located in time. In this view, the examples of Klein in which tense markers do not comply with the temporal location of the entire event (see 3.4.2), can still be accounted for. With respect to states, for example, this view explains why it is possible to use a past tense for a state that still holds at the moment of speaking. Consider (29):

(29) He didn’t come to the party. He was ill.

In the light of his not coming to the party in the past, it is only relevant to talk about the interval of the state of his being ill that held at the same time interval of his not coming to the party, even though he might still be ill.

So, Comrie’s view on tense can with slight modifications readily account for the problems that Klein has pointed out. I therefore propose to define the function of tense markers as follows:

D2. **Tense** markers locate the part of the event that is relevant to the discourse on the time axis in relation to a reference time interval, in general the speech time interval.

---

\footnote{12}{Make your contribution [to the conversation] as informative as required and do not make your contribution more informative than is required.}
In cases of absolute tense marking, the use of a present tense indicates that the relevant part of the event overlaps ST. Whenever the past tense is used, the relevant part of the event is anterior to ST. Whenever the future tense is used, the relevant part of the event is posterior to ST. ‘Relative’ tense marking can in many cases be explained by the interference of perspectival aspect and tense, like Klein claims (section 3.4.2). However, there are languages that encode real anterior or posterior tenses, such as Japanese (Bohnemeyer, personal communication). In those cases the relevant part of the event is not located with respect to the speech time interval, but with respect to another reference time interval. This other reference time interval is comparable to Comrie’s RT.

In addition to the location of the relevant part of the event in time, some languages specify the length of the period between the location of the event and the reference time interval. These languages have different encodings for events that are situated relatively close to or relatively far away from RT or ST. This leads to categories like remote past, recent past, immediate future or remote future. Some languages have a different encoding of what happened today (hodiernal) or before today (prehodiernal), or they distinguish between past events that took place today, past events that took place yesterday (hesternal) and past events that took place before yesterday (prehesternal) (Comrie 1985: 83-101).

3.4.4 Scope of tense

Where should grammatical tense expressions be located in the model of FG? Although I have modified the semantic definition of tense compared to the traditional FG view, its communicative function has remained the same. Tense contributes to the situating function of the utterance, that is, it situates the state of affairs in time. Tense expressions do not change the description of the event, such as aspect expressions do. They therefore belong to the class of predication-operators (π2), located at the second layer of the semantic representation.

FG thus assumes expressions of aspect and tense to modify different parts of a clause. This has already been noted by Jakobson (1971), Foley & Van Valin (1984, and subsequent work on RRG), and Bybee (1985). These authors all recognize that aspect is only concerned with the temporal structure of the action or property designated by the predicate, ‘without involving its participants and without reference to the speech act.’ (Jakobson 1971: 134). The participants are unaffected by a marker of aspect. Tense on the contrary is concerned with locating the relevant part of the event in time in its entirety, i.e., including the participants that are involved in that event.

Although RRG and FG agree that the clause can be divided into layers and that TMA expressions should be considered operators that modify a layer, the
exact layered structure is not identical in both theories. In RRG (Foley & Van Valin 1984: 208) no distinction is made between predication and proposition. A main point of difference with FG is that tense in RRG is considered a peripheral operator that modifies the whole clause (similar to the proposition in FG), whereas in FG it only modifies the predication.

3.4.5 Tense and aspect

What are the implications of the definition of tense for the relation between tense and aspect markers? Aspect markers select the part of the temporal structure of a property or relation that is relevant to topic time. Only this part helps building up the proper state of affairs description. Tense markers on the other hand locate the part of the state of affairs that is relevant to the conversation on the time axis in relation to the speech time interval. In the figures below, the relevant part of the event is indicated by a dotted oval, the topic time interval by "TT" and speech time by "ST". For example, in Mary has written a letter in Figure 3-14, the relevant part of the event is the post-state, selected by the perfect. The present tense has indicates that this post-state is overlapping the speech time interval (indicated by ST). Consider Figure 3-14:

A past tense marker indicates that the relevant part of the event, the post-state in case of perfect aspect, is located anterior to speech time. See Figure 3-15:

---

**Figure 3-14.** Representation of present tense and perfect: *Mary has written a letter*

**Figure 3-15.** Representation of past tense and perfect: *Mary had written a letter*
A future tense marker indicates that the relevant part of the state of affairs, the post-state in this case, is located posterior to speech time. There are three possible temporal interpretations for an utterance like *Mary will have written a letter.*\(^{13}\) Consider Figure 3-16 for the representations of these interpretations:

*Figure 3-16. Representations of future tense and perfect: Mary will have written a letter*

When a prospective, instead of a perfect, is used, it is the pre-state that is located on the time line with respect to ST, when an ingressive is used, it is the initial boundary that is located on the time line with respect to ST, etcetera.

There are close relationships between certain aspects, tenses and situation types, mainly due to the implicatures of aspectual expressions. First, when the pre-state of an event is selected (prospective aspect) or the initial boundary (ingressive), it may be inferred that the event will indeed take place (although

\(^{13}\) There is a fourth and even a fifth possible interpretation, in which *will* is not interpreted as a future tense marker, but as a prediction marker, in which case the post-state may overlap the moment of speech or even precede it. This will be discussed in Chapter 4.
this implicature may be denied). There is thus a close relationship between prospective or ingressive and with future tense. Second, when TT is in the present, and the predicate refers to a dynamic event, it is quite “natural” to select the middle part of the temporal structure, by using a progressive, continuative or imperfective aspect marker, as it happens very rarely that a complete event occurs exactly at the speech time interval. Cases where this might occur in English will be discussed in section 6.2.3.2. When the speaker wants to refer to a stative event at the present time, it will in most cases overlap S; when it is a permanent state, an imperfective marker might be used to select the relevant interval of the state for the conversation, and when it is a resultant state, following from a telic event, the speaker may use a perfect or resultative marker to select the post-state of the event. There are thus close relations between states, present tense and imperfective aspect, between dynamic events, present tense and imperfective or progressive aspect and between telic events, present tense and perfect or resultative aspect. Finally, the selection of the final boundary or the post-state by an egressive, perfective, perfect or resultative aspect may often imply that the event has taken place, or in other words, that it is located in the past tense. Telic events seem to be prime candidates for ‘perfective’ aspects and past tense: these operators imply that the inherent endpoint of such an event type is reached, which is a rather crucial piece of information. Similar claims are made by Chung and Timberlake (1985):

The different temporal locations of an event—past, present, and future—are inherently correlated with differences in mood and aspect. An event that will occur after the speech moment is non-actual and potential. Hence there is a correlation between future tense and non-actual potential mood and, by implication, between non-future tense and actual mood. An event that is ongoing at the speech moment has not been completed. Hence there is a correlation between present tense and incompletive (imperfective or progressive) aspect and, by implication, between past tense and completive (perfective or non-progressive) aspect. (p.206)

The relation between aspect, tense and situation type will be a returning topic in later chapters, especially Chapter 9.

3.4.6 Summary

Tense expressions contribute to relating the set of possible events to the concrete event the speaker has in mind. They locate the part of the event that is relevant to the discourse on the temporal axis in relation to some \( t \). Tense expressions leave the description of the state of affairs intact and do not change the internal dynamics of the state of affairs nor any of the other elements that build up the description of the state of affairs (the arguments, the property or relation, the additional participants). They therefore function as \( \pi_2 \)-operators in
the second layer of the semantic representation, which is the standard approach in FG.

3.5 QUANTIFICATION

3.5.1 Defining quantification

The third domain to be discussed is quantification. In FG, it is traditionally referred to as a specific category of aspectuality, but since quantification has a different function than aspect, I will refer to it as a separate domain.

There are two types of quantification. Both Klein (1994) and Anstey (2002) distinguish between expressions that indicate the frequency of a property or relation and expressions that indicate the frequency of an entire event. Anstey shows that some languages have expressions that mark the iteration or intensity of the predicate. For example, Biblical Hebrew employs phonological modifications of the verbal stem to indicate that a relation or property is intense or iterated. For example, הָעָבָר "break" may be modified to הָעָבָר, meaning 'break into pieces' (Anstey 2002: 6). I will refer to these types of functions as property quantification. This domain is defined in D3:

D3. Property quantification modifies the internal structure of the property or relation, ascribed to the arguments, by adding a sense of iteration or intensity.

Expressions of property quantification modify the temporal structure of the property or relation designated by the predicate. They thus function as π1-operators, with scope over the predicate only.

The second type of quantification is event quantification (Anstey 2002: 6) that indicates the frequency of an entire event, for example a habitual or frequentative. Only this type of quantification was recognized in the standard approach of FG. The definition presented here is a slight adaptation of Dik's definition (1997a: 236):

D4. Event quantification situates the event (in different ways) with respect to its frequency, without entering into the definition of the event itself.

Markers of event quantification function as π2-operators. They do not change the description of the event and contribute to the situating function of the utterance: there are several instances of an event that are relevant to the discourse. The speaker does not refer to one specific event in the real or
imaginary world, but to a sequence or group of events. In section 3.5.2 and 3.5.3 the different operators within the domain of quantification will be discussed in more detail.

### 3.5.2 Property quantification

Possible categories of property quantification ($\pi_1$) are:

- **Iterative**: the property or relation ascribed to the arguments denotes a repeated action (on one occasion),
- **Intensity**: the property or relation ascribed to the arguments denotes an intense action.
- **Semelfactive**: the property or relation ascribed to the arguments denotes a single action.

In English, these categories are not expressed by grammatical means. Lexical equivalents for the iterative meaning in English would be *jump (one jump)* vs. *jump repeatedly*, *hit* vs. *hit several times*. For the intense meaning, one may think of the distinction between *look around* and *look around thoroughly* or *read and read through and through*. For the semelfactive aspect, it is explicitly specified that some action is performed only once instead of the default iterative interpretation, such as *He bounced the ball (repeatedly)* vs. *He bounced the ball once*. Modifications of iterative and intensity markers on the structure of the property or relation may be represented as in Figure 3-17:

![Figure 3-17](image)

**Figure 3-17.** Representations of an unmodified (left), an iterative (middle) and an intense (right) property or relation: *jump, jump up and down, make big jumps*

Property quantification falls within the scope of aspect: the former modifies the property or relation and the latter selects the part of this modified property or relation that is relevant to topic time. They both specify what is predicated of the arguments, therefore, they both function as $\pi_1$-operators.
3.5.3 Event quantification

Possible operators of event quantification ($\pi_2$) are:

- Habitual: the same state of affairs recurs on different occasions, potentially due to a habitual propensity of the participants involved, (adapted from Dik 1997: 236)
- Repetitive: the same state of affairs occurs again on a different occasion,
- Frequentative: the same state of affairs occurs several times on different occasions,
- Distributive: the same state of affairs occurs several times, with different participants.

What is most crucial to this thesis is first, that the internal structure of the event is not changed by expressions of event quantification and second, that event quantification relates the description of the set of possible events to several concrete events the speaker has in mind. Consider Figure 3-18 for a general representation of event quantification:

![Figure 3-18. General representation of event quantification](image)

The scope relation between aspect and event quantification is opposite to the relation between aspect and property quantification: aspect falls within the scope of event quantification; the complete event may be described as occurring frequently, but also the initial or terminal boundary, the pre- or post-state or the interval in the middle: He frequently draws a still life, He frequently started to draw a still life, He was frequently drawing a still life, He has drawn a still life frequently, etcetera. Furthermore, event quantification falls within the scope of tense, ‘since for an event to occur at all, it must occur at some time’ (Anstey 2002: 20).

What is the relation between the set of events and topic time for a habitual? There are different proposals. In Klein’s analysis (1994: 47) a speaker chooses to speak about a series of topic times, and for all of these topic times, the event
holds. The representation of a habitual is presented in Figure 3-19, for an event with boundaries. The event is repeated. Note that the topic time may overlap the entire event, as in this representation, but it may also overlap only part of the event, which is the case when aspect is also expressed.

![Figure 3-19. Representation of a habitual following Klein’s analysis](image)

An alternative way to represent a habitual is provided by Bohnemeyer (1998: 338). He assumes that the topic time is an extended interval, which is included in the interval of a sequence of the target event. Consider Figure 3-20:

![Figure 3-20. Representation of a habitual following Bohnemeyer’s analysis](image)

In my view, Bohnemeyer’s account is the best way to represent a habitual. The topic time interval is mutually understood by the interlocutors and a habitual expression indicates that the event referred to consists of a repeated sequence of events. I assume that the same kind of representation could be used for a frequentative. In the representation of a distributive, the events should occur simultaneously or overlap partly, as they may occur at the same time, whereas in the representation of a repetitive, there are only two events involved, that occur sequentially. The second event overlaps topic time.

### 3.5.4 Summary

I made a distinction between property quantification (π1) and event quantification (π2). Expressions of property quantification specify that the property or relation ascribed to the argument(s) is constituted of a repeated or intense action. They function as π1-operators. Expressions of event quantification specify that the speaker does not have one concrete event in
mind, but a set of concrete events. They thus contribute to situating the state of affairs and they function as $\pi_2$-operators.

Finally, the scope-relations between the different domains were discussed. Operators may have scope over another operator at the same layer since several operators may contribute to the different communicative functions of an utterance at the same time. Event quantification has scope over aspect and falls itself within the scope of tense. Property quantification falls within the scope of aspect. The outcome of this discussion is presented in (30):

\[(30) \quad [\pi_2 \text{ tense } [\pi_2 \text{ event quantification } [\pi_1 \text{ perspectival aspect } [\pi_1 \text{ (im)perfective / phasal aspect } [\pi_1 \text{ property quantification [Pred]]]]]]]]
\]

The scope relations between operators of the same type are not further relevant to this thesis.

3.6 Conclusion

In this chapter, I considered the semantics of grammatical expressions of aspect, tense and quantification. The semantics are crucial in deciding where the grammatical categories should be located in the underlying representation of the clause. They determine the communicative function of the expressions, whether they contribute to the description of the property or relation ($\pi_1$) or to situating the event ($\pi_2$). I proposed alternative definitions for aspect, tense and quantification for those presented in the standard FG-approach (Dik 1997a). Therefore, the notion of topic time was introduced; ‘the time span to which the speaker’s claim on this occasion is confined’ (Klein 1994: 4).

Firstly, the definition of aspect used in this thesis is: aspect selects the relevant part of the temporal structure of a property or relation, including the pre- or post-state (D1). The relevant part is the part that holds at topic time. The speaker only ascribes this selected part of the temporal structure of the property or relation to the argument(s). Aspect markers modify what is ascribed to the argument(s) and thus function as $\pi_1$- or predicate operators. It was also shown that there is no difference in the communicative function of perspectival aspect, phasal aspect and (im)perfective aspect. Although perspectival aspect may have scope over other aspectual markers, the resulting interpretation is still a selection of the temporal structure of the property or relation designated by the predicate. There is furthermore no difference between the communicative function of phasal aspect operators and (im)perfective operators: the latter express a less fine-grained binary opposition (either the temporal structure with or without the boundaries is selected) whereas the former can make more specific selections.
Secondly, the definition of tense was refined: tense locates the part of the event relevant to topic time on the time axis in relation to the speech time interval (D2). It does not change the internal structure of the event: tense does not modify the descriptions of the designated entities nor of the designated property or relation. It relates the description of the set of possible events temporally to the concrete event the speaker has in mind. Tense expressions thus contribute to situating the event and they function as $\pi_2$-operators.

Thirdly, a distinction was made between expressions of quantification that specify the frequency or intensity of a property or relation (property quantification) and operators that specify the frequency of the state of affairs (event quantification). Property quantification modifies the internal structure of the property or relation that is ascribed to the arguments, by making it iterated or intense (D3). Property quantification helps to build up a proper description of the event and belongs to the category of $\pi_1$-operators. Event quantification relates the description of the set of possible events to a set of concrete events in the real or imaginary world (D4). Operators of event quantification contribute to the situating function within an utterance and are predication operators ($\pi_2$): they do not change the description of the event, but situate the event in terms of frequency of occurrence. The resulting categorization is presented in Table 3-3.

The TMA domains in this table all belong to the representational level in FG. There is, however, a significant difference between the domains that function as $\pi_1$-operators and those that function as $\pi_2$-operators, because $\pi_2$-operators have scope over $\pi_1$-operators. In general, $\pi_1$-operators can be interpreted independently of contextual information, whereas $\pi_2$-operators can only be interpreted by pragmatic inferences about topic time, based on context, speech situation and world knowledge.

Table 3-3. Classification of aspect, tense and quantification according to scope

<table>
<thead>
<tr>
<th>TMA Domain</th>
<th>Scope over</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspect</td>
<td></td>
</tr>
<tr>
<td>prospective, ingressive,</td>
<td>Predicate ($\pi_1$)</td>
</tr>
<tr>
<td>continuative, progressive,</td>
<td></td>
</tr>
<tr>
<td>egressive/</td>
<td></td>
</tr>
<tr>
<td>completive, (im)perfective,</td>
<td></td>
</tr>
<tr>
<td>perfect</td>
<td></td>
</tr>
<tr>
<td>Property quantification</td>
<td></td>
</tr>
<tr>
<td>iterative, intensity,</td>
<td></td>
</tr>
<tr>
<td>semelfactive</td>
<td></td>
</tr>
<tr>
<td>Tense</td>
<td></td>
</tr>
<tr>
<td>past, present, future,</td>
<td>Predication ($\pi_2$)</td>
</tr>
<tr>
<td>non-past, non-future, ± specification of</td>
<td></td>
</tr>
<tr>
<td>temporal distance</td>
<td></td>
</tr>
<tr>
<td>Event quantification</td>
<td></td>
</tr>
<tr>
<td>habitual, frequentative,</td>
<td></td>
</tr>
<tr>
<td>repetitive, distributive</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 4
Modality, Evidentiality and Irrealis

4.1 INTRODUCTION
In Chapter 3 the domains of tense, aspect and quantification were discussed. The topic of this chapter will be the domains of modality, evidentiality and irrealis. There is a crucial distinction between the former and the latter domains. Chafe (1995) states that there are:

various paths by which ideas enter consciousness, among them the direct perception of current states and events, the remembering of previously experienced states and events, and the imagining of states and events that are judged not to accord with current objective reality. (p. 349)

Grammatical expressions of tense, aspect and quantification are mainly used for communication about perceptible and remembered events, whereas grammatical expressions of modality mainly concern the communication about imagined states and events. In section 4.2, I will discuss the major subcategories within the domain of modality. I will argue that three parameters are needed, sense, scope and source, to define different categories of modality and I will show that this leads to a sound classification. In section 4.2.5 the areas related to modality, future tense, prospective aspect, irrealis and evidentiality, will be discussed.

4.2 MODALITY
Distinctions in the domain of modality are illustrated in (1)-(4), again concerning the event of Mary writing a letter:

(1) Mary has to write a letter.
(2) Mary could write a letter.
(3) Mary might write a letter.
(4) Mary will be writing a letter.

In (1), it is stated that Mary is somehow obliged to write a letter and in (2), that Mary is able or has the opportunity to write a letter; in (3) the speaker supposes
that there is a possibility that Mary will write a letter; finally, one of the interpretations of (4) is that the speaker predicts that Mary is writing a letter. The examples in (1)-(4) illustrate that the nature of modality can hardly be covered by one general definition; the meanings within the domain are very diverse. There is much controversy about how to define the domain of modality. A rather successful attempt to formulate the general semantics of modality is made by Perkins. He states that ‘human beings often think and behave as though things might be, or might have been, other than they actually are, or were.’ By marking modality, a speaker speaks ‘in terms of ‘things being otherwise” and conceives of something ‘being true or real in some non-actual world, or true or real in some state of the actual world at a point in time other than the present moment’ (Perkins 1983: 6-7).

Another debate concerns the question which meanings should be included in the domain of modality and which ones should not. In this thesis I follow the standard approach in FG that sentence types or basic illocutions (such as declarative, interrogative and imperative) and different moods, such as subjunctive or indicative, are excluded from the domain of modality, since they do not belong to the representational level, but to the interpersonal level (see 2.2). On the other hand, modal expressions that concern characteristics of participants, such as ability or volition, are included in the domain of modality.

Although there is no general agreement on the boundaries of the domain of modality, some modal distinctions are generally acknowledged. First, it is often claimed that modality expresses notions of either necessity or possibility: this opposition will be further elaborated as the parameter sense. Second, a general division in the domain of modality is between epistemic, deontic and dynamic or root meanings: general definitions are that epistemic meanings have to do with the knowledge of the speaker about the truth of the propositional content, as in (3) and (4), deontic meanings have to do with notions of permission and obligation, as in (1) and dynamic meanings with internal characteristics of the participants, as in (2). These distinctions will be refined as the parameter source. Finally, I will add a third parameter, scope, which was defined in 2.1. This notion is much less commonly recognized, although notions like objective, subjective and agent-oriented modality are related to scope. I will show that together with sense and source, scope may account for the specific modal meanings and the close relationships that exist between them.

4.2.1 Sense: potentiality, disposition, weak necessity and necessity

Modal expressions are often used with different meanings. A particular phenomenon is that in very diverse languages, such as Indo-European, Semitic, Philippine, Dravidian, Mayan, and Finno-Ugric, single modal expressions cover
MODALITY, EVIDENTIALITY AND IRREALIS

the same meanings (cf. Steele et al. 1981; Sweetser 1990: 49). For example, many languages use a single form for expressing obligation (5) and certainty of the speaker (6), like in English:

(5) You must go now.
(6) You must be joking.

In (5) the participant you is obliged to go, whereas in (6) the speaker is certain that the addressee is joking. The fact that in both cases must is used indicates that there is probably a close semantic relationship between the notions of obligation (or requirement) and certainty. A similar relation seems to exist between permission and possibility. Palmer (2001) rightly asks why both connections exist.

There is no immediately obvious reason why the same forms should be used for expressing the speaker’s degree of commitment to truth and for getting other people to do things. It is by no means obvious that permission is a related notion to possibility, or requirement to certainty. (p.98)

A plausible view is that there is one general meaning that is involved in the specific meanings of possibility and permission and that there is one general meaning that is involved in the meanings of obligation and certainty. Most often it is assumed that permission has something to do with possibility—which I will refer to as potentiality—and that obligation and certainty both have to do with necessity. The opposition between potentiality and necessity derives from modal logic (Lyons 1977: 787; Von Wright 1951), but it would be mistaken to speak of a binary opposition between the two. In language there is rather a gradient scale between potentiality and necessity. Perkins (1983) adds a third distinction of disposition, which means ‘with a tendency to …’. Bybee, Perkins & Pagliuca (1994) make a further distinction between necessity and weak necessity. This leads to a gradual scale of senses from potentiality to necessity, which is represented in Figure 4-1:

| potentiality | disposition | weak necessity | necessity |

Figure 4-1. Scale of senses

The degree of modal strength1 can run from potentiality on the one extreme via disposition and weak necessity to necessity on the other extreme. I assume that the scale of senses represents a universal conceptual continuum: in how many

1 A term borrowed from Harder (1998).
parts the scale is divided is a language-specific characteristic. There may be less or there may be more than four distinctions. English uses different expressions for the four general senses given in Figure 4-1. Potentiality is expressed by can, could, may and might, and the meaning of these expressions can be paraphrased as ‘not precluding’. Disposition is expressed by will, would and wanna and the meaning of these expressions can be paraphrased as ‘disposed towards’. Weak necessity is expressed in English by shall, should, ought to and be supposed to, which can be paraphrased by ‘expected to’. Finally, necessity is expressed by need to, have (got) to, got to and must. The general sense of expressions that belong to this domain could be paraphrased as ‘forced to’ or ‘entailing’. The expressions for one basic sense differ in formality, politeness, and possible contexts of use. As a matter of fact, the basic sense of a modal expression is only part of the resultant meaning. The basic sense of a modal marker in combination with the two other parameters, source and scope, leads to more specific meanings. First, the source of modality will be addressed.

4.2.2 Source: internal, external and epistemic

The second parameter that is needed to describe the semantics of modality is the source of the modality. This is a refinement of the distinction between epistemic, deontic and root or dynamic\(^3\) modality (Bybee 1985; Bybee et al. 1994; Coates 1983; Lyons 1977; Palmer 1986; Perkins 1983; Sweetser 1982). Epistemic modality is ‘concerned with matters of knowledge, belief’ (Lyons 1977: 793), with ‘opinion rather than fact’ (p.681-82), whereas deontic modality ‘is concerned with the necessity or possibility of acts performed by morally responsible agents’ (p.823). Obligation (must, have to) and permission (may, can) form the major types of deontic modality. The term ‘root modality’ is used by Hofmann (1976), Sweetser (1990) and Coates (1983) to refer to all types of non-epistemic modality, but I will reserve this term for modality which has its source in participant-external, but non-deontic circumstances (see later this section) like in Bybee et al. (1994: 320) and Van der Auwera & Plungian (2001: 84). Another term that is sometimes used for non-deontic and non-epistemic modality is dynamic modality. This term will not be further used in this thesis.

Several authors, among whom Sweetser (1990) and Perkins (1983), have already connected sense and source, although they used different labels for it. Sweetser has formulated the senses of specific modal expressions in English and assumes that they apply to different ‘worlds’, a sociophysical and a world of

---

\(^2\) The paraphrases for potentiality, disposition and necessity (entailed to) are adopted from Perkins (1983), who used them for specific modal expressions. The paraphrase forced to is adopted from Sweetser (1990).

\(^3\) A term of Von Wright (1951: 28).
reasoning. Sweetser formulates the basic senses for each modal auxiliary in terms of Talmy’s force dynamics (1988a): May for example indicates the force of an absent potential barrier. This absent potential barrier may apply to the concrete, external sociophysical world, in which case it is interpreted deontically or dynamically: the participant x is not barred from doing something. By metaphorical extension, it may apply to the internal, mental world of reasoning, in which case it is interpreted epistemically: ‘the speaker is not barred by the available premises from the conclusion that …’ (Sweetser 1990: 61). Similarly, the deontic use of must is paraphrased as follows: ‘the direct force of y’s authority compels x to do something’, or in its epistemic use: ‘the available (direct) evidence compels S to the conclusion that …’ (1990: 61).

Perkins (1983) too describes the semantics of the English modals by a basic meaning such as ‘entailing X’ for must, ‘not precluding X’ for can and may and ‘disposed towards X’ for will and shall. Again, these basic meanings may apply to different worlds, that Perkins describes in terms of different sets of laws or principles to which the truth or actuality of propositions or events is relativized. He distinguishes three types of modality: the first type, epistemic modality is concerned with rational laws of inference and deduction. It relates to the ‘interpretation of the world via the laws of human reason’ (p.10). The second type, deontic modality, is defined in terms of social or institutional laws. These laws can be either installed by a legal authority or institution or they can relate to the less formal rules of social status, to the authority one person has over another. The third type, dynamic modality, relates to the set of natural laws. This type of modality is concerned with the relationship between empirical circumstances or states of affairs and non-actualized events, which are the result from natural laws of e.g. physics, chemistry, biology, etcetera (p.11).

The (...) rational laws (or the laws of reason), social laws (or the laws of society), and natural laws (or the laws of nature) define three different types of possible worlds in which the truth/actuality of propositions/events may be assessed (...). (Perkins 1983: 12)

Although the metaphorical descriptions of epistemic, deontic and root modality as different worlds or different sets of laws are appealing, a more lucid, less symbolic approach is presented by Van der Auwera & Plungian (1998: 80-81), who state that modality is ascribed to different sources, namely, participant-internal characteristics, participant-external circumstances, either deontic or non-deontic, and knowledge. This approach links the semantic distinctions in a consistent and plausible way to one another and it will serve as the basis for the

---

4 Talmy systematically describes meanings in terms of entities that exert forces on each other towards rest or towards action. Modal auxiliaries in Sweetser’s terms can be understood as expressing forces.
parameter of source in this thesis. With some slight adaptations, the different sources are defined as follows:

- **Participant-internal** modality ascribes the source of the modality to characteristics internal to the participant.
- **Participant-external** modality ascribes the source of the modality to circumstances external to the participant.
  - **Deontic** modality identifies the external circumstances as a deontic source.
  - **Non-deontic** modality identifies the external circumstances as something other than a deontic source.
- **Epistemic** modality ascribes the source of the modality to knowledge.

The combinations of basic senses with source lead to different meanings. This is illustrated for English, for the domain of potentiality expressed by *can*. Each example ascribes the potentiality to a different source. The reference of the example is given between brackets, C for Coates (1983) and P for Palmer (1986), followed by the page number.

(7) I can only type very slowly as I am quite a beginner. (C92)
(8) Can I pinch a ciggie? – Course you can. Would you like a menthol or a plain? (P71)
(9) Yes, we can send you a map, if you wish. (P86)
(10) Can they be on holiday? (P62)

In (7), the source of the potentiality is ascribed to characteristics internal to the participant, i.e. the (limited) typing skills. This raises an interpretation of ability. In (8) the source of the potentiality is ascribed to circumstances external to the participant, in this case the deontic source of the addressee and later the speaker. This raises an interpretation of permission. In (9) the source of the potentiality is ascribed to circumstances (mainly) external to the participant: there are maps available, there is a postal service, and the speaker is physically and mentally able to send a map. This raises an interpretation called root-possibility. Finally, in (10), the source of the potentiality is ascribed to knowledge, in this case about potential events. This raises the interpretation of epistemic possibility.

Different sources can be distinguished by using different paraphrases. For example, the root-possibility interpretation is preferably paraphrased by ‘have the opportunity’, whereas the epistemic possibility interpretation is preferably paraphrased by ‘it could occur that’ or ‘it could be the case that’. Compare (11)
and (12), in which the exclamation marks indicate the preferred interpretation and the question marks the less plausible interpretation.

(11) Yes, we can send you a map, if you wish.
   a. !We have the opportunity to send you a map.
   b. ?It could occur that we send you a map.
   → external non-deontic source, root-possibility

(12) Can they be on holiday?
   a. ?Do they have the opportunity to be on holiday?
   b. !Could it be the case that they are on holiday?
   → epistemic source, epistemic possibility

Similar distinctions hold for the other basic senses. The possible combinations between sense and source are presented in Table 4-1, with some provisional labels. When the source of the modality lies in internal characteristics of the participant, the interpretations that arise are ability, volition (will, 'want to'), weak internal need and internal need. (Weak) internal need means that the participant is forced to do something because of inherent properties, as in I need nine hours sleep every day. In combination with external non-deontic sources, the root-interpretations arise. With external deontic sources, the meanings for each sense are permission, desirability, weak obligation and obligation. Whether each possible combination is in fact attested in languages will be discussed later.

Table 4-1. Combinations of sense and source

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>Potentiality</th>
<th>Disposition</th>
<th>Weak necessity</th>
<th>Necessity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal</td>
<td>ability</td>
<td>volition</td>
<td>weak internal</td>
<td>internal need</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>need</td>
<td></td>
</tr>
<tr>
<td>External</td>
<td>root-</td>
<td>root-disposition</td>
<td>weak root-necessity</td>
<td>root-necessity</td>
</tr>
<tr>
<td>Non-deontic</td>
<td>possibility</td>
<td>dispositional</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deontic</td>
<td>permission</td>
<td>desirability</td>
<td>weak obligation</td>
<td>obligation</td>
</tr>
<tr>
<td>Epistemic</td>
<td>epistemic</td>
<td>epistemic</td>
<td>epistemic probability</td>
<td>epistemic necessity</td>
</tr>
</tbody>
</table>
4.2.3 Scope: participant-, event- and proposition-oriented

Although many semantic distinctions can be accounted for by the parameters sense and source, a third parameter is needed to describe finer distinctions. This is the parameter of scope. The third parameter stems directly from FG, in which expressions of modality are classified according to their scope, that is, the part of the utterance that is modified by the modal expression. The scope of a grammatical TMA expression may be the predicate, the predication or the proposition (see Chapter 2) (Dik 1997a; Hengeveld 1988, 1989, 2004b). If the scope is the predicate, the description of the relation or property predicated of the argument(s) is modified. In FG, this is called participant-oriented or inherent modality. It defines ‘relations between a participant and the realization of the [state of affairs] in which he is involved.’ (Dik 1997a: 241). If the scope is the predication, the event is situated in the real or imaginary world. This is called event-oriented or objective modality. It expresses the likelihood or actuality of the event. Finally, when the scope is the proposition, the truth of the propositional content is evaluated. This is called proposition-oriented or subjective modality. The speaker expresses his personal attitude towards the content of the proposition. Table 4-2 presents an overview of the different types of modality, based on scope.

Although the notion of scope is not generally used in the literature on modality, some authors do use the term subjectivity (among others) in relation to the speaker’s attitude or degree of commitment towards the propositional content (Coates 1983; Langacker 1990; Lyons 1977; Traugott 1982, 1989, 1995, 1996). Furthermore, Foley & Van Valin (1984) distinguish modality types on the basis of scope. They divide modality in four categories, which they call: (i) modality: ability, obligation, intention; (ii) status: actuality, realis-irrealis; (iii) evidential and (iv) illocutionary force (1984: 224). The first category is similar to participant-oriented modality in FG. However, the categories (ii), (iii) and (iv) are all considered to have scope over the proposition, whereas in FG, event-oriented modality (ii) is assumed to operate at the predicational layer, proposition-oriented modality (iii) at the propositional layer and illocutionary force (iv) is located at the interpersonal and not at the representational level.

Table 4-2. Different types of modality according to scope

<table>
<thead>
<tr>
<th>Type of modality</th>
<th>Scope over</th>
<th>Specification of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant-oriented modality (π1)</td>
<td>Predicate</td>
<td>Realization of property/relation</td>
</tr>
<tr>
<td>Event-oriented modality (π2)</td>
<td>Predication</td>
<td>Actuality of event</td>
</tr>
<tr>
<td>Proposition-oriented modality (π3)</td>
<td>Proposition</td>
<td>S’s attitude towards the proposition</td>
</tr>
</tbody>
</table>
Bybee also distinguishes between different types of modality on the basis of relevance to the verb: ‘A category is relevant to the verb to the extent that the meaning of the category directly affects the lexical content of the verb stem’ (Bybee 1985: 15). She makes a rough distinction between two types of modality. The first type includes FG’s proposition-oriented modality and basic illocutions. According to Bybee it has the ‘whole proposition in its scope, and does not only modify the verb. Furthermore, since it expresses the speaker’s attitude, it does not have a direct effect on the situation described by the verb’ (1985: 22). Bybee further distinguishes agent-oriented modality (similar to participant-oriented modality in FG), but she does not indicate what the relevance of this category is. From the layered model of FG a more fine-grained distinction follows: participant-oriented modality is most relevant to the verb as it directly modifies the action or property designated by the predicate (or verb); event-oriented modality is less relevant to the predicate, since it modifies the situating of the event, designated by the whole predication, and finally, proposition-oriented modality is least relevant to the predicate since it modifies the presentation of the propositional content.

An important point to stress is that the oppositions in scope do not coincide with the oppositions in source or in sense. On this point, I deviate from Van der Auwera (2001: 240) who states that there is a one-to-one correspondence between source and scope of modality, i.e., participant-internal modality would

<table>
<thead>
<tr>
<th>SCOPE OVER</th>
<th>SENSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Potentiality</strong></td>
<td><strong>Disposition</strong></td>
</tr>
<tr>
<td>Predicate (π1)</td>
<td>x is not precluded from PRED-ing</td>
</tr>
<tr>
<td>Predication (π2)</td>
<td>e is not precluded from occurring</td>
</tr>
<tr>
<td>Proposition (π3)</td>
<td>S does not preclude p’s truth</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SCOPE OVER</th>
<th>SENSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Weak necessity</strong></td>
<td><strong>Necessity</strong></td>
</tr>
<tr>
<td>Predicate (π1)</td>
<td>x is expected to PRED</td>
</tr>
<tr>
<td>Predication (π2)</td>
<td>e is expected to occur</td>
</tr>
<tr>
<td>Proposition (π3)</td>
<td>S expects p to be true</td>
</tr>
</tbody>
</table>

*Notes: x = argument; PRED = predicate; e = event; S = speaker; p = proposition*
be participant-oriented ($\pi_1$), participant-external modality would be event-oriented ($\pi_2$) and epistemic modality would be proposition-oriented ($\pi_3$). I will, however, show that scope and source are different aspects of modal meanings and that both parameters are needed to describe the full array of modal distinctions. First, the meanings will be discussed that arise when scope and sense are combined. The basic senses may apply to the participant (x), the event (e) or to the propositional content (p). This results in different meanings, listed in Table 4-3.

The combination of scope and sense already results in twelve different meanings, but each of these meanings can be made more specific in language use as they also reveal the source of the modality. For example, there may be different sources that account for the fact that ‘x is expected to PRED’ or that ‘e is forced to occur’. Participant-oriented modality ($\pi_1$) that specifies the realization of the property or relation by the participants may have its source in participant-internal or participant-external factors. Event-oriented modality ($\pi_2$) that specifies the actuality of the event, may have its source in general participant-external circumstances, deontic circumstances, such as moral, legal, or social norms or in terms of knowledge about events. Finally, proposition-oriented modality ($\pi_3$) may have its source in the certainty of the speaker, or in his personal attitude about the content. The latter will be conceived of as a special case of deontic modality. Not every combination between source and scope is possible. The logically possible combinations are presented in Table 4-4.

If the source of the modality is participant-internal, then the scope can only be participant-oriented. But if the source is participant-external or epistemic, then different scopes are possible. This will be illustrated in the next sections.

Table 4-4. Possible combinations of source and scope

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>Scope over</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predicate</td>
<td>Predication</td>
<td>Proposition</td>
</tr>
<tr>
<td></td>
<td>($\pi_1$)</td>
<td>($\pi_2$)</td>
<td>($\pi_3$)</td>
</tr>
<tr>
<td>Participant-internal</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant-external</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-deontic</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Deontic</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Epistemic</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
4.2.3.1 External, non-deontic modality

Modal expressions that ascribe the source of the modality to participant-external, non-deontic circumstances can either have scope over the predicate (participant-oriented) or over the predication (event-oriented). In the former case, a specific participant is not precluded from, disposed to, expected to or forced to do something, because of external circumstances. In the latter case, however, it is not the specific participant, but the event that is not precluded, disposed, expected or forced to occur because of external factors. This difference will be shown with the basic sense of potentiality. If the scope is the predicate, then a specific participant is not precluded from doing something mainly because of extrinsic factors, and this raises the meaning of root possibility. Consider (13) and (14):

(13) Can you pick your own trousers up? (C94)
(14) I am afraid this is the bank’s final word. I tell you this so that you may make arrangements elsewhere if you are able to. (C141)

If the scope is the predication, then the event is not precluded from occurring, and the occurrence of the event is independent of the involved participants. These meanings arise when the participant is non-specific, as in (15) and (16), or not mentioned at all such as in passive constructions. See (17) and (18).

(15) I know the place. You can get all sorts of things here. (P84)
(16) National pressure groups cannot exist without full time staffs and a regular income. (P91)
(17) Salts can easily be separated from the solid residue by dissolving them. (C98)
(18) Well, I’ll see what can be done and give you a ring. (P84)

4.2.3.2 External, deontic modality

Modality expressions that ascribe the source of the modality to a deontic source can have scope over the predicate, the predication or the proposition. If the scope is the predicate, a specific participant is not precluded, disposed to, expected to or forced to do something because of an individual deontic authority. If the scope is the predication, an event is not precluded, disposed to, expected to or forced to occur because of a general deontic authority, such as a general rule or law. The occurrence of the event is independent on the involved participants. If the scope is the proposition, the speaker expresses his/her attitude towards the propositional content. These differences will be illustrated for the basic sense of necessity. Consider the distinction between participant-oriented (19) and event-oriented modality (20).
(19) You must tell me how to get to it. (C34)
(20) All students must obtain the consent of the Dean of the faculty concerned before entering for examinations. (C35)

In (19), a specific participant you is obliged to do something by an individual deontic authority, the speaker. In (20) the event of all students obtaining consent of the Dean before entering examinations is forced to occur because of a general deontic authority, the rules that hold at the faculties. The distinction between participant- and event-oriented deontic modality can be made clear by adding an adverb of location that indicates the location where the permission or obligation holds. If the deontic source is an individual, it is often impossible to add an adverb that specifies the location of appliance of the obligation or permission\(^5\), see (21) and (23), whereas this is possible if the deontic source is a general rule, see (22) and (24):

(21) Can I borrow your pen (*here)? \(\rightarrow\) participant-oriented (\(\pi_1\))
(22) Can I pay by credit card (*here)? \(\rightarrow\) event-oriented (\(\pi_2\))

(23) (*In this office) you mustn’t put words into my mouth, Mr. Williams. \(\rightarrow\) participant-oriented (\(\pi_1\))
(24) (In England) you must drive on the left side. \(\rightarrow\) event-oriented (\(\pi_2\))

The general rule or law can apply to specific participants as in (22). In most cases however the participants are non-specific ((20), (24) (25)) or not mentioned at all (26).

(25) In the library you can take a book out and keep it out for a whole year unless it is recalled. (P103)
(26) No vehicle may be left in the University grounds during vacation. (C132)

Finally, the scope of deontic modality can be the complete proposition. Consider (27):

(27) The government must act. It must make up its mind about priorities – offices or houses, housing estates or luxury buildings. (P106)

---

\(^5\) The location of the permission or obligation needs to be distinguished from the location of the property or relation. For example, you can park you car here (at this particular spot) does not yield an event-oriented reading since the adverb here modifies the action of parking, not the permission.
In (27), in contrast to (19) and (20), there is no actual obligation applied to the participants or the event. The speaker does not in fact lay his authority on the government, and there is no general rule that forces the event of the government acting to occur. With participant- and event-oriented deontic modality, obligation or permission belongs to the descriptive level just like is the case with participant-internal and participant-external modalities. In (19), the participant you is indeed obliged by the speaker to do something. It can be paraphrased by ‘x is obliged by the deontic authority S to do …’. The same holds for event-oriented deontic modality, where the general law or rule does indeed apply to the occurrence of the event. With proposition-oriented deontic modality on the other hand, the modal marker does not belong to the descriptive level: the speaker expresses that in his opinion, the proposition that the government acts and makes up his mind should be true. It is an imagined obligation, a thought or wish of the speaker, not a factual obligation that is ascribed to the participants or that holds for the event. The utterance in (27) cannot be paraphrased by ‘I oblige the government to …’, or ‘there is a general rule that obliges the government to …’ but rather with ‘It is my opinion that the government should be obliged to …’. The speaker thus talks about a mental construct.

Notice that it is not a difference of commitment that distinguishes proposition-oriented modality from the other scope types. Both in (19) and in (27) the speaker is personally committed to the obligation, since he or she is the deontic source, but with participant-oriented modality it is not the main purpose of the modal to express the speaker’s commitment, but rather to lay an obligation upon the addressee. The commitment of the speaker follows from conversational implicatures, not from using the modal. With proposition-oriented modality on the other hand, it is the main goal to express the speaker’s attitude or commitment. With event-oriented deontic modality, it is in general irrelevant whether the speaker is committed to what is stated by the general rule.

4.2.3.3 Epistemic modality

The final category of modality in which a distinction in scope exists is epistemic modality. Here, the source of the modality is knowledge. An epistemic expression can have either scope over the predication or over the proposition. If the scope is the predication, the speaker presents the modal information as an objective epistemic statement about the actuality of the event. This use could be paraphrased as ‘e is not precluded/ disposed/ expected/ forced to occur because of objective knowledge about events’. If the scope is the proposition, the speaker presents the modal information as a subjective epistemic statement about his personal commitment to the propositional...
content or the amount of responsibility he takes for the truth of the proposition. A paraphrase would be ‘S does not preclude / expects p’s truth’ or ‘S is disposed / forced to conclude p because of the available premises’. The distinction between objective epistemic statements about the actuality of an event and subjective epistemic statements that report the stance of the speaker is not always clear-cut and many authors have conflated both uses. Notable exceptions are Lyons (1977) and especially Papafragou who refers to ‘objective’ modality as alethic (or logical) modality and says that alethic modality concerns ‘mind-independent abstract entities’ (Papafragou 2000: 81, emphasis mine).

In distinguishing between event-oriented epistemic modality and proposition-oriented epistemic modality, it matters most how S presents his utterance in the given context, as mainly objective or mainly subjective. Both interpretations are in fact closely related: if S presents his utterance as objective—it can occur that e—it may imply that S does not preclude the truth of p. And if S presents his utterance as subjective—S does not preclude p’s truth—it probably implies that the event e is not objectively precluded from occurring. Although there are no clear tests to distinguish between the two interpretations, it is possible to reveal the communicative intention by using paraphrases and/or by adding adverbs. This will be illustrated for the basic sense of potentiality. First of all, with objective epistemic modality it is acceptable to add the (subjective) adverb certainly, whereas this is strange for subjective epistemic modality, since subjective certainty and subjective potentiality (uncertainty) express a contradictory stance (Hengeveld 1988). Secondly, utterances with objective epistemic modality are preferably paraphrased by: ‘there is a possibility that …’, or ‘it can occur that …’, whereas utterances with subjective epistemic modality are preferably paraphrased by: ‘I suppose that …, but I’m not sure’ or ‘I doubt whether …’. The application of the tests are illustrated in (28) and (29):

(28) The only snag is that it has been raining (much later than is usual) and I could get held up for anything up to a week. (C108)
   a. !I certainly could get held up….
   b. !There is a possibility that I’ll get held up for anything up to a week.
   c. ?I suppose that I’ll get held up …., but I’m not sure.
   → could is used OBJECTIVELY, event-oriented (π2)
January I suppose, there may be an interview round about January.

(a) there certainly may be an interview round about January.
(b) there is a possibility that there will be an interview round about January.
(c) I suppose that there will be an interview round about January but I'm not sure.

→ *may* is used SUBJECTIVELY, proposition-oriented (π3)

In sum, there are two types of epistemic modality: epistemic modality that describes the actuality of the event and epistemic modality that indicates how certain the speaker is about the truth of the proposition. There are some remarks to be made about the expression of the actuality of events in combination with sense. When the basic sense is disposition, epistemic event-oriented modality expresses that e is disposed to occur (from an objective point of view). This can be considered a specific instantiation of the basic modal sense of disposition, but at the same time, the resultant meaning is closely related (or maybe even identical) to what a future tense marker expresses: it is presented as an objective fact that the event is disposed to occur. The relation between modality and future tense will be worked out in section 4.3.1.

Furthermore, within the sense of necessity, event-oriented epistemic necessity does not seem to exist, at least not without negation. Objective necessity denotes that ‘of the only possible e, which is e₁, e₁ is the case’. This could be regarded the default situation for every utterance. It seems communicatively unimportant to express that e is objectively necessary, since that is the normal, unmarked communicative situation. There is no need to specify that only e₁ is a possible actual event, unless the actuality status of e₁ is somehow special, because there are other possible e’s. There is, however, negated epistemic objective necessity which expresses ‘not necessarily e₁’, with the implication that ‘possibly e_{not1}’. It is semantically closely related to epistemic possibility. In English, it can be expressed by *need not* or *doesn’t have to* + negation. Examples are:

(30) Oh gosh, getting married is an awfully complicated business. [other speakers argue] Actually it *needn’t* be – it can be very straightforward.

(31) He *doesn’t have to* be at home: he could have gone straight to Caroline.

These utterances can be paraphrased as: ‘e does not necessarily occur’.
4.2.4 Classification of modality

Three parameters have been discussed on which the meaning of a modal expression may differ: sense, source and scope. All three parameters are necessary for the classification of modality as there are minimal differences between modal expressions on each of the parameters. This will be shown for English.

Firstly, the meaning of modal expressions can differ only on the parameter of sense. Consider the examples in (32) and (33):

(32) I can jump very high.
(33) I wanna jump very high.

Both sentences have scope over the predicate and the modal source of *can* and *wanna* is participant-internal. However, the sense is different: in (32) an expression is used (*can*) that indicates a sense of potentiality, whereas in (33) an expression is used (*wanna*) that indicates a sense of disposition. The paraphrase for (32) is: ‘x is not precluded from jumping because of intrinsic factors’, which raises the specific meaning of ability. In (33), x is disposed to jump, because of intrinsic factors, which raises the specific meaning of volition.

Secondly, the meaning of modal expressions may differ only on the parameter of source. Consider the examples in (34) and (35):

(34) Wow, you can jump five meters! Unbelievable.
(35) Everything is ready. You can jump now.

*Can* in (34) and in (35) has scope over the predicate and its general sense is potentiality. The only parameter on which *can* in both utterances differs is the source. In (34) the source of the potentiality is participant-internal: x is not precluded from jumping because of intrinsic factors. This raises the meaning of ability. In (35) the source of the potentiality is participant-external: x is not precluded from jumping because of external factors. This raises the meaning of root-possibility.

Finally, the meaning of modal expressions may differ only on the parameter of scope. See (36) and (37):

(36) You may borrow my book for three weeks if you want.
(37) In the library, you may borrow a book for three weeks.

The general sense of both utterances is potentiality and the source of the potentiality is deontic. This yields the interpretation of permission in both utterances. What is different, however, is that the permission in (36) concerns
the specific participant you and in (37) the event of borrowing a book. The scope of may in (36) is the predicate; this is participant-oriented modality, meaning x is not precluded from borrowing because of an individual deontic authority. In (37) the permission concerns the event of borrowing a book for three weeks by any participant (non-specific you). The scope of may is the predication; this is event-oriented modality, meaning e is not precluded from occurring because of a general deontic authority.

The examples in (32)-(37) illustrate that sense, source, and scope operate independently to a certain extent. Therefore, all parameters are necessary to account for the semantic distinctions of modal expressions. The combination of the parameters is presented in Table 4-5 and leads to the specific modal meanings listed in the individual cells, such as root-necessity or desirability. Within a basic sense, each cell is semantically related to the adjacent cells in that they only differ on one of the parameters. Furthermore, a cell at a specific location within a basic sense is related to the cell at the similar location in the adjacent basic sense: for instance, ability is related to volition. The dotted line between the two subsources of participant-external modality indicates that both categories are semantically related to the previous row of participant-internal modality and to the next row of epistemic modality. In English, subtypes of potentiality are expressed by can, could, may or might; subtypes of disposition by will, shall, would, or wanna; subtypes of weak necessity by should, ought to or be supposed to; and subtypes of necessity by must, need to, have (got) to, or got to.

The classification in Table 4-5 could be considered a representation of a conceptual space. Which categories are actually expressed and distinguished linguistically is a language-specific characteristic. In section 6.2.5 the modal system in English will be discussed in detail and in Chapter 7 crosslinguistic variation in modal systems will be examined.

Whether this classification of modality should be regarded as a polysemy account, in which a single expression has separate related meanings, or as a monosemy account, in which a single expression has one underdetermined meaning that in combination with pragmatic inferences gives different contextual interpretations, is not a very relevant issue in my opinion (see Papafragou 2000 for a discussion). Speakers use expressions in different contexts with different intentions: it is beyond the scope of this thesis to determine whether language users have stored one item together with the pragmatic conventions for the possible specific uses, or several items for which the semantic relations are specified. Possibly, they use both strategies.
Table 4-5. Classification of the domain of modality along the parameters sense, scope, and source

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>SCOPE</th>
<th>POTENTIALITY</th>
<th>DISPOSITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant-internal</td>
<td>ability</td>
<td>volition</td>
<td></td>
</tr>
<tr>
<td>Participant-external</td>
<td>root-possibility</td>
<td>root-disposition</td>
<td></td>
</tr>
<tr>
<td>Deontic</td>
<td>permission</td>
<td>desirability</td>
<td></td>
</tr>
<tr>
<td>Epistemic</td>
<td>possibility</td>
<td>disposition/</td>
<td>prediction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
*Note:* or = oriented; nec. = necessity; obl. = obligation

**SENSE**

<table>
<thead>
<tr>
<th>WEAK ENTAILMENT</th>
<th>ENTAILMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant-or.</td>
<td>Event-or.</td>
</tr>
<tr>
<td>weak need</td>
<td>weak root-nc.</td>
</tr>
<tr>
<td>weak root-nc.</td>
<td>weak oblig.</td>
</tr>
<tr>
<td>weak oblig.</td>
<td>probability</td>
</tr>
<tr>
<td>weak certainty</td>
<td>necessity</td>
</tr>
</tbody>
</table>
4.2.5 Summary

Modality can best be defined on the basis of three parameters: **sense** conveys the basic meaning of a modal expression, on a scale from potentiality to necessity; **source** concerns the factors to which the modality is ascribed, participant-internal, participant-external or epistemic source; the **scope** of modality is the part of the utterance that is modified. The parameter of scope is most important to the remainder of this thesis. Grammatical modal expressions can have scope over the predicate; these expressions function as \( \pi_1 \)-operators (participant-oriented modality). Second, grammatical modal expressions can have scope over the predication; these expressions function as \( \pi_2 \)-operators (event-oriented modality). Third, grammatical modal expressions can have scope over the proposition; these expressions function as \( \pi_3 \)-operators (proposition-oriented modality).

4.3 Related Domains

4.3.1 Future tense and prospective aspect

As was already mentioned briefly in section 4.2.3.3, there is a close relationship between future tense and modality. On the one hand, modal expressions very often have future time reference: they do not describe or refer to an event that is remembered or currently perceived, but to an event that is imagined. There is often a strong implicature that the event described will take place in the (near) future. On the other hand, future reference has a sense of modality. This is a commonly accepted stance; Lyons, for example, states that future tense is partly temporal and partly modal in that there is always an element of prediction involved. (Lyons 1977: 677). Comrie (1985), in a similar vein, claims that:

There is a sense in which the future is clearly different from the past. The past subsumes what may already have taken place and, barring science fiction, is immutable, beyond the control of our present actions. The future, however, is necessarily more speculative, in that any prediction we make about the future might be changed by intervening event, including our own conscious intervention. (p.43)

Mithun (1995: 378) claims that languages that systematically mark irrealis (see 4.3.2), often include future events in the irrealis domain. Finally, Palmer states that ‘the future is not fully known and it is always no more than a reasonable assumption that a future event will ensue’ (2001: 104-05).
The distinction between past and present tense is a pure distinction of tense, whereas the distinction between future and non-future tense often involves also a difference in the definiteness or factuality of the event.

What is conventionally regarded as the future tense (in languages that are said to have a future tense) is rarely, if ever, used solely for making statements or predictions, or posing and asking factual questions, about the future. It is also used in a wider or narrower range of non-factive utterances, involving supposition, inference, wish, intention and desire. (Lyons 1977: 816)

Similar claims can be made for prospective aspect, as in Mary is going to write a letter. When a participant is engaged in the pre-state of an event, it is probable, but not necessarily the case that the event will be realized. Furthermore, the being in a pre-state of an action may consist of having the intention to do that action, which is a modal notion and not only an aspectual one. In general, the inference that the event will in fact take place is much stronger for future tense than for prospective aspect.

Within the category of disposition will and would can express epistemic, event-oriented modality. This can be paraphrased by ‘e is disposed to occur’, which can hardly be distinguished from a future tense reading. There is thus a close relationship between the modal category of disposition and the future tense in English. Bybee and colleagues (Bybee, Pagliuca, & Perkins 1991: 279-80; Bybee et al. 1994) suggest that there is a crucial bridge from volition (and other categories) to future, which they refer to as intention. When will is combined with a first person, and the event is under the control of the first person, then this indicates not only that the subject is willing to act, but it also stresses that the event will be realized in the future, if not cancelled by unexpected circumstances. See (38) and (39):

(38) I'll get you a map. (C174)
(39) But I will bring you more today I promise. (C174)

The label ‘intention’ suggests a participant-internal source, but that is a too narrow conception as the notion also implies that there are no external circumstances that prevent x from doing something. Where should intention be located in the classification of modality in Table 4-5? The semantics of ‘intention’ lie in between volition, which is participant-internal participant-oriented modality, and future, which may be conceived of as epistemic, event-oriented modality. Such utterances can function as a promise. This meaning might occupy the cell of root-disposition, probably with scope over the predicate, but it may as well be interpreted as a future tense restricted to first person.
I do not think that a strict decision has to be made whether future tense belongs to the domain of tense or of event-oriented modality, and whether prospective belongs to aspect or to participant-oriented modality. Both categories may be conceived of as strong links between different conceptual domains and there may be uses in which the tense or aspect senses are most prominent and uses in which the modal senses are most prominent. For English, the involved expressions (will and be going to) seem to denote most prominently tense and aspect. They are used most frequently in opposition to tense and aspect markers, respectively, rather than in opposition to modal markers. For this thesis, the main point is that future tense and possible modal implications is analyzed as $\pi_2$-operator and prospective aspect and possible modal implications as $\pi_1$-operator.

### 4.3.2 Irrealis

A second domain related to modality is the domain of irrealis. Evidence for its linguistic relevance is found in several languages, among others Central Pomo (Mithun 1995), Tok Pisin (Romaine 1995), Burmese and Dyirbal (Comrie 1985: 50-52), Caddo and Northern Iroquian languages (Chafe 1995). Irrealis is not crosslinguistically uniform, but in general it includes notions like future, negatives, yes-no questions, modality (necessity, possibility), conditions, imperatives, and prohibitives. Languages that mark irrealis differ in the constructions they mark as such, but ‘counterfactuals and conditionals, are generally classified as Irrealis in all languages with a grammaticized Irrealis/Realis distinction.’ (Mithun 1995: 376). Within this thesis, the label of irrealis will be used to refer to expressions that have hypothetical (40) and counterfactual (41) interpretations:

(40) If she had time, Mary would write a letter.
(41) If she had had time, Mary would have written a letter.

The definition of hypothetical is adopted from Bybee et al. (1994: 322) and here presented in D5:

D5. A **hypothetical** statement indicates that ‘the situation is unreal or imagined, but one that could be true’.

The speaker is thus neutral about whether the conditions for realization of the event are or will be fulfilled.

The definition of counterfactuals is also adopted from Bybee et al. (1994: 322) and here presented in D6:
D6. A **counterfactual** statement describes an ‘unreal or imagined situation that could have been true but was not.’

Counterfactuals express that the condition(s) for realization of the event are not fulfilled.

Many languages that do not have a systematically marked opposition between reals and irreals do have special expressions for hypothetical and counterfactual statements. In several languages, the same morphemes are used for marking past tense and hypothetical or counterfactual events. This led Steele (1975) and others to claim that the past tense and hypothetical events share the meaning ‘distant from present reality’. However, Bybee stresses that it is never the past tense marker alone that accounts for the hypothetical reading, but the combination of past tense morphology and a modal verb, subjunctive mood, a hypothetical marker (such as if), or imperfective aspect. Diachronically, it is not the remoteness of the past tense that develops a hypothetical sense, but ‘rather the fact that past combined with modality leaves open the possibility that certain conditions on the completion of the predicate have not been met.’ (Bybee 1995: 514). Not all languages make a distinction between hypothetical and counterfactual events, but, instead, they use only one general irrealis marker that covers both notions. Hypothetical and counterfactual expressions are closely related to event-oriented (objective) epistemic modality ($\pi_2$). Event-oriented modality situates the event in actuality in terms of the scale of senses: the event occurs potentially or necessarily in this world. Irreals expressions relate the event to a hypothetical or counterfactual world. In particular hypotheticality and epistemic possibility are closely related and in some languages expressed by a single element. Other languages use one and the same expression for epistemic possibility and future tense. (See 7.5.2.1).

In FG, hypothetical and counterfactual expressions are analyzed as $\pi_2$-operators that indicate the actuality of the event: the event the speaker has in mind is related to an unreal, non-actual world. These expressions modify the predication and contribute to the situating function of the utterance.

### 4.3.3 Evidentiality

The final domain to be discussed is the domain of evidentiality. The definition used in this thesis (D7) is based on Dik (1997a: 242) and Palmer (2001: 8):

D7. **Evidential** expressions indicate the source of knowledge or the kind of evidence the speaker has for the proposition.

Consider the following examples of lexical (42)-(43) and grammatical (44) expressions of evidentiality:
Chafe (1986: 262) and Mithun (1986: 89) both define evidentiality as marking the reliability of the information. Common distinctions within the domain of evidentiality are between sensory (42) and non-sensory evidence or between direct and non-direct evidence. Within the category of sensory evidence there may be a distinction between visual, non-visual and auditory evidence. Another evidential category that is often expressed is the function of hearsay (also referred to as quotative or reportative), see (43), which indicates that the source of the speaker’s knowledge is the utterance of someone else. Finally, there are expressions that indicate that the speaker’s knowledge is an inference made from non-sensory evidence. See (44).

It may be clear from this discussion that evidentiality is analyzed in FG as an operator with scope over the proposition (π3): it expresses the reliability of the propositional content by indicating the source of the information. How did the speaker acquire this information? Evidential markers thus modify the presentation of the propositional content.

Evidentiality is related to modality in that the notion of inference occurs both as a specific distinction within the system of proposition-oriented (subjective) epistemic modality and within evidentiality. Furthermore, in some languages, modal auxiliaries have come to express evidential notions. For example, German uses sollen (‘shall’) to express hearsay. See (45):

(45) *Und das soll gut sein?*

And that shall good be

‘Is it said to be good?’

4.4 CONCLUSION

In this chapter I discussed the semantics of modality and related notions. The domain of modality was described using the parameters of sense, source and scope. The minimal differences in the interpretation of modal expressions can only be accounted for when all three parameters are taken into account, as there is no one-to-one relation between the parameters.

Grammatical expressions of modality can have scope over the predicate, the predication or the proposition, depending on their specific function. The domain of irrealis covers notions of hypothesis and counterfactuality indicating

---

6 Cf. Willett (1988) for an extensive study on evidential functions that occur crosslinguistically.
that a situation is non-actual. Expressions of irrealis function as $\pi_2$-operators. Expressions of evidentiality encode the source of knowledge or the kind of evidence the speaker has for the proposition to indicate the reliability of the proposition. These expressions function as $\pi_3$-operators.

Table 4-6 presents an overview of the TMA domains and their subcategories discussed in this chapter. There are several semantic relations within and between the domains in Table 4-6. An additional relationship exists with future tense and prospective aspect.

In Chapter 5 the hierarchical relations between different TMA domains will lead to concrete predictions about linguistic implications of the Scope Hierarchy.

**Table 4-6.** Classification of modality, irrealis and evidentiality according to scope

<table>
<thead>
<tr>
<th>TMA domain</th>
<th>Scope over</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant-oriented modality</td>
<td>Predicate ($\pi_1$)</td>
</tr>
<tr>
<td>- internal</td>
<td>ability, volition, (weak) need;</td>
</tr>
<tr>
<td>- external, non-deontic</td>
<td>root-possibility, root-disposition, (weak) root-necessity;</td>
</tr>
<tr>
<td>- external, deontic</td>
<td>permission, desirability, (weak) obligation</td>
</tr>
<tr>
<td>Event-oriented modality</td>
<td>Predication ($\pi_2$)</td>
</tr>
<tr>
<td>- external, non-deontic</td>
<td>root-possibility, root-disposition, root-(weak) necessity;</td>
</tr>
<tr>
<td>- external, deontic</td>
<td>permission, desirability, (weak) obligation;</td>
</tr>
<tr>
<td>- epistemic</td>
<td>epistemic possibility, (future,) probability, necessity</td>
</tr>
<tr>
<td>Irrealis</td>
<td>hypothetical, counterfactual</td>
</tr>
<tr>
<td>Proposition-oriented modality</td>
<td>Proposition ($\pi_3$)</td>
</tr>
<tr>
<td>- epistemic</td>
<td>uncertainty, prediction, (weak) certainty, inference;</td>
</tr>
<tr>
<td>- deontic</td>
<td>permission, desirability, (weak) obligation</td>
</tr>
<tr>
<td>Evidentiality</td>
<td>sensory evidence, inference, hearsay, etc.</td>
</tr>
</tbody>
</table>
Chapter 5

Implications of the Scope Hierarchy

5.1 Introduction

In section 2.5 it was argued that operators with wider scope modify more complex semantic units that are conceptually more abstract, and that operators with wider scope are communicatively less motivated than operators with narrower scope. I expect that semantic complexity, conceptual abstractness and communicative motivation play an important part in the limits on language variation. What is conceptually more abstract, communicatively less needed and semantically more complex occurs probably less frequently in human language systems. From the viewpoint of FG, operators with wider scope are more marked than operators with narrower scope. It is thus assumed that there is an implicational hierarchy as in (1):

\[(1) \quad \pi_1\text{-operator} \subset \pi_2\text{-operator} \subset \pi_3\text{-operator},\]

in which ‘\(\subset\)’ means ‘is implied by’ or ‘is less marked than’. Consequently, it is to be expected that operators with wider scope behave differently from operators with narrower scope both in languages of the world and in stages of child language. Which differences may be expected will be discussed in this chapter.

Each category of operators in (1) represents specific domains of TMA. In Chapters 3 and 4 I discussed the analysis of different TMA domains with respect to their scope. A summary of these findings is presented in Table 5-1, which unites Table 3-2 and Table 4-6. This table shows to which class of operators the different TMA domains belong. It should be stressed that \(\pi_1\)-operators with scope over the predicate, i.e., aspect, property quantification and participant-oriented modality, are considered to serve a similar communicative function in the utterance, namely modification of the predicate. Therefore, they are conceived of as a single category of modifiers. The same holds for different expressions of tense, event-oriented modality, event quantification and irrealis that all function as \(\pi_2\)-operators that modify the predication and for different expressions of proposition-oriented modality and evidentiality that all function
### Table 5-1. Classification of TMA domains according to scope

<table>
<thead>
<tr>
<th>TMA Domain</th>
<th>Scope over</th>
<th>Operator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Property quantification</td>
<td></td>
<td>$\pi_1$</td>
</tr>
<tr>
<td>Participant-oriented modality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event quantification</td>
<td></td>
<td>$\pi_2$</td>
</tr>
<tr>
<td>Event-oriented modality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irrealis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposition-oriented modality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evidentiality</td>
<td></td>
<td>$\pi_3$</td>
</tr>
</tbody>
</table>

as $\pi_3$-operators that modify the proposition. A specification of (1) is presented in (2):

(2) \{aspect, property quantification, participant-oriented modality\} $\subseteq$ \{tense, event quantification, event-oriented modality, irrealis\} $\subseteq$ \{proposition-oriented modality, evidentiality\}

The implicational hierarchy is assumed to have reflections in adult and child language. This chapter will discuss the specific linguistic implications of the Scope Hierarchy. As this research is about *grammatical* markers of TMA, first a closer look is needed into the evolution of grammatical markers in language, i.e. into grammaticalization processes. This will lead to specific hypotheses about diachronic development, frequency, synchronic configurations, and morphology and syntax of TMA expressions. It will also lead to the formulation of a specific hypothesis about the order of acquisition of TMA expressions in first language development.

### 5.2 Grammaticalization

In many theoretical models of language, including FG, a sharp distinction is made between lexical and grammatical items, or, similarly, between content words and function words. Sapir (1921: 25) spoke of ‘material content’ (or lexical meaning) and ‘relational content’ (or grammatical meaning). Although languages universally distinguish between the two types of elements 'a given
concept may in one language be treated as though it were material, while in another language the same concept appears as relational’ (Givón 1984: 48).

Lexical elements are independent entities, whereas grammatical elements are dependent entities; they relate lexical elements with one another, or modify their connotation. In general, the differences between a typical lexical and typical grammatical element can be described as in Table 5-2. Haspelmath (1999: 1056) adds that there is also a correlation with modes of mental processing: lexical items are processed deliberately and consciously, whereas grammatical items are processed automatically and unconsciously.

Although most theoretical models suggest otherwise, the distinction between the lexical and grammatical extremes is not sharp, but rather continuous. Diachronically, ‘grammatical morphemes develop gradually out of lexical morphemes or combinations of lexical morphemes with lexical or grammatical morphemes.’ (Bybee et al. 1994: 4) This process is referred to as grammaticalization (or grammaticization). Grammatical items or grams (after Bybee et al.) evolve from lexical items and during this process they will increasingly take on the characteristics in the right column in Table 5-2. It is, however, hard to define necessary and sufficient characteristics, so that it is difficult to say at which point exactly an item should be defined as grammatical. There are many elements in a language that show some, but not all of the properties of either the lexical or the grammatical pole. For example, prepositions show characteristics of both sides and periphrastic constructions are also notoriously difficult to classify (Muysken 2006; Olbertz 1996).

---

Categories of lexical and grammatical items are thus imprecise: it is impossible to define the categories in such a way that ‘for every member x of its domain (that is, the set of things to which the category can be meaningfully applied), the definition determines a truth-value to the statement that x belongs to the category in question’ (Dahl 1985: 3). Section 6.2.1 discusses the impreciseness of grammatical elements in this thesis.

In order to predict what the influence of scope may be on TMA systems, it is important to understand the processes by which grammatical TMA expressions arise in languages. Grammaticalization results from semantic or functional change, morphosyntactic processes, and phonetic reduction. Firstly, the meaning of an item becomes more general and abstract and correspondingly, the range of contexts in which it can be used becomes wider. This process is referred to as desemanticization (Heine 1993), semantic generalization (Bybee & Pagliuca 1985), bleaching (Givón 1975) or erosion (Lehmann 1995). Because of the increasing generality of meaning and loss of specificity, the selection restrictions on the arguments or type of main predicate with which the item can combine are loosened. For example, lexical verbs that are on their way towards becoming an auxiliary, typically change from combining only with a human agentive subject, to combining with all kinds of subjects, and from combining with a complement that expresses a concrete object or location to a complement that expresses a dynamic situation (Heine 1993: 54). For example, a common semantic change is that of a lexical verb meaning want changing to a morpheme denoting future tense. Verbs of wanting typically combine with two arguments: an animate subject and a concrete object as for the Dutch wil ‘want’ in ik wil een koekje ‘I want a cookie’. In a next stage, the verb may come to combine with verbal complements ik wil lezen ‘I want to read’ or with an inanimate being de auto wil niet starten ‘the car does not want to start’. When it further develops into a future meaning, the selection restrictions on the arguments gradually disappear completely, as has happened with will in English: any subject is allowed and any type of complement, as can be seen in examples such as the stone will fall or economy will rise.

As the applicability of an item increases, it will become much more frequent in discourse: there are far more contexts in which it may be used. In the next stage, the use of the item is no longer restricted to contexts where it adds semantic content. It may now be used when the semantics are in fact redundant, whenever it is compatible with the context. This may lead to a stage in which it becomes meaningful not to express the item: the absence of item X comes to denote ‘not X’. For example, the absence of a past tense marker may come to mean ‘non-past tense’, so that a past tense marker becomes obligatory in every context that concerns the past tense, which again, results in an increase in frequency. Grammatical items may develop this systematic, obligatory use in
contrast to lexical items: it is never the case that the absence of a lexical item Y means ‘not Y’, for example, it is not the case that every sentence without the adverb yesterday means ‘not yesterday’. Note, however, that not every grammatical marker is obligatorily used.

Secondly, grammaticalization is characterized by morphosyntactic change. A grammaticalizing morpheme or construction becomes phonologically and semantically more dependent on surrounding material, which leads to an increasing rigidification of the syntactic position of the morpheme (Bybee et al. 1994: 7). Its syntactic position is less and less free. Furthermore, decategorialization of the source item takes place (Heine, Claudi, & Hünnemeyer 1991a): it loses the characteristics of a typical verb, noun, adjective or adverb. For example a verb on its way to becoming an auxiliary loses verbal properties, such as the ability to form imperatives, to be nominalized, to have an infinitival form, to passivize, or to be negated separately. It may no longer have a noun as its complement nucleus, but instead a (nominalized or nonfinite) verb.

Thirdly, grammaticalization is characterized by phonological reduction. There is quantitative reduction: forms become shorter, because phonemes erode or fuse with other phonemes or because the articulation is compressed so that the temporal duration of the item is decreased. There is also qualitative reduction: phonological segments stem more and more from the unmarked set of phonemes of the specific language. (Hopper & Traugott 1993: 145), (Bybee et al. 1994: 6). Heine (1993: 56) mentions that a verb loses its ability to carry distinctive tone or stress.

During grammaticalization there is a stage in which a single item is used both in contexts where it shows mainly lexical characteristics and in contexts where it shows mainly grammatical characteristics. The ‘grammatical’ use is at first often restricted to a specific construction and it is this specific construction that forms the input for the grammaticalization process. In the process of grammaticalization the grammatical use gradually becomes more frequent, possibly, but not necessarily, at the cost of the lexical use. In sum, the more grammaticalized an item is:

- the more it loses in semantic complexity, functional significance and/or expressive value
- the more it loses in pragmatic and gains in syntactic significance
- the more reduced is the number of members belonging to the same morphosyntactic paradigm
- the more its syntactic variability decreases, the more its position within the clause becomes fixed
- the more its use becomes obligatory in certain contexts and ungrammatical in others
• the more it coalesces semantically, morphosyntactically, and phonetically with other units
• the more it loses in phonetic substance (Heine, Claudi, & Hünnemeyer 1991b: 15-16)

When an item is grammaticalized, the development does not stop: the item tends to generalize further in meaning and to reduce further in size. A common development is that several items generalize in meaning and these generalized meanings come to overlap. For example, there may be different items, for example a prospective, a volition and an obligation marker, that all develop towards the more general meaning of future tense. This leads to diachronic competition between the different elements and one or more of them may become extinct.

The remarkable thing about grammaticalization is not so much that it exists, but rather that many analogous developments are attested in unrelated languages. Just to mention a few, sources for functional items originate over and over again in general terms for movement, postures, locations and body parts (Bybee 2003a: 151-52); The word for ‘say’ developed into a marker of direct quotation in several languages, among others Khmer, Buru, Saramaccan and Twi (Heine & Kuteva 2002: 267-68); A completive or perfect marker has developed from a lexical verb meaning *finish* in e.g. Bantu languages, Cantonese, Tok Pisin, Lao, Engenni, Tucano, and American Sign Language, (Bybee 2003a: 149). Heine & Kuteva (2002) collected several hundred universal grammaticalization paths, universal in the sense that they have independently occurred in unrelated languages. The actual goal of linguistics, however, is not only to collect universals but rather to understand why they exist. We want to know what are the ‘true universals’, the ‘mechanisms of change that propel grams along these paths of development.’ (Bybee et al. 1994: 302). These mechanisms that create the changes should be understood ‘in terms of more basic cognitive and interactive processes.’ (Bybee 2003a: 151). What are the universal cognitive and communicative mechanisms that underlie grammaticalization?

5.3 COGNITIVE MECHANISMS AND COMMUNICATIVE NEEDS

5.3.1 Pragmatic inference

When we want to understand universal patterns, a closer look is needed into the cognitive processes involved in grammaticalization. Particularly the process of semantic change needs some elaboration. The change in meaning that morphemes undergo in the process of grammaticalization is not arbitrary or random. Sweetser (1990: 9) states that ‘(...) new senses are acquired by
cognitive structuring’ and that ‘multiple synchronic senses of a given word will normally be related to each other in a motivated fashion.’ What exactly is this motivated relation between different senses?

5.3.1.1 Metaphor and metonymy

There are two important mechanisms or motivations that may play a part in semantic change: metaphor and metonymy. Metaphor is a figure of speech whereby a certain (abstract) concept is understood as if it were another (concrete) concept. Both concepts are related in the mind of the speaker by analogy or iconicity. Sweetser claims that the most important mechanism of semantic change is **metaphorical transfer**. She states that:

- we model linguistic expression itself not only (a) as a description (a model of the world), but also (b) as action (an act in the world being described), and even (c) as an epistemic or logical entity (a premise or a conclusion in our world of reasoning). (1990: 21).

Human beings understand epistemic or logical entities in terms of real-world entities. They transfer concepts and vocabulary from the more accessible external sociophysical world to the less accessible abstract internal world of reasoning (1990: 31). As discussed in 4.2.2, Sweetser claims that the basic sense of modal auxiliaries is by metaphorical extension applied to the internal mental world. Logical necessity is ‘the mental analogue of sociophysical force, while logical possibility is the mental (or epistemic) analogue of permission or ability in the real world.’ (1990: 30). The same metaphor is used in other areas of the linguistic system, such as with verbs of sensory perception (see, grasp) are used for mental perception (meaning ‘know’, ‘understand’). Metaphorical transfer thus involves the application of a certain grammatical element or gram to a more abstract world, while the sense of the gram is maintained. A serious problem with the metaphor model is that it involves ‘a discrete “jump” from one domain to another’ (Heine 1993: 97), whereas the process of grammaticalization supposes a gradual change.

A second mechanism of semantic change is **metonymy**, also referred to as pragmatic or context-induced inference, or conventionalization of implicature (see e.g. Bybee 1988; Bybee 2003a; Heine et al. 1991b; Hopper & Traugott 1993; Traugott 1989, 1996). It is defined as:

- a figure of speech whereby the name of one entity e1 is used to refer to another entity e2 which is contiguous to e1. (…) The essence of metonymy resides in the possibility of establishing connections between entities which co-occur within a given conceptual structure. (J. R. Taylor 1989: 122-24)

---

2 Notice the similarity in formulation with the layers in FG that function to describe, to refer and to transfer propositional content.
Pragmatic inference is a basic principle of communication. Speakers make inferences when they encode the intention they want to communicate and addressees make inferences when they want to decode the communicated intention. They both use their knowledge of the preceding context, the current speech situation and their general knowledge of the world as they infer what the pragmatic information of the interlocutor is. Speakers anticipate on the interpretation of addressees by estimating the pragmatic information of the addressee and encode their intention in such a way that addressees are probably capable of reconstructing their intention by inference. Addressees make an estimation of the pragmatic information of the speaker when they try to infer what the speaker has tried to communicate by the linguistic forms he or she has chosen to use. The utterance in itself does not establish the relation between the intention of the speaker and the interpretation of the addressee but rather mediates it and in this sense communication is purely a matter of pragmatic inference.

Pragmatic inference plays an important role in semantic change. Within a certain context, a specific expression may have certain implicatures about related senses, for example, when someone says I need to go now, it only encodes that it is somehow necessary for the speaker to go, but it is also very probable that the speaker will indeed leave soon. If an expression occurs frequently in a context in which these implicatures are present, then the expression becomes frequently associated with the inference and language users come to ‘assume that the inference is a necessary part of the meaning of the form’ (Bybee et al. 1994: 197). The implicature may become conventionalized and form part of the meaning: the item has now acquired a secondary sense, besides its primary sense. After this, the use of the item may become restricted to contexts in which the new meaning is appropriate: the secondary sense becomes the primary sense, whereas the original meaning becomes the secondary sense. Eventually, the original meaning may disappear completely. For example, a perfect aspect marker, which selects the post-state of an event (see 3.3.3), implies that an event has taken place, although this implicature may in most cases be irrelevant to the conversation. If however, the perfect is often used in contexts where not only the post-state of the event is relevant, but also the fact that the event has taken place, or in contexts where the post-state is even less relevant than the fact that the event has taken place, then the implicature of the past tense is highlighted and gradually becomes part of the meaning. The perfect construction may at first develop a past tense use besides its perfect aspect use and gradually lose its perfect aspect use.

Metaphor and metonymy are not necessarily mutually exclusive mechanisms. Both Goossens (1995) and Heine et al. (1991b) argue that they complement one another and this position is now widely accepted. Although metaphorical
transfer involves discrete, discontinuous steps from one cognitive domain to another, there is also a gradual change: there is a continuous extension from using the item in linguistic contexts where only the ‘concrete’ meaning is appropriate, to contexts where there is ambiguity between the concrete and more abstract reading, and from there, to contexts where only the more abstract reading is appropriate. The metaphorical interpretation is in fact an implicature within a certain context. This may be illustrated with an example from Traugott & König (1991) for the modal auxiliary *must*:

*Must* in the epistemic sense of ‘I conclude that’ [is] derived from the obligative sense of ‘ought to’ by strengthening of conversational inferences and subjectification. If I say *She must be married* in the obligation sense, I invite the inference that she will indeed get married. This inference is of course epistemic, pertaining to a state of affairs that is anticipated to be true at some later time. (p.209)

### 5.3.1.2 Polysemy and ambiguity

Within the process of semantic change, ambiguity or polysemy of a specific morpheme is a common (and maybe even necessary) phase: there is a stage in which the morpheme is used with different interpretations. Coates (1983: 15-17) distinguishes between two kinds of ‘overlapping’ senses (polysemy): ambiguity and merger. Ambiguity yields an ‘either/or’ interpretation: either sense A or sense B is a possible interpretation, but the two interpretations are mutually exclusive. Consider (3):

(3) He must understand that we mean business.

*Must* in (3) expresses either ‘surely he understands that we mean business’ or ‘it is essential that he understands that we mean business’ (Coates 1983: 16). The second type of polysemy is merger, which yields a ‘both/and’ interpretation: both sense A and sense B are possible interpretations and both interpretations are mutually compatible. Consider (4):

(4) Rutherford suggested to Marsden that he should follow this up.

*Should* in (4) gets the interpretation of weak obligation, but also of a quasi-subjunctive. (Coates 1983: 17).

---

3 According to Sweetser, ‘no historical shift of meaning can take place without an intervening stage of polysemy’ (1990: 9).
Schematically, semantic change can be represented as in (5) in which the sense in between brackets is the less common interpretation. The secondary sense, B, can arise through metaphor or metonymy:

(5) Ambiguity: \( A \rightarrow \{A \text{ (or B)}\} \rightarrow \{A \text{ or B}\} \rightarrow \{B \text{ (or A)}\} \rightarrow B \)

Merger: \( A \rightarrow \{A \text{ (and B)}\} \rightarrow \{A \text{ and B}\} \rightarrow \{(A \text{ and) B}\} \rightarrow B \)

The semantic change of a linguistic item can result in an extra sense which is in general more abstract, referred to as specialization of meaning (\( A \rightarrow AB \)), or it can result in the loss of a sense, called generalization of meaning or bleaching (\( AB \rightarrow B \)). The more specialized and the more generalized interpretation may be used at the same time in a synchronic stage of language, in which case we speak of polysemy.

It should be stressed that different types of semantic change are possible and that the above representations may be too simplistic. For example, a specific item could trigger different implicatures in different contexts. As a result, the item may develop more than one secondary sense, yielding a 'split' in the semantic development, as represented in (6):

(6) \( A \rightarrow AB \rightarrow B \)

\( \downarrow \)

\( AC \rightarrow C \)

Other scenario’s are conceivable: two senses may be involved, the first sense may remain present for a long time, secondary or third senses may become primary senses, etcetera. Bear in mind that an earlier interpretation does not have to disappear immediately when a new interpretation gets conventionalized: different interpretations may co-exist in a single stage of language.

### 5.3.2 Automaticity

For the different processes of grammaticalization (semantic change, morphosyntactic rigidification and phonological reduction), frequency of use is an essential factor. It is, however, impossible to tell the cause from the result. On the one hand, frequent use is a necessary condition for grammaticalization to take place; on the other hand, the process of grammaticalization induces frequent use. Frequent repetition of a linguistic construction leads to greater ease of processing, that is, automatization of cognitive processes. When a single task is executed repeatedly, one needs less time and effort to execute it. The duration of distinct sequential acts within this task decreases and several acts come to overlap or even fuse into one. Automatization is a driving force in grammaticalization. Only if an item is used frequently in a context that triggers a specific implicature, this implicature may become conventionalized as part of
the meaning. Furthermore, repetition of an item causes semantic bleaching through the cognitive process of habituation: if a stimulus is repeated frequently, the level of response of an organism tends to decrease (Bybee & Hopper 2001a: 13; Haiman 1994). Linguistic elements that are used frequently, gradually lose their original strength and their contribution to the information value of the utterance decreases. This leads to ‘inflation’ of meaning (Dahl 2001). Grams add less specific information to the discourse than lexical elements. Automatization also plays a part in syntactic rigidification. If two or more elements co-occur frequently in sequence, the accessibility for this specific string improves and pronunciation takes less time. The construction will become tighter and a string of morphemes may even fuse into one single morpheme (Bybee & Scheibman 1999). Automatization, then, also underlies phonological reduction, as the magnitude of articulatory gestures is reduced and distinct articulatory acts come to overlap (Bybee & Hopper 2001a: 11).

5.3.3 Communicative needs

Why does grammaticalization occur at all? Bybee et al. (1994: 298) argue that grammaticalization cannot be explained simply in terms of communicative need. There are several reasons why it cannot be maintained that languages need a specific grammatical item: no single gram type is universal, two or more markers may express the same function, inflectional markers are often redundant and some grams create rather than solve ambiguity. Discourse is structured and organized, but regularities and structures also coincidentally result from discourse. At first, in the stage where speakers regularly, but not yet consistently, make pragmatic inferences, these inferences are not yet part of the meaning, and ‘conceptual manipulation is still largely provisional, unstable, “epiphenomenal”’. (Heine et al. 1991b: 77) In the next phase, when inferences of a certain construction become adopted in the meaning,

the amount of morphological and syntactic repetition increases, there are recurrent strategies for building discourses, and there are emergent grammatical regularities. This is the phase where grammatical concepts are stabilized, or conventionalized, and develop into distinct ‘senses’. (Heine et al. 1991b: 77)

Grammar, then, would be mechanistic or emergent (Bybee et al. 1994: 298, 300; Hopper 1987, 1998) rather than functional or goal-directed. That grammar emerges does not imply that grammar has no function: it is just not created for a specific purpose. When grammar has emerged, it allows speakers to organize and present information without foregrounding it, as lexical expression does (Bybee et al. 1994: 299). Besides that, grammar facilitates production through automatization. Because morphemes and word order are supplied automatically, the speaker’s attention can be fully directed to the propositional
content. ‘The advantages of automation can explain why categories or structures become obligatory despite the fact that they are not always strictly necessary for communication’ (Bybee et al. 1994: 300).

Although I agree to a certain extent with the view that grammar is emergent and not goal-directed, I have strong objections to Hopper’s claim that grammar is continuously created on the spot, in discourse. In my view, a construction that is conventionalized forms quite a stable instrument in the repertoire of speakers for encoding their communicative intentions. I furthermore disagree with the strong rejection of the role of communicative need in the development of a gram. Although it is beyond doubt that speakers do not consciously innovate a construction for the purpose of creating an extra gram, the context in which a new gram develops is in language use; the grammaticalizing construction is continuously used for fulfilling a certain communicative need, with a certain communicative intention. It is true that no single gram type is universal and that every language has its own set of grammatical expressions, but the total set of grammaticizable notions is restricted. Aspect, tense and modality are typically grammaticalized in languages of the world, but for example the color, material or symmetry of a participant or geometric concepts are never grammatically expressed (Talmy 1985, 1988b). Certain concepts are and others are not grammatically expressed in languages and certain lexical elements do and others do not function as the source for grams. This means that certain meanings of linguistic units are universally much more basic to speaking than others (e.g. ‘possession’, ‘instrument’ are more basic than ‘bicycle’ or ‘moon’ in this sense)’ (Haspelmath 1999: 1055-56). Only general meanings, meanings that are needed in many contexts, will be used so frequently that they may enter the grammaticalization path. In my view, communicative need is a driving force in grammaticalization: not in the sense that language users need new grams but in the sense that communicative need of a construction determines whether that construction may enter the grammaticalization path.

Communicative need is closely related to general rules of communication. Haspelmath (1999), based on Keller (1994), presents five ‘Maxims of action’ that play a part in communication. Compared to Grice (1975), Haspelmath shifts the accent to sociological processes:

4 In Haspelmath’s words: ‘It is difficult to prove that grammatical items are so frequent because their meanings are needed more often than those of other items, but it seems plausible to me’ (1999: 1056).

5 Cf. also Enfield (2003) for an elaborate exposition on the role of sociolinguistic factors on language change and language contact.
1. Hypermaxim: talk in such a way that you are socially successful, at
   the lowest possible cost.
2. Clarity: talk in such a way that you are understood.
3. Economy: talk in such a way that you do not expend superfluous energy.
4. Conformity: talk like the others talk.
5. Extravagance: talk in such a way that you are noticed. (Haspelmath 1999:
   1055)

The five maxims are often competing. Speakers may use an innovative
construction when they want to be noticed (fifth maxim), but this maxim goes
against the fourth maxim and perhaps also against the third and second. Other
speakers may adopt this innovation according to the Maxim of Conformity,
first only in contexts where they want their utterance to be noticed (Maxim of
Extravagance). Whether a new construction spreads or not depends on the
social influence of the primary and secondary innovators and on whether the
meaning of the construction is basic to communication. When the innovation is
used frequently and ever more speakers incorporate the new construction in
their repertoire, the ‘extravagancy’ of the new construction weakens (inflation)
and it now spreads throughout the linguistic community only because of the
maxim of conformity. If the construction is used frequently, it will become
more predictable for the addressee and its processing will become more
automatized. This may cause phonological reduction (Maxim of Economy)
without violating the Maxim of Clarity.

What may be added to this picture is that different individuals within one
linguistic community may independently invent the same construction, which
probably increases the speed of grammaticalization, or they may invent
different constructions for a similar communicative intention, which results in
competition of constructions. The fact that there may be two or more markers
to express the same function is easily understood if grammaticalization is
viewed as a process (unconsciously) instigated by individual language users who
continuously seek new ways to make themselves noticed.

5.4 QUESTIONS AND HYPOTHESES

Operators are the result of grammaticalization. In what way is the scope of
operators related to grammaticalization? How is the Scope Hierarchy in fact
reflected in linguistic domains? In this section specific hypotheses will be
derived with respect to the diachronic development of TMA expressions
(5.4.1), their crosslinguistic and language internal frequency (5.4.2), their
possible synchronic configurations (5.4.3), their expression form (5.4.4) and
expression order (5.4.5). From the hypotheses on synchronic configurations
further hypotheses follow concerning the acquisition order of TMA
expressions (5.4.6).
5.4.1 Diachronic development

The first domain in which the scope of TMA may be reflected is in diachronic development. TMA expressions develop through grammaticalization which is a primarily unidirectional process: linguistic elements develop towards the grammatical pole, but not towards the lexical pole (Givón 1975; Haspelmath 1999). Although the semantics of a specific item may become more or less specified, the combination of semantic generalization and morphosyntactic and phonological changes describes a one-way-direction, an in general irreversible process. Some counterexamples are attested, but they are very scarce (Campbell 1991; Joseph & Janda 1988; Newmeyer 1998; P. Ramat 1992). It should be noted that a cyclic development is not in contradiction to unidirectionality: a grammatical item may re-enter the grammaticalization process in a new construction, on a new path.

In Figure 5-1 a possible, unidirectional, semantic path for a single linguistic item is represented. In this path, Meaning 3 develops into two different meanings, Meaning 4 and Meaning D, which in turn develop further into Meaning 5 and Meaning E, respectively.

How do the meanings on the left and on the right end of the path differ? The process of grammaticalization is regarded as a development from less to more abstract meanings (e.g. Bybee et al. 1994; Heine et al. 1991b; Sweetser 1982), from referential to interpersonal functions (Heine et al. 1991b; Traugott 1982) and from specific to general meanings. Meaning 1 will thus be most concrete and specific and the meanings 5 and E will be most abstract, general and interpersonal.

Regarding TMA expressions in FG, the wider their scope is, the more abstract, general, and interpersonal their meanings are. Predicate operators help in building up a proper description: they specify the action or property ascribed to the first argument, designated by a verbal or non-verbal predicate. These operators serve a fairly concrete and specific function and they may apply selection restrictions on the arguments, since not every argument combines with every modification of a property or relation. Predication operators help
locating the event in time and actuality. They play a part in relating the event to
the real world or an imaginary world and they probably do not or less often
apply selection restrictions on the arguments or predicate. Proposition
operators have the most abstract, interpersonal and general function in that
they express the speaker’s attitude or belief with respect to the propositional
content. They modify the abstract function of the ‘presentation of the content’.
TMA expressions with narrow scope (π1) have the least abstract and most
specific meanings and are thus more to the left side on the grammaticalization
path, whereas TMA expressions with wide scope (π3) have the most abstract
and general meanings and are on the right side of the path. TMA expressions
with medial scope (π2) are in between those two.

Metaphor and metonymy are the driving mechanisms in semantic change of
TMA expressions. A gram that has narrow scope could get used in contexts
where a wider scope interpretation is a possible inference, for example, from a
perfect aspect expression past tense reference may be inferred. By the process
of metaphor our minds make sense of a wider scope interpretation. The core
sense of the gram, at first only used to modify the property or relation, is used
metaphorically to modify the event or the propositional content. If the context
in which the wider scope interpretation is possible occurs frequently, this
interpretation can get conventionalized and the gram can become polysemous
and come to function as two different operators. Both the mechanism of
metaphor and the mechanism of metonymy are relevant to a change in scope.
It is, however, also possible that the sense of an operator changes without a
change in scope. In that case metaphor is not a relevant mechanism: it is only
the conventionalization of implicature that can account for specialization or
generalization of meaning.

It is to be expected that the semantics of a morpheme only change towards
generality and towards abstractness and not the other way around. TMA
expressions should therefore diachronically develop from π1-operator to π2-
operator to π3-operator and not in the other direction. FG, then, predicts that
Note that this hypothesis concerns semantic scope, not structural scope.
Lehmann (1995) has claimed that structural scope decreases in
grammaticalization whereas Tabor & Traugott (1998) have presented evidence
that structural scope increases.

The research question and hypothesis on the influence of the Scope
Hierarchy on diachrony of TMA expressions is formulated in Q1 and H1:

Q1: What is the diachronic development of TMA expressions?
H1: Diachronically, operators show an increase in scope, and develop in
the direction from π1-operator to π2-operator to π3-operator.
5.4.2 Frequency and variation

A second area in which the Scope Hierarchy is probably reflected is in the frequency of operators. There are different types of frequency: token frequency of language-specific TMA expressions, the number of expressions that belong to a specific class of operators and crosslinguistic incidence of operators.

The predictions about the frequency of operators are primarily based on two factors that both play a part in grammaticalization: diachronic competition and communicative motivation. These notions have to be explained first. Diachronic competition occurs when different grams generalize in meaning and come to cover (partially) overlapping semantic spaces. So, at later stages of development, different diachronic paths converge to more general meanings. This leads to competition between the different grams with the result that only one or a few of the grams involved prevails for the more abstract function while the others gradually disappear (Bybee et al. 1994: 8, 15). Furthermore, grams may lose phonetic saliency in the course of time and their meaning may inflate and newer, more salient grams, may gradually replace the older grams (Dahl 2001). Thus, although the course of development of a gram runs from less to most abstract and general meanings, grams may also completely die out and never reach the most abstract level. As was argued in 5.4.1, π₁-operators are expected to be earlier on grammaticalization paths than π₂-operators, and π₂-operators are in turn earlier on these paths than π₃-operators. This means that developmental paths of different grams with narrow scope converge when scope increases, and that in the course of time younger grams may replace older grams in a bottom-up direction. What this means for the different types of frequency will be discussed below.

A second factor that plays a role in frequency is the communicative motivation of the function of an operator. Dahl argues that the communicative motivation of items decreases when they grammaticalize further:

Both semantic bleaching and the diminished reliance on relevance considerations lead to a general decrease in communicative motivation of an item. Thus, when an item is grammaticalized, its content becomes less significant to the communication. (…) The communicative motivation for a tense morpheme marking past time reference is arguably less the more early the time reference is derivable from the context (Dahl 2000: 9).

The communicative motivation of TMA expressions depends, in other words, on their predictability or specificity in meaning. In FG, the function of operators with wider scope is communicatively least motivated. Most motivated is the function of predicate operators (π₁) with narrow scope. These operators specify the description of the event, which takes place in nearly every utterance. The modifications that predicate operators make involve such specific changes in this description that this information has to be expressed linguistically. It is
impossible to predict or infer from context, and it is impossible to deduce this information from intonation or non-verbal clues. Predication operators ($\pi_2$) are communicatively less motivated than predicate operators. Predication operators help to situate the event in the real or imagined world: here, the context often helps the addressee to establish the referential point, sometimes by inference, sometimes by adverbial markers, so that operators involved with tense, event quantification or actuality, more often than predicate operators present information that is to a certain extent redundant. Finally, proposition operators ($\pi_3$) seem to be communicatively least relevant of fundamental. The commitment of the speaker to the propositional content is irrelevant in most utterances, since speakers are by default committed to the truth of the propositional content. Furthermore, both intonation and facial expression may indicate the speaker’s attitude or commitment, or the commitment may be inferred from predication operators for actuality. In sum, it is assumed that if the scope is wider, the operator has a less fundamental role in communication.

Diachronic competition and communicative motivation influence the different types of frequency. With respect to token frequency, it is in general stated that an item in the process of grammaticalization shows a dramatic increase in use (Bybee 2003b; Bybee et al. 1994: 8; Krug 2000: 74). When its semantics become more general, the gram is applicable in more contexts. It comes to be used redundantly and eventually obligatorily, so that the discourse frequency keeps increasing long after grammatical status has been reached. This sketch of development is, however, too simplistic, since it applies only to a certain stage in the development and only to certain grams. The token-frequency of a lexical element will indeed increase dramatically when it becomes grammatical, but once fully part of the grammar, many expressions gradually die out because of diachronic competition. Those grams that do survive the sorting out, however, are not necessarily used more frequently either because their communicative motivation may decrease. It all depends on the obligation of expression: certain grams may be obligatorily expressed in a specific language and these grams will in fact be used very generally and frequently, like tense in English. Other grammatical expressions, however, remain optionally expressed and this scenario seems to take place more frequently. These non-obligatory expressions are in continuous competition with lexical expressions, for example the grammatical expression for uncertainty in English might competes with the lexical construction I suppose that. Depending on whether an expression becomes obligatory, the discourse frequency of a gram may either continuously increase until it is used in every sentence with which its meaning is compatible, or its frequency may remain stable, or even slowly decrease so that the gram dies out. The token frequency of grams in discourse, then, is in general not expected to increase when they are further on the grammaticalization path.
Because many grams die out, it may be expected that after their first stage as a “real” grammatical element, the frequency of grams on average becomes lower.

In terms of the Scope Hierarchy, it is expected that those expressions that are more at the beginning of a grammaticalization path, with narrower scope, are on average more frequently used than those that are at the end of a grammaticalization path, with wider scope. This is partly due to a decrease in communicative motivation, and partly due to diachronic competition. These factors will reduce the token frequency of operators with wider scope, if they have not become obligatory. This hypothesis is in line with Bybee’s claim (1985: 23) that the relative frequency of inflectional markers depends partly on relevance, that is, the more relevant a marker is with respect to the verb, the more frequently it is expressed.

Not only the token frequency of language-specific grams, but also the number of expressions that belongs to a specific class of operators is expected to correlate with scope. The disappearance of grams because of converging semantic paths probably leads to a reduction in the number of different grams upon the grammaticalization cline; since n1-operators are assumed to be located at the starting point of grammaticalization, it is to be expected that this group has most members; there is not much overlap in meaning as the meanings are still rather specific. When items further generalize in meaning and become n2- or n3-operators, there is more overlap in meaning and more competition with other expressions so that fewer expressions remain. Secondly, a decrease in communicative motivation leads to a reduction in variation for wider scope operators. The information added by n1-operators is most specific, hence, the number of expressions needed for the different modifications of the description of the event should be largest. The information added by n2-operators is more general and hence, the number of expressions needed for the different modifications of location of the event is probably lower. Finally, the information added by n3-operators is most general and the number of expressions needed for the different modifications of the content is probably lowest. The classes of grams with wider scope, that express more general meanings, are thus expected to be smaller than the classes of grams with narrower scope, that express more specific meanings.

Finally, it may be expected that the crosslinguistic occurrence of operators correlates with scope. Since the function of n1-operators is communicatively most motivated, most languages will have grammatical expressions that fulfil this function, whereas grammatical expression of the less motivated functions of n2- or n3-operators should be less widespread.

In sum, the research questions and hypotheses on frequency are:
Q2a: *What is the token frequency of TMA expressions?*
H2a: The token frequency of $\pi_1$-operators is higher than or equal to the frequency of $\pi_2$-operators and the frequency of $\pi_2$-operators is higher than or equal to the frequency of $\pi_3$-operators.

Q2b: *What is the size of operator classes?*
H2b: The class of $\pi_1$-operators has more members than or an equal number of members as the class of $\pi_2$-operators and this class has more members than or an equal number of members as the class of $\pi_3$-operators.

Q2c: *What is the crosslinguistic incidence of operator types?*
H2c: The crosslinguistic incidence of $\pi_1$-operators is larger than or equal to the incidence of $\pi_2$-operators, which in turn is larger than or equal to the incidence of $\pi_3$-operators.

5.4.3 Synchronic configurations
What are the possible synchronic configurations of operators? First, a general hypothesis can be formulated with respect to possible variation in TMA systems. A basic assumption in this thesis is that $\pi_3$-operators are more marked than $\pi_2$-operators and $\pi_2$-operators are more marked than $\pi_1$-operators because the function of operators with wider scope is communicatively less motivated and more complex. For many linguistic features, it is the case that languages only have more marked features if they also have the unmarked features. With respect to operators, this leads to the expectation that modification with more abstract meaning and wider scope can only be expressed grammatically in a language as long as modification with less abstract meaning and narrower scope can also be expressed by grammatical means. For the synchronic inventory of operators within a single language this implies that a language does not allow grammatical modification of the proposition if it does not also allow grammatical modification of the predication and this is in turn not allowed if the language does not also allow grammatical modification of the predicate. In other words, there will be $\pi_3$-operators only if there are also $\pi_2$-operators, and $\pi_2$-operators only if there are also $\pi_1$-operators. This implication does not hold the other way around. This results in hypothesis 3a:

Q3: *What are possible synchronic configurations for TMA expressions?*
H3a: The presence of operators with wider scope depends on the presence of operators with narrower scope according to the hierarchy: $\pi_1$-operator $\subset$ $\pi_2$-operator $\subset$ $\pi_3$-operator.
Second, a hypothesis may be formulated that is restricted to single linguistic elements. As already stated in section 5.2, similar developments in grammaticalization are reported crosslinguistically. Human beings that speak different languages make similar inferences; they relate new meanings to earlier meanings in similar ways. Apparently, there are universals in what human beings consider contiguous senses.

The claim is that the abstract meanings of grammatical constructions arise from common patterns of inference. The types of meanings that arise in this way suggest that hearers are commonly working to infer as much as possible about the relations of narrated descriptions to the current speech situation and to the speaker’s subjective evaluation of it. (Bybee 2003a: 156)

This leads to the prediction that certain semantic changes should be possible or probable, while others will not occur. In order to clarify relations between senses, universal semantic or conceptual spaces are described (see 1.2):

A semantic map is a geometric representation of meanings or, if one likes, uses, and of the relations between them. Meanings/uses and their connections thus constitute a semantic space. As employed in linguistic typology, the map describes and constrains languages that venture their grammars and/or lexicons into this space, both with respect to diachrony and synchrony (Van der Auwera & Plungian 1998: 86).

I assume that the semantic development of a gram is determined by conceptual continua. It is language-specific which categories are expressed by the same item, but the division is not arbitrary: conceptual spaces are universal. I adopt the hypothesis of Van der Auwera & Plungian (1998: 113) that if a single gram expresses more than one meaning, it may only express adjacent meanings or fields on a single developmental path. This is illustrated in Figure 5-2, where two possible synchronic configurations are presented. The bold meanings are the interpretations that are currently used in the language. The dotted meanings on the left side have died out; the dotted meanings on the right side have not (yet) emerged. It is predicted that meanings to the right of a single path can only develop out of their adjacent meanings on the left. Furthermore, a meaning will not die out if its adjacent meanings on the right and on the left side are still in use.

Both configurations in Figure 5-2 are expected to occur: In the top configuration, one expression covers the adjacent meanings 2, 3, 4, and 5 that are on a single path. In the bottom configuration, meaning 4 and 5 are on different grammaticalization paths than meaning D and E. So here as well one expression only covers adjacent meanings on single paths. Meaning 3 has died out.
Figure 5-2. POSSIBLE synchronic configurations of polysemy (adapted from Figure 20, 21, Van der Auwera & Plungian 1998)

Figure 5-3. IMPOSSIBLE synchronic configurations of polysemy (adapted from Figure 22, Van der Auwera & Plungian 1998)

Figure 5-3 presents two configurations that are expected to be impossible. In the first configuration, the absence of Meaning 4 is disallowed: there is a gap in
the path from Meaning 2 to Meaning 3 to Meaning 4 to Meaning 5. In the second configuration, the presence of Meaning 1 in combination with the absence of Meaning 2 and 3 is disallowed. There is no continuous line of adjacent meanings.

What does the adjacency hypothesis imply with respect to the scope of operators? The hierarchical structure of the clause and the location of operators in the FG model are supposed to play a role in the semantic or conceptual continua that relate to TMA. Expressions for adjacent notions on a semantic path will have similar scope or show a minimal difference in scope. Single TMA expressions will only express multiple semantic notions that are adjacent on a path and as semantic paths probably show a gradual increase in scope, there will be no single gram that can modify the predicate and the proposition but not the predication. This prediction not only holds for polysemous items, but also for portmanteau expressions: two forms will only fuse into a single form expressing multiple meanings if they co-occur very frequently. It is assumed that the categories with adjacent scope (π1- and π2-operators or π2- and π3-operators) are conceptually stronger related than categories with non-adjacent scope (π1- and π3-operators); categories with adjacent scope will, therefore, more often co-occur than categories with non-adjacent scope and, as a consequence, have more chance to fuse.

In sum, this leads to a second hypothesis with respect to synchronic configurations:

Q3: What are possible synchronic configurations for TMA expressions?
H3b: A single expression can only cover adjacent regions in semantic space. As a result, a polysemous or portmanteau expression will only have semantic functions with similar or adjacent scopes.

5.4.4 Expression form

The Scope Hierarchy may also be reflected in the morphology of TMA expressions. It was discussed in section 5.2 that grammaticalizing items undergo various morphosyntactic and phonological changes. They tend to become phonologically reduced (in length and in phonological contrasts), to lose properties of the original lexical category (verb, noun etcetera, ‘decategorialization’) and to get a more fixed position with respect to other items or even fuse with other items (rigidification). It may be hypothesized that an item further on a grammaticalization path exhibits an equal or higher degree of formal grammaticalization than an item earlier on this same path (cf. Van der Auwera & Plungian 1998: 115-16).

The most common grammatical expression forms are periphrases, auxiliaries, particles, inflection, stem-or tone changes and reduplication. How are these
expression forms related to the degree of formal grammaticalization? Although it is claimed that more grammaticalized items are more fused (Bybee et al. 1994: 40), the general language type, whether it is isolating on the one hand or fusional or agglutinative on the other hand, may be of great influence on the highest possible degree of formal grammaticalization. It may be expected that within isolating languages the major development runs from periphrasis or auxiliary to particle, whereas in fusional or agglutinative languages, the major development is from periphrasis or auxiliary to inflection, the latter types being more reduced, having a more rigid position and showing least lexical properties.

The above hypothesis cannot straightforwardly be translated in terms of scope. Although the diachronic development of operators is expected to run from narrow to wide scope and not the other way around, not every grammaticalization path necessarily runs from $\pi_1$-operator up to $\pi_3$-operator. There is thus no one-to-one relationship between age and scope of grams. First, when lexical elements become grammatical they may be used as $\pi_2$- or $\pi_3$-operator immediately, without going through a stage in which they function as $\pi_1$-operator. Furthermore, grammatical elements may be very stable in a language: for example, past and present tense markers in English are among the oldest grams, and they have been used to encode tense for centuries. In this function, they have medial scope and as the diachronic paths in Chapter 7 will show, these markers are not expected to develop senses with wide scope. It is also not the case that an increase in generality of meaning necessarily correlates with scope, since the meaning of an item may generalize while the scope remains the same. For example, a progressive marker may develop into an imperfective marker, which is more general in meaning (it also includes continuous interpretation) but has the same scope. In spite of these objections, I pose as a working hypothesis that wider scope correlates with a higher degree of formal grammaticalization. By formal grammaticalization I understand a more rigid syntactic position, less lexical properties and phonologically more reduced items.

To conclude, this leads to question and hypothesis 4:

Q4: What is the expression form of TMA expressions?

H4: Operators with wider scope show a higher or equal degree of formal grammaticalization than operators with narrower scope.

### 5.4.5 Expression order

A final domain in adult languages in which the scope of operators may be reflected is the syntactic order of TMA expressions. In FG it is supposed that constituent ordering adheres to certain universal and potentially competing
principles. The first general principle that is relevant here is the Principle of Iconic Ordering (cf. Haiman 1983, 1985):

Constituents conform to [the Principle of Iconic Ordering] when their ordering in one way or another iconically reflects the semantic content of the expression in which they occur. (Dik 1997a: 399)

A second principle relevant to TMA expressions is the Principle of Centripetal Orientation:

Constituents conform to [the Principle of Centripetal Orientation] when their ordering is determined by their relative distance from the head, which may lead to “mirror-image” ordering around the head. (Dik 1997a: 401)

The constituents x, y, and z comply with this principle if they are expressed with respect to a Head constituent as for example zyxH, zyHx, zHxy or Hxyz. Both ordering principles predict that the hierarchical relations of operators will be represented in their relative ordering with respect to the predicate, as follows: \( \pi_3 \pi_2 \pi_1 \) Predicate \( \pi_1 \pi_2 \pi_3 \) (cf. Hengeveld 1989: 141). This hypothesis is a refinement of prediction II in Bybee (1985: 24): ‘If linguistic expression is iconic, then we would predict that the categories that are more relevant to the verb will occur closer to the stem than those that are less relevant.’

Q5: What is the expression order of TMA expressions?
H5: The expression order of TMA expressions iconically reflects the scope relations as follows: \( \pi_3 \pi_2 \pi_1 \) Predicate \( \pi_1 \pi_2 \pi_3 \)

5.4.6 Language acquisition

Hypothesis 3a and 3b lead to specific expectations with respect to language acquisition. The general question to be investigated in this thesis is whether the limits on possible synchronic configurations of linguistic properties in languages also hold for stages of child language. It is expected that this is in fact the case, since there are similar underlying “true” universals, that is, universal cognitive and communicative factors. As was already argued in 1.3, only if the acquisition of linguistic elements follows the order of a markedness scale, from least marked to more marked, the language system of a child is in each stage in accordance with the predicted limits on synchronic variation. With respect to TMA, this leads to Q6 and H6:

Q6: What is the acquisition order of TMA expressions?
H6: The acquisition order of TMA operators follows the order of the Scope Hierarchy, \( \pi_1 \)-operator \( \subset \pi_2 \)-operator \( \subset \pi_3 \)-operator. Because of increasing complexity and decreasing communicative need, the
acquisition order of TMA expressions is $\pi_1$-operators before $\pi_2$-operators, and $\pi_2$-operators before $\pi_3$-operators.

Furthermore, as argued in 5.4.3, it is assumed that there are universal conceptual spaces that lead to similar diachronic paths in unrelated languages. For adult languages, it is expected that single linguistic items can only simultaneously express those meanings that are adjacent on such a path, so that at all times, the semantic relatedness is reflected in a synchronic stage of language. It may be expected that stages in child language will show the same picture, since conceptual maps are universal. This implies that also in child language, single items may only express meanings that are adjacent on a path. With respect to the acquisition order of the different meanings of polysemous items, this implies that children will acquire meanings in an order that reflects the semantic relatedness. In sum, this is formulated as:

Q7: *In what order are the different functions of polysemous items acquired?*

H7: *At any stage in language acquisition, a single expression can only cover adjacent regions in semantic space. As a result, a polysemous or portmanteau expression will only have semantic functions with similar or adjacent scopes.*

**5.5 Conclusion**

The main question of this thesis is: are the limits on variation in adult languages of the world identical to the limits on variation in stages of language acquisition? It is expected that markedness scales or implicational hierarchies are reflected both in adult language and in stages of child language. Whether this is indeed the case will be tested in the domain of TMA and a markedness scale based on scope is expected to have reflections in language variation. Thus, the more specific question in this thesis is: does the hierarchy $\pi_1$-operator $\subseteq \pi_2$-operators $\subseteq \pi_3$-operator in the domain of TMA account for the limits on variation in adult languages and in stages of child language? In this chapter, several hypotheses were derived, mainly on the basis of grammaticalization theory, with respect to the kind of linguistic reflections that may be expected. Questions 1 to 5 will be examined in one particular language, English, and in a crosslinguistic survey of 76 unrelated languages in Chapter 6 and Chapter 7, respectively. Questions 6 and 7 will be investigated in detail in English (Chapters 8 and 9) and in a more general crosslinguistic study on the acquisition of TMA expressions in 24 different languages (Chapter 10).
Part II  Typology

Introduction to Part II

Now that the theoretical foundations of TMA have been defined in Part I, language-specific systems can be analyzed. Although variation is expected in the TMA systems of different languages, it was argued in 1.2 that it is also expected that this variation is not arbitrary. One factor that is supposed to define limits on variation is the scope of operators. Operators modify different parts of an utterance and, as a consequence, they serve different semantic and communicative functions. According to the Scope Hierarchy, operators with wide scope ($\pi_3$) are more marked than operators with medial scope ($\pi_2$) which in turn are more marked than operators with narrow scope ($\pi_1$): $\pi_1$-operator $\subset$ $\pi_2$-operator $\subset$ $\pi_3$-operator. Wide scope correlates with more complex and more abstract functions (see section 2.5).

The general hypothesis in this thesis is that scope accounts for the possible variation in grammatical TMA systems both of adult languages and of stages in child language (5.5). Part II investigates the limits on variation in the TMA systems of adult languages, or, in the standard terminology, the limits on the typological variation in the domain of TMA. The question is whether the Scope Hierarchy is actually reflected in several domains of language, diachrony, frequency, synchronic configurations, expression form and expression order, such as formulated in 5.4. Chapter 6 will test the hypotheses H1 to H5 in the TMA system of English. There, the role of scope is investigated in detail. Chapter 7 will compare the TMA systems of 76 different languages to one another. This chapter will test H1 to H5 across languages and examine more generally the influence of scope on the possible variation in TMA systems.
Chapter 6
TMA in English

6.1 INTRODUCTION
The TMA system of English has been the subject of an enormous number of studies. There is no other TMA system to which so many monographs, book chapters, and articles have been devoted. Yet it cannot be claimed that the definitive description has yet been achieved. The domain of TMA is very complex. Inevitably, this chapter cannot cover all the literature on this subject. The focus here will be on the role of scope in English TMA, and many other interesting topics will be ignored. The question is: is the Scope Hierarchy reflected in English?

In Chapter 5 hypotheses were derived on the basis of the Scope Hierarchy with regard to diachrony, frequency, synchronic configurations, expression form and expression order. In this chapter, these hypotheses will be explored to see if they make the correct predictions for English. As a first step, an inventory is made of the grammatical TMA expressions in English. Criteria are given as to which expressions should be considered grammatical and belong to the set of TMA operators. For each operator in English, the semantic functions and scope will be discussed (6.2). The inventory of TMA expressions serves as the basis to examine whether English indeed shows the predicted linguistic patterns.

The hypotheses derived in Chapter 5 are tested for English. Section 6.3 discusses the diachronic development of English TMA expressions and will test H1 (see 5.4.1). In section 6.4 the frequency of TMA expressions is examined. H2a (see 5.4.2) about token frequency is tested in corpus data of adult conversations and H2b (see 5.4.2) about the size of operator classes is tested on the basis of the inventory presented in 6.2. Not only is the Scope Hierarchy expected to be reflected in diachrony and frequency, the possible synchronic configurations of a TMA system are also supposed to be restricted by the Scope Hierarchy. Hypotheses H3a and H3b (see 5.4.3) will be tested in 6.5. The inventory of TMA expressions forms the basis for testing H3a and the semantic paths that result from the diachronic survey for H3b. In section 6.6 the English expression forms of TMA expressions will be examined to see whether they are in line with H4 (see 5.4.4). Finally, section 6.7 is dedicated to the expression
order of TMA expressions, which is expected to reflect scope as formulated in H5 (see 5.4.5).

6.2 INVENTORY OF TMA EXPRESSIONS

Which TMA functions are grammatically encoded in English? Before the configuration of TMA functions in English can be determined, the expressions to be considered grammatical have to be defined. This will be the subject of 6.2.1.

In 6.2.2-6.2.6 the specific expressions and functions within the different TMA domains are discussed. As a functional model of language should account for actual language use (in contrast to sentences made up by linguists), the inventory of TMA expressions is almost entirely based on corpus data. It relies heavily on earlier work on TMA in English, in particular the corpus-based studies on modality of Coates (1983) and Palmer (1990) and the general corpus-based grammar of English of Biber, Johansson, Leech, Conrad & Finegan (1999). Examples from these sources are referred to with C for (Coates 1983), P for (Palmer 1990) and B for (Biber et al. 1999), followed by the page number. Additional data were collected from the Collins’ Wordbanks Online English Corpus (titania.cobuild.collins.co.uk/form.html), abbreviated as CWO, and from the World Wide Web (indicated by the URL). Every attempt was made to use examples from conversational data, or else, from informal registers. Although the corpus examples formed the majority of data, a few examples had to be made up in order to illustrate a certain use or interpretation. Most of the literature and corpus data available are on British English. However, as the data in this chapter will later be compared to child American English (Chapters 8 and 9), some specific properties of American English will be discussed.

6.2.1 Criteria for grammaticality

Before the TMA system of English can be defined, it must be made clear what the limitations of this study are. Firstly, this study is restricted to the domains of aspect, tense, and modality and the related domains quantification, irrealis and evidentiality, such as defined in Chapters 3 and 4. This excludes, for example, imperative and negation and constructions like try to, happen to and dare (to).

Secondly, this thesis is restricted to grammatical expressions, as opposed to lexical expressions. Section 5.2 showed that the boundaries between grammatical and lexical elements are not clear-cut. However, for this research, it has to be decided which items should be included and which excluded. Clear lexical items are adverbs, nouns and adjectives, such as for example the words probably, maybe, expectation, capacity, certain, and possible. Clear grammatical items
are the tense inflections and irregular forms for past and present tense. More difficult constructions, however, are periphrastic constructions (periphrases), with two verb-like elements, a finite verb and a non-finite verb, such as an infinitive with or without to, a past participle (-ed) or a gerund (-ing). Examples are must go, be able to see, appear to snow, stop running, be closed and have talked. Which of these expressions are grammatical and which are lexical?

Olbertz (1996: 26) defined a grammatical periphrastic construction as consisting of the inseparable combination of an auxiliarized finite verb with a specific non-finite verb in which the finite verb agrees with the first argument of the non-finite verb.¹ There are two important aspects to this definition: the finite verb is an auxiliarized verb, which means that syntactically it functions as a modifier of the main, non-finite verb. When the construction is lexical, the finite verb is the main verb and the non-finite verb functions as part of an argument or adjunct of the main verb. The second important feature is that a periphrasis functions as an inseparable construction: its meaning is not identical to the interpretation of its components. Because of both properties, a grammatical periphrasis shows different uses in language than a lexical combination. This makes it possible to distinguish between the two.

Firstly, an important characteristic of a lexical verb (in contrast to an auxiliary) is its potential to occur on its own, without a complement. Constructions in which the finite verb with non-finite complement can be replaced by the finite verb only (b-sentences), or by a finite verb and a nominal complement (c-sentences) are therefore considered lexical (Olbertz 1996: 34). Consider (1) and (2):

(1) a. I stopped working at eight.
   b. I stopped at eight.
   c. I stopped it.
(2) a. I was going to work at eight.
   b. *I was going at eight.
   c. *I was going it.

In (1), stop may occur on its own, without the complement working (1b) or with a nominal complement it (1c). It is therefore a lexical element. In contrast, be going to in (2) cannot occur without a complement (2b) or with the nominal complement it (2c) without changing its meaning and it should therefore be considered a grammatical construction. Many English verbs that express an aspectual notion, for example, stop, start, begin, finish, continue, quit, and give up, can

¹ This definition excludes causatives, since in those constructions the finite verb does not agree with the first argument of the non-finite verb (cf. Olbertz 1996: 31). For example: I make him go, He let her sleep, They got the car fixed.
occur on their own, as full lexical, intransitive verbs. They are excluded from this research.

A second reason to consider an expression lexical is if part of the construction can be modified by an adverb. If this is possible, it indicates that the syntagmatic position of the separate elements is not completely fixed and that at least the element that is modified has meaning of its own. This holds for example for the modal idioms (ha)d better and (wo)uld rather, and also for be willing to and be able to. Consider the examples of natural speech (3)-(6):

(3) Oh, you’d much better go alone. (CWO)
(4) You’d much rather play with shared memory programming. (CWO)
(5) The farmers were only too willing to go back to a price system. (CWO)
(6) How well will they be able to use this? (CWO)

The modal expressions in (3)-(6) can be modified partly. They are considered lexical and will therefore be excluded from this research.

Thirdly, a specific type of construction that looks like a periphrasis, but is not, is a raising construction. In a raising construction the subject of an embedded clause is raised to the subject position of a matrix clause (Dik 1997b: 344-46). An example is Jack seems to work, which is an instance of raising related to the construction it seems that Jack is working. Jack, the subject of the embedded clause with the main predicate work, is raised to the subject position of the matrix clause with the main predicate seem. The relation between the two predicates seem and (to) work is therefore a matter of subordination and not of periphrasis, in which the relation is one of modification. Because of this criterion, the constructions seem, appear and turn out are considered lexical and they are excluded from this research.

Fourthly, passive constructions must be distinguished from periphrases (cf. Olbertz 1996: 29). An active counterpart and the possibility to add a by-phrase are clear indicators of passive constructions. This holds for expressions like be allowed to, be obliged to, or be expected to, as shown in (7). These expressions are therefore considered lexical.

---

2 A verb that is mainly used intransitively is much less likely to grammaticalize than a verb that is mainly or only used transitively. Only a transitive construction in which a verbal phrase functions as a complement could be reanalyzed as a periphrastic construction.

3 An expression is not necessarily lexical if the construction can be modified as a whole. For example, the construction be about to can only be modified in its entirety. In the sentence He is just about to go the adverb just does not modify about, but be about to.
(7) a. I'm allowed (by him) to go.
   b. He allowed me to go.

An expression that looks like a passive, but is not, is be supposed to. It is impossible to combine it with a by-phrase (8a) or to make it active (8b).

(8) a. He’s supposed (*by them) to go
   b. *They supposed him to go.

Whether be supposed to should be considered grammatical is dependent on the final criterion, selection restrictions.

In addition to the formal criteria mentioned above, there is a fifth, semantic, criterion that is used here to distinguish lexical from grammatical constructions, i.e. selection restrictions. Since grammatical elements have general meanings, they will freely combine with different lexical elements, that is, they hardly impose selection restrictions on the elements with which they may co-occur, in contrast to lexical elements (Hengeveld 1992: 30f). However, since grammatical elements are not meaningless, they are also not compatible with every other lexeme of the language. For example, present tense inflection does not combine with a time adverb that indicates past time, an ingressive is rare with a punctual event and a modal marker expressing ability will in general not combine with an inanimate subject. On the basis of their general meaning, it is thus predictable for grammatical elements what kind of selection restrictions occur, again in contrast to the selection restrictions of lexical elements.

Which selection restrictions may be expected within the different domains? Firstly, aspect operators select part of the internal temporal structure of an event to focus on. Since the internal temporal structure of stative and punctual events is rather undefined compared to other event types, it may be expected that not all aspect operators combine freely with these types of events, or that, if they do, it results in a specific interpretation. On the contrary, it is not expected that an aspectual operator imposes selection restrictions on the first argument, since the first argument does not interfere with the aspectual semantics. The same holds for whether the predicate designates a controlled or a non-controlled event. In a controlled event the first argument has the power to determine whether or not the event will obtain, such as in John opened the door or John was sitting in the garden, versus the non-controlled events in the substance was red or the tree fell down (Dik 1997a: 112). There is no reason to assume that aspectual operators interfere with the semantic feature of control. In conclusion, only aspectual expressions that can combine with both animate and inanimate first arguments, and occur in controlled and non-controlled events are considered grammatical. The aspectual expressions that apply to this
criterion are be –ing, be going to, keep –ing, be about to, have –ed and be –ed. Not all of these expressions combine freely with every event type—*be knowing, *be jumped—but they do not show unexpected selection restrictions on the arguments or on the feature controlled /non-controlled events. Both with regard to their formal behavior and their semantic behavior, these expressions can be considered grammatical and they are included in this research. Most other aspectual verbs in English (stop, start, finish, continue, begin) were already excluded based on their independent main verb use, but moreover, it holds for most of these verbs (except for begin) that they also apply selection restrictions on their infinitival complement or first argument.

Secondly, expressions of event quantification may be expected to impose selection restrictions on the event type with which they occur as the event type has to be compatible with a notion of ‘repetition’, which is not the case for permanent states or for specific telic events (?he died repeatedly). Furthermore, the notion of habituality seems in principal to be restricted to animate arguments: only animate arguments can be involved in habitually repeated events. Other expressions of event quantification are not expected to impose selection restrictions on their arguments. In English, there are three expressions of habituality, would, will, and used to: these expressions have grammaticalized to such an extent that they are compatible with all kinds of event-types and with both animate and inanimate arguments (see 6.2.4.1 for examples and further discussion of the habitual in English). The construction keep –ing, which is later explained to be an expression for continuative aspect and for frequentative event quantification (see 6.2.2 and 6.2.4.2), cannot combine with every event-type (the event has to be repeatable), but it does not impose any selection restriction on the arguments. All these constructions can therefore be considered grammatical.

Tense operators and operators of irrealis (hypothetical and counterfactual) are not expected to impose any selection restrictions on the predicate or argument since their semantics are compatible with any type of event or argument. This is in fact the case for the past and present tense inflection and for the future tense markers will/shall. It also holds for the hypothetical constructions ‘past modal + V^’ and ‘past tense form’ and the counterfactual constructions ‘past modal + have –ed’ and ‘past perfect form’.

For expressions of modality, the matter of selection restrictions is more complicated. Event-oriented or proposition-oriented modality should neither impose selection restrictions on the predicate nor on the arguments. However, participant-oriented modality, in particular modality that ascribes the source of the modality to participant-internal characteristics, is expected to impose certain restrictions on the arguments. The notions of (weak) obligation, desirability, permission, ability, internal need (x is forced because of internal characteristics)
and volition (will) are restricted to animate beings and, except for the latter two, restricted to controlled events. Most modal expressions in English have event-oriented or proposition-oriented uses besides their participant-oriented uses and they are clearly grammaticalized. This holds for all the central modals—must, will, would, shall, should, can, could, may, might—and the constructions have to, have got to, got to, need (to), ought (to) and be supposed to. In their participant-oriented use, these constructions do impose selection restrictions, as is to be expected, but they have also much more general uses, characteristic for wider scope operators. All these expressions are therefore considered real grammatical markers. The so-called marginal modal dare is excluded as it has a questionable modal meaning. Moreover, it is very infrequent and its use is mainly limited to fixed expressions, which are both characteristics of lexical items. Both the constructions dare (to) and try to can only combine with animate first arguments and controlled events: their meaning is still very specific.

A construction that needs some closer inspection is want to. The modal use of this construction is confined to expressing will or volition. It combines only with animate first arguments, although it can be combined with controlled and non-controlled events. The question is whether want to is a lexical predicate with a clausal argument or a modal operator (still) restricted to participant-oriented modality. Krug (2000) has carried out an in-depth analysis of the historical development of want and want to in English. He gives several arguments why want to, or at least the contracted construction wanna V, has grammatical status in (American) English. First, the frequency of the verb want has tremendously increased from Middle English to present day American English (Krug 2000: 119), in particular the string want to is much more frequently used (it makes up about 60% of all instances of want in present day written American English) (p.134). In spoken language the frequency will probably be even much higher, which is also the case in spoken British English compared to written British English (p.136). Second, the string want to is often phonologically reduced to wanna, particularly in American English and by younger speakers (up to 45%) (p.161). The phonological reduction indicates that speakers analyze the construction no longer as a two-verb phrase [want] + [to V], in which want is a lexical verb, but rather as a one-verb phrase [wanna [V]], in which wanna functions as an operator. An indication that speakers apply the latter analysis is that ellipsis of to V, as in you can wait outside if you want is less common than ellipsis of only the verb: ...if she wants to, she can get in touch with you. (p.156). A third argument that want to is grammaticalizing is its syntactic behavior, which in certain respects resembles more the syntactic behavior of the central modals.

---

4 Martine Taeymans, personal communication.
Table 6-1. English TMA expressions included in this research

<table>
<thead>
<tr>
<th>Expression</th>
<th>Function</th>
<th>Function</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>–ed/irregular</td>
<td>be going to</td>
<td>might</td>
<td>will</td>
</tr>
<tr>
<td>–Ø/-s</td>
<td>be supposed to</td>
<td>must</td>
<td>would</td>
</tr>
<tr>
<td>be –ed</td>
<td>can</td>
<td>need (to)</td>
<td>wanna</td>
</tr>
<tr>
<td>be –ing</td>
<td>could</td>
<td>ought (to)</td>
<td></td>
</tr>
<tr>
<td>have –ed</td>
<td>have to</td>
<td>shall</td>
<td>and possible</td>
</tr>
<tr>
<td>keep –ing</td>
<td>(have) got to</td>
<td>should</td>
<td>combinations</td>
</tr>
<tr>
<td>be about to</td>
<td>may</td>
<td>used to</td>
<td>thereof</td>
</tr>
</tbody>
</table>

(can, could, may, must, etcetera) than that of lexical verbs with a similar token frequency as want to. It appears that ‘full verbs in general combine far more often with modals than the emerging modal want to, the difference being even more drastic when this occurs in its contracted form.’ (p.158). Furthermore, wanna is sometimes used with third person, i.e., without a third person inflection –s (p.160). Finally, the past tense of want to can be used as a polite form, such as in I wanted to ask you a question, where it does not indicate the pastness of want (Bybee 1995). This is similar to the past forms of the central modals that also yield a polite reading, whereas past tenses of lexical verbs do not get this interpretation (Krug 2000: 155). To conclude, Krug suggests that we are possibly ‘witnessing the very initial stages of a functional split between modal use and nonmodal use, very similar in type to historical changes leading to modalization.’ (p.153). In this thesis, all instances of contracted wanna are analyzed as grammatical while all other instances of want (to) are considered lexical.

In sum, Table 6-1 presents all the TMA expressions in English that are considered grammatical and that are included in this research. In the next sections, the semantic functions of these expressions will be discussed. Corpus examples will illustrate the functions of each expression form. As the definitions of specific TMA categories were already discussed in Chapters 3 and 4, the discussion here will be limited to the characteristics specific for English. The final inventory of the TMA expressions in English and the functions for which they are used serves as the basis for testing the hypotheses.

### 6.2.2 Aspect

Which aspectual functions are grammatically encoded in English? The constructions be –ing, keep –ing, be going to, be about to, have –ed and be –ed are used
to express progressive, continuative, prospective, immediate prospective, perfect and resultative aspect (see 3.3.3).

The construction *be –ing* expresses progressive aspect. It limits the predicated property of the participant to the inner part of an event and can be used in contexts of past time reference (9), present time reference (10) or future time reference (11):

(9) That’s why I was thinking I might hang on to the Volvo. (B460)
(10) What’s she doing? (B470)
(11) But she’s coming back tomorrow (B471)

A continuative selects the initial boundary and the internal part of the event, but not the terminal boundary. In English the construction *keep –ing* has continuative as one of its meanings. Especially in combination with atelic events, this construction is interpreted as a continuative. *Keep –ing* in combination with telic events may encode either a continuative or a frequentative meaning. The latter meaning will be discussed in 6.2.4.2. Clear examples with continuative interpretations are presented in (12) and (13):

(12) I don’t want to keep living with my mum. (CWO)
(13) She kept thinking while she was smiling. (CWO)

The construction *be going to* marks prospective aspect. It limits the predicated property to the pre-state of an event. See (14)-(17):

(14) Well I was going to say just about the same as that. (CWO)
(15) I was going to be called Kate if I was a girl. (B456)
(16) And he’s going to see it. (B456)
(17) When are you going to see your parents? (P145)

*Be going to* is by some authors considered a future tense marker in English, whereas I have classified it as aspect. This issue will be discussed below.

The expression *be about to* marks immediate prospective. Like *be going to* it selects the pre-state of an event, but it also indicates that the pre-state is very close in time to the initial boundary of the event. See (18) and (19):

(18) Now my colleagues were about to throw that one out as irrelevant. (CWO)
(19) This was about to melt. (CWO)
The expression *have–ed* limits the predicated property of the participant to the post-state of the event. This construction marks perfect aspect, see (20)-(23).

(20) We've been to a lot of seminars too. (B464)
(21) He’s gone home. (B465)
(22) They’ve done so much. (B465)
(23) We’ve had it since last January. (B468)

The perfect is mainly restricted to present time reference, especially in conversations. Except for *had been*, past perfect forms are very rare in conversation (Biber et al. 1999: 468). The perfect in English selects a post-state, but does not indicate by itself whether the event and the post-state of this event are close in time (*I have just arrived*) or not (*I have been in Paris once*) and it also does not specify the length of the post-state or of the event leading up to the post-state (*I've lived here all my life*). This is a matter of contextual interpretation.

The final expression in English that marks aspect is *be–ed* for resultative aspect. Resultative indicates that an ‘action in the past produces a state that persists into the present.’ (Bybee et al. 1994: 318). Like a perfect, a resultative selects a post-state, though the focus on the post-state seems to be stronger and the implication of a preceding event that led up to the post-state seems to be much weaker. Examples are presented in (24)-(26):

(24) The door was closed.
(25) He is gone.
(26) The stick is broken.

The resultative construction and the passive are equal in form in English, but only the passive construction can be combined with a *by*-phrase. If there is no *by*-phrase, then the two constructions can be distinguished using an adverb of time: with a resultative construction, the time adverb refers to the time of the resultant state, whereas with a passive, the time adverb refers to the time of the event leading up to the result (Nedjalkov & Jaxontov 1988: 49). This can be shown by an example from Jespersen (1924: 247) in (27):

(27) When I came at five, the door *was shut*, but I do not know when it *was shut*.

The first *was shut* is a resultative construction; at five, the door was in the state of being shut. The second *was shut on* the other hand is a passive construction: *when* refers to the time of the action of shutting the door.
Notice that all aspectual forms in English are in fact combined tense-aspect markers. *Be*, *keep* and *have* bear the tense marking, whereas the non-finite verb form expresses the aspectual component. The auxiliaries *be* and *have* are in spoken English mostly contracted to, for example, *'m* or *'s* or to *'ve* or *'s*.

The classification of *be going to* as prospective aspect is not uncontroversial. Biber et al. (1999) consider *be going to* as one of the possible markers for future time in English. However, in using this expression, the focus is on a current activity or state or on a present decision or intention leading to a future event. According to Bohnemeyer (personal communication) prospectives crosslinguistically are associated frequently with the expression of intentions, since intention is one of the many possible pre-states of an event: intention is, however, not entailed by the prospective. Intentional readings in English admit the specification of the time of the event, so that the argument can be described as being in the pre-state of ‘doing something at a specific time’. The temporal specification of the event is thus inside the scope of the aspectual operator and there is no entailment of realization of the event at that time, contrary to a real future tense. It is this possibility of specifying the event-time that is responsible for the tense-like appearance of *be going to*. Quirk et al. (1985: 214) show that *will* in the main clause of a conditional can in most cases not be replaced by *be going to*, which proves that the prospective is not a future tense marker:

(28) If you leave now, you'll never regret it.
(29) *If you leave now, you are never going to regret it.

The classification of *have –ed* as perfect aspect is not uncontroversial either. For example, Elsness (1997) claims that this construction marks indefinite past tense. In my view, however, the indefiniteness of the temporal location of the event is the result of the fact that it is not the event that is located in time, but the post-state of the event. It is an implicature that the event occurred prior to the post-state, but its temporal location is not specified and therefore indefinite.

In sum, both the prospective and the perfect in English have temporal sequentiality as an implicature, whereas future and past tense encode temporal sequentiality. When the notion of current relevance were in most cases not required for the use of *be going to* or *have –ed*, then the constructions would express the temporal sequentiality and they would have grammaticalized into tense-markers. But since in English the notion of current relevance is obligatory for both constructions, the encoded meaning is here considered aspectual and not temporal.
6.2.3 Tense

In most contexts, tense is obligatorily marked on the main verb in English. Only imperative and non-finite clauses are not marked for tense. The inflection on the finite verb and the auxiliary will distinguish between past, present and future tense. This will be further discussed below. The relation between tense and modality and between tense and irrealis will be discussed in 6.2.5 and 6.2.6.

6.2.3.1 Past

There are several expressions in English that mark past tense. The so-called simple past or preterit is the suffix –ed on the main verb or irregular verb forms. Examples are presented in (30)-(35):

(30) I saw him yesterday. (B467)
(31) She just shrugged her shoulders. (B460)
(32) Well I rang them up yesterday. (B453)
(33) Hey, did you read through this yet? (B463)
(34) We already gave him a down payment. (B463)
(35) Then they said: ‘Well,’ and then I realized that it was Fennite. (B467)

In non-verbal predicates the past form of the copula be marks past tense and in periphrastic constructions, the past forms of the auxiliaries be, have or keep or of modal auxiliaries mark past tense. Past tense indicates that the relevant part of the event occurred anterior to the moment of speech (ST). It is used to refer to events in the immediate and in the remote past. The simple past is also used in contexts when a past event has led to a present resultant state, especially in American English conversation (see (33)-(34)) (Biber et al. 1999: 463).

The simple past in English makes no explicit selection of the event structure, i.e., it is aspectually neutral and the exact interpretation depends on the event type with which it occurs. In combination with temporary events, past tense leads in general, but not necessarily, to a perfective interpretation: the entire temporal structure is ascribed to the argument(s) and located anterior to ST. In combination with permanent states, the past tense only locates the part that is relevant to the conversation in time with respect to ST (see 3.4.3).

6.2.3.2 Present

The morphologically unmarked (zero) verb form in English, except for the third person singular that has a suffix –s, indicates present tense5. The general

---

5 Strictly, the present tense in English is a non-past tense as in certain contexts it may be used to refer to future events. However, since future reference of the present tense is very limited I will speak of present tense.
meaning of the present tense is that the relevant part of the event overlaps the moment of speech. It is claimed, among others by Comrie (1976: 66), that the present tense is essentially imperfective. But is this indeed the case? In the proposed account it is assumed that the simple present, like the simple past in English, does not make a selection of the temporal structure of the property or relation. In other words, the simple present is aspectually neutral. Which part of the event is located in time is dependent on the topic time and the event structure. As discussed in 3.4.2 topic time is the time interval to which an assertion is confined. It determines which part of the state of affairs is relevant to the discourse, and it is only this part of the event that is located in time. The interpretation of the simple present therefore depends on the relation between topic time, event type and speech time. What are the possible interpretations of the simple present in English?

First, present tense can describe a state existing at the speech time (Biber et al. 1999: 452-54), as in (36)-(38):

(36) I want a packet of crisps. (B453)
(37) I know her very well.
(38) I think you might be wrong. (B453)

In this interpretation, the topic time interval overlaps speech time and the interval of the event overlaps topic time. In Figure 6-1, the relation between topic time (TT), speech time (ST) and event structure (solid line on the time line) is depicted. As states are conceived of as having no temporal boundaries, the event is represented as a solid line without an initial or terminal boundary.

![Figure 6-1](image)

**Figure 6-1.** Representation of simple present in combination with a stative event in English

The present forms in (36)-(38) make no selection of the temporal structure, but language users know that only the part of the state that is relevant to the discourse is located in time. The topic time interval selects the relevant part of the state and as this part overlaps ST, the present tense is used. It is a matter of
pragmatic knowledge that not the complete state is located in time, for indicating the time location of an entire state would be virtually impossible or at least pragmatically very inconvenient. Logically, it holds before, at and after the moment of speaking, but to continuously encode this fact—*I knew, know and will know her very well*—is communicatively irrelevant. Language users infer that the state may extend past the boundaries of the speech time and even the topic time interval. There is no reason to assume that the present tense marks imperfective aspect in this use.

A second use of the present tense in English is to describe present habitual behavior (39) or generic statements (40)-(41):

(39) She’s vegetarian but she *eats* chicken. (B453)
(40) When the prey is dead, the tiger drags it to a safe place and eats it.⁶
(41) Tigers eat deer, wild pigs and wild cattle.⁷

These readings generally arise with dynamic events. The time interval for which the speaker wants to assert something is in these cases much larger than ST. Since most events do not overlap the entire topic time the interpretation is that the state of affairs is repeated, such that it fills the topic time. The interpretation may be either habitual / frequentative or distributive which depends on the number and specificity of the argument(s). For example, (39) is interpreted as habitual, since a specific individual is involved; (40)-(41) are interpreted as distributive, since the first argument is generic. The state of affairs is ‘repeated’ because it holds for several individuals. Consider Figure 6-2 for a representation of the habitual or distributive⁸ interpretation of the present tense:

![Diagram](https://www.kidcyber.com.au/topics/tigers.htm)

**Figure 6-2.** Representation of habitual or distributive interpretation of simple present in English

---


⁸ For a proper representation of a distributive a three dimensional picture is necessary, since several events occur simultaneously or overlap partly.
In this interpretation the boundaries of the temporal structure of one event are included in topic time, but at the same time, the boundaries of the repeated event-sequence are not included in topic time. Note that it is not necessarily the case that one of the repeated events actually takes place at the same time interval as ST. The resulting interpretation is that the situation of the repeated events holds at an interval that overlaps the speech time interval. A habitual or distributive interpretation therefore resembles a permanent state. The relation between topic time and event structure is very similar to the one in Figure 6-2 and it does indeed yield a ‘stative’ interpretation: it is a general statement, a situation that holds at all times. The topic time determines that the sequence of events that overlaps ST is relevant to discourse, hence the present tense. The difference with a real state is that in a habitual or distributive reading there are time intervals at which the state of affairs is not actually taking place.

A third interpretation of the present is a report on an ongoing action. Examples are presented in (42)-(44):

(42) Here comes your mother. (B454)
(43) Situation: talking about a toddler.
   Oh, my goodness. There he goes. Look at him walk. (B454)
(44) Spreads it wide. (football commentary, from Mackenzie (2005))

In this interpretation the topic time is more or less confined to ST. This interpretation is possible if the interval of the state of affairs indeed more or less coincides with ST and this is only the case for momentaneous events. The effect is a reportative interpretation, very common in sports reports. In Figure 6-3 the reportative interpretation is represented:

![Figure 6-3. Representation of reportative interpretation of simple present in English](image)

A fourth possible interpretation of the simple present in English is future reference in cases of scheduled events, such as in the train leaves at five tomorrow but also in temporal adverbial phrases such as when he comes here, I’ll tell him (from Comrie (1976: 68)). In both structures the present tense cannot be interpreted as habitual. Why is the simple present used in these cases? I follow Klein’s
analysis that the topic time stretches from the present into the future and includes the time of the complete temporal structure. Apparently, the speaker wants to assert something about a time interval for which the state of affairs does not (yet) hold up to the interval for which the event holds in its entirety. The complete temporal structure of the event including the preceding temporal span is relevant to discourse and located on the time axis. Since this interval overlaps speech time, it is possible to use a present tense. The resultant reading has a sense of a scheduled or planned event. See Figure 6-4 for a representation.

![Figure 6-4. Representation of scheduled event interpretation of simple present in English](image)

The counterpart of the scheduled event interpretation is that the topic time extends into the past. The speaker wants to assert something for a time span that stretches from some point in the past up till the moment of speech. This may account for the use of the simple present for narrating past events, the so-called historic present (also Klein’s analysis, subm.). The simple present occurs in narratives to create a more vivid style. It is especially common with speech-act verbs. See (45):

(45) No. He says, are you going home tonight? He thought I was going home to my parents. (B455)

The possible interpretations of the present tense do not support the view that the present tense in English is imperfective. In combination with a punctual event, a reportative interpretation is most plausible, which is similar to a perfective view. The most common use, however, is that the topic time extends ST and that the state of affairs is either a state or a repeated dynamic event. In these uses the simple present has a sense of imperfective aspect, since only a part of the event is located in time. However, in the case of habitual and

---

9 This is not identical to a pre-state, since the event is not in any way the consequence or purpose of the preceding time interval.
distributive, it also has a sense of perfective aspect, since the temporal structure is in its entirety ascribed to the argument(s). Note that the simple present does not mark one of the above interpretations: it only marks present tense. The different aspeccal interpretations result from implicatures or inferences. I have shown that the relation between topic time, the event structure and speech time result in different interpretations of the simple present.

The difference between the possible interpretations of simple tenses in English and an imperfective and perfective marker is that the latter explicitly select a subinterval or the entire temporal structure, whereas with simple tenses, no selection is explicitly encoded. And exactly because there is no specific selection, different interpretations are possible, depending on the event type.

6.2.3.3 Future

English has a specific marker for future tense, which indicates that the relevant part of the event takes place posterior to the moment of speech. It is expressed by will or by shall for first person. In the case of indirect speech with the main clause marked for past tense, future tense is expressed by would or should (for first person). Consider (46)-(49):

(46) Well, I'll be back tomorrow. (C179)
(47) It will be lovely to see you. (C179)
(48) My babe-in-arms will be fifty-nine on my eighty-ninth birthday … the year two thousand and fifteen when I shall be ninety. (P137)
(49) The judge in the mail train robbery # eh trial # said today that it was unlikely that the jury would be able to retire to consider their verdict until late next Tuesday. (C210)

Will and would can also be used to indicate volition or prediction. These meanings will be discussed in 6.2.5.2.1 and 6.2.5.2.3. Will and would are furthermore used for the expression of habituality (6.2.4) and would for the expression of irreals (6.2.6).

6.2.4 Event quantification

6.2.4.1 Habitual

English has three markers that specifically encode habituality: used to and would are past habitual markers and will is a present habitual marker, in one of its uses. Note that the present tense can yield a habitual interpretation, but it is not a specific marker of habituality. In 6.2.1 it was discussed that habituals are in principle expected to impose selection restrictions on their first argument and
the predicate, for a habit can only be ascribed to entities that can have habits (*Houses have the habit of being big) and habits are only properties or relations that are controlled (*He has the habit of being tall). Are there still selection restrictions of this kind to the expression of used to? See (50)-(53) for some possible uses of used to:

(50) I used to read a lot. (CWO)
(51) Everyone used to call her smelly. (CWO)
(52) The flags used to be in the colors of the “Home” Club.10
(53) There used to be a playground here.11

The examples show that there are no selection restrictions on the arguments or predicates with which used to can combine: it combines with animate and inanimate subjects and with controlled and uncontrolled properties. It may therefore be concluded that used to is highly grammaticalized. The specific ‘habitual’ sense of used to in English is marginalized and the expression has become closer to a more general frequentative. It is probably a matter of world knowledge or inference that if a specific animate entity is involved in a repeated and controlled event, the most likely interpretation is that the repetition of the event is due to the habit or propensity of this entity. The representation of a past habitual in combination with an event with boundaries is presented in Figure 6-5 (see also 3.5.3):

![Figure 6-5. Representation of a past habitual in combination with an event with boundaries in English](https://www.meliorism.com/coh/sucks/8.htm)

The examples in (52) and (53), however, show that used to can also combine with permanent states, such as be in the colors of the “Home Club” and be here. In Klein’s analysis (1994: 47) of a habitual a speaker chooses to speak about a series of topic times when using a habitual marker, and for all of these topic

---

10 [www.kenaston.org/KenAston/KenAston48.htm](http://www.kenaston.org/KenAston/KenAston48.htm)
times, the event holds. It is thus not necessarily the case that the state of affairs indeed recurs, but the time interval relevant to the discourse is a sequence of intervals in the past. When the event is a (semi-)permanent state, such as in (52) and (53), the state of affairs does not recur, but the state holds at all the topic time intervals. This is represented in Figure 6-6, in which the solid line indicates a state, that apparently ended before ST. Note that the event holds at all times in the past, but since the topic time intervals determine which parts of the state are relevant to the discourse, it creates a sense of repetition. I assume that this use of a habitual expression is later in diachronic development and inferred from the interpretation in Figure 6-5. Only a generalized habitual expression may combine with states.

![Figure 6-6. Representation of a past habitual in combination with a state with a terminal boundary in English](image)

In English, it is not necessarily the case that the state of affairs is actually recurring when a habitual expression is used: in that case, the time interval relevant to the discourse is a sequence of intervals.

*Would* and *will* also freely combine with different types of predicates and arguments. These expressions are, however, less frequently used than *used to*. Examples of *would* for expressing habitual or frequentative are presented in (54)-(55):

(54) His blue eyes gazing seriously through a wisp of fair hair which *would* keep falling across his eyes. (C207)

(55) The thing was he *would* # he *would* concoct anecdotes and he’d *tell* them to me over and over again you know obviously not realising that he’d *told* them to me before. (C209)

Finally, *will* can be also used with a sense of habituality, or typical behavior (56):

(56) So one kid will say to another, one kid will make a suggestion to another, he’ll say the moon’s further away from the earth than the sun. (P136)
6.2.4.2 Frequentative

A second subcategory of quantification, the frequentative, is in English expressed by *keep –ing*. A certain event is repeated several times, not habitually and not necessarily on one occasion. The frequentative reading may arise when the auxiliary is combined with a telic event, as in (57)-(60):

(57) I keep saying to him Oh hang on a bit don’t go to XXX. (CWO)
(58) Gizzmo keeps trying to persuade me to go with her. (B741)
(59) She kept running out of the room. (B364)
(60) He kept bringing it up so in the end I said … (CWO)

In section 6.2.2 it was already shown that *keep –ing* is used as a marker of continuative aspect (\(\pi\)). This is the most common interpretation when the construction is combined with a state, compare the examples earlier given in (12)-(13). When the construction is combined with a dynamic event, however, it is often ambiguous as to which interpretation is intended. The two interpretations of *keep –ing* can be distinguished by adding an adverb of specified duration, such as *for an hour*, which is preferred with the continuative interpretation, but dubious with the frequentative reading. The frequentative, on the contrary, can be easily combined with an adverbial phrase of repetition, such as *over and over again*. Consider the following examples:

(61) He kept running for an hour /?over and over again.
(62) He kept bringing it up ?for an hour / over and over again.

*Keep running* in (61) cannot be interpreted as a frequentative, but only as a continuing action. *Kept bringing it up* in (62) is not completely impossible with *for an hour*, but then the repeated activity is continuous and uninterrupted. In particular momentaneous or punctual events in combination with *keep –ing* are ambiguous between a continuative or a frequentative reading. They allow both interpretations. Consider (63) and (64):

(63) His brake lights kept flashing on *for an hour / over and over again*. (modified example from B746, *keep > kept*)
(64) I kept doing garlic burps *for an hour / over and over again*. (modified example from B361, *keep > kept*)

The expression *keep –ing* can thus express a continuative aspect, in which case it selects part of the temporal structure to ascribe to the argument(s). It can also express a frequentative, in which case it indicates that the state of affairs was repeated, not necessarily on one occasion or with regular intervals. In that
case it indicates that the speaker does not refer to one event in the real or imaginary world, but to a range of events. The interpretation is dependent on the situation type: telic events may be interpreted as frequentative, if the intervals between the repetitions are quite long.

6.2.5 Modality

English has an elaborate modal system. As a consequence, this section is fairly long, even though the discussion is restricted to the most important aspects. This section is structured according to the subcategorization for modality discussed in 4.2. There, it was shown that the domain of modality can be described by the parameters of sense, scope and source. The different functions of the modal expressions in English are primarily discussed on the basis of their basic sense: potentiality, disposition, weak necessity or necessity. For each sense, it will be discussed which specific meanings arise according to the scope of the modal expression. If the scope is the predicate (participant-oriented modality, \( \pi_1 \)), the modal marker specifies the relation between the predicate and the argument. If the scope is the predication (event-oriented modality, \( \pi_2 \)), the actuality of the event is expressed. If the scope is the proposition (proposition-oriented modality, \( \pi_3 \)), the commitment or attitude of the speaker to the propositional content is expressed. Within each scope type, examples of different sources of modality are presented. For participant-oriented modality, the source of the modality can lie in participant-internal characteristics or in external circumstances that are either non-deontic or deontic. For event-oriented modality, the source of the modality can lie in general or deontic external circumstances, or in general (objective) knowledge about events. For proposition-oriented modality, the source can lie in the (subjective) knowledge of the speaker or in the attitude of the speaker (the speaker as a deontic source). For each specific semantic function, examples will be presented. If the example contains different modal expressions, then the modal expression(s) that illustrates the semantic function is underlined. In general, English has a few expressions for the same functions, but the dominant contexts of use differ for each expression. Distributional properties of each expression are not discussed in detail.

A general point to discuss in advance is the relation between tense and modality. The past modal forms, could, might, would and should, started out as past tense forms of can, may, will and shall. In non-epistemic contexts, these forms can still express past tense, but it is also possible that they do not refer to past time anymore, but rather express a more tentative (and therefore more polite) sense than their present modal counterpart or indicate irrealis in combination with the semantics of their present modal counterpart. The same development is taking place in the past tense forms of the newer modals, had to, wanted to,
etcetera. They are still used to express past time reference beside their modal meaning, but they are also used to express irrealis or tentativeness. In epistemic and irrealis contexts, modal markers are tenseless. Past modal forms in general carry the meanings of the present forms, but they may have developed uses that the present form lacks. For example, *should* is used epistemically, whereas *shall* is not, and *could* is used freely as an uncertainty marker whereas *can* has a very restricted use as a marker of certainty (see 6.2.5.1.3).

### 6.2.5.1 Potentiality

The first sense to be discussed is potentiality, ‘not precluded from’. The English forms used to express this sense are *can* / *could* and *may* / *might*. The combination of potentiality with the parameter scope, results in three subcategories, which will be discussed in the following sections. The application of the parameter of source leads up to a further refinement of these modal subcategories.

#### 6.2.5.1.1 Participant-oriented potentiality

When the scope of the forms *can*, *could*, *may* and *might* is the predicate, the resultant meaning is that the argument x is not precluded from PREDing. The reason why x is not precluded from PREDing, may be because of participant-internal characteristics, which raises the specific meaning of ability. *Can* and *could* are used to express ability. Core examples concern animate subjects with dynamic verbs, (65) and (66), but inanimate subjects (67), and combinations with cognition (68) and perception verbs (69) can also be interpreted as expressing ability.

(65) I can only type very slowly as I am quite a beginner. (C92)
(66) They can’t speak a word of English, of course, not a word, but, you know, they can say what they like. (P85)
(67) The plane has a built in stereo tape recorder which can play for the whole four hours it will take to fly to Majorca. (C92)
(68) They asked me and I just couldn’t refuse. (B493)
(69) I couldn’t feel my hand. (B493)

A second source for x not being precluded from PREDing may lie in participant-external non-deontic circumstances. This creates the meaning of root-possibility: x is not precluded from PREDing (mainly) because of external circumstances. *Can* and *could*, and less frequently *may* and *might* as more formal variants, encode root possibility. Consider (70)-(73):

(70) You can always say it’s just not your style. (P85)
(71) Can you pick your own trousers up? (C94)
(72) I could have gone straight there but I just couldn't get there. (C121)
(73) I am afraid this is the bank's final word. I tell you this so that you may make arrangements elsewhere if you are able to. (C141)

The third possible source of participant-oriented potentiality are participant-external, deontic circumstances. This leads to the specific meaning of permission: x is not precluded from doing something because of a deontic source. Permission is expressed by all four potentiality markers, although can and could are most frequent. See (74)-(77):

(74) Can I pinch a ciggie? – Course you can. Would you like a menthol or a plain? (P71)
(75) Could I go just briefly to describe the other two areas? (C118)
(76) If you want to recall the doctor, you may do so. (P71)
(77) May one taste? You said I might. (C156)

6.2.5.1.2 Event-oriented potentiality
When the scope of potentiality markers is the predication, the potential occurrence of an event is expressed: the event e is not precluded from occurring. The first possible source for event-oriented potentiality are general external circumstances. This category is distinct from the participant-oriented counterpart (π1) in that it concerns a non-specific or generic participant or the participant is not mentioned at all, such as in bare passives. The potentiality does not apply to a specific argument; it applies to the entire event, independently of the involved arguments. It is mainly expressed by can, but may and might can also express this meaning, as in (78)-(82). No examples were found of could in this interpretation.

(78) I mean, you can travel from Belgium to France with much less palaver than you can travel from the North to the South of Ireland. (P84)
(79) National pressure groups cannot exist without full time staffs and a regular income. (P91)
(80) Well, I'll see what can be done and give you a ring. (P84)
(81) … and will examine ways in which this may be more effectively safeguarded. (C144)
(82) In 1814 the completion of the Mons-Conde canal increased the ease with which Mons Coals might be sent to Nord. (C156)

The second possible source for event-oriented potentiality are external circumstances that are deontic. Deontic modality with scope over the
predication is distinct from deontic modality with scope over the predicate in that the deontic source is not an individual, but a general law or rule of conduct, that is independent of the participants involved. *Can, could, may and might* express that e is not precluded from occurring because of external circumstances, more specifically a general deontic source. Deontic event-oriented modality mainly combines with non-specific or generic participants, or the participants are not mentioned at all, as in bare passives. Consider (83)-(85):

(83) In the library you *can* take a book out and keep it out for a whole year unless it is recalled. (P103)
(84) Men cannot enter.
(85) No vehicle *may* be left in the University grounds during vacation. (C132)

The general law can also apply to specific participants, which was discussed in section 4.2.3.2. The third possible source of event-oriented potentiality is general knowledge (objective epistemic modality). Within the domain of potentiality *can or could* and *may or might* express epistemic possibility: the occurrence of e is not precluded because it is logically not excluded. It is objectively possible that e occurs or that e is the case. Examples of epistemic possibility are presented in (86)-(90):

(86) …and this can mean, it doesn’t always mean, it can mean, that the students are restructuring their learning … (P108)
(87) You can get quite lost in that, I think, you see. (P84)
(88) Everything they said *could* be quite true and yet it *could* still remain a good book. (P101)
(89) They *may* or may not come and connect the television on Saturday. (C134)
(90) On the other hand he may say ‘My dear fellow, of course we understand this problem and we would arrange it this way’. (P52)

The difference between epistemic event-oriented possibility and root-possibility was discussed in 4.2.2.

6.2.5.1.3 Proposition-oriented potentiality

When the scope of potentiality markers is the proposition, the general meaning arises that the speaker S does not preclude the truth of the propositional content. The source may lie in the attitude of the speaker towards the propositional content (deontic modality) or in the knowledge of the speaker
(subjective epistemic modality). Proposition-oriented deontic modality expresses the attitude of the (reported) speaker towards the proposition. The distinction between proposition-oriented deontic modality and participant-oriented deontic modality was discussed in 4.2.3.2. The principal difference is that with proposition-oriented deontic modality the participants are not in fact influenced by the deontic source (mostly the speaker). No directive speech act is performed; no permission is actually applied, no rule of conduct is active. It is only the attitude of the (reported) speaker towards the propositional content that is expressed. This use is restricted to the potentiality marker *can* in negative contexts. It is used to express that S does not appreciate that p is true. S precludes the truth of p because of his values and opinions. See (91):

(91) You *can't* have just given up painting completely, not if you had that kind of talent. (C101)

A specific use of *might* and *could* can be considered proposition-oriented deontic modality. In particular with deontic modality, expressions can have a stronger or weaker basic sense than the same expressions in other uses (cf. Van der Auwera & Plungian 1998). In (92), *might* is used to express proposition-oriented deontic modality: no actual permission is applied and the speaker only expresses his attitude towards the proposition:

(92) I don’t mind getting pin money for typing someone’s thesis but they *might* tell me so beforehand. (C162)

The interpretation of (92) is, however, different from what is to be expected in that the speaker does not so much permit that the proposition would be true, but actually desires that it should be true. This use of a potentiality marker is ironic: it is an indirect speech-act to express obligation in a mitigated way. A similar use is possible with *could*.

When the source of the potentiality lies in the knowledge of the speaker, it expresses that S does not preclude p’s truth on the basis of his or her personal knowledge, or, in other words, the speaker is uncertain about the truth. The distinction between proposition-oriented epistemic modality and event-oriented epistemic modality was discussed in 4.2.3.3. The potentiality markers *may, might* and *could* encode the speaker’s uncertainty or doubt about the truth of the proposition. See (93)-(97) for some examples:

(93) As chairman of our court he may perhaps sometimes feel that… (C138)
(94) I may have put it there # out of the way. (C137)
(95) I think I might have walked out too, from all accounts. (P64)
(96) I’m not very good on mechanical language, but one of the cylinders might be just missing a bit from the sound of it. (C150)
(97) Mambo, from Peter Ashworth’s Treadwell stables, could be the one to give King’s Probity most trouble. (C165)

When epistemic potentiality is used with negation, it indicates that the S precludes p from being true, he or she is certain that p is not true. This can be expressed by can’t (98) or couldn’t (99):

(98) I almost phoned them up and said come a bit later – and then I thought oh they’ve probably left by now – so I didn’t and – twelve thirty, now that …. can’t be them, and it was. (Coates 1995: 63)
(99) Well, it couldn’t have been in April, Milord. (Pl62)

The preclusion of the truth of the proposition is based on inference, so that it might be considered as the negative counterpart of the certainty marker must (Coates 1983: 101), that will be discussed in 6.2.5.4.3.

A specific use of may is the concessive “speech-act” use, as in (100):

(100) On the ferries I was always being told # you may have a degree but you’re no better waiter than I am. (C137)

This use falls outside the scope of this thesis as may does not function as a π1-, π2- or π3-operator, but rather as a marker of illocution, which belongs to the interpersonal level in FG (see 2.4). Table 6-2 presents the different semantic functions with the sense of potentiality.

**Table 6-2. Subtypes of potentiality, expressed by can, could, may or might**

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>Participant-internal</th>
<th>Participant-external</th>
<th>Epistemic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SCOPE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant-oriented (π1)</td>
<td>ability</td>
<td>root-possibility</td>
<td>permission</td>
</tr>
<tr>
<td>Event-oriented (π2)</td>
<td>root-possibility</td>
<td>permission</td>
<td>possibility</td>
</tr>
<tr>
<td>Proposition-oriented (π3)</td>
<td>permission</td>
<td>uncertainty</td>
<td></td>
</tr>
</tbody>
</table>
6.2.5.2 Disposition

The second general sense is disposition, ‘disposed to’. Forms that express this sense are wanna, will, shall and would. In the next sections, the different subcategories are discussed along the lines of the parameters of scope and source.

6.2.5.2.1 Participant-oriented disposition

When the scope of disposition expressions is the predicate, the meaning is that the argument x is disposed to PRED. The source for this disposition in general lies in participant-internal characteristics, which results in the meaning of volition. It is primarily expressed by wanna, and sometimes by will or would. Especially won’t can be combined with inanimate beings, by metaphor. See (101)-(105) for some examples:

(101) I wanna go home.\(^{12}\)
(102) Psst, wanna have a better puppy?\(^{13}\)
(103) I mean I don’t think the bibliography should suffer because we can’t find a publisher who will do the whole thing. (C170)
(104) The trouble is # oh for god’s sakes # the key won’t go in the lock. (C173)
(105) I rang up and said oh I’ve finished now# would you come and get the machine? (C213)

Disposition can also have its source in a deontic authority, expressed by will. However, in these uses, its sense is closer to obligation than to desirability. First, it may be used in questions as an indirect expression of deontic modality. Consider (106):

(106) Will you listen to me and stop interrupting? (C172)

In its literal meaning (106) questions the volition of the addressee. However, by convention, it functions as an indirect expression of obligation. Like might in (92) the basic sense of the form is indirectly used to express a stronger sense.

Furthermore, will can be used deontically in declarative sentences with second person. In this use, it also does not express a mild sense of desirability, but a strong sense of obligation. See (107):

(107) You will listen to me, do what I say, and keep quiet!\(^{14}\)

\(^{12}\) [www.imc.org/atom-syntax/mail-archive/msg15517.html]
\(^{13}\) [www.cyberpet.com/cyberdog/articles/general/rotq5.htm]
In my view, deontic *will* might have originated from its use as a future tense marker. To describe an event in which the addressee is involved as (definitely) taking place in the future, implies that the addressee has no choice, hence a strong deontic sense of obligation. At the present time, though, I have no concrete evidence to support this idea.

### 6.2.5.2.2 Event-oriented disposition

Event-oriented disposition expresses that the event e is disposed to occur. As argued in 4.3.1 this is nearly similar to the meaning of future tense, expressed by *will* (or *shall* for first person). Examples of this use were discussed in 6.2.3.3.

A special use intermediate between volition (participant-oriented) and future tense (event-oriented) is intention. When the first person singular is used in combination with *will*, it may at the same time express that x is disposed to PRED and that e is disposed to occur. See (108)-(110):

(108) I’ll keep an eye open for you (C174)
(109) But I will bring you more today I promise (C174)
(110) I’ll take the old car and then you’ve got the other one (C175)

According to Coates (1983: 173), with intention uses, the focus is not on the ‘subject’s state of mind’, but on the predication. Accordingly, intention uses are in this thesis analyzed as future tense, although, strictly, they may be considered in between modality and future tense. There are no event-oriented modal uses of *will* and *would*, except for its future tense use.

### 6.2.5.2.3 Proposition-oriented disposition

Finally, *will* and *would* can express proposition-oriented disposition. This indicates prediction: S is disposed to conclude p. Future tense can be distinguished from prediction in that future tense indicates that an event will take place in the future, whereas prediction indicates that a proposition will turn out to be true in the future, whereas the event itself can take place in the present or future, or even in the past, in which case it is marked for past by the perfect, (114) and (115). *Would* indicates a more tentative prediction than *will*. Consider (111)-(115):

(111) Tell him Professor Cressage is involved – he will know Professor Cressage. (P57)
(112) They’ll probably be bored with me anyway. (C179)

---

(113) He wouldn't know exactly where it came down, but he might well have a rough idea. (P68)
(114) And my mother was not [drunk]. Several people in the house will have said that to you. (C179)
(115) Let me see # when would he have been born. (C217)

When the event itself takes place in the future, utterances can be ambiguous between a future and a prediction interpretation. See (116) and (117). This ambiguity is comparable to event-oriented and proposition-oriented modality: does the speaker present the information as an objective or subjective statement? The correct interpretation can only be given in the context.

(116) He'll burn himself out if he goes on at this rate. (C227)
(117) It is a fairly safe bet that one of the guests will want to take the empty flask home; they make delightful lamp bases. (C178)

Table 6-3 summarizes the different semantic functions with the basic sense disposition.

Table 6-3. Subtypes of DISPOSITION, expressed by will, would, or wanna

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>Participant-internal</th>
<th>Participant-external</th>
<th>Epistemic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-deontic</td>
<td>Deontic</td>
<td></td>
</tr>
<tr>
<td>Participant-oriented ((\pi_1))</td>
<td>volition</td>
<td>-</td>
<td>“desirability”</td>
</tr>
<tr>
<td>Event-oriented ((\pi_2))</td>
<td>-</td>
<td>-</td>
<td>(future)</td>
</tr>
<tr>
<td>Proposition-oriented ((\pi_3))</td>
<td>-</td>
<td>prediction</td>
<td></td>
</tr>
</tbody>
</table>

6.2.5.3 Weak necessity

The third basic sense that is discussed is weak necessity. This sense could be paraphrased as ‘expected to’. English has different forms to express this sense: should, ought to, and be supposed to are the most important ones. One use of shall may also be covered under the sense of weak necessity.
6.2.5.3.1 Participant-oriented weak necessity

When the scope of weak necessity is the predicate, it expresses that x is expected to PRED. The source for this expectation can only lie in participant-external circumstances. So English has no expressions for ‘weak internal need’. The external circumstances can be deontic or non-deontic, but especially within weak necessity, this is not always clear. Examples (118)-(121) can be interpreted as non-deontic circumstances; x is expected to PRED because of general external circumstances:

(118) A: What you been doing? B: Well, I shouldn't be here. I ought to be on holiday today. (C82)
(119) We ought to return the cassette some time fairly soon you see. (C71)
(120) I left a note in my pigeon hole which should be in the student's pigeon hole, so I'd be grateful if you could move it …(C65)
(121) She's supposed to be coming in. (B500)

In particular for expressions of weak necessity it is not always clear whether the external circumstances should be considered a deontic source or not. In the above examples (118) and (120) the participants seem to be forced by a sort of organizational plan or schedule, which is closely related to a general deontic source, such as a rule of conduct or a law (which is analyzed as event-oriented deontic modality, see 6.2.5.3.2). However, the specific participants in (118) and (120) cannot be replaced by a non-specific participant, so that the modal expressions are analyzed as participant-oriented.

Examples (122)-(125) below can be interpreted as expressing weak obligation: x is expected to PRED because of a deontic source. This category is intermediate between permission and obligation. Dependent on the status difference between the deontic source and the participant—thus not on the expression chosen—the strength of the desirability runs from a mere advice or invitation to a weak or even rather strong obligation. Compared to the obligation markers in the necessity domain, this type of obligation is a mitigated form. When it is used with speaker-reference the speaker expresses self-exhortation.

(122) You should walk round the ramparts of the old city too. (C58)
(123) A: If I put a Marvel [tin] in his wastepaper basket will he think it unaesthetic. 
      B: I shouldn’t worry. (C222)
(124) So perhaps I ought to ask you some further questions. (P123)
(125) You really ought to be buying something a bit more modern and a bit more expensive. (C72)
6.2.5.3.2 Event-oriented weak necessity

When the scope of weak necessity is the predication, it comes to mean that the event e is expected to occur. The source of the necessity may first of all lie in general non-deontic circumstances. This may apply to utterances with non-specific subjects or passives, such as ought to in (126) and should in (127), although both examples leave open the possibility of a deontic interpretation. Be supposed to can also be interpreted in the sense that e is expected to occur because of general, unspecified circumstances. It means something like ‘it is meant to’. Consider (126)-(129):

(126) The job here ought to be finished in a matter of days. (C73)
(127) This should be done before the pollen is ripe. (P123)
(128) My class is supposed to start in fifteen minutes. (CWO)
(129) Are you supposed to be going with him on holiday? (B500)

The source of event-oriented weak necessity may also be a general deontic source, such as a rule of conduct or a law. Examples are presented in (130)-(133):

(130) I just insisted very firmly on calling her Miss Tillman but one should really call her President. (C59)
(131) I don’t think you ought to … I’ve a sort of feeling you shouldn’t ask. (C81)
(132) But they can’t decide yet whether the new man should be appointed by Freeman (…) or whether the appointment should be made by whoever takes his place. (C59)
(133) As Americans, we believe that one is supposed to vote and participate in the democratic process.15

With first person subjects it is not always clear whether the speaker is the deontic source and the utterance is an instance of self-exhortation or whether the speaker cites a general rule of conduct, which also applies to him/herself. This is illustrated in (134) and (135):

(134) I ought to be ashamed to say so, but I can’t. (P123)
(135) I think I ought to say something about my step father. (C71)

The final possible source of event-oriented weak necessity is general knowledge. Should, ought to and supposed to express probability: e is expected to

15 foucault.info/Foucault-L/archive/msg08657.shtml
occur or to be the case because of general knowledge. Objectively, the occurrence of the event is probable. Consider (136)-(138):

(136) Lad says if it is on the delivery note it should have been delivered. (C226)
(137) It ought to be I suppose in those two # those double grey filling cabinets you remember. (P74)
(138) But medical supplies are supposed to be getting through. (CWO)

6.2.5.3.3 Proposition-oriented weak necessity

Finally, when weak necessity modifies the proposition, it expresses that S expects p’s truth or expects p to be true. The first possible ground for expecting the truth of the proposition is that the speaker infers the propositional content on the basis of his or her personal knowledge (subjective epistemic modality). This is mainly expressed by should, as in (139)-(142):

(139) Have sent off my diary a couple of days ago – you should get it soon. (C58)
(140) So he should be around sort of between half past two and half past three. (P60)
(141) The trip should take about sixteen days. (C64)
(142) That shouldn’t be difficult. What do you want to know? (P62)

A second ground for expecting the truth of the proposition are the speaker’s own values and opinions about how the world should be. This is proposition-oriented deontic modality. The weak obligation is not actually applied to the participants; it is only the personal opinion of the speaker that the proposition should be true (4.2.3.3). Consider (143)-(144):

(143) The unemployed, they should be made to do some some work and not scrounge off the state. (C59)
(144) I think people ought to be better informed about what marriage entails. (P123)

A use of should which falls outside the scope of this thesis is its quasi-subjunctive reading, as in (145)-(146):

(145) It is most necessary that we should have the funeral bill. (P68)
(146) It was inevitable that Peter Ustinov should join the exclusive four-star club by writing, producing, directing and starring one film. (P68)
This use belongs to the interpersonal level in FG (see 2.4) and is further excluded.

To resume, the specific semantic functions with the basic sense of weak necessity are presented in Table 6-4.

Table 6-4. Subtypes of weak necessity, expressed by should, ought to, or be supposed to

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>Participant -internal</th>
<th>Participant-external</th>
<th>Epistemic</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCOPE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-deontic</td>
<td>Deontic</td>
</tr>
<tr>
<td>Participant-oriented (π₁)</td>
<td>-</td>
<td>weak necessity</td>
<td>weak obligation</td>
</tr>
<tr>
<td>Event-oriented (π₂)</td>
<td></td>
<td>weak necessity</td>
<td>weak obligation</td>
</tr>
<tr>
<td>Proposition-oriented (π₃)</td>
<td></td>
<td>weak obligation</td>
<td>weak certainty</td>
</tr>
</tbody>
</table>

6.2.5.4 Necessity

The final basic modal sense is necessity, expressing ‘forced to’. There are several expressions in English with this basic sense: need (to), have to, (have) got to, must and shall.

6.2.5.4.1 Participant-oriented necessity

When the scope of necessity is the predicate, it expresses that the argument x is forced to PRED. The source of this force may be internal characteristics of the argument, which expresses the meaning of internal need: x is forced by internal/intrinsic factors to PRED. It is expressed by need to, have to and have got to. See (147)-(150):

(147) Boris needs to sleep ten hours every night for him to function properly. (Van der Auwera & Plungian 1998: 80)
(148) I need to be left in peace today. (Van der Auwera & Plungian 1998: 83)
(149) I’ve got to eat something, otherwise I’ll just die.₁⁶
(150) I’m going to make this short since it’s way past midnight and I really have to sleep.₁⁷

More commonly than purely by internal characteristics, it is expressed that \( x \) is forced to PRED (mainly) by external circumstances. These circumstances may be non-deontic. This use is illustrated by (151)-(153):

(151) I’ve got to be at London airport at fourish. (P114)
(152) We had to make a special trip down to Epsom to collect the bloody thing. (P120)
(153) I may need to stay a couple of nights at Minna before I can find transport for the last 60 miles or so. (P127)

It is also possible that the external circumstances are deontic: \( x \) is forced by a deontic source to PRED. The necessity markers \textit{have to}, \textit{have got to}, and sometimes \textit{must} or \textit{need} express obligation. Consider (154)-(157):

(154) Mum says I have to wait. (Krug 2000: 92)
(155) Have we got to go on this hike? (C52)
(156) You must tell me how to get to it. (C34)
(157) “I’m very grateful to you.” “You needn’t be. I told you. I’m glad to do it.” (C50)

English makes a distinction in deontic expressions that indicate that the speaker is the deontic source or that it is some other deontic source. In \textit{You must come and see me tomorrow} (Palmer 2001: 75), it is the speaker who is the deontic source, whereas with \textit{you have to come and see me tomorrow} there is an obligation that is independent on the speaker. The same distinction holds for the opposition between \textit{should} and \textit{ought to} on the one hand, where the speaker is the responsible deontic source in contrast to \textit{be supposed to}, where the speaker is not responsible for the weak obligation.

Finally, the expression \textit{shall} can be used as a deontic expression of obligation. See the example in (158):

(158) You shall go, I insist on it. (Sweetser 1990: 55)

The notion of necessity or obligation is one of the oldest uses of \textit{shall} and it grammaticalized to a future meaning (see 6.3.3). \textit{Shall} can also be used deontically in questions such as (159), in which case the addressee is the deontic

\footnotesize
\textsuperscript{17} \url{www.infektia.net}
\textsuperscript{18} Contrary to Bybee et al. (1994: 320) I use the term obligation only if the external circumstances are identified as a deontic source and not for all types of external circumstances.
source. However, in these constructions its sense is much weaker, and closer to weak obligation or desirability than to strong obligation.

(159) What shall I choose for Ann if she wins the raffle? (CWO)

These uses are the only remnants of the modal uses of *shall*. Its past counterpart *should* developed to a marker of weak necessity that is fairly frequent.

6.2.5.4.2 Event-oriented necessity

When the scope of necessity is the predication, it means that the event $e$ is forced to occur. It may be forced to occur because of general external circumstances. It occurs in contexts with non-specific participants or passives. Consider (160)-(163):

(160) Clay pots must have some protection from severe weather. (C35)
(161) A really healthy effective opposition which you’ve got to have if you’re going to shake the government. (C53)
(162) He’s going on the 7.40 tomorrow morning and everything must be packed tonight. (C35)
(163) He fully understands that the thing has to be reprogrammed every year. (C55)

It is also possible that $e$ is forced to occur because of a deontic source, such as a general rule or law. This is in general expressed by *have to*, *have got to* or *must*: $e$ is forced to occur because of a general deontic source. Examples are presented in (164)-(168):

(164) If you commit murder, Charlotte, you must be punished. (C34)
(165) In the whole of southern Africa one has to drive on the left side of the road.\(^{19}\)
(166) You haven’t got to park on double yellow lines – it’s against the law. (Perkins 1983: 62)
(167) You have to listen carefully and look for clues from three sources.\(^{20}\)
(168) There is already a great imbalance between what a student has to pay if he’s in lodgings and what he has to pay if he is in a hall of residence. (C55)

*Shall* can also be used in this sense, although it is infrequent. See (169):

---

19 www.namibia-travel.net/namibia/a_z.htm
20 www.pinkmonkey.com/electricdesk/studysmart/ssmart5.asp
(169) (The law decrees that) all citizens shall constantly carry violet parasols from 3/9/83 on. (Sweetser 1990: 55)

The third possible source of event-oriented necessity can be knowledge, although this use is restricted to negated necessity: e is not forced to occur. It seems communicatively irrelevant to encode that the event is necessarily actual, since that is the unmarked communicative situation (see 4.2.3.3). Negated epistemic necessity is semantically the opposite of epistemic possibility: ‘not necessarily e’ is similar to ‘possibly not e’. This meaning can be expressed by need not or don’t have to. See (170)-(171):

(170) Oh gosh, getting married is an awfully complicated business. [other speakers argue] Actually it needn’t be – it can be very straight forward. (C50)
(171) He doesn’t have to be at home: he could have gone straight to Caroline.

6.2.5.4.3 Proposition-oriented necessity

Finally, when the scope of necessity is the proposition, it means that the speaker S is forced to conclude the propositional content p. This is most commonly expressed by must, but also by have to or have got to. S may be forced to conclude p because of the available premises. In that case it expresses certainty about the truth of the proposition. Consider (172)-(175):

(172) She must have been such a pain in the neck to her Mum and vice versa. (C44)
(173) This has to be the biggest ant-hill ever seen. (Perkins 1983: 61)
(174) You’ve got to be joking. (P56)
(175) It had to be there – there wasn’t anywhere else it could have been. (P65)

Note that event-oriented epistemic necessity is always negated (see (170)-(171) and 4.2.3.3), whereas proposition-oriented epistemic necessity on the other hand is always positive. The speaker states that from the available premises, he is forced to conclude p. If the speaker wants to state that he is certain that p is not true, can’t is used as in (98) above.

It is also possible that S may be forced to conclude p because of his personal opinions and values. In that case the speaker expresses that he is certain that p should be true. In this interpretation the modal expresses proposition-oriented deontic modality (4.2.3.3). There is no obligation actually laid on the
participants: it is only the speaker’s attitude towards the proposition that is expressed. S states that p must be true in his opinion. See (176) and (177):

(176) The government *must* act. It *must* make up its mind about priorities – offices or houses, housing estates or luxury buildings. (P106)
(177) This I think is something on which universities *have got to* begin now to take a stand on. (C53)

Table 6-5 presents an overview of the different functions that can be expressed with the basic sense of necessity.

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>Participant-internal</th>
<th>Participant-external</th>
<th>Epistemic</th>
</tr>
</thead>
<tbody>
<tr>
<td>NON-DEONTIC</td>
<td>Deontic</td>
<td>Deontic</td>
<td>(no) necessity</td>
</tr>
<tr>
<td>Participant-oriented (π₁)</td>
<td>internal need</td>
<td>root necessity</td>
<td>obligation</td>
</tr>
<tr>
<td>Event-oriented (π₂)</td>
<td>root necessity</td>
<td>obligation</td>
<td>(no) necessity</td>
</tr>
<tr>
<td>Proposition-oriented (π₃)</td>
<td>obligation</td>
<td>certainty</td>
<td>certainty</td>
</tr>
</tbody>
</table>

### 6.2.6 Irrealis

English has no simple morpheme that marks irrealis. There are, however, specific syntactic constructions that indicate irrealis. The construction to mark hypothesis consists of a subordinate conditional clause (*if*-clause) and a main clause. The conditional clause contains the condition-marker *if* and a verb with past tense-morphology: the past form does however not mark past tense, but indicates irrealis in this construction. The condition is often explicated, but can also be left implicit. The main clause contains a secondary or ‘past’ modal, which also does not indicate past tense, but irrealis. The most neutral one is *would* in most cases it only marks irrealis, but sometimes it also has a sense of volition. The other secondary modals—*could, should, might*—can also be used. Apart from marking irrealis, they bear the semantics of their primary modal
counterparts: *can*, *shall*²¹ or *may*, discussed in the previous section. Examples of the hypothetical construction are presented in (178)-(181):

(178) If I acted like that in front of him I # I mean that would be # that would be inexcusable. (C213)
(179) If he came to live with us then she would have to give up work. (C229)
(180) Anybody could produce at least two test teams of English writers who would be much better worth celebrating than Burns. (C215)
(181) That would apply to Swift too # wouldn’t it? (C217)

Counterfactual constructions in English consists of a conditional clause with a main verb with ‘past perfect’-morphology, and of a matrix clause that contains a secondary modal plus a main verb with perfect-morphology. Once again, *would* is the most neutral modal, in most cases only indicating irrealis, whereas the other secondary modals—*should*, *might*, *could*—also add the semantics of their primary modal counterpart. Consider (182)-(184):

(182) In fact, I *would have said* that it looks as though London would be worth going through. (P173)
(183) I *would have been* very surprised really if you *had got* a British Academy award. (C214)
(184) If I *could have thought* of that quotation I *would have used* it. (C214)

6.2.7 Summary

In Chapters 3 and 4 the possible semantic functions within different TMA domains and their scopes were established. Here it was examined which English forms cover which semantic functions. It appeared that there are several expressions that are used for different functions and that similar functions are often expressed by different expressions. Although it is sometimes difficult to distinguish between semantic functions and although certain utterances may allow different interpretations, this is not an argument for rejecting the distinct semantic functions all together. On the contrary, it explains how the semantics of a specific linguistic item—as understood by language users—can change or differentiate and why processes of grammaticalization are possible. Only by virtue of implicatures and ambiguity

²¹ The overlap of the semantics of *shall* and *should* is not so neat as the overlap of the semantics of *can* and *could* or of *may* and *might*. Both *shall* and *should* have uses that the other form cannot have. There already are considerable differences between *sceall*/*scealt* and *sceold* in Old English (Goossens 1985).
between related meanings can language users change the use and meaning of a linguistic item.

This section provides a summary of the discussion by giving a list of the English operators that currently belong to the TMA system. In Table 6-6 all the TMA operators with narrow scope (π1-operators) are presented: the operators of aspect and operators of participant-oriented modality. Table 6-7 presents the TMA operators with medial scope (π2-operators): operators of tense, event quantification, event-oriented modality and irrealis. Finally, Table 6-8 presents the operators with wide scope (π3-operators), expressing proposition-oriented modality. These tables serve as the basis for examining the research questions in the next sections.

### Table 6-6. English TMA expressions with narrow scope (π1-operators)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Semantic function</th>
<th>Expression(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progressive</td>
<td></td>
<td>be –ing</td>
</tr>
<tr>
<td>Continuative</td>
<td></td>
<td>keep –ing</td>
</tr>
<tr>
<td>Prospective</td>
<td></td>
<td>be gonna / be going to</td>
</tr>
<tr>
<td>Immediate prospective</td>
<td></td>
<td>be about to</td>
</tr>
<tr>
<td>Perfect</td>
<td></td>
<td>have –ed, have + irregular form</td>
</tr>
<tr>
<td>Resultative</td>
<td></td>
<td>be –ed, be + irregular form</td>
</tr>
<tr>
<td>Ability</td>
<td></td>
<td>can, could</td>
</tr>
<tr>
<td>Root possibility</td>
<td></td>
<td>can, could, may, might</td>
</tr>
<tr>
<td>Permission</td>
<td></td>
<td>can, could, may, might</td>
</tr>
<tr>
<td>Volition</td>
<td></td>
<td>wanna, will, would</td>
</tr>
<tr>
<td>Weak root necessity</td>
<td></td>
<td>should, ought to, be supposed to</td>
</tr>
<tr>
<td>Weak obligation</td>
<td></td>
<td>should, ought to, be supposed to</td>
</tr>
<tr>
<td>Internal need</td>
<td></td>
<td>need (to), have to, have got to</td>
</tr>
<tr>
<td>Root Necessity</td>
<td></td>
<td>must, have to, have got to</td>
</tr>
<tr>
<td>Obligation</td>
<td></td>
<td>must, have to, have got to, need to, shall</td>
</tr>
</tbody>
</table>
Table 6-7. English TMA expressions with medial scope ($\pi_2$-operators)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Semantic function</th>
<th>Expression(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TENSE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past</td>
<td>−ed, irregular form</td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>−o, −s</td>
<td></td>
</tr>
<tr>
<td>Future</td>
<td>will, shall, would</td>
<td></td>
</tr>
<tr>
<td><strong>EVENT-ORIENTED MODALITY</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root possibility</td>
<td>can, could, may, might</td>
<td></td>
</tr>
<tr>
<td>Permission</td>
<td>can, could, may, might</td>
<td></td>
</tr>
<tr>
<td>Weak Necessity</td>
<td>should, ought to, be supposed to</td>
<td></td>
</tr>
<tr>
<td>Weak Obligation</td>
<td>should, ought to, be supposed to</td>
<td></td>
</tr>
<tr>
<td>Necessity</td>
<td>must, have to, have got to</td>
<td></td>
</tr>
<tr>
<td>Obligation</td>
<td>must, have to, have got to, shall</td>
<td></td>
</tr>
<tr>
<td>Epistemic possibility</td>
<td>could, may, might</td>
<td></td>
</tr>
<tr>
<td>(negated) epistemic necessity</td>
<td>need not, don't have to</td>
<td></td>
</tr>
<tr>
<td>Probability</td>
<td>should, be supposed to</td>
<td></td>
</tr>
<tr>
<td><strong>IRREALIS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypothetical</td>
<td>−ed, irregular form, past modal</td>
<td></td>
</tr>
<tr>
<td>Counterfactual</td>
<td>had −ed, past modal + −ed</td>
<td></td>
</tr>
<tr>
<td><strong>EVENT QUANTIFICATION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Habitual</td>
<td>used to, would, will</td>
<td></td>
</tr>
<tr>
<td>Frequentative</td>
<td>keep −ing</td>
<td></td>
</tr>
</tbody>
</table>

Table 6-8. English TMA expressions with wide scope ($\pi_3$-operators)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Semantic function</th>
<th>Expression(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PROPOSITION-ORIENTED MODALITY</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncertainty / certainty that not</td>
<td>could, may, might / can't, couldn't</td>
<td></td>
</tr>
<tr>
<td>Prediction</td>
<td>will, would</td>
<td></td>
</tr>
<tr>
<td>Weak certainty</td>
<td>should</td>
<td></td>
</tr>
<tr>
<td>Certainty</td>
<td>must, have to, have got to</td>
<td></td>
</tr>
<tr>
<td>“Permission” / no permission</td>
<td>might / can’t</td>
<td></td>
</tr>
<tr>
<td>Weak obligation</td>
<td>should</td>
<td></td>
</tr>
<tr>
<td>Obligation</td>
<td>must, have to, have got to</td>
<td></td>
</tr>
</tbody>
</table>
6.3 **DIACHRONY**

This section will examine whether English shows the hypothesized pattern of diachronic development: on the basis of inference, speakers will widen the semantic scope of operators to more abstract functions. The expected diachronic development of TMA expressions is formulated in H1:

H1: Diachronically, operators show an increase in scope, and develop in the direction from $\pi_1$-operator to $\pi_2$-operator to $\pi_3$-operator.

A brief discussion will be presented of the diachronic development of each TMA expression in English. Inevitably, this discussion involves strong simplifications and the reader is referred to more elaborate works on the history of modal auxiliaries and English syntax in general (Hogg 1992-2001; Lightfoot 1979; Plank 1984; Traugott 1972; Visser 1963-1973).

Reference will be made to the different stages of English that may not be familiar to all readers. The earliest stage recognized as ‘English’ is Old English, from approximately AD 450 till 1150. The Middle English period extends from about 1150 till 1500, Early Modern or New English from about 1500 till 1750 and Late Modern English or Present Day English from circa 1750 onwards. In this section, asterisks are used to indicate that a historical form is reconstructed.

### 6.3.1 Tense inflection

The tense inflection and irregular forms are the oldest TMA operators in English. The strong, irregular past forms in English stem from regular Proto-Indo-European ablaut formation (root vowel alternation). The origin of the past tense inflection –*ed* is not known (Bammesberger 1992: 59; P. Ramat 1998: 405; Shields 1992: 43). It has a long history, since all Germanic languages, in contrast to other Indo-European languages, have a dental preterit marker.

The present tense inflections, -o and –s, are the remnants of a more elaborate system of present tense person/number inflection in Proto-Indo-European. In Old English it was a general non-past marker, being the default marker for present and future time reference, whereas in Present Day English reference to future time by the simple present tense is very limited. The history of both tense markers cannot be tested to the hypothesis, as there is not enough information available.

### 6.3.2 Aspectual constructions

The resultative be –*ed* and perfect have –*ed* have their origin in the Old English lexical constructions *wesan/beon* (‘be’) + adjectival participle of intransitive verbs and *habban* (‘have’) + adjectival participle of transitive verbs. The adjectival
participle agreed with the subject (intransitive) or object (transitive) and the verbs still had their lexical meaning of existence ('be') and possession ('have'). This raised meanings like ‘they were in the state of having come’ or ‘I had him in a state of being bound.’ More and more often, the inflection on the participle was dropped which made it possible to analyze the construction as an auxiliary with a past participle. Furthermore, a word order change took place so that formerly ‘has the fish caught’ gradually became ‘has caught the fish.’ Finally, *have takes over more and more from *be as perfect auxiliary and in the 19th century it has become the default perfect auxiliary (Fischer 1992; Traugott 1972: 91-94). In sum, the developmental paths are presented in Figure 6-7:

![Figure 6-7. Developmental paths to resultative and perfect](image)

The history of the resultative and perfect constructions is in accordance with the hypothesized order: lexical constructions have developed into aspect markers.

The origin of the progressive *be –ing is not completely clear. According to Traugott (1992: 187-89) two separate constructions are underlying the later progressive. Old English had a construction *wesan/beon + present participle – ende, as in *he wæs huntende and a locative construction of a preposition *on / *an in combination with a nominal derived from a verb, ending in -ung/-ing, as in he wæs on huntunge. In several contexts, the constructions were equivalent in that they created the same meaning. Furthermore, the inflectional endings began to be confused in late Old English/Early Middle English and possible the two constructions became very similar in Middle English, i.e., *He was huntynge(e) and He was on/an/in/a huntynge(e). In the end they may have coalesced to what is the Present Day English progressive construction. Although this hypothesized development is plausible, the actual language data do not prove it (Fischer 1992). Nevertheless, what is important to this thesis is that the progressive construction *be –ing has its immediate origin in a lexical construction. See Figure 6-8:
This path supports the hypothesis that narrow scope operators do not derive from wider scope operators, but rather from lexical constructions.

The prospective construction *be going to* started as the progressive construction of the directional lexical verb *go* and a purposive clause ‘to V’, meaning *going in order to*. This led to the common inference that the subject intended to V at a later time. In the next stage, the directionality sense of *go* was demoted and the inference of intention of event realization at a later time was promoted. In the stage thereafter, the purposive sense was also demoted and *be going to* could refer to any pre-state of a coming event or state and no longer applied selection restrictions on the subject or the predicate with which it could combine (Hopper & Traugott 1993: 61, 81-83). In sum, the development is:

![Developmental path to progressive](image)

This path supports the hypothesis again: the aspect operator directly derives from a lexical construction.

There was hardly any information found about the history of the two other aspectual constructions, *keep –ing* and *be about to*. The latter construction occurs from Late Middle English (Fischer 1992), refers to planned action and has aspectual implications from the start (Rissanen 1999: 233). In all likelihood, both constructions derive from lexical constructions, not from other grammatical operators. This may be considered evidence in support of the hypothesis.

### 6.3.3 Central modals

For modal expressions, in most cases there is no indication in the literature whether a root- or deontic sense can be combined with specific and non-specific participants (the distinction between participant-oriented and event-oriented modality) and whether an epistemic meaning should be considered expressing an objective epistemic stance or the speaker’s commitment. If there is not enough information available, the root- and deontic meanings will be
conceived of as $\pi_1$- and $\pi_2$-operators and the epistemic meanings as $\pi_2$- and $\pi_3$-operators.

The forms will and would go back to the Old English verb willan (present wile, past wolde), which was the lexical verb for ‘want’. Late Old English willan developed into a grammatical marker of volition. It is occasionally used to express typical or general occurrences, close to a habitual or epistemic interpretation. Sometimes, wile is used to express senses of futurity, but this is restricted to contexts of generalizations, prophetic or inevitable future or relative future tense. These uses occur in impersonal or passive contexts. Gradually, during Middle English will (and shall) becomes a real future marker, and replaces the present tense in this use (Traugott 1989: 38-43; 1992: 195). The sense of desire is lost in the development of Early Modern English into Present Day English and the sense of volition has become much less frequent (Warner 1993: 181) The prediction use of will and would, in the sense of ‘the speaker is disposed to conclude …’, occurs only in the 19th century. (Traugott 1989: 43). In sum, the development is:

\[
\text{‘want’} \quad \rightarrow \quad \text{volition (}\pi_1\text{)} \rightarrow \quad \text{future (}\pi_2\text{)} \rightarrow \quad \text{prediction/inference (}\pi_3\text{)}
\]

**Figure 6-10. Developmental path of will**

in which the symbol ‘†’ indicates that this use has died out. The development in Figure 6-10 shows an increase in scope and is therefore in accordance with the predicted development.

The origin of shall and should lies in the Old English *sculan* (present sceal, past sceolde), a lexical verb meaning ‘owe, be necessary’, both in moral and financial respect (Traugott 1992: 195). In Old English, it developed into an auxiliary marking obligation and necessity of various kinds. At the end of Old English it acquired a sense of intention. According to Traugott (1992: 197) and Warner (1993: 170) sceal was used occasionally for future of arrangement or inevitable future, especially in impersonal constructions. During the Middle English stage, the forms shal and wil become the normal markers for future time reference. Shal is the predominant form, but wil is more typical in informal usage. Will as a future marker gradually replaces shal in Early Modern English with third and later second person subjects. In Present Day English shall is entirely restricted to use with first person and nowadays, will has even replaced shall in first person contexts in certain dialects or in informal speech (Warner 1993: 181). The development of shall supports the hypothesis. It is presented in Figure 6-11:
Some specific developments took place in the use of the past form, *sceolde*. *Should* develops at first similar to *shall*, although it grammaticalizes a bit later as a future marker, in Early Modern English (Rissanen 1999: 235). In contrast to *shall*, *should* develops further into a marker of epistemic probability in Present Day English (Bybee et al. 1994: 200). There is no detailed information available about the scope of the epistemic modality. Consider Figure 6-12:

This development is consistent with the hypothesis. There is a developmental order from lexical to $\pi_1$-operator, to $\pi_2$-operator to $\pi_3$-operator.

*Should* furthermore had a special use in Old English in indirect speech. It indicated hearsay (evidential use). This use existed into Modern English (Traugott 1989: 41-42; Warner 1993: 171, 78). It is not known how this use came into being. Therefore, it will not be used to test the hypothesis. It is however clear that this is an independent development from the one described in Figure 6-12.

The form *must* has its origin in the Old English verb *mōtan* (present *mōt*, past *mōste*). It was an auxiliarized expression marking permission and root-necessity. However, in late Old English, it developed a meaning of root-necessity and obligation, which quickly became the predominant use (13th century). The contradictory interpretations of *mōt* may have been caused by the basic sense of *mōt* ‘to be allotted’, which allows both interpretations. Another possibility is that the sense change took place in negative contexts, as ‘not possible’ means ‘necessarily not’ (Warner 1993: 160). The permission meaning gradually lost ground during the Middle English period and *may* took over the function of *mōt* in those contexts. During late Middle English the form *must* became the default form, which may have its origin in the subjunctive preterite of Old English, *mōste*, and in the second person singular present form of Old English, *mōst*. The present form *mōt* disappears in the 16th century.
In Middle English *moot* begins to be used in contexts of inevitable or expected future, with a sense of epistemic necessity. Strong epistemic interpretations are only possible if *moot* is accompanied by epistemic adverbs. From around the 17th century, *must* is established as a common marker of epistemic certainty (Traugott 1989: 42). Consider Figure 6-13:

![Figure 6-13. Developmental path of *must*](image)

The development of *must* is clearly in accordance with the hypothesis: it follows the direction of the Scope Hierarchy.

*May* and *might* stem from the Old English verb *magen* (present *mæg*, past *meahte*), expressing physical ability or power. Toward the end of this stage it developed a general ability meaning. During Middle English *may* takes over from *mōt* in contexts of deontic uses of permission and root-possibility. The use of *may* for expressing physical power was lost in the 15th century. During the 17th century the sense of general ability is also lost (Rissanen 1999: 237; Warner 1993: 180-81).

The development of epistemic meaning is not very well documented. In Old English, *may* is occasionally used with marginal epistemic senses (Traugott 1992: 197). In Middle English this use becomes more firmly established but even in Early Modern English it is less commonly used epistemically than with a general root-possibility sense. (Rissanen 1999: 237). Epistemic uses of *may* seemed to have developed before the permission uses of *may*; however, the permission meaning did not derive from the epistemic meaning, but directly from the ability meaning (Bybee & Pagliuca 1985). It is not clear from which meaning the epistemic meaning has derived, the ability or root-possibility sense. It is furthermore not clear whether the first epistemic uses were event-oriented or proposition-oriented. I assume that they were event-oriented as Traugott states that pre-modals in Old English ‘with possible epistemic meanings concern possibility in a world independent of the speaker and can be considered to be only weakly subjective.’ (1989: 42, emphasis mine). As for today, *may* and especially *might* are losing their deontic and dynamic uses and become more and more restricted to epistemic use, especially in American English (Denison 1998: 165; Fischer 2003: 23-24). In sum, the two developmental paths are:
Both paths agree with the predicted order: they develop from narrow scope operator to wide scope operator.

*Can* developed from an Old English lexical verb *cunnan* (present *cann*, past *cūđe*) that expressed ‘know’/ ‘have the mental ability’ (Warner 1993). In Middle English it has become an auxiliary expressing general ability and root-possibility and this is still the predominant use in Early Modern English. In the 16th and 17th century it ousts *may* in these functions. The meaning of permission is late, not before the 19th century (Rissanen 1999: 237; Warner 1993: 177). Epistemic meanings of *can* are very infrequent before the Modern English period and *can* and *could* are probably used epistemically from the 19th century, especially in negations (not possible). *Can* gradually loses the meaning of ‘know’ in the period from Early Modern English to Present Day English (Rissanen 1999: 237). As with *may*, epistemic uses of *can* seem to have developed before the permission uses of *can* and the permission meaning has directly derived from the ability meaning, not from the epistemic meaning. They are parallel, independent semantic changes (Bybee & Pagliuca 1985). The use of *can* as a marker of uncertainty is very restricted in Present Day English (see 6.2.5.1.3), which suggests that it developed later than objective epistemic possibility. See Figure 6-15:
Like the other modals, the semantic changes of *can* and *could* support the hypothesis.

### 6.3.4 Other modals

This section will discuss the diachronic development of all other modal markers. They have gained grammatical status much more recently. *Be supposed to* is the first expression to be discussed. The source of this construction is the French verb *supposer* meaning ‘believe’ or ‘hypothesize’ and from the 15th century also ‘expect’. In the early stages in which the construction with *be* was used, it occurred in statements describing time-stable truths or generic situations, most frequently in the constructions *it’s supposed to be X* or *it is supposed to have –ed*. Ziegeler (2003: 60) claims that these uses are evidentials of hearsay. In a later stage there has been a change from mainly expressing belief or hypothesis of an unnamed source to intention or expectation imposed on the subject (Ziegeler 2003).

If Ziegeler’s reconstruction of the history of *be supposed to* is correct, then there has been an evidential use (π3) before a weak obligation or weak root-necessity use (π1). In fact, this development would semantically be plausible: If the general opinion is that something is the case, then it may be inferred that there are probably (external) circumstances that cause this something to be the case. What is however unclear from Ziegeler’s description is whether the evidential reading should be considered a grammatical construction. It seems to have been the normal passive form of the active lexical verb, which is, according to my criteria (6.2.1) not a grammaticalized construction. In the weak obligation or necessity reading, however, the construction has no longer the resultant meaning of its components and this is probably the first real grammatical use. The development would then look like Figure 6-16, which is in line with the predicted order:

![Figure 6-16. Assumed developmental path of supposed to](image)

The construction *have to* stems from the lexical verb *habban* ‘have’. In Old English this verb underwent the following generalization in meaning: ‘to hold in hand’ developed into ‘to have in one’s immediate personal possession (physically present)’ which developed into ‘to have or own as a possession (not physically present)’ which in turn came to mean ‘to have as an abstract
possession, such as time, an idea, an education, a debt’ (Bybee & Pagliuca 1985: 72).

Additional steps in the grammaticalization process are described in Krug (2000: 55), based on Heine (1993: 41f): 

I have a letter, expressing possession 
became 
I have a letter to mail, expressing possession and purpose, which in turn developed into 
I have a letter to write in which the possessive meaning of have to has bleached out. This further developed into the meaning 
I have to write a letter, in which have to expresses obligation and this finally came to mean 
I have to write, in which the object can be omitted. The sense of obligation or necessity derives from the semantics of have itself. In particular the gradually more frequent construction 
I have to say/do (something unpleasant) seemed to have given rise to the inference of obligation or necessity (Krug 2000: 97ff). In Middle English the first interpretations as obligation or necessity arise, although it remains rare until about 1850 when there is a dramatic frequency increase of have to. Epistemic senses only occurred from the 20th century, but these uses are still rare (Krug 2000: 74, 89). See Figure 6-17 for the complete developmental path:

![Figure 6-17. Developmental path of have to](image)

Related to have to is have got to, originally the perfect construction of get, that combined only with a noun phrase. It entered the English language in the 16th century and expressed possession at first. The rise of the construction have got to could have been motivated by the fact that have was very often reduced to the contracted forms ‘ve, ‘s or ‘d. Got may have been inserted in cases where have was reduced to make the expression perceptually more salient. In the 19th century, the necessity and obligation reading emerged and came to be used very frequently in the 20th century. Have got to did not go through all the stages that have to did. It seems that have got to could become used with obligation and necessity meanings by analogy to the construction have to (Krug 2000: 61ff). In Present Day English have got to, often reduced to got to, can be used epistemically. This use must be very recent and is probably rather infrequent. The developmental path of have got to is presented in Figure 6-18.
Both Figure 6-17 and Figure 6-18 correspond to the hypothesis.

In 6.2.1 it was argued that the contracted construction wanna is grammaticalized. The verb want (would) is first attested in early Middle English, meaning ‘lack’. This remains the core sense until the end of Early Modern English (Krug 2000: 127, 41). However, already in Middle English there are a few examples in which a sense of desire can be inferred in contexts of want, for a human being often desires what he lacks (p.129). This volitional sense gradually changed from an implicature to the basic sense of want (p.142). At the same time, it became possible to express the desired entity not only by a noun (I want a car) but also by an infinitive (I want to go) (p.144). Since Middle English there has been a general increase in infinitival complements. In the 19th and especially the 20th century the frequency of the verb want dramatically increases, both with a nominal and with a verbal complement (p.131). In Present Day English (American and British) examples are attested in which wanna is used with other senses than volition, even in epistemic uses. Consider (185) and (186), in which wanna indicates desirability, or (187) where it is used to express probability (from Krug 2000: 147ff):

(185) You’ve got tooth ache? You wanna see a dentist.
(186) You wanna turn right at the next corner.
(187) Customer: Do you have Coolers?
Assistant: Coolers? They wanna be on one of the topshelves somewhere. They only arrived this morning.

The development of wanna is summarized in Figure 6-19. It is in accordance with the hypothesis.

The final modal constructions to be discussed are ought (to) and need (to). Ought to comes from the Old English verb agan (present ah, past ahbe), which meant ‘have’. In Middle English the verb changed to owe and it developed a past
form *ought* (Traugott 1992; Warner 1993). In late Middle and Early Modern English it develops auxiliary properties and combines with an infinitive (with and without *to*). The past form has no longer past time reference and becomes the normal form (Rissanen 1999: 232). As well as *ought* (*to*), *need* (*to*) develops auxiliary properties in late Middle and Early Modern English. In the 17th century it often combines with an infinitive without *to* and the third person *s* is often not realized (Rissanen 1999: 232). Neither construction can be used for testing the hypothesis, since there is not enough information available on their exact semantic development.

6.3.5 Habitual

As discussed in 6.3.3, the forms *will* and *would* go back to a lexical verb *willan* meaning ‘want’ that developed into an auxiliary expressing volition. Already in Old English, *will* and *would* are occasionally used to express characteristic or habitual behavior in the present and past, respectively. This use is closely related to the earlier meaning of these verbs, volition, as one can infer that ‘what one wants to do, one is inclined or disposed to do’ (Bybee et al. 1994: 157). The developmental path is presented in Figure 6-20. This path is independent of the development of *will* and *would* as irrealis and future marker.

![Developmental path of will and would](image)

This development is in line with the expected order. There is increase in scope.

The habitual expression *used to* is semantically close to its origin, the lexical verb *use*, which had as one of its meanings ‘to follow a usage or custom’. Around 1400 it became very frequent in a construction with a *to*-infinitive, which raised the meaning ‘be accustomed to.’ The construction was at first restricted to human subjects but from around 1600 it could also combine with inanimate subjects. It has furthermore extended to the combination with states (Bybee et al. 1994: 155-56). Still in early Present Day English, *use to V* could be used to express habituality in present and past. Probably in the 19th century the use as a present habitual marker died out (Denison 1998: 175). See Figure 6-21.
This development is in accordance with the hypothesis regarding the direction of change. Its first grammatical use, however, is immediately with medial scope. This is possible probably because the lexical construction is so close in meaning to the grammatical operator and the change is more a change in formal status than in semantic status.

### 6.3.6 Irrealis

In Old English, conditional clauses have the same general structure as in Present Day English, a subordinate *if*(gif)-clause (protasis) and a main *then*(ponne)-clause (apodosis). In general, the verbs have indicative inflection. However, the past subjunctive may be used to express imaginary and unreal conditions, in both the main and the subordinate clause (Traugott 1992: 257). In Middle English, the subjunctive becomes more frequent in conditional clauses instead of the indicative and the past subjunctive becomes the rule when the condition is hypothetical. The subjunctive pluperfect is used to indicate a counterfactual event. In the main clause, the subjunctive is often substituted by a past modal auxiliary (Fischer 1992: 350). The same situation holds in Early Modern English (Rissanen 1999: 308). Because of the loss of inflection, the past subjunctive forms become identical to the past indicative forms, except for *be*. Subjunctive past and pluperfect are still fairly common in the apodosis to encode hypothetical and counterfactual but more frequently a periphrasis with the modal auxiliaries *should* or *would* in combination with the indicative past and pluperfect are used to indicate irrealis. The replacement of the past occurs earlier than of the pluperfect (Rissanen 1999: 228-30). In Present Day English the unreal conditional apodosis—the *then*-clause—requires a past modal verb. As a modal verb is followed by an infinitive, past tense can only be marked by a perfect construction (Denison 1998: 137, 300). In American English, modal verbs are also commonly used in the protasis, such as in (188):

(188)  If he *would* have left him he wouldn’t have gone through all that trouble.\(^{22}\)

\(^{22}\)http://www.storiestogrowby.com/stories/kids_say_elephpit.html, children’s spelling and grammar is supposed to be revised.
The semantic changes of the components of the irrealis constructions, past tense forms, perfect forms and past modal forms, are not easy to describe, since they did not change their meaning in isolation, but within the construction of conditional clauses. The past and perfect forms replaced the subjunctive past and perfect that indicated irrealis, but to suggest that the meanings past tense or perfect aspect developed into the meanings of irrealis would be an oversimplification. The same holds for the irrealis senses of would and should, of which it cannot be claimed that they developed out of future senses: according to Bybee (1995) their irrealis sense can only arise because of the combination of past tense and modality, because this offers 'two areas of vagueness: (i) whether or not the predicate event was completed; and (ii) whether or not the modality remains in effect.' (p.506). A past modal may thus imply that certain conditions on carrying out the desired predicate were not met in the past and, as an inference, may not be met at all. This results in a hypothetical reading if the construction is used in a conditional context or in a polite reading if the conditions are implicit. As the genesis of the irrealis constructions is such a complex matter, in which several TMA expressions are involved, it will not be used as a test case for the hypothesis.

6.3.7 Discussion and conclusion

There is overwhelming support for the hypothesis. In as far as the diachronic developments of TMA expressions are known, they are in accordance with H1: Diachronically, operators show an increase in scope, and develop in the direction from \(\pi_1\)-operator to \(\pi_2\)-operator to \(\pi_3\)-operator. The aspectual expressions all stem directly from lexical constructions, the habitual markers originate in a lexical source and from participant-oriented modality (volition), and modal and future markers have begun their grammatical career as participant-oriented modal markers, climbing up to event-oriented modality or future tense and finally to epistemic, proposition-oriented modality. The Scope Hierarchy has thus made the correct predictions for English. TMA expressions show an increase in scope. They become more abstract and complex in time.

With respect to modal constructions, the combination with impersonal constructions, passives or generic subjects appears to have been a crucial step in the development of wider scope. Warner (1993: 172-73) suggests that modal markers in these contexts impose no selection restrictions on its subject and can therefore act as sentence modifiers, i.e., with wider scope. In a similar vein, Bybee et al. (1994: 287) claim that the first epistemic inferences of may occurred in contexts where the participant is non-specific (general pronouns) or absent (passive and stative sentences) as 'in these cases, the agent is practically devoid of semantic content'. Finally, Ziegeler (2003: 42) claims that there is a clear relationship between generic statements and prediction: a generic statement
may form the evidential justification for a predictive statement. All these claims support the view presented in Chapter 4 that event-oriented root- and deontic modality, which combine with non-specific participants, are important links to epistemic interpretations in the conceptual space. Especially in impersonal or generic contexts, epistemic inferences can arise which make the ‘metaphorical leap’ from the sociophysical to the mental world much more comprehensible.

6.4 FREQUENCY

This section will investigate the frequency of TMA operators in American English, more specifically, the questions 2a: What is the token frequency of TMA expressions?, and 2b: What is the size of different operator classes?

6.4.1 Token-frequency

The expected token frequency of operators is formulated in H2a:

\[ \text{H2a: The token frequency of } \pi_1 \text{-operators is higher than or equal to the frequency of } \pi_2 \text{-operators and the frequency of } \pi_2 \text{-operators is higher than or equal to the frequency of } \pi_3 \text{-operators.} \]

It is assumed that wider scope operators are less relevant to communication than narrower scope operators and therefore, there discourse frequency will be lower.

6.4.1.1 Data-selection

Question 2a is examined in spontaneous speech of adults addressed to other adults in normal day conversation. The data are collected from the Santa Barbara Corpus of Spoken American English (Dubois 2000). The speech of eight different speakers in everyday conversation was selected (see Table 6-9). I have used several criteria for the selection of speakers from the corpus. Only speakers were selected who took part in conversations that were characterized as face-to-face conversation, preferably recorded in private homes. Only Phil is recorded in an office, but it is an informal face-to-face conversation. Every speaker was a monolingual and a native speaker of Standard American-English. No more than five participants took part in a single conversation. I furthermore tried to make the sample as representative as possible for adult Standard American English. Therefore, an equal number of female and male speakers were selected with a more or less equal average age in both sex groups. As a result of these criteria, only those speakers are selected whose age was known.
Table 6-9. Selected data for the adult conversations sample (SBCSAE, Dubois 2000)

<table>
<thead>
<tr>
<th>Speaker No.</th>
<th>Speaker Name</th>
<th>Sex</th>
<th>Age</th>
<th>File(s)</th>
<th>No. of utterances</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001</td>
<td>Lenore</td>
<td>f</td>
<td>30</td>
<td>Sbc0001 + Sbc0006</td>
<td>237</td>
</tr>
<tr>
<td>0003</td>
<td>Lynne</td>
<td>f</td>
<td>19</td>
<td>Sbc0001</td>
<td>585</td>
</tr>
<tr>
<td>0007</td>
<td>Pete</td>
<td>m</td>
<td>36</td>
<td>Sbc0002 + Sbc0003</td>
<td>370</td>
</tr>
<tr>
<td>0009</td>
<td>Marilyn</td>
<td>f</td>
<td>33</td>
<td>Sbc0003</td>
<td>475</td>
</tr>
<tr>
<td>0016</td>
<td>Darryl</td>
<td>m</td>
<td>33</td>
<td>Sbc0005</td>
<td>201</td>
</tr>
<tr>
<td>0028</td>
<td>Phil</td>
<td>m</td>
<td>30</td>
<td>Sbc0010</td>
<td>448</td>
</tr>
<tr>
<td>0044</td>
<td>Marci</td>
<td>f</td>
<td>50</td>
<td>Sbc0013</td>
<td>311</td>
</tr>
<tr>
<td>0046</td>
<td>Kevin</td>
<td>m</td>
<td>26</td>
<td>Sbc0013</td>
<td>268</td>
</tr>
</tbody>
</table>

Preferably only one speaker was selected from each conversation so that the topic of the conversations would not have too much influence, but this was not always possible because of the other criteria. The data of Lynn and Lenore and of Marci and Kevin were as a result taken from the same conversation. All speakers in the sample happen to be Caucasian, partly as a result of the distribution in the corpus, since the majority of monolingual speakers is Caucasian. The age of the few African American monolinguals was not available.

Table 6-9 presents the selected data from the corpus, specific information about the speakers and the number of utterances for each speaker. The complete sample consists of 2895 utterances but the number of utterances for each speaker is different.

6.4.1.2 Coding of utterances

The samples were coded for TMA expressions. Each TMA expression that was interpretable in the context was coded for its semantic function and scope, according to the classifications discussed in 6.2. Several utterances were coded for two or more TMA expressions. For example, aspectual periphrases also contain tense marking on the auxiliary *be, have* or *keep*, so that both aspect and tense are coded. The same holds for many (non-epistemic) modal expressions and the past habitual forms. Past modals that do not refer to past time but instead express politeness, irrealis or epistemic modality were not coded for tense. In non-verbal predicates (nominal, adjectival, adverbial or phrasal) the copula *be* was coded as a tense marker.
6.4.1.3 Results

The complete sample contains 2502 TMA expressions. As the samples of the different speakers are not equal in size, the number of TMA expressions for each speaker is also different (range: 167-630). In order to give each speaker an equal weight, the average percentages are not based on the overall absolute token frequency but rather on the relative frequencies within each speaker’s sample. The data from each speaker contribute one-eighth to the average distribution.

It appears that in American English adult-adult conversations the proportion of $\pi_1$-operators (16%) is much smaller than the proportion of $\pi_2$-operators (83%). The proportion of $\pi_3$-operators (1%) is smallest. This distribution is in contradiction to the prediction that $\pi_1$-operators will be more frequently used within a specific language than $\pi_2$-operators.

Which categories account for the percentages? In Table 6-10 the classes of operators are divided according to the broad TMA domains. From this table it appears that the frequency of $\pi_2$-operators is for the greater part the result of the predominance of tense operators in English, whereas other $\pi_2$-operators are far less common than $\pi_1$-operators. As discussed in 6.2.3 tense is an obligatory category in English, and many non-tense markers in English obligatorily co-occur with a tense marker. For example, an aspectual expression cannot occur without tense, whereas a tense marker can occur without an aspect marker. The obligatory expression of tense, in contrast to most other categories (except perhaps for progressive and perfect aspect) is such a strong factor of influence on the frequency, that the possible influence of scope is obscured. A second analysis was therefore performed on categories that are more similar to each other. Modal markers were selected for this, since the occur with all scopes, are semantically related, have similar expression forms

<table>
<thead>
<tr>
<th>TMA domain</th>
<th>N = 2502</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspect ($\pi_1$)</td>
<td>10.4</td>
</tr>
<tr>
<td>Participant-oriented modality ($\pi_1$)</td>
<td>5.6</td>
</tr>
<tr>
<td>Tense ($\pi_2$)</td>
<td>78.5</td>
</tr>
<tr>
<td>Event-oriented modality ($\pi_2$)</td>
<td>1.6</td>
</tr>
<tr>
<td>Event quantification ($\pi_2$)</td>
<td>0.4</td>
</tr>
<tr>
<td>Irreals ($\pi_2$)</td>
<td>2.4</td>
</tr>
<tr>
<td>Proposition-oriented modality ($\pi_3$)</td>
<td>1.1</td>
</tr>
</tbody>
</table>
and are not obligatory. For this category the influence of scope on frequency should probably be more directly visible. It appears that for this restricted comparison, the hypothesis holds: of all modal expressions (N = 228), participant-oriented modality accounts for 67%, event-oriented modality for 20% and proposition-oriented modality for 13%.

In sum, the hypothesis is too strong in that obligatory expression of a certain category has an enormous impact on the token frequency of such a category. However, when the comparison is restricted to expressions with related semantics, related expression form and related syntactic status (+/- obligatory), then the hypothesis does hold and the token frequency of operators with narrower scope is higher than of operators with wider scope.

6.4.2 Size of classes
The second question that relates to frequency of TMA markers is whether the size of operator classes correlates with scope. It is expected that:

H2b: The class of $\pi_1$-operators has more members than or an equal number of members as the class of $\pi_2$-operators and this class has more members than or an equal number of members as the class of $\pi_3$-operators.

There are two reasons to expect that an increase in scope correlates with a decrease in the size of an operator class. During the process of grammaticalization (increase in scope), semantic paths tend to converge because their meanings generalize. This causes competition between different items with overlapping functions and probably leads to the use of expressions dying out. Secondly, it is expected that operators with narrower scope have more basic communicative functions. Because it is communicatively more important to make descriptive modifications ($\pi_1$) than to make modifications in situating the event ($\pi_2$) or modifications in the presentation of the content ($\pi_3$) there should be more different devices for doing so.

Whether this correlation between size of operator class and scope indeed exists for English can easily be seen from Table 6-6, 6-7 and 6-8. There are two ways in which this question can be answered: it is possible either to compare the number of semantic functions within an operator class or to compare the number of different expressions within an operator class. Both counts were made. In the analysis of expression forms, contracted forms were not counted as separate expression forms, but past forms of the central modals were considered as separate forms, since the present and past forms have sometimes developed different uses. Different tense or irrealis forms (irregular, regular, do-


Chapter 6

did, forms of be, etcetera) are considered as one expression form, as their use is not dependent on semantic grounds, but purely on morphosyntactic grounds. Expression forms that have different semantic functions within the same domain are counted as one expression form. For example, ability, root-possibility and permission can as narrow scope modality are counted as one expression form. Similar expression forms in different domains (would as habitual, future and hypothetical marker) are counted as different forms. Table 6-11 presents the number of expression forms and semantic functions for each operator class.

Table 6-11. Number of expression forms and semantic functions per operator class in English

<table>
<thead>
<tr>
<th>Operator</th>
<th>π1</th>
<th>π2</th>
<th>π3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expression forms</td>
<td>20</td>
<td>23</td>
<td>10</td>
</tr>
<tr>
<td>Semantic functions</td>
<td>15</td>
<td>16</td>
<td>6</td>
</tr>
</tbody>
</table>

It appears that H2b is not completely supported by the data. The class of π2-operators is slightly larger than of π1-operators. The class of π3-operators is on the other hand smaller than the other two classes, which is in accordance with the hypothesis. This suggests that the communicative relevance and/or the cognitive complexity of the function of π2-operators—modifying the relation of the event to the real world—is about equal to the function of π1-operators—modifying the description of the event—whereas the function of π3-operators—modifying the presentation of the propositional content—is in fact communicatively less relevant and cognitively more complex. In Chapter 7 this idea will be further examined to see if it is crosslinguistically valid.

6.5 Synchronic Configurations

What are possible synchronic configurations for TMA expressions? H3a predicts that the presence of more marked operators implies the presence of less marked operators.

H3a: The presence of operators with wider scope depends on the presence of operators with narrower scope according to the hierarchy: π1-operator ⊆ π2-operator ⊆ π3-operator.
The Scope Hierarchy predicts that wider scope operators only occur in a language if there are also narrower scope operators. In other words, it is only possible to modify the presentation of the content by grammatical means ($\pi_3$-operators) if it is also possible to modify the situating of the event by grammatical means ($\pi_2$-operators), and this in turn is only possible if it is also possible to modify the description of the property or relation by grammatical means ($\pi_1$-operators). As Tables 6-6 to 6-8 show, the configuration in English is clearly in accordance with this hypothesis, as there are operators of all three types.

The second hypothesis on the possible synchronic configurations of a TMA system concerns polysemous items. According to H3b polysemous expressions can only cover adjacent regions in semantic space.

H3b: A single expression can only cover adjacent regions in semantic space. As a result, a polysemous or portmanteau expression will only have semantic functions with similar or adjacent scopes.

The steps in the diachronic paths discussed in 6.3 are the presumed adjacent regions in semantic space. Along these paths it was indicated which meanings have become extinct, and in all cases, these extinct meanings were one or two at the beginning of the path. It was never the case that the meaning of an English TMA expression in the middle of the path became extinct with the earlier meanings being retained.

With respect to scope, the hypothesis implies that a particular item can only be used with similar or adjacent scopes, i.e., with scope over the predicate ($\pi_1$) and over the predication ($\pi_2$) or with scope over the predication and over the proposition ($\pi_3$), but not with scope over the predicate and over the proposition, but not over the predication. All the TMA expressions in English that are used with different scope interpretations are presented in Table 6-12. Uses that are very infrequent are presented in between brackets.

The table shows that the predicted correlation between scope and polysemy holds. There is no grammatical TMA expression in English that may have scope over the predicate and the proposition, but not over the predication.

With respect to synchronic TMA systems, hypotheses H3a and H3b make correct predictions for English. Polysemous items cover adjacent meanings on a semantic path, which are assumed to be reflections of underlying universal conceptual relations. English has $\pi_1$-operators for the least marked function of description, $\pi_2$-operators for the more marked function of situating, and $\pi_3$-operators for the most marked function of expressing the speaker's evaluation of the propositional content.
Table 6-12. Combinations of scopes of polysemous TMA expressions in English

<table>
<thead>
<tr>
<th>Operator</th>
<th>Operator</th>
<th>Expression</th>
<th>π1</th>
<th>π2</th>
<th>π3</th>
</tr>
</thead>
<tbody>
<tr>
<td>be supposed to</td>
<td></td>
<td>be supposed to</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>can</td>
<td></td>
<td>can</td>
<td>X</td>
<td>X</td>
<td>(X)</td>
</tr>
<tr>
<td>could</td>
<td></td>
<td>could</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>have to</td>
<td></td>
<td>have to</td>
<td>X</td>
<td>X</td>
<td>(X)</td>
</tr>
<tr>
<td>have got to</td>
<td></td>
<td>have got to</td>
<td>X</td>
<td>X</td>
<td>(X)</td>
</tr>
<tr>
<td>keep—ing</td>
<td></td>
<td>keep—ing</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>may</td>
<td></td>
<td>may</td>
<td>(X)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>might</td>
<td></td>
<td>might</td>
<td>(X)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>must</td>
<td></td>
<td>must</td>
<td>(X)</td>
<td>(X)</td>
<td>X</td>
</tr>
<tr>
<td>need to</td>
<td></td>
<td>need to</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>ought to</td>
<td></td>
<td>ought to</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>shall</td>
<td></td>
<td>shall</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>should</td>
<td></td>
<td>should</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>will</td>
<td></td>
<td>will</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>would</td>
<td></td>
<td>would</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

6.6 EXPRESSION FORM

The fourth question—what is the expression form of TMA expressions?—will be examined in this section. The expected relation between form and scope is formulated in H4:

H4: Operators with wider scope show a higher or equal degree of formal grammaticalization than operators with narrower scope.

Roughly speaking, inflection is the most grammaticalized form, auxiliaries are less and periphrases the least grammaticalized (Bybee et al. 1994: 40). However, especially within the group of periphrases, finer distinctions are possible, such as the degree of reduction and of fusion of the component parts of the construction. In general, periphrases that contain bound elements are more grammaticalized than periphrases that only contain non-bound elements, and
periphrases of which the different components are phonologically reduced to one form are more grammaticalized than periphrases where this is not the case. As these finer grades of grammaticalization are hard to establish objectively, they will only be taken into account with great caution. Consequently, the question will only be examined in an exploratory way.

With respect to English, the TMA expressions with the highest degree of formal grammaticalization are the past and present tense inflection. The central modals have the second highest level of formal grammaticalization as they have become real auxiliaries. All other expressions are considered periphrases and are least grammaticalized. However, within this group, the aspectual constructions progressive, perfect and resultative are considered more grammaticalized than other constructions as they contain a suffix and their auxiliary is very often reduced. Furthermore, some of the periphrases are frequently contracted, such as be going to to gonna, have got to to gotta, have to to hafta, used to to useda, and want to to wanna. In the reduced forms, these constructions are closer to auxiliaries than to periphrases. However, as they are not in all contexts and not by all speakers reduced to such a high extent (see Krug 2000), they are still treated as (highly grammaticalized) periphrases and not as auxiliaries. Constructions with the lowest degree of formal grammaticalization are probably be about to, keep –ing and be supposed to.

Does scope indeed correlate with expression form? Table 6-13 shows the scope with which each expression is used, in relation to its degree of formal grammaticalization. For polysemous items, the least frequent use is put in between brackets. The table shows that in general, there is a tendency for wider scope operators to have a higher degree of grammaticalization, but there are quite some exceptions to this tendency. Most importantly, the items that are formally the most grammaticalized—tense inflection—function as operators with medial scope, whereas many auxiliaries or even periphrases function as operators with wide scope. Furthermore, two items that are formally the least grammaticalized can have medial scope (be supposed to, keep –ing), whereas some more grammaticalized items only have narrow scope.

A provisional conclusion may be that the hypothesis is too strong. When, however, only related meanings are compared, the picture becomes different. Within the related domains of tense, aspect and event quantification, the markers with the widest scope (tense inflection, habitual used to, will and would) have a higher or equal degree of formal grammaticalization than the markers with narrow scope (the aspectual periphrases). The only exception is keep –ing: it has the lowest degree of formal grammaticalization but can get a frequentative interpretation with scope over the predication (π2). Within the domain of modality, it is in general the case that operators that can have wider scope are
Table 6-13. Relation between expression form and operator type in English

<table>
<thead>
<tr>
<th>Degree of formal grammaticalization</th>
<th>Operator</th>
<th>( \pi_1 )</th>
<th>( \pi_2 )</th>
<th>( \pi_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>High (Inflection)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium (Auxiliary)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>may</td>
<td>(X)</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>might</td>
<td>(X)</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>must</td>
<td>(X)</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>will</td>
<td>(X)</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>would</td>
<td>(X)</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>should</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>could</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>can</td>
<td></td>
<td>X</td>
<td>X</td>
<td>(X)</td>
</tr>
<tr>
<td>shall</td>
<td>(X)</td>
<td>(X)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium-Low (Periphrasis)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ought to</td>
<td></td>
<td>X</td>
<td>X</td>
<td>(X)</td>
</tr>
<tr>
<td>used to / usta</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>have to / hafta</td>
<td></td>
<td>X</td>
<td>X</td>
<td>(X)</td>
</tr>
<tr>
<td>have got to / gotta</td>
<td></td>
<td>X</td>
<td>X</td>
<td>(X)</td>
</tr>
<tr>
<td>need to</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>wanna</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>be going to / gonna</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>be –ing</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>have –ed</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>be –ed</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (Periphrasis)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>be supposed to</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>keep –ing</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>be about to</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

more grammaticalized (auxiliaries). However, there are two counterexamples to this picture: the auxiliary ***shall*** is restricted to narrow scope (desirability /
obligation, π1) or to future tense (π2), whereas some of the less grammaticalized periphrastic constructions (have to, ought to, have got to) can have wide scope. However, shall is very infrequent and probably on its way to become extinct. With the exception of shall, all modal expressions with wide scope are formally grammaticalized to an equal or higher degree than modal expressions with medial or narrow scope. So, if only semantically related expressions are compared, there is indeed a strong correlation between scope and level of formal grammaticalization, although even then, there is one counterexample.

Although the domain restricted analysis supports the hypothesis, an alternative explanation might be possible, namely, the age of the grams. As was discussed in 5.4.4 there is probably a relation between age of grams and level of formal grammaticalization. Since it was also hypothesized that operators increase in scope, hypothesis 4 was stated in terms of scope and not in terms of age. However, the relation between scope and age of a gram is not straightforward. If grams change their scope, they will increase in scope, but it is not necessarily the case that grams keep changing their scope. Although the hypothesis abstracted away from this feature, for English, it may be worthwhile to investigate whether it is in fact age rather than scope, that is the prime predictor for formal grammaticalization. Scope may have only indirect influence, as it correlates to a certain extent with age.

What is the English situation? The oldest grams are the past and present tense. However, the meanings of these inflections have not drastically changed for centuries. The scope is the predication (π2) from the earliest records on English. The next oldest expressions are the central modals, which have grammaticalized in Old English, with the exception of can/could that grammaticalized in Middle English. Aspectual expressions are younger, especially the prospective and immediate prospective be about to. In Table 6-14 the relation between age and expression form is presented. The first period in which a particular construction clearly came to be used grammatically is taken as the age of the gram.

Table 6-14 shows that there is a very strong relation between age and expression form. For a few expressions, however, there is not enough information available on their diachronic development: from which period should they be considered grammatical? Be about to is said to occur from late Middle English, but it is unknown whether it was a grammatical construction at that time. The earliest lexical uses of the active verb suppose are from the 15th century, but it is unknown when the passive construction with this verb became grammatical. According to Ziegeler (2003), the passive construction occurs late, but she gives no exact dates. Furthermore, there is not enough information
Table 6-14. Relation between age and expression form in English

<table>
<thead>
<tr>
<th>First Important Period of Grammaticalization</th>
<th>Degree of Formal Grammaticalization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Before Old English</td>
<td></td>
</tr>
<tr>
<td>Past</td>
<td>X</td>
</tr>
<tr>
<td>Present</td>
<td>X</td>
</tr>
<tr>
<td>Old English</td>
<td></td>
</tr>
<tr>
<td>must</td>
<td>X</td>
</tr>
<tr>
<td>may / might</td>
<td>X</td>
</tr>
<tr>
<td>will/ would</td>
<td>X</td>
</tr>
<tr>
<td>shall / should</td>
<td>X</td>
</tr>
<tr>
<td>Middle English</td>
<td></td>
</tr>
<tr>
<td>could</td>
<td>X</td>
</tr>
<tr>
<td>can</td>
<td>X</td>
</tr>
<tr>
<td>Early Modern English</td>
<td></td>
</tr>
<tr>
<td>have—ed</td>
<td>X</td>
</tr>
<tr>
<td>be—ed</td>
<td>X</td>
</tr>
<tr>
<td>be—ing</td>
<td>X</td>
</tr>
<tr>
<td>ought to</td>
<td>X</td>
</tr>
<tr>
<td>need to</td>
<td>X</td>
</tr>
<tr>
<td>used to</td>
<td>X</td>
</tr>
<tr>
<td>Present Day English</td>
<td></td>
</tr>
<tr>
<td>be going to</td>
<td>X</td>
</tr>
<tr>
<td>have to</td>
<td>X</td>
</tr>
<tr>
<td>have got to</td>
<td>X</td>
</tr>
<tr>
<td>wanna</td>
<td>X</td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>be supposed to</td>
<td>X</td>
</tr>
<tr>
<td>be about to</td>
<td>X</td>
</tr>
<tr>
<td>keep—ing</td>
<td>X</td>
</tr>
</tbody>
</table>

available on keep—ing. So for the formally least grammaticalized constructions the information on age is lacking. This makes it impossible to test the relation between age and expression form thoroughly. Interestingly, the fact that the history of these constructions is not studied in much detail and the fact that
they are not highly grammaticalized may have the same cause: they are not frequently used. Probably, the level of formal grammaticalization is not only dependent on age, but also on the frequency of use during its history (cf. Bybee & Hopper 2001b).

In conclusion, there is a correlation between scope and expression form, but it does not account for all the facts. A different analysis, whereby the age of the expressions is related to the expression form, reveals a very strong relation for the expressions of which the age is known with certainty. The apparent relation between scope and expression form may therefore be a tertiary relation resulting from a relationship between age and scope and between age and expression form. As was already discussed in 5.4.4 there is no straightforward relation between scope and age of a gram: younger grams are earlier at the grammaticalization paths but not every step on a path leads to widening of scope.

6.7 Expression order

The final question to be examined in this chapter is: what is the expression order of TMA expressions? H5 predicts that:

H5: The expression order of TMA expressions iconically reflects the scope relations as follows: $\pi_3 \pi_2 \pi_1 \text{Predicate} \pi_1 \pi_2 \pi_3$

As it is impossible to consider all the possible combinations of TMA markers in English, the discussion will be restricted to a few core examples.

In aspectual periphrases, the auxiliaries that carry tense, be and have, are farther away from the predicate than the part that signals aspect, -ing, -ed, going to/gonna and about to, consider (189)-(191):

(189) He was about to read.        $> \pi_2$ [immediate prospective $\pi_1$ [Pred]]
(190) He’s gonna read.            $> \pi_2$ [prospective $\pi_1$ [Pred]]
(191) He will be reading.         $> \pi_2$ [[Pred]-progressive $\pi_1$]

Foley & Van Valin (1984: 213) present some examples that combine different modal expressions. One is presented in (192):

(192) John may have to leave.

In (192), may can only be interpreted as a marker of objective ($\pi_2$) or subjective ($\pi_3$) epistemic modality and have to as a marker of participant-oriented necessity ($\pi_1$). It could be paraphrased as it is possible that John is forced to leave, or I doubt that
John is forced to leave. This meaning cannot be expressed by the opposite order: *John has to may leave. In the dialect of the southern United States, central modals can be combined, such as in the following examples taken from Foley & Van Valin (1984: 231):

(193) John might could do it.
(194) Don’t get so far ahead - I may not could make it.

The only possible meaning of (193) is: ‘it is possible that John is able to do it.’ The only possible meaning of (194) is ‘it is possible that I will not be able to make it.’ Might and may express epistemic modality (π2 or π3) and could expresses ability (π1). This reflects the predicted order π3 π1 Pred or π2 π1 Pred. When modal expressions are combined, then, the expressions closer to the predicate have narrower scope than the expressions farther away from the predicate.

A further example is provided by Hengeveld (fc.: ex.13), in which a tense marker is combined with a π3 modal and a lexical aspect marker:

(195) The tree must have begun to grow.

The order in (195) is the only possible order of ‘certainty’ must (π3), past tense have (π2) and ingressive aspect begin (lexical) and it iconically reflects the scope: π3 π2 Pred. All other combinations are impossible, such as must begin to have grown, has begun to must grow, begins to have must grow, etcetera.

Finally, when the habitual (π2) and progressive (π1) are combined, again the predicted order arises. See (196)-(197):

(196) Drug using prostitutes now complain about so-called low-life prostitutes who used to be bumming cigarettes who are now bumming condoms. (CWO)
(197) He used to be talking about whether he should run for president. (CWO)

The opposite order, be is being used to talk about … raises a completely different reading, in which be is literally used for the purpose of talking about ….

The only cases in which the expression order is not a clear reflection of scope relations seem to be when tense is expressed on modal markers or on the verb keep in keep –ing, because then, tense occurs in the middle of the expression. Consider (198):
Both the ability and the past tense meaning are expressed by the one full form could, therefore, no order can be established between the two. In forms like needed to, had to or kept -ing, the tense marker belongs to the inner part of the periphrastic expression and there is no reflection of the Scope Hierarchy. In a strict interpretation of the hypothesis, they might be considered counterexamples. However, a real counterexample would be the opposite order, where π1-operators would be farther away from the predicate than π2-operators. In sum, the general picture is that H5 holds for English. The expression order of TMA operators indeed reflects the scope relations.

6.8 CONCLUSION
Scope is clearly reflected in the English TMA system. The Scope Hierarchy makes correct predictions for the diachrony of grams, for the synchronic system, and for expression order. First, the developmental paths of TMA expressions only show an increase in scope. Second, polysemous expressions currently reflect the semantic relatedness of the steps in their developmental path and have only adjacent scopes. Third, the synchronic set of TMA expressions adheres to the implicational hierarchy between the operators. Finally, the expression order of operators reflects scope relations: operators with scope over the predicate are expressed closest to the predicate, operators with medial scope are expressed farther away and operators with scope over the proposition are expressed farthest away.

With respect to frequency and expression form, the hypotheses need some adjustments. Firstly, wider scope correlates with lower token frequency, but this only holds for TMA operators with related semantics and grammatical status. Obligatoriness of expression has more influence on frequency than scope. Secondly, the class of π2-operators is slightly larger than of π1-operators, whereas the opposite pattern was predicted. This suggests that operators that modify the description of the set of possible events and operators that modify the situating of the event are equally relevant categories in English. In contrast, π3-operators seem to be less relevant, as was predicted, since there are far less expression forms and semantic functions with wide scope. Thirdly, with respect to the expression form of operators, there seems to be only an indirect relation between expression form and scope, in that scope correlates to a certain extent with age of grams and the age of a gram, probably in combination with

23 ultimatehorse.com/temp/brokenbackranch/puppypage.html
frequency, seems to be a primary determining factor in the degree of formal grammaticalization.

This chapter has presented an overview of the English TMA system. The TMA system was continuously illuminated from the viewpoint of scope. Although some studies on English TMA have mentioned the role of scope, especially the work of Traugott (1982; 1989; 1995; 1996), it has not been taken into account systematically. Furthermore, the scope distinctions have only been broadly specified, for example a distinction between sentence or proposition scope versus scope over the verb or predicate. The definition of scope in FG is exceptional in that it makes a more fine-grained analysis of possible scopes than is done in other frameworks and specifies a clear link to cognitive and communicative factors. This approach has been shown to be fruitful in that it accounts for different characteristics of the TMA system. Needless to say that scope is not the only factor involved; it is only part of the TMA story.

One of the requirements of a functional theoretical model of language is that it can provide accurate descriptions of a specific language. For English, as shown in this chapter, the Scope Hierarchy has proved to be a rather reliable predictor. However, a further requirement is that a theoretical model is typologically adequate. It has to define the boundaries on possible language systems. Implicational hierarchies, such as the Scope Hierarchy, in general describe statistical tendencies, not necessarily absolute universals. The predictional power of the Scope Hierarchy therefore needs to be tested in a large sample of languages. The next chapter examines whether the assumed correlations between scope and properties of TMA systems have crosslinguistic validity.
Chapter 7

TMA in Languages of the World

7.1 INTRODUCTION
As argued in 1.5 a functional theoretical model of language should be able to describe any possible language and it should predict crosslinguistic patterns of morphosyntax, distribution and diachrony. The Scope Hierarchy, based on the layered structure of the clause in the FG model (2.3.1), should account for typological variation and diachronic development within the domain of TMA. In this chapter the crosslinguistic variation in TMA systems will be discussed with respect to the Scope Hierarchy. The hypotheses 1-5 (see 5.4) will be tested in this chapter in a crosslinguistic perspective. The first hypothesis on diachrony will be tested in 7.2. In the previous chapter, the diachronic paths of the specific TMA expressions in English were examined but here the focus lies on universal paths of development. For testing hypotheses 2-5, an inventory has been made of the TMA systems in a representative sample of 76 languages. In 7.3 the methodology used is outlined. Section 7.4 examines the frequency of operators in the language sample. Section 7.5 considers the possible and impossible synchronic configurations of TMA systems. Finally, the expression form (7.6) and expression order (7.7) of TMA expressions are examined.

7.2 DIACHRONY

7.2.1 Introduction
It has been argued in 5.4.1 that TMA expressions will probably show an increase in scope in diachronic development. In the process of grammaticalization, expressions develop from specific to general, from concrete to abstract and from referential to subjective meanings. Grammaticalization paths are in general unidirectional. In the theory of FG wider scope implies a more abstract and subjective meaning, so that diachronic development coincides with an increase in scope. Operators may acquire wider scope in time and not narrower scope. In sum, the research question and hypothesis of this section are:
Q1: *What is the diachronic development of TMA expressions?*

H1: Diachronically, operators show an increase in scope, and develop in the direction from \( \pi_1 \)-operator to \( \pi_2 \)-operator to \( \pi_3 \)-operator.

### 7.2.2 Method

The diachronic development hypothesis is tested in particular on the basis of the results of the seminal study of Bybee et al. (1994); this study focused on the evolution of grams. The present study examines the sources of TMA expressions in a sample of 76 languages, the GRAMCATS sample, which will also be the basis for testing hypothesis 2-5. The languages in the GRAMCATS sample are presented in Appendix A. The languages were chosen from the genetic classification of more than four thousand languages in Voegelin & Voegelin (1978). The languages were evenly selected from this list with the aim to avoid genetic bias. Of the 55 minimal language groups in the classification, i.e. isolates and groups up to 8 members, two were selected (Basque and Inuit). Of the pidgins and creoles, one language was picked (Tok Pisin). Of the 24 other phyla languages were randomly selected. On the condition that grammars were available, a phylum with less than forty members contributed one language to the sample. Larger phyla contributed more than one language, the number depending on the size and stratification of the subgroups within the phylum, for details, see Bybee et al. (1994: 27-32, 303-10). The same GRAMCATS sample also served as the basis for the study of Bybee et al. (1991) on grams expressing future tense and related senses.

The studies based on the GRAMCATS sample are supplemented with the results of the pioneering studies of Bybee (1985) and Dahl (1985). These studies were independently undertaken and based on different language samples. Bybee’s study on morphology is based on a sample of 50 languages, not related genetically or areally.² The data were collected using reference grammars. Dahl’s study on tense and aspect is based on a 64-language sample, which, unfortunately, shows a strong bias to Indo-European languages (21 of the 64). The data were obtained by the use of a questionnaire, filled out by native informants. The results of Bybee (1985) and Dahl (1985) are remarkably similar: ‘not only are the mechanisms by which grammaticization is implemented the same across languages, but also the actual semantic material that is molded by this process appears to be very similar across languages’ (Bybee & Dahl 1989: 96). The studies conveyed that there is a restricted set of lexical sources for each gram. For the most frequent tense and aspect categories, there appear to be three major paths of development:

---

1 A phylum is the maximal group of languages that does not belong to any larger linguistic group.

2 There is hardly any overlap between the sample in Bybee (1985) and the GRAMCATS sample.
a. expressions with a copula or possession verb plus a past participle, or verbs meaning ‘finish’, ‘come from’ or ‘throw away’, develop into grams marking anterior or perfect, which in turn develop into perfectives or pasts;
b. expressions with a copula, locative or movement verb develop into progressives which in turn develop into imperfectives;
c. expressions with a verb meaning ‘desire’, ‘movement towards a goal’ or ‘obligation’ develop into grams expressing intention and future. (Bybee & Dahl 1989: 57)

In Bybee et al. (1994) these diachronic paths are investigated in more detail. By comparing the original lexical sources of grams and the synchronic uses, common grammaticalization paths were established. The semantic paths within the domain of modality were investigated in more detail by Van der Auwera & Plungian (1998). In this section the results from Bybee et al. and Van der Auwera & Plungian will be discussed and compared to the hypothesis. Do the semantic paths indeed show an increase in scope?

7.2.3 Results

The first major developmental path that was found by Bybee et al. (1994) involves the semantic notions of resultative, completive, perfect, perfective and past. This path is represented in Figure 7-1, in which the FG-classification according to scope is added.

[Figure 7-1: Developmental paths leading to perfective aspect and past tense (adapted from Bybee et al. 1994: 105)]

In Figure 7-1 lexical elements meaning ‘be’, ‘have’, ‘come’, ‘finish’ and (other) directional items develop into specific aspectual expressions with narrow scope (π1), such as resultative, perfect or completive. A completive and resultative may further develop into perfect aspect, which in turn may develop into a more general perfective aspect (same scope) or a past tense marker (increase in scope). Bybee & Dahl (1989) state that the semantic change from a perfect to a
past tense implies that a perfect must restrict its point of reference to the moment of speech and ‘the part of its meaning that specifies that the past event is especially relevant to the current moment must be lost’ (1989: 74). All developmental steps towards perfective aspect or past tense are in accordance with the predicted direction. If the scope of a marker changes, there is increase in scope.

A further path for the meanings related to perfective aspect is a resultative or perfect that develops a sense of inference from results, and later on becomes a marker that really indicates evidentiality. See Figure 7-2:

![Figure 7-2. Developmental path leading to evidentiality](Bybee & Dahl 1989: 73; Bybee et al. 1994: 95ff)

This development is attested in languages in the Balkan and neighboring areas of the Middle East (Turkish, Kurdish, Bulgarian), in the Tibeto-Burman language Newari, in Udmurt (west-central Russia), Inuit (Greenland) and Tucano (Colombia) (Bybee et al. 1994). An explanation for this development is given in Bybee & Dahl (1989):

Perfect is used to describe past actions or events with present results. If the focus of meaning is on the idea that the present results are connected to and perhaps attest to past actions or events, then the notion of an action known by its results can be extended to actions known by other indirect means, such as by inference (from reasoning in addition to inference from results) and by reports form other parties. (p.73-74)

This development, that is, from an operator with scope over the predicate immediately into an operator with scope over the proposition, is not directly expected from the hypothesis, although it is also not in conflict with the hypothesis. The developmental direction is from narrow to wide scope. The inferential relation between the two meanings of the perfect can explain why this large extension in scope is a possible development.

A second major developmental path concerns the semantic notions of progressive, imperfective and present tense. The developmental path is from progressive into imperfective or present tense (see Figure 7-3). According to Bybee & Dahl (1989: 82-83) a progressive can become an imperfective if it loses some of its original specific semantics, such as specifying a limit on the period of time in which an activity is ongoing, and implying involvement of the subject. An imperfective signals that a situation is simply ongoing at reference time.
As a criterion for the development from a progressive into an imperfective or present tense marker, Bybee et al. (1994) take the expression of habitual meaning. Once a progressive starts to mark habituality, they consider it an imperfective or present tense marker. In my view this is not the correct approach: an habitual sense may arise as an implicature for progressive, imperfective or present tense markers in certain contexts, but then, they do not by itself mark habituality. Only when there is clear evidence that the implication of habituality is conventionalized and has become the basic meaning of the expression, it should be analyzed as such. Consequently, in the current analysis of the languages, there may be fewer habituals recognized than in Bybee et al.'s analysis.

The developmental path in Figure 7-3 is consistent with the predicted development: a lexical verb or construction can begin to express progressive aspect and the progressive gram may generalize towards an imperfective marker (same scope) or to a present tense marker (increase in scope). There is only increase in scope, no decrease.

A third common developmental path found by Bybee et al. (1994: 165-66) is on the boundary of quantification and aspect (See Figure 7-4). There is evidence that continuatives and frequentatives can both develop from lexical meanings, ‘keep on’ and ‘often’, respectively, but probably also from an earlier iterative marker. An iterative marker is often restricted to telic, punctual verbs, whereas a continuative sense may arise when an iterative marker is combined with atelic verbs. A further generalization is that the continuative marker develops into a progressive. When an iterative develops frequentative use, it extends from expressing repetition on one occasion to repetition on different occasions. A frequentative may further develop into a habitual. Finally, it is possible that a habitual marker derives directly from a lexical source (like in English, Figure 6-21). All the proposed developments are in accordance with the predicted order. Increase in scope is possible (iterative to frequentative), but no decrease in scope.
Bybee et al. (1994: 172) tentatively suggest that in particular reduplication may develop from a habitual sense into an imperfective. Note that this would conflict with the predicted diachronic order in H1. However, this apparent counterexample is the result of the definition of the imperfective marker used by Bybee et al. (1994) that also covers the sense of habituality, whereas, according to my analysis, an imperfective and habitual are distinct semantic functions. The only example in the GRAMCATS sample that would support the step from a habitual marker to an imperfective is from Nakanai (46):

(46) **Nakanai** (R. L. Johnston 1980: 131)

a. *Eia o-io sa-sapa*
   3S at-there REDUP-sweep
   ‘She is there sweeping.’

b. *Eia sa-sapa te la kavikoki.*
   3S REDUP-sweep PREP NCL morning
   ‘She sweeps in the mornings.’

However, (46a) and (46b) are different constructions: in (46a) the reduplicated verb is at the clause-final position and combines with a locative expression whereas in (46b) the reduplicated verb is at the second position and co-occurs with a temporal expression. This suggests that it is incorrect to consider the two uses of the reduplicated verb as one polysemous expression to mark imperfective and habituality. Furthermore, the possibility to combine a specific
form—here, reduplication—with a phrase indicating habituality—here ‘in the mornings’—does not mean that this form expresses habituality by itself. Therefore, I assume that there is no evidence for a developmental path for reduplication from habitual to imperfective aspect.

Bybee et al. (1994) thus found evidence for four common diachronic paths within the domains of aspect, tense and quantification. These paths all support the hypothesis: there are changes from lexical elements to π₁- or π₂-operators and changes from π₁-operator within the category to another, more general π₁-operator or to a π₂- or even π₃-operator. There is no evidence for development in the opposite direction. The only development that is not directly expected from the hypothesis is the change from a perfect aspect marker (π₁) by way of pragmatic inference to an evidential marker of inference (π₃). This involves an increase in scope in the predicted direction, but the extension in scope skips an intermediate step.

With respect to the development from perfect and progressive into either perfective and imperfective aspect or into past and present tense, Bhat (1999: 182) suggests that this is determined by the language type. In his view there are aspect-prominent languages, in which aspeсtual distinctions are very important and frequent. In those languages, a perfect and progressive will develop into the general aspeсtual markers perfective and imperfective. There are also tense-prominent languages, in which tense oppositions are obligatorily marked, and in those languages the perfect and progressive will develop into a past and present tense marker.

Besides the above discussed semantic paths, Bybee et al. (1994) found developmental paths for modal and future markers. The first major development occurs within the field of possibility, again with the classification according to scope in FG added. See Figure 7-5:

![Figure 7-5. Developmental paths of items with a basic sense of potentiality (adapted from Bybee et al. 1994: 240, 66)](adapted_from_Bybee_et_al_1994_240_66)

Although this point was not investigated by Bybee et al., I assume that a root-possibility marker and a permission marker will first express participant-
oriented meanings ($\pi_1$) and only later event-oriented meanings ($\pi_2$). An epistemic possibility marker probably expresses first objective epistemic modality ($\pi_2$) and later on notions of uncertainty (subjective modality, $\pi_3$). Both developments were supported for English (6.3.3), but there should of course be a more thorough investigation to find out whether this is a universal development. The steps in the developmental path in Figure 7-5 are in accordance with the predicted order: there is only increase in scope, from $\pi_1$ to $\pi_1/\pi_2$ to $\pi_2/\pi_3$.

A second major development within the domain of modality and related senses is of modal markers and lexical sources into future grams, which turn into epistemic modal markers. Consider Figure 7-6:

![Figure 7-6. Developmental paths leading to future tense and prediction (adapted from Bybee et al. 1994: 240, 53ff)](image)

Different types of participant-oriented modality markers—obligation, desire, ability—may go through a stage in which they can signal intention and eventually, they may turn into future tense markers. Future tense markers may again turn into epistemic markers of probability or prediction. Why future tense clusters with modal meanings was already discussed in 4.3.1. According to Bybee et al., the ‘central functions in future grams are intention and prediction. It follows from this that future is less a temporal category and more a category resembling agent-oriented and epistemic modality, with important temporal implications.’ (1994: 280). In my view, however, this statement is too strong:

---

3 Intention is not considered a specific semantic function in FG but rather represents a stage in which intention occurs as a strong implicature for modal markers in combination with first person subjects. It may lead to a future tense interpretation.
intention is an important precursor and prediction an important implication of future tense expressions, but there is also a rather objective future time reference, rather similar to objective epistemic meanings. In my view, the expression of intention may lead to the implicature that the event will take place in the future and this may lead to the conventionalization of future time reference. At that stage, the core function of a future gram is indicating future tense, not intention. The sense of prediction is also not the core function of a future gram but only a matter of mutual knowledge about future events. By implicature and conventionalization, this notion of prediction may develop into a separate meaning. The development from a participant-oriented modal marker or a movement verb towards a future gram involves the loss of an obligatory animate agent and an active predicate (Bybee & Dahl 1989: 90). A further development into an epistemic probability or prediction marker involves the loss of the restriction that the event occurs in the future. Bybee & Dahl (1989: 93) state that it is ‘typically used when verification of present prediction lies in future.’

The developmental paths in Figure 7-6 support the hypothesis that grammatical expressions only show an increase in scope, and never a decrease in scope. They develop from concrete meanings to more abstract and subjective meanings.

The developmental paths for potentiality and for future and related senses are in line with the predictions that follow from FG. Participant-internal modality (π1) may develop into participant-external modality (root- and deontic meanings, π1 or π2), which may develop into epistemic modality (π2 or π3) or future tense (π2). Future may develop into epistemic modality (π2/π3). These developments can all be described as an increase in scope, in the direction from π1-operator to π2-operator to π3-operator.

7.2.4 Conclusion

Of the semantic functions included in this research (Table 4-6), many are part of one of the attested developmental paths, presented in Figures 7-1 to 7-6. If these paths are analyzed according to their scope, it appears that there are only changes from narrow to medial scope, or from medial to wide scope and not the other way around. The hypothesis that TMA expressions only change in the direction from π1 to π2 to π3-operator is clearly supported. The only sequence that is not really expected from the layered model, but also not in conflict with it, is a resultative or perfect (π1) directly developing into an inferential / evidential marker (π3). Apparently, descriptive categories (perfect aspect) may in certain contexts give rise to pragmatic inferences relating to the evaluation of the content (evidentiality), skipping the layer of situating. The assumption that semantic relatedness always involves meanings with adjacent or
similar scopes is thus not absolutely true. Finally, for a few TMA categories the possible origins have not yet been established. For example, more investigation is needed of possible universal developments of irrealis expressions.

7.3 Methodology for H2-H5

The Scope Hierarchy is clearly reflected in universal developmental paths of TMA operators. However, it was predicted in Chapter 5 that the Scope Hierarchy is also reflected in the frequency, synchronic configurations, expression form and expression order of TMA. The research questions and hypotheses relevant to this chapter are repeated here:

Q2b: What is the size of operator classes?
H2b: The class of $\pi_1$-operators has more members than or an equal number of members as the class of $\pi_2$-operators and this class has more members than or an equal number of members as the class of $\pi_3$-operators.

Q2c: What is the crosslinguistic incidence of operator types?
H2c: The crosslinguistic incidence of $\pi_1$-operators is larger than or equal to the incidence of $\pi_2$-operators, which in turn is larger than or equal to the incidence of $\pi_3$-operators.

Q3: What are possible synchronic configurations for TMA expressions?
H3a: The presence of operators with wider scope depends on the presence of operators with narrower scope according to the hierarchy: $\pi_1$-operator $\subset$ $\pi_2$-operator $\subset$ $\pi_3$-operator.
H3b: A single expression can only cover adjacent regions in semantic space. As a result, a polysemous or portmanteau expression will only have semantic functions with similar or adjacent scopes.

Q4: What is the expression form of TMA expressions?
H4: Operators with wider scope show a higher or equal degree of formal grammaticalization than operators with narrower scope.

Q5: What is the expression order of TMA expressions?
H5: The expression order of TMA expressions iconically reflects the scope relations as follows: $\pi_3 \pi_2 \pi_1$ Predicate $\pi_1 \pi_2 \pi_3$

In order to test H2-H5, many different TMA systems have to be compared. As a starting point, I have used the study on TMA by Bybee et al. (1994) that
presents an extensive inventory of TMA expressions in the languages of the GRAMCATS sample discussed in 7.2.2 (see Appendix A). Furthermore, more specific data on future tense and related meanings in this sample are presented in Bybee et al. (1991). The resulting inventory of TMA expressions is completely checked, adjusted if necessary, and reanalyzed according to the FG-based classification presented in Chapters 3 and 4. This means that for each TMA expression the scope is determined and its semantic function is reconsidered.

For this purpose reference grammars were used. A shortcoming of the use of secondary sources in linguistic research is that the information presented in the grammar is sometimes not specific enough for the research question and that one has to rely on the information presented by the author. Nevertheless, it is a quick way to compare a large number of languages without the time-consuming task of having to consult native speakers. If the grammars are consulted with care, the objections mentioned above can be overcome to a large extent.

In order to make this study as comparable as possible to the studies of Bybee et al. I have consulted the same reference grammars for each language in the sample that were used as the main sources of information in Bybee et al. (1994: 325-55), if they were accessible. For a few languages, however, the reference grammars were not the same. For Basque, I used Hualde & Urbina (2003), in addition to the reference grammar by Saltarelli et al. (1988) that was used by Bybee et al. For Cocama, I used Faust Wille (1972) as well as (Faust 1971) that was used in Bybee et al. An additional grammar was also used for the language Ikung or Ixí, namely Köhler (1981), as it contained more specific information on the TMA system than Snyman (1970), used in Bybee et al. For nine languages the reference grammar used in Bybee et al. was not available in the Netherlands. For five of these languages I used another reference grammar instead: this was the case for Tucano (Ramirez 1997), Guugu-Yalanji (Patz 2002), Dakota (Riggs 1893), Udmurt (Winkler 2001), and Latin (Pinkster 1990). For four languages—Jivar, Tem, Abipon and Haka—I had no access to any reference grammar of the language and I had to rely on the analysis presented in Bybee et al. For Jivar, Tem and Abipon, Bybee et al.’s analysis could easily be adapted to the FG-analysis. However, Bybee et al.’s analysis of the Tibeto-Burman language Haka raised too many questions and could not be reanalyzed according to my own criteria. Therefore, it was replaced by another language of this family, Meithei (Chelliah 1997).

4 The reference grammar used in Bybee et al. (1994), Sorensen (1969), apparently mentions a polysemous marker of resultative and evidentiality, whereas the alternative reference grammar (Ramirez 1997) does not.
7.3.1 Selection of forms

For each language an inventory is drawn up of all the grammatical TMA expressions. Both formal and semantic criteria are used for the selection of grammatical expressions. I follow the formal criteria of Bybee et al. (1994: 37-39) and have added a fifth criterion:

i) The gram must belong to a closed class,
ii) The gram must have a fixed position in relation to the verb,
iii) The gram must be lexically general, i.e. combine with all verbs or with all members of a large semantic class of verbs,
iv) The gram must have a predictable meaning in most contexts.

v) The use of the gram is not restricted to subordinate clauses.

Because of the last criterion, purposive and subjunctive markers included in Bybee et al. are excluded here.

The analysis is furthermore restricted in a semantic way in that the gram has to have scope over the predicate, the predication or the proposition. Such operators belong to the first, second or third layer in FG at the representational level. Recall that the conceptions of scope in FG (2.4) and of relevance in Bybee (1985) do not coincide. Bybee (1985: 166) defines mood as ‘a marker on the verb that signals how the speaker chooses to put the proposition into the discourse context.’ Both markers of sentence type or illocution (indicative, imperative, hortative) and markers of the degree of commitment of the speaker to the truth of the proposition (like dubitative) come under her class of mood. ‘What all these markers of the mood category have in common is that they signal what the speaker is doing with the proposition, and they have the whole proposition in their scope’ (p.166). In FG, however, operators that indicate sentence type or illocution do not belong to the representational level but rather to the interpersonal level, since they are concerned with the level of performing a speech act, not with the constitution of the content of a clause. This leads to a further difference between the Bybee-studies and the current study.

Although the criteria for selecting forms to code do not differ fundamentally, a considerable number of forms are included in my study that are not documented in the inventories in Bybee et al. (1991; 1994) and vice versa. This may be explained partly by the fact that Bybee et al. (1994: 47) have concentrated on the meaning labels that occur more than twenty times in the sample whereas I do not make this restriction. However, it does not explain all the differences. For Slave for example there are many modal markers that are not documented in Bybee et al. and, to give another example, for Margi, Bybee et al. do not mention cítú as an ability and root-possibility marker, reduplication
as a marker of iterative (π1) or frequentative (π2) and ḫānā as a marker of repetition. Compared to the overview in Bybee et al., I often added expressions of modality, sometimes of hypothetical or counterfactual, evidential markers and zero-expressions.5

7.3.2 General coding

The selected grams are coded for form and function, again in a similar way as in the Bybee-studies. The first coding involved the expression form. The prime mode of expression, the prime allomorph, was coded for its expression form. A major distinction was made between bound and non-bound expression, according to the following criteria:

If a morpheme is inseparable from the stem, and/or occurs in a fixed order contiguous to the stem, or with only closed class items intervening between it and the stem, it is considered bound. If in the language descriptions this information is not given explicitly, then the author’s decision to write the morphemes as separate or bound is taken as an adequate indication of their status. (Bybee 1985: 27)

Bound expression is coded as inflection (prefix, suffix, infix), reduplication, stem-, stress- or tone-change or zero expression. Derivational expression is excluded, since in FG it does not have operator status. The criteria to distinguish derivational from inflectional markers are again based on Bybee (1985: 27):

i) an inflectional category is obligatorily marked every time a stem category to which it applies appears in a finite clause. It must be combinable with every stem;

ii) the meaning of the category must be predictable with every verb (be the same with all classes of verbs);

iii) in inflectional expression, the lack of a marker is meaningful.

In spite of the clear formulation of criteria, the distinction between derivation and inflection is not always clear in practice. If there is any doubt and if the gram expresses any of the notions summed up in Table 4-6, it is included in this study.

Non-bound grams are coded as an auxiliary or particle, according to the following criterion:

---

5 The term zero expression is used when the absence of a marker is meaningful and used in opposition with one or more non-zero expressions. This occurs when concepts are obligatorily expressed, such as e.g. past tense, and the non-use of this marker implies that this meaning is absent, e.g. non-past tense.
Inflected auxiliaries are easy enough to identify, since their inflection is indicative of verb-like behavior. Uninflected non-bound grams were coded as auxiliaries if they exhibited any verb-like behavior at all or if they appeared to bear an etymological relationship to a verb. Other uninflected forms were coded as particles. (Bybee et al. 1994: 42)

Finally, if an expression form consists of more than one element it is coded as a periphrastic expression. If it consists of only bound forms, for example a prefix and a suffix, it is considered a bound expression. If it consists of only non-bound forms, for example an auxiliary and a particle, or two auxiliaries, it is considered an unbound expression and if it consists of a bound and a non-bound form, for example an auxiliary and a non-finite form of the verb or a stem change, it is considered a combined expression.

In addition to coding the expression form, the gram is coded for its semantic function. The work of Bybee (1985), Dahl (1985), and Bybee et al. (1991; 1994) established the most prevailing TMA notions expressed in languages of the world. I have described the different gram types according to scope and used this list for assigning meaning labels to language-specific grams. This approach is based on the idea that language-specific grams cover parts of a more universal semantic space. And although the exact part of the semantic space covered may differ in detail from language to language, it is possible to define gram types that are frequently expressed with crosslinguistically comparable functions (Bybee et al. 1994: 48). Only grams are included that come under the general semantic domains of aspect, tense, quantification, irrealis, modality or evidentiality. This criterion excludes notions as negation and negative polarity. The list of meaning labels is presented in Table 7-1.

The semantic classification of morphemes is based on the definitions and descriptions presented by the author of the grammar, and on the examples and translations that are presented throughout the grammars. My classification is not necessarily the same as the author’s, since the labels used by different authors do not always cover the same semantic domain.

In general, the classification in Table 7-1 covers most semantic notions found in the grammars. In a few cases, the description in a grammar of a specific gram is too vague to label the gram with a specific semantic function; in those cases the gram is assigned a more general label such as aspect, property quantification or participant-oriented entailment. A few notions had to be added, for unusual semantic functions, such as inability and impossibility and tense markers that indicate distinctions in temporal distance without specifying whether the location in time is in the past or in the future.
### Table 7-1. TMA domains and semantic functions included in the study

<table>
<thead>
<tr>
<th>TMA Domain + specific semantic functions</th>
<th>Scope over</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aspect</strong></td>
<td></td>
</tr>
<tr>
<td>prospective, ingressive (inchoative),</td>
<td></td>
</tr>
<tr>
<td>continuative, progressive,</td>
<td></td>
</tr>
<tr>
<td>egressive / completive, imperfective,</td>
<td></td>
</tr>
<tr>
<td>perfective, perfect</td>
<td></td>
</tr>
<tr>
<td><strong>Property quantification</strong></td>
<td></td>
</tr>
<tr>
<td>iterative, intensity, semelfactive</td>
<td></td>
</tr>
<tr>
<td><strong>Participant-oriented modality</strong></td>
<td></td>
</tr>
<tr>
<td>potentiality: ability, root-possibility,</td>
<td></td>
</tr>
<tr>
<td>permission</td>
<td></td>
</tr>
<tr>
<td>disposition: volition, desirability</td>
<td></td>
</tr>
<tr>
<td>weak necessity: weak root-necessity,</td>
<td></td>
</tr>
<tr>
<td>weak obligation</td>
<td></td>
</tr>
<tr>
<td>necessity: internal need, root-necessity,</td>
<td></td>
</tr>
<tr>
<td>obligation</td>
<td></td>
</tr>
<tr>
<td><strong>Predicate (π1)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Tense</strong></td>
<td></td>
</tr>
<tr>
<td>past, present, future, non-past, non-future,</td>
<td></td>
</tr>
<tr>
<td>(± specification of temporal distance)</td>
<td></td>
</tr>
<tr>
<td><strong>Event quantification</strong></td>
<td></td>
</tr>
<tr>
<td>habitual, frequentative, repetitive,</td>
<td></td>
</tr>
<tr>
<td>distributive</td>
<td></td>
</tr>
<tr>
<td><strong>Event-oriented modality</strong></td>
<td></td>
</tr>
<tr>
<td>potentiality: root-possibility,</td>
<td></td>
</tr>
<tr>
<td>permission, epistemic possibility</td>
<td></td>
</tr>
<tr>
<td>disposition: desirability, (future)</td>
<td></td>
</tr>
<tr>
<td>weak necessity: weak root-necessity,</td>
<td></td>
</tr>
<tr>
<td>weak obligation, probability</td>
<td></td>
</tr>
<tr>
<td>necessity: root-necessity, obligation,</td>
<td></td>
</tr>
<tr>
<td>epistemic necessity (negative)</td>
<td></td>
</tr>
<tr>
<td><strong>Predicate (π2)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Irreals</strong></td>
<td></td>
</tr>
<tr>
<td>hypothetical, counterfactual</td>
<td></td>
</tr>
<tr>
<td><strong>Proposition-oriented modality</strong></td>
<td></td>
</tr>
<tr>
<td>potentiality: uncertainty, permission</td>
<td></td>
</tr>
<tr>
<td>disposition: prediction, desirability</td>
<td></td>
</tr>
<tr>
<td>weak necessity: weak certainty, weak</td>
<td></td>
</tr>
<tr>
<td>obligation</td>
<td></td>
</tr>
<tr>
<td>necessity: certainty, obligation</td>
<td></td>
</tr>
<tr>
<td><strong>Proposition (π3)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Evidentiality</strong></td>
<td></td>
</tr>
<tr>
<td>sensory evidence, inference, hearsay,</td>
<td></td>
</tr>
<tr>
<td>etc.</td>
<td></td>
</tr>
</tbody>
</table>

In particular in the domain of quantification, definitions are very inconsistently used across grammars. However, the translations and glosses are often similar and these were the main basis for classification in this domain. A translation as ‘V again’ is analyzed as a repetitive; ‘V again and again’ or ‘V over and over’ is analyzed as an iterative if the examples describe repetition on one occasion and/or as frequentative if the examples describe repetition on different occasions. In cases of doubt, it was analyzed as both. The glosses ‘Vs continuously’ or ‘Vs intensely’ are encoded as an iterative; ‘continues/keeps –
ing’ as a continuative; ‘Vs habitually’/ ‘uses to V’/ ‘Vs by custom/customarily’ as a habitual. If the gram indicates that the action is performed by different participants and/or at different places, it is coded as a distributive. In cases of doubt, the expression may be coded for more than one use or as a general marker of quantification.

Even though the bulk of grams could be classified without much trouble, a serious problem arose with the coding of modality in terms of FG. Remember that FG distinguishes modality types according to scope: root-modality, which indicates that the source of the modality lies in general external circumstances, can have scope over the predicate or over the predication (see 4.2.3.1); the same holds for deontic modality, which indicates that the source of the modality lies in a deontic authority (see 4.2.3.2). When the scope is the predicate (π₁), the modal marker specifies the relation between the argument and the predicate: X is not precluded from / is disposed to / is forced to ... When the scope is the predication (π₂), the modal marker specifies characteristics of the event that are independent of the participants: the event is not precluded from occurring / is disposed to occur / is forced to occur. An event-oriented interpretation in general arises when the modal marker is used in a passive or impersonal construction or with a non-specific participant. See the examples in (47)-(49):

(47) I know the place. You can get all sorts of things here. (P84)
(48) National pressure groups cannot exist without full time staffs and a regular income. (P91)
(49) Well, I’ll see what can be done and give you a ring. (P84)

These utterances all express that the event is not precluded from occurring independently of the participants involved in the event. Compare these to (50), in which can describes the root-possibility of the specific participant you:

(50) Can you pick your own trousers up? (C94)

However, since scope is not a notion generally applied in linguistics, most reference grammars lack the information that is necessary to determine whether a modal marker may be combined with non-specific participants. It is therefore questionable whether a marker of root-necessity or obligation and of root-possibility or permission in a specific language should be coded either as π₁- and π₂-operator or only as π₁-operator. Nevertheless, I have assumed that in many languages it will be possible to use participant-oriented markers in an event-oriented way, although this is not explicitly stated in the grammar used. In the sample, there are 65 markers of deontic and root modality that might be polysemous in having narrow scope (π₁) and medial (π₂) scope: there are 36
expressions for obligation, 23 for potentiality, 2 for disposition and 4 for more general modal markers. I have arbitrarily assumed that 50% of these potential event-oriented markers in each language can indeed be used as such and therefore of all the uncertain modal π2-operators 50% is included in the count. With this procedure the analysis should not result in the overclassification of one particular type. The tables in the following sections that include information about frequency, the figures for event-oriented modality (π2) always include half of the number of operators potentially used with medial scope, unless it is stated otherwise. Although deontic modality can also be used with scope over the proposition (π3), this use is probably much more restricted and marginal. I have only encoded a marker for this use if there was clear evidence in the grammar that it could be used this way.

A comparable, but slightly different, problem arose with epistemic modality: FG makes a distinction between event-oriented (π2) and proposition-oriented (π3) epistemic modality. Event-oriented epistemic modality expresses a general, objective fact about the actuality of the event. Proposition-oriented epistemic modality is a personal, subjective evaluation of the speaker about the reliability or truth of the propositional content. Since it is in general the context that forces the one or the other interpretation (see 4.2.3.3), it is often difficult on the basis of the isolated examples in the grammars to establish whether a marker should be considered as event-oriented, proposition-oriented, or both. When translations and definitions by the author allowed both interpretations, it was coded for both interpretations. This held for 41 expressions. They are all encoded as polysemous markers of event-oriented (π2) and proposition-oriented (π3) modality. As a result, there could be too many modality markers in my corpus, but this does not favor the hypotheses in any way.

Finally, some more refined modal distinctions in FG could not be applied to the grammars, because the examples were too scarce and not detailed enough. The distinction between the root-modal and deontic meanings of volition and desirability and of necessity and obligation had to be abandoned because grammars in general do not describe this distinction systematically.

7.3.3 Details of the analysis

Although most reference grammars used for the inventory of TMA expressions in each language were the same as in Bybee et al. (1994), and, in general, the gram types are defined similarly, there are many differences between the analyses of TMA systems in Bybee et al. (1991; 1994) and in this study. As an illustration, consider Table 7-2 in which the two analyses of the TMA system in Chacobo are compared:
<table>
<thead>
<tr>
<th>No.*</th>
<th>Form</th>
<th>Meaning</th>
<th>Form</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td><em>ki</em></td>
<td>perfective</td>
<td><em>kï</em></td>
<td>perfective (π1)</td>
</tr>
<tr>
<td>41</td>
<td><em>ki</em></td>
<td>completive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td><em>ki</em></td>
<td>present</td>
<td><em>ki</em></td>
<td>imperfective (π1)</td>
</tr>
<tr>
<td>35</td>
<td><em>ya</em></td>
<td>past hodiernal</td>
<td><em>ya</em></td>
<td>past hodiernal (π2)</td>
</tr>
<tr>
<td>36</td>
<td><em>tïta</em></td>
<td>past hesternal</td>
<td><em>tïta</em></td>
<td>past hesternal (π2)</td>
</tr>
<tr>
<td>37</td>
<td><em>yami</em></td>
<td>past immediate</td>
<td><em>yami</em></td>
<td>past immediate (π2)</td>
</tr>
<tr>
<td>38</td>
<td><em>ni</em></td>
<td>past remote</td>
<td><em>ni</em></td>
<td>past remote (π2)</td>
</tr>
<tr>
<td></td>
<td>ZERO</td>
<td>non-past</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>šari</em></td>
<td>future (tomorrow)</td>
<td><em>šari</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>pari</em></td>
<td>prospective</td>
<td><em>pari</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>kas</em></td>
<td>volition</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td><em>ca</em></td>
<td>progressive</td>
<td><em>ca / ci</em></td>
<td>progressive (tr./intr.) (π1)</td>
</tr>
<tr>
<td>20</td>
<td><em>ci</em></td>
<td>progressive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td><em>pao</em></td>
<td>past habitual</td>
<td><em>pao</em></td>
<td>past habitual (π2)</td>
</tr>
<tr>
<td>9</td>
<td><em>rabi</em></td>
<td>iterative</td>
<td><em>rabi</em></td>
<td>iterative (π1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>rabi</em></td>
<td>frequentative (π2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>ria</em></td>
<td>iterative / intensive (π1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>tïki</em></td>
<td>repetitive (“again”) (π2)</td>
</tr>
<tr>
<td>48</td>
<td><em>mica</em></td>
<td>epistemic possibility</td>
<td><em>mica</em></td>
<td>epistemic possibility (π2)</td>
</tr>
<tr>
<td>23</td>
<td><em>tiari</em></td>
<td>potential</td>
<td><em>tiari</em></td>
<td>probability (π2)</td>
</tr>
<tr>
<td></td>
<td><em>kara</em></td>
<td>evidential</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* *Every separate linguistic form in a language was given a number in Bybee et al. (1994).*

A first difference is that *kï* is analyzed as a perfective and a completive in Bybee et al. and only as a perfective in this study. *Ca* and *ci* are analyzed as two different markers of progressive aspect in Bybee et al. and as two variants for the progressive in this study, one for transitive verbs, one for intransitive verbs. The marker *ki* is analyzed as a present tense marker by Bybee et al., but here as
an imperfective, whereas zero marking is coded as a non-past tense. Rabi is not only used as a marker of iterative (repetition at one occasion, π1) but also as a frequentative (repetition at different occasions, π2). Several forms are not documented by Bybee et al.: a marker for the prospective, the future tense, a second iterative or intensity expression, a marker of repetition, volition and an evidential (Prost 1962: 116-18).

As Table 7-2 shows, there are several points on which I did not agree with the analyses in Bybee et al. (1991, 1994). This section discusses some details of the analysis carried out in this study. It concentrates on those aspects of the methodology here used that deviate from Bybee et al. (1991, 1994). For the definitions of the specific semantic functions, I refer to Chapters 3 and 4.

When making an inventory of TMA expressions, a first point on which one has to decide is the number of different uses or functions that are separately coded. In Bybee et al. (1994) the number of uses coded is determined by the following criteria:

(1) Two uses of a single morpheme are distinguished if these uses would be expressed by different grams in another language. (…)
(2) Two distinct uses of a gram will have two different paraphrases. (…)
(3) A gram might take on a different sense when it occurs with other grammatical morphemes or with verbs of a specific semantic class. (p.44-45)

Criteria 2 and 3 are taken as a guideline here, but I consider criterion 1 very problematic. It would imply that first the language with most grams for a certain semantic domain has to be discovered before it can be objectively established how many uses a marker in other languages has. If for example, a language were to be discovered that makes eight or even ten remoteness distinctions in the past tense, it would imply that in every other language the expression for the past tense would in fact have to have these same eight or ten different uses. However, within the specific language, the remoteness distinctions might be completely irrelevant to the grammatical system, and it would be very undesirable to treat this language as if this distinction was relevant.

The approach taken here is the opposite: if one expression covers different senses within the same domain (e.g. aspect) and with the same scope, only the most general semantic function is coded. This is in accordance with the analysis advocated in an earlier work of Bybee (1985):

If a morpheme appeared in a variety of functions that are usually associated with the imperfective (i.e. continuous, habitual, durative), then the morpheme was labeled an imperfective. If it appeared to be restricted to one of these functions, it was given a more specific label (not necessarily the label used by the author). (p.141)
This difference in approach explains why Bybee et al. listed a completive and a perfective for *ki* in Chacobo (Table 7-2), whereas I only listed the most general aspectual meaning for this marker, the perfective (which may also be used with a completive sense). If, however, different uses of a single expression correlate with a distinction in scope, I considered them as separate uses. If the scope of an expression may vary this means that the communicative functions of the expression vary and that was considered an important criterion to speak of different uses. As a matter of fact, a difference in scope has not been a criterion in Bybee et al. It explains why *rabi* is listed only as an iterative in Bybee et al., whereas I analyzed it as an iterative with scope over the predicate and as a frequentative with scope over the predication.

As well as there being a principal difference in the number of encoded uses, differences in the definitions of specific grammatical types sometimes lead to deviating analyses. In what follows, the details are presented of the analyses of specific grammatical types.

**Perfective and perfect.** One of the possible uses of a general perfective or a perfect is the combination with a stative verb, which yields an inceptive or ingressive meaning (Comrie 1976). Contrary to the analysis in Bybee et al. (1994), this was not distinguished here as a separate use. Only if a grammeme could only express an ingressive, was it coded as an ingressive. The same holds for a perfect, which can be used in some languages to indicate a resultant state or an immediate past ('hot news') (Dahl 1985). Only when a specific morpheme could only express a resultant state or an immediate past, was it coded as such. For Palantla, for example, Bybee et al. (1994: 79, T.3.7) coded the marker *ma*- (Merrifield 1968: 25) as an anterior (= perfect) together with another use, ‘state exists’. I, however, coded it only as a perfective. For Inuit, Bybee et al. (1994: 53, T.3.1) coded *sima* (Fortescue 1984: 265-66) as a marker of perfect, resultative and past tense. In this thesis, it was only coded as a perfect, since a perfect also covers the sense of a resultative and since the sense of past tense is only an implicature, not a basic meaning of the verb.

**Non-past and non-future.** In Bybee et al. (1994) the semantic functions non-past and non-future are not recognized. An expression that indicates both present and future tense or both past and present tense is listed two times for both uses. In contrast to this, I coded these markers only once, with a broader semantic sense of non-past or non-future.

**Pluperfect.** If the combination of markers results in a predictable meaning, this combination is not considered in my analysis as a separate expression or semantic category, so if a pluperfect meaning is expressed by a combination of a perfect marker and a past tense marker—like in English—the pluperfect is not considered a separate expression or semantic category in this language. See for a discussion of separate versus non-separate grammatical types Dahl (1985: 67).
Portmanteau expressions of tense and aspect. Quite often, a perfective-imperfective distinction is only relevant for reference to past situations. It is questionable whether this phenomenon should be interpreted as basically a temporal distinction between present and past tense with a subdivision of the past in perfective and imperfective aspect (the “traditional” view), or else, as basically an aspectual distinction between perfective and imperfective with a subdivision of the imperfective in past and present tense. See Figure 7-7:

Dahl (1985: 78-83) argues for a classification that is primarily based on aspect, which is also the analysis adopted in Bybee et al. (1994). Dahl argues that the traditional distinction of a past perfective and past imperfective on the one hand and a present tense on the other hand is incorrect, since there is no common expression form for the past perfective and the past imperfective, but rather a common expression form for the present and the past imperfective, as for example in many Indo-European systems. The prototypical use of a perfective should be to refer to single, completed events, located in the past. This makes further past time reference redundant and the basic distinction then, would be between perfective and imperfective aspect, whereas the imperfective aspect is divided for tense. However, Dahl also observes that there is crosslinguistic variation in the extent to which perfectives can refer to non-past situations and that a perfective is not necessarily restricted to past time reference (cf. the analysis of the present tense in English presented in 6.2.3.2):

It is said that a Perfective verb form must refer to a ‘total’ event and that what happens at the moment of speech is necessarily on-going rather than completed. A Perfective Present would therefore be a contradiction in terms. However, there are candidates for ‘total’ events at the moment of speech: notably the so-called reportive present and performatives. In fact, PFV is used in such cases in several
Dahl’s main argument is that the formal expression in many languages is used for both a present tense and a past imperfective. However, my sample shows that crosslinguistically, this is hardly ever the case. Still, if it would have been the case, it is no argument for the ‘primary aspect’ analysis, on the right in Figure 7-7. The developmental paths of grams account for the expression forms used, but what matters most to the analysis, is what is the contrastive use of the expression forms from a synchronic perspective. Since a perfective is not in principle restricted to past tense, I maintain the traditional tripartite analysis of a basically temporal distinction and a secondary aspectual distinction in languages where the perfective is indeed restricted to past tense reference.

**Future, prospective and prediction.** The conception of the future in Bybee et al. (1991, 1994) is not similar to the one in FG. Bybee et al. (1994: 244) regard ‘the focal use of future as equivalent to a prediction on the part of the speaker that the situation in the proposition, which refers to an event taking place after the moment of speech, will hold.’ Admittedly, the use of a future tense marker implies prediction by the speaker because intervening events may lead to a change, but signaling prediction by the speaker is not the essential function of the future tense marker. The essential function of a future marker is indicating that an event will take place in the future, whereas the uncertainty of future events belongs to mutual common knowledge of language users. This is not something that has to be marked linguistically. Future tense markers can undergo further subjectification, so that real epistemic senses develop (also recognized in Bybee et al. (1991)). The function of such a marker can develop into a true marker of prediction on the part of the speaker, as discussed in 6.2.5.2.3, see example (51):

(51) A commotion in the hall: “That will be Celia,” said Janet. (C177)

Contrary to a marker that indicates future tense, the situation in (51) holds at the present time: Celia is already present and makes noises at that time, but it is only heard and not seen. *Will* does not indicate that the event will take place in the future, but that the proposition will turn out to be true in the future.

A further difference in the approach of the future tense is in the stages before a marker expresses a future meaning. I agree with Bybee et al. (1994: 254) that each marker that develops into a future tense marker goes through a stage in which it expresses the intention of the speaker, and later of the animate agent of the main verb. The stage of intention seems to lie somewhere in between participant-oriented modality like volition or obligation and future tense. On the way to becoming a future tense marker, the gram loses the
restriction to an animate agent or a controlled situation. A possible precursor of
a future tense marker not recognized in the study of Bybee et al. (1991, 1994), is
a prospective aspect marker (*be going to* in English). This is a category distinct
from a future tense: it does not just express immediate future, which is also
recognized in Bybee & Dahl (1989):

In each case where *be going to* is used, the entity involved is interpreted as already
on the path leading to the goal expressed by the main verb (…) whereas the
statements with *will* make predictions for some unspecified future with no
implication of connection to present states. (p.92)

Just as a perfect can develop into a past tense marker, a prospective can
develop into a future tense marker, when it loosens its connection with the
present moment.

**Quantification.** The definitions here used to (sub)categorize the domain of
quantification are not entirely the same as the definitions used in Bybee et al.
(1994). The major difference is that I make a distinction between quantification
of the predicate and quantification of the predication, the event. Quantification
of the predicate may be iteration or intensity of a certain action at one moment
by the same participant(s), which expresses differences like *jump one jump* versus
*jump a lot of jumps*, or *think* versus *think over*. According to Bybee (1985: 100)
iterative markers are usually derivational because ‘they are applicable primarily
to semelfactive verbs (although they may apply to activity verbs as well) but not
to stative verbs.’ The meaning of iteration or repetition is in her view only
applicable to ‘active verbs, and then only to certain types of active verbs, i.e.
those which describe telic events, events that have identifiable endings.’ (1985:
150) However, the restriction to a certain class of verbs is not in itself a
criterion to classify an expression form as derivational: the use of a progressive
or a continuative is also often restricted to non-stative verbs and a perfective
may be primarily used with telic, punctual verbs. These categories are, however,
not considered as derivational. I furthermore doubt the restrictive applicability
of iterative markers. In several languages, the same marker can express iteration
and intensity depending on the types of verb with which it combines. I
therefore include iterative markers in my research: they potentially develop into
more general markers of quantification, a habitual or frequentative marker.

A further difference within the quantificational domain lies in the approach
of the habitual. Bybee et al. (1994) consider the habitual a subcategory of
imperfective aspect, similar to Comrie (1976: 25). Bybee (1985: 143) claims that
a progressive gram may loose its selection restrictions and develop habitual
uses. When that happens, the gram has become an imperfective or present
tense marker. In FG, however, an imperfective is considered an aspectual
marker in the sense described in 3.3.3, covering progressive and continuative
uses, without selection restrictions; a habitual marker on the other hand is
considered a marker with medial scope, that indicates the quantification of the event. An imperfective marker may indeed develop habitual inferences, which may conventionalize into the real meaning of this marker. If that happens, the gram is coded for both uses in my analysis: imperfective aspect ($p_1$) and habitual ($p_2$). A habitual use is thus not considered a subcategory of an imperfective, but a separate semantic function, that may also arise directly from lexical sources (like used to in English).

**Modality.** Although the subcategories within the domain of modality used here clearly differ from the study of Bybee et al. (1994), there are also many similarities. The participant-oriented notions used in this study are practically similar to the agent-oriented uses in Bybee et al. (1994), such as ability, root-possibility, permission, volition, (weak and strong) obligation/necessity. The same holds for epistemic modal markers: uncertainty, certainty, epistemic possibility and probability. The main difference is due to the analysis according to scope, which results in a more fine-grained distinction within the different modal senses and sources. As discussed in 7.3.2, the reference grammars of the languages in the sample do not always provide enough information to determine the possible scopes of an expression, in particular in the case of deontic and root-modal meanings. This problem was overcome by including potential event-oriented expressions for 50% in the counts (see 7.3.2).

Although many of the differences between the analyses in Bybee et al. (1991, 1994) and the analysis here result from the distinctions set out above in the method of approach, some are a matter of different insights and interpretations of the examples and definitions of the authors. In some cases my interpretation of the reference grammar deviates quite fundamentally from the analysis in Bybee et al., as, for example, in the case of the tense/aspect system in Chacobo (see Table 7-2).

A final remark on the analysis is necessary with respect to the language Abipon. This is one of the three languages for which the analysis of Bybee et al. (1994) is adopted because there was no reference grammar available. However, a crucial gram appears to be missing in the list of all the TMA expressions in the sample (their Appendix C). This list mentions only two TMA expressions for Abipon—a habitual and a future marker. No aspectual markers are presented. However, the written text does mention the existence of a progressive marker in Abipon, expressed by a suffix (1994: 119). I added this progressive marker to my inventory of TMA expressions in the GRAMCATS sample.

A complete overview of all the morphemes in the languages according to my analysis is presented in Appendix G. It served as the basis for testing the hypotheses on frequency, synchronic configurations, expression form and expression order.
7.4 Frequency

As stated in 5.4.2 the Scope Hierarchy predicts a relation between the scope of TMA markers and their frequency. It is to be expected that at the start of a grammaticalization cline (narrow scope) there is a wider variety of expressions and semantic notions, because i) during grammaticalization, developmental paths tend to converge and ii) the function of narrow scope operators (modifying the description of a state of affairs) seems to be communicatively more motivated than of medial scope operators (modifying reference to the real world) or of wide scope operators (modifying the presentation of the content) (see 5.4.2). With respect to crosslinguistic frequency, this leads to two specific hypotheses on the size of different operator classes (H2b) and on the crosslinguistic incidence of operators (H2c).

**H2b:** The class of $\pi_1$-operators has more members than or an equal number of members as the class of $\pi_2$-operators and this class has more members than or an equal number of members as the class of $\pi_3$-operators.

**H2c:** The crosslinguistic incidence of $\pi_1$-operators is larger than or equal to the incidence of $\pi_2$-operators, which in turn is larger than or equal to the incidence of $\pi_3$-operators.

In this section the hypotheses will be tested in the GRAMCATS sample. The crosslinguistic frequency (H2c) will be discussed first (7.4.1).

### 7.4.1 Crosslinguistic frequency

The crosslinguistic frequency of operator types (H2c) is determined on the basis of the inventory of TMA expressions in the GRAMCATS sample. Table 7-3 presents the number of languages that have at least one expression that functions as a certain type of operator. Nearly all languages in the sample have

<table>
<thead>
<tr>
<th>Type of Operator</th>
<th>$\pi_1$</th>
<th>$\pi_2$</th>
<th>$\pi_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>76</td>
<td>73</td>
<td>37</td>
</tr>
<tr>
<td>%</td>
<td>100 %</td>
<td>96 %</td>
<td>49 %</td>
</tr>
</tbody>
</table>
expressions that function as $\pi_1$-operators and expressions that function as $\pi_2$-operators. Only half of the languages have expressions that function as $\pi_3$-operators. Whether there is an implicational relation between the presence of the different operators in single languages will be examined in 7.5.

The crosslinguistic frequency of operators supports the hypothesis that narrow and medial scope operators are much more frequent than wide scope operators. The frequency of medial scope operators is about equal to the frequency of narrow scope operators, which is in accordance with H2c. The distribution suggests that medial and narrow scope operators are equally relevant to communication.

### 7.4.2 Size of operator classes

#### 7.4.2.1 Method

The size of operator classes (H2b) is tested in different ways: the number of different operators is compared within and across languages and the number of TMA markers for a specific class are counted in two different ways. The size of an operator class may be determined by the number of different expression forms within that class or by the number of different semantic functions within that class. The numbers of different expression forms and of different semantic functions are of course not always identical; different forms may express the same semantic notion or one form may express different semantic notions (for example in polysemous or portmanteau expressions).

How the expression forms and the semantic functions are counted needs some further discussion. The number of different expression forms is based on what the author of the grammar considers different forms. If different forms are presented as allomorphs, they all count as one expression form. Furthermore, the different expression forms are counted within one semantic domain, such as tense, event quantification or irrealis. If the same expression is used as a marker of tense, of quantification and of irrealis, it is counted three times. Determining the number of different semantic functions within a language is a more difficult issue. Again, the description in the reference grammar is decisive. If the author describes the meaning of two or more forms in identical terms, then the forms are considered as expressing the same semantic function. For example, the expressions *malila*, *wandicila* and *pinga* in Mweru are all described as markers of prospective, without any further semantic distinction mentioned for the three expressions (Harries 1950: 109-10). These three forms are thus counted as three separate expression forms and as one separate semantic function. Remoteness distinctions in temporal or
aspectual categories are considered as indicating different semantic functions. Mwera, for example, has a marker for immediate perfective, *ci*, and for remote perfective, *ile* (Harries 1950: 79ff, 94ff). Since the use of *ci* and *ile* is mutually exclusive, they are counted as two different semantic functions (and of course also as two different expressions).

A complex category is formed by portmanteau expressions. A portmanteau expression denotes two or more semantic functions by only one morpheme, such as the –s suffix in English, which expresses present tense, third person and singular. A certain semantic function may only occur in a portmanteau expression or both in a separate expression and as part of a portmanteau expression. For example, Agau has a general past marker (–γwà) and a past marker that occurs in the portmanteau expression *šiŋ* that indicates past habitual and past imperfective (Hetzron 1969: 13ff, 28ff). This is counted as two expression forms for past (–γwà and *šiŋ*), but only one semantic function ‘past tense’. The portmanteau expression *šiŋ* is furthermore counted as an expression form for the habitual and an expression form for the imperfective and as the semantic functions ‘habitual’ and ‘imperfective’.

Another example may be taken from Guaymi that expresses most tense and aspect oppositions by eight portmanteau expressions (see Table 7-4). The suffix –*ri* for example indicates perfective immediate non-future while the suffix –*e* or –*a* indicates imperfective immediate non-future (Kopesec 1975: 22). This results in eight expressions indicating aspect as part of their semantics, and eight as indicating tense as part of their semantics. There are, however, only two different aspectual functions—perfective versus imperfective—while there are four different temporal functions: immediate future, remote future, immediate non-future and remote non-future.

Table 7-4. Eight portmanteau expressions in Guaymi (Kopesec 1975: 22)

<table>
<thead>
<tr>
<th>Immediate</th>
<th>Non-future</th>
</tr>
</thead>
<tbody>
<tr>
<td>Future</td>
<td>perfective</td>
</tr>
<tr>
<td>Non-future</td>
<td>perfective</td>
</tr>
</tbody>
</table>

In the domain of modality, the most general semantic function of an expression is counted per operator type. So, if an expression with narrow scope can be used to indicate ability (*π1*), root-possibility (*π1*) and permission (*π1*), it is only counted as one (general) marker of potentiality (*π1*). Secondly, if there are two different grams with that for example both express ability (*π1*) and root-possibility (*π1*), then this is counted as one semantic function “potentiality”. If, however, different expressions cover different subcategories
of the same sense, then they are considered separate semantic functions. For example, in Cantonese, *nàng gāu* expresses only root-possibility, *dak* expresses ability and root-possibility and *bó yī* expresses root-possibility and permission (Kwok 1971: 73, 77-78). Since these forms all cover different categories of the general sense potentiality, they are counted as three different semantic functions. Thirdly, if there are two expressions for obligation and the author indicates that one expresses weak obligation and the other one strong obligation, then they are considered two distinct semantic functions. Fourthly, if negation is part of the modal meaning (scope of negation is modal sense), for instance, *im* possibility or *in* ability, it is counted as a category separate from the marker expressing the positive counterpart. Finally, every evidential expression is considered a separate semantic function.

As for the domain of irrealis, expressions that indicate hypothetical or counterfactual, or a more general notion of irrealis, I tried to include only those expressions that can be used in main clauses. Negative and affirmative forms of an irrealis are not considered as different semantic notions, contrary to negative and affirmative forms of modal markers, since here the scope of the negation is not the irrealis marker, but the predication. This holds for Zuni, in which – *anna* and – *šuk* "a both express hypothetical, the latter including negation (S. Newman 1965: 37-38). If there is one form that expresses both hypothetical and counterfactual meaning, then it is counted as one broad semantic function irrealis. This is the case for Bongu, in which the suffix – *dun* can be used to express hypothetical and counterfactual meaning (Hanke 1909: 45, 87).

### 7.4.2.2 Results

In Table 7-5 the number of expression forms and semantic functions for the complete sample are presented. The numbers are the cumulative numbers of the different expressions forms and the different semantic functions within each specific language. Table 7-5 clearly shows that in the whole sample, there are more expression forms and semantic functions with narrow scope than with medial scope. Forms and semantic functions with wide scope are by far the

| Table 7-5. Number of expression forms and semantic functions per type of operator |
|---------------------------------|---|---|---|
| Type of Operator               | σ1 | σ2 | σ3 |
| Expression forms               | 451 | 413 | 89 |
| Semantic functions             | 387 | 352 | 79 |
The distinction between the classes of $\pi_1$- and $\pi_2$-operators is, however, very small. This suggests again that, contrary to the expectation expressed in 5.4.2, $\pi_1$- and $\pi_2$-operators may be about equal in communicative relevance.

As well as a comparison across languages, the hypothesis can be tested by investigating the number of languages in which the class of $\pi_1$-operators is larger than or equal to the class of $\pi_2$- and $\pi_3$-operators and the number of languages in which the class of $\pi_2$-operators is larger than or equal to the class of $\pi_3$-operators. This analysis is presented in Table 7-6. If a language lacked expressions of both scope types that were necessary for the comparison, then the language was excluded from analysis.

**Table 7-6.** Percentage of languages (N = 76) that are in accordance with H2b

<table>
<thead>
<tr>
<th></th>
<th>$\pi_1 \geq \pi_2$</th>
<th>$\pi_1 \geq \pi_3$</th>
<th>$\pi_2 \geq \pi_3^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expression forms</td>
<td>59%</td>
<td>96%</td>
<td>99%</td>
</tr>
<tr>
<td>Semantic distinctions</td>
<td>61%</td>
<td>97%</td>
<td>99%</td>
</tr>
</tbody>
</table>

*Three languages lacked both $\pi_2$- and $\pi_3$-operators: for 72 out of the remaining 73 languages, $\pi_2 \geq \pi_3$.

Whether the languages are compared on the basis of their semantic functions or on the basis of their expression forms, makes hardly any difference. If the expression forms are compared, the number of languages that have more $\pi_1$- than $\pi_2$-operators or an equal number is 59%. In 34 of the languages the group of $\pi_1$-operators is larger than of $\pi_2$-operators, in 10 languages both classes have equal size and in 32 languages the group of $\pi_2$-operators is larger than of $\pi_1$-operators. This shows again that the classes of $\pi_1$-operators and of $\pi_2$-operators are not significantly different in size. This is not in conflict with the hypothesis—the frequency of $\pi_1$-operators is not lower than that of $\pi_2$-operators—but it also does not support the idea that $\pi_1$-operators are communicatively more motivated than $\pi_2$-operators and therefore show more variation. However, H2b is strongly supported by the fact that in the vast majority of languages the group of $\pi_1$- and $\pi_2$-operators is larger than of $\pi_3$-operators. This is completely in accordance with the predicted distribution.

### 7.4.2.3 Summary

H2b on the size of operator classes has been examined in different ways. The different analyses show that the class of $\pi_1$-operators is slightly larger than of $\pi_2$-operators and the class of $\pi_3$-operators is much smaller. In general, this supports the hypothesis. The Scope Hierarchy makes the correct predictions.
However, although the class of $\pi_1$-operators is larger than of $\pi_2$-operators, the difference is very small. This suggests that the communicative relevance of both groups is about equal, whereas it was assumed initially that $\pi_1$-operators are more relevant for communication than $\pi_2$-operators. The results clearly confirm the assumption that $\pi_3$-operators are communicatively less relevant: there are not as many semantic distinctions within this class as in the other two classes and the number of expression forms is also very small compared to the other classes of operators.

### 7.4.3 Reconsidering H2b and H2c

Although the figures in 7.4.1 and 7.4.2 support H2b and H2c to a certain extent, it is questionable whether scope is indeed the decisive factor. The figures might be the result of the relatively high or low frequency of specific semantic functions and it is possible that the effects are not directly the result of scope differences, but rather of purely semantic factors. The hypotheses are both reconsidered in this section by examining the role of semantic functions.

Which semantic functions are the most or the least widespread in the sample? First, a general overview is presented of the occurrence of larger semantic domains in Table 7-7. This table shows that there may indeed be a large influence of specific semantic categories. The vast majority of languages have one or more expressions of aspect and tense. Event quantification is also present in nearly two-thirds of the languages. The influence of these functions on the frequency of operator classes is thus very large. Different types of modality are present in about one half of the languages. The grammatical

<table>
<thead>
<tr>
<th></th>
<th>Aspect ($\pi_1$)</th>
<th>Tense ($\pi_2$)</th>
<th>Property quantification ($\pi_1$)</th>
<th>Event quantification ($\pi_2$)</th>
<th>Irrealis ($\pi_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>76</td>
<td>68</td>
<td>26</td>
<td>46</td>
<td>21</td>
</tr>
<tr>
<td>%</td>
<td>100</td>
<td>89</td>
<td>34</td>
<td>61</td>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Participant-oriented modality ($\pi_1$)</th>
<th>Event-oriented modality ($\pi_2$)</th>
<th>Proposition-oriented modality ($\pi_3$)</th>
<th>Evidentiality ($\pi_3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>42</td>
<td>41</td>
<td>36</td>
<td>13</td>
</tr>
<tr>
<td>%</td>
<td>55</td>
<td>54</td>
<td>47</td>
<td>17</td>
</tr>
</tbody>
</table>
In Table 7-8 the exact numbers of expression forms and different semantic functions are presented for each semantic domain. It appears that there are by far the most expressions and semantic notions for aspect and tense and fewest for quantification, irrealis and evidentiality. The groups of $\pi_1$- and $\pi_2$-operators are indeed for the majority made up of expression forms for aspect and tense. This makes it more difficult to see the influence of scope on frequency clearly.

The study on English in Chapter 6 showed that a more precise way of testing the hypotheses on frequency is to eliminate the possible influence of specific semantic functions. The distribution of scope within a single semantic domain will therefore be examined. If the frequencies of the related domains of aspect and tense in Table 7-8 are compared, the hypothesis is supported: aspect markers are more frequent than tense markers, both within and across languages. However, the opposite result is yielded when property quantification is compared to event quantification, as here, operators with narrower scope occur less frequently. Modality expressions are the only ones that occur with all types of scope. Hence, they are the best category to examine the influence of scope. See Table 7-9, which only includes modal expressions.

---

Table 7-8. Number of expressions and semantic functions for each TMA domain

<table>
<thead>
<tr>
<th></th>
<th>Aspect ($\pi_1$)</th>
<th>Tense ($\pi_2$)</th>
<th>Predicate-qnt ($\pi_1$)</th>
<th>Event-qnt ($\pi_2$)</th>
<th>Irreals ($\pi_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expression forms</td>
<td>306</td>
<td>228</td>
<td>33</td>
<td>67</td>
<td>33</td>
</tr>
<tr>
<td>Semantic functions</td>
<td>259</td>
<td>188</td>
<td>29</td>
<td>65</td>
<td>27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Participant-oriented modality ($\pi_1$)</th>
<th>Event-oriented modality ($\pi_2$)</th>
<th>Proposition-oriented modality ($\pi_3$)</th>
<th>Evid. ($\pi_3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expression forms</td>
<td>112</td>
<td>86</td>
<td>64</td>
<td>25</td>
</tr>
<tr>
<td>Semantic functions</td>
<td>99</td>
<td>72</td>
<td>54</td>
<td>25</td>
</tr>
</tbody>
</table>

Notes: $\text{qnt} = \text{quantification} ; \text{evid.} = \text{evidentiality}$

expression of property quantification, irrealis and evidentiality is the least frequent.

6 Bhat (1999: 98-99) warns against a temporal bias in most reference grammars: ‘because English is the language of description in most reference grammars and in English tense is—in contrast to aspect and modality—obligatorily marked, researchers are biased to defining and translating other languages in temporal terms. ‘It is something like trying to understand the colour of various objects around us while looking at them through a red-coloured glass.' This warning is probably not an exaggeration and it thus has to be taken into account that especially the number of tense markers might be lower if languages were reanalyzed with the present-day understanding of modality and aspect.
Table 7-9. Number of modal expression forms and semantic functions per type of operator

<table>
<thead>
<tr>
<th>Type of Operators</th>
<th>(\pi_1)</th>
<th>(\pi_2)</th>
<th>(\pi_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expression forms</td>
<td>112</td>
<td>85</td>
<td>64</td>
</tr>
<tr>
<td>Semantic functions</td>
<td>99</td>
<td>72</td>
<td>54</td>
</tr>
</tbody>
</table>

When only modality is taken into consideration, the hypothesis that there are more operators with narrow scope than with medial scope, and more operators with medial than with wide scope is borne out: the group of \(\pi_1\)-operators is larger than of \(\pi_2\)-operators and the group of \(\pi_2\)-operators is larger than of \(\pi_3\)-operators. The distinctions in size between the three classes are similar. This conclusion also holds within languages: the average size of the \(\pi_1\)-class (2.6 expression forms) is larger than the average size of the \(\pi_2\)-class (2.1 expression forms), which is again larger than the average size of the \(\pi_3\)-class (1.8 expression forms).

7.4.4 Conclusion

The analyses demonstrate that operators with wide scope are far less frequent than the other two categories, whatever analysis is applied. Their crosslinguistic incidence is lower than of the other two categories and the class of \(\pi_3\)-operators is smallest, both across and within languages. Operators with narrow and medial scope are about equally represented, both across and within languages, and their classes have about equal size. Narrow scope operators are slightly more frequent, but the difference with \(\pi_2\)-operators is small. However, when the specific semantic domains are compared, it appears the frequency of operator classes is largely affected by a few semantic categories, aspect and tense, which are overwhelmingly present in the GRAMCATS sample. In order to investigate the role of scope more clearly, the analysis was restricted to the domain of modality. When only modality is taken into consideration, there does seem to be a difference in frequency that is dependent on scope: the number of participant-oriented modal markers is larger than of event-oriented modal markers and the number of proposition-oriented modal markers is smallest. This distribution holds both across and within languages.

The figures support the view that scope plays a role in the frequency of operators, both in crosslinguistic incidence and in the size of operator classes. However, there is also a large effect of semantic functions on frequency. The hypotheses were based on the assumption that the function of \(\pi_1\)-operators,
modifying the descriptive function of an utterance, is communicatively more important than the function of \( \pi_2 \)-operators, modifying the situating function of an utterance. The function of \( \pi_3 \)-operators, modifying the presentation of the content, was assumed to be communicatively least relevant.

The figures support the view that description is in fact a very important communicative function. Every language in the sample has at least one expression to modify the description of the event. These modifications concern for the major part aspectual specifications, but participant-oriented modality also occurs in 55% of the languages. It is however questionable whether the function of relating the description of the event to the concrete event the speaker has in mind (\( \pi_2 \)-operators) is indeed less relevant to communication, since most languages (except two) also have grammatical expressions to modify the situating function of an utterance. For the majority, this concerns tense markers, but 61% of the languages also have an expression for event quantification and 54% for event-oriented modality.

To conclude, the Scope Hierarchy makes correct predictions for the frequency of operators, but the difference between \( \pi_1 \)-operators and \( \pi_2 \)-operators is much smaller than the difference between these two categories as opposed to \( \pi_3 \)-operators. The frequency of \( \pi_1 \)- and \( \pi_2 \)-operators is about equal, whereas the frequency of \( \pi_3 \)-operators is considerably lower.

### 7.5 Synchronic Configurations

This section will examine the hypotheses regarding synchronic configurations of TMA expressions. The first hypothesis, H3a, is based on the markedness of different operators and concerns the limits on possible synchronic configurations of all TMA expressions. The second hypothesis, H3b, concerns the limits on possible meanings that can be expressed by a single item. This concerns polysemous and portmanteau expressions.

#### 7.5.1 Implicational presence

As already set out in 5.4.3 the Scope Hierarchy, which reflects a distinction in markedness, predicts that if a language has grammatical means to make modifications with wide scope (\( \pi_3 \)-operators), it also has grammatical means to make modifications with medial scope (\( \pi_2 \)-operators) and if it has grammatical means to make modifications with medial scope, it also has grammatical means to make modifications with narrow scope (\( \pi_1 \)-operators). The motivation for this implicational relation is that a language supposedly does not allow more abstract modifications by grammatical means without the possibility of more
concrete modifications by grammatical means. The more marked functions can only evolve if the less marked functions are present. This is formulated in H3a:

H3a: The presence of operators with wider scope depends on the presence of operators with narrower scope according to the hierarchy: $\pi_1$-operator $\subset \pi_2$-operator $\subset \pi_3$-operator.

H3a is tested in two ways. First, it will be examined which configurations occur in the GRAMCATS sample. Second, a more detailed discussion will follow on TMA systems in creole languages, in 7.5.1.1.

For each language in the GRAMCATS sample the types of operators that co-occur are examined on the basis of the inventory of TMA expressions. Table 7-10 presents the configurations that occur and the number of languages that show this configuration. Although there are eight possible configurations, only three are attested. Of the four configurations that the Scope Hierarchy precludes, none is in fact attested. There are no languages in the sample in which $\pi_1$-operators are absent, but $\pi_2$-operators or $\pi_3$-operators are present, and there are no languages in which $\pi_2$-operators are absent, but $\pi_3$-operators are present. The hypothesis is evidently borne out by the distribution. More marked operators are only present if there are also less marked operators. It is remarkable that there are no languages in the sample that have no grammatical TMA expressions at all and there are only three languages, !Kung, Car and

| Table 7-10. Frequency of different synchronic configurations |
| --- | --- | --- | --- |
| Configurations | Frequency | |
| $\pi_1$ | $\pi_2$ | $\pi_3$ | $N$ | % |
| Predicted | | | | |
| - | - | - | 0 | 0 |
| + | - | - | 3 | 4% |
| + | + | - | 36 | 47% |
| + | + | + | 37 | 49% |
| Not predicted | | | | |
| - | + | + | 0 | 0 |
| - | + | - | 0 | 0 |
| - | - | + | 0 | 0 |
| + | - | + | 0 | 0 |
Tojolabal, that have only $\pi_1$-operators. This stresses the fact that TMA expressions are quite an important linguistic category.

As well as possible implicational relations between operator types, there could be implicational relations between broad semantic domains or more specific semantic functions within a specific language. No such relationships were found in my study or in the studies of Bybee (1985) and Dahl (1985) (Bybee & Dahl 1989: 95). Grams develop, at least in part, independently of other grams.

### 7.5.1.1 Creole languages

The GRAMCATS sample contains only one creole language. However, H3a is very interesting to inspect in pidgins and creoles because these languages may shed some light on the question which linguistic elements evolve first in language systems. It is assumed that creole speakers seek expressions first for notions that are communicatively most needed so that the operators in creole languages may be considered as highly communicatively motivated. In this section, I will discuss in more detail the configurations of creole languages.

A well-known viewpoint on TMA in Creole languages is the one of Bickerton (1981): he studied the TMA systems in Sranan, Guyanese, Haitian Creole French and Hawaiian Creole English and arrived at the conclusion that a creole TMA system consists of three preverbal markers: anterior (relative) tense, irrealis mood and non-punctual aspect. With respect to these three markers, Givón (1982) suggests that they are natural to be found as the core of TMA systems, both in creoles and in non-contact languages:

The special marking for *out-of-sequence* [= anterior, AB], non-punctual and irrealis merely underscores the fact that in discourse—at least in the *narrative* style—they are conceptually and in frequency the marked case. That is, we do tend to talk about events *in-sequence*, events as *actions* and events that *did happen*. (Givón 1982: 155, italics his)

According to Bickerton creole languages do not have an absolute tense-marking system. Labov (1990) notes in this respect that although a system without tense marking may seem primitive, it is not inadequate.

If time is expressed with optional adverbs, then it is only necessary to signal the time once at the beginning of a narrative or for as long as we are in the same sequence of events. But with a tense system, we have to use the tense marker over and over again. […] It is the tense system of the standard languages that is normally redundant. The pidgin system can be cumbersome, but in most discourse situations it proves to be quite efficient to designate time relations: specifically when called for, otherwise not. (1990: 18)
A point of critique on Bickerton’s prototypical system is that it is based solely on creoles with Indo-European superstrate languages⁷ (cf. Singler 1990b). In Singler (1990c) aspect and tense systems in eight Pidgin and Creole languages are discussed. The lexifying languages are English, French, Dutch, Portuguese, Spanish and the Bantu language Kikongo and geographically, the creoles are spoken in West and Central Africa, the Caribbean, and the Pacific. One of these languages is usually unmarked for tense and aspect (18th Century Nigerian Pidgin English, Fayer 1990); four languages have markers for aspect but no systematic tense marking (Papiamentu, Andersen 1990; Hawaiian Creole, Labov 1990; Berbice Dutch, Robertson 1990; Kru Pidgin English, Singler 1990a); the language Kituba has aspect and relative tense markers (Mufwene 1990); the language Haitian Creole has markers for aspect and for future tense (Spears 1990); and finally, Capeverdean Crioulo has both aspect and several tense markers (Silva 1990). There are thus no languages in the sample with tense markers but without aspect markers, whereas there are languages with aspect markers and without tense markers. From the TMA systems of these languages it seems that absolute tense is diachronically not among the first semantic notions to be systematically marked by grammatical expression, whereas aspect is. This strongly supports the idea that modification of the predicate (narrow scope) is more relevant and basic to communication than modification of the predication or the proposition.

7.5.1.2 Conclusion

There are no languages in the GRAMCATS sample in which π3-operators are present without π2- or π1-operators or in which π2-operators are present without π1-operators. However, whether the presence of wider scope operators is indeed dependent on the presence of narrow and medial scope operators could not be established, since narrow and medial scope operators are nearly always present. The configurations of TMA expressions in creoles, however, support the idea that narrow scope operators evolve earlier than medial scope operators and that wide scope operators appear last. The Scope Hierarchy is indeed reflected in synchronic configurations of TMA expressions.

7.5.2 Polysemous expressions

As explained in 5.4.3, grammaticalization paths are universal. For the domain of TMA, several paths were discussed in 7.2.3. It is assumed that the semantic changes of linguistic items in these paths reflect underlying universal semantic or conceptual space. This leads to the prediction that synchronically, a single

⁷ The major language on which the creole is based.
gram can only express different semantic notions that are related: they will be adjacent on a grammaticalization path, or, in other words, in semantic space. With respect to scope, it is to be expected that these notions are either of the same scope or of adjacent scope levels. Consider H3b:

H3b: A single expression can only cover adjacent regions in semantic space. As a result, a polysemous or portmanteau expression will only have semantic functions with similar or adjacent scopes.

This prediction is first tested for the polysemous grams (7.5.2.1) and then for the portmanteau expressions in the sample (7.5.2.2).

7.5.2.1 Polysemous grams

The 76 languages in the GRAMCATS sample contain 159 expressions that cover more than one semantic function. This includes all the modal expressions that are potentially polysemous between denoting participant-oriented modality and event-oriented modality. As discussed in 7.3.2, it is not always clear on the basis of the reference grammars whether a modal expression with narrow scope may also have scope over the predication to express event-oriented modality (π2). Counting only half of the potential event-oriented modal operators solved this problem. In contrast to the figures in the previous sections, this section includes all and not half of the expressions that potentially express event-oriented modality. In the tables, they are given in between brackets with question marks.

In this section, the combinations of semantic functions attested for single items are first discussed. According to H3b these semantic functions should be adjacent on a semantic path, and whether this is indeed the case, will be established on the basis of the semantic paths presented in 7.2.3. The focus is, however, on the question whether the Scope Hierarchy applies to polysemous grams. Is it indeed the case that single items only have semantic functions with similar or adjacent scope?

Which semantic functions are expressed by single items? Although numerous combinations of semantic functions are logically possible, the majority of combinations falls within a restricted set. There are about six large sets of combinations to which the majority of polysemous expressions belong.

A domain in which there are many polysemous grams is the domain of modality. This is partially a consequence of the analysis. Grams that express participant-oriented modality, with a source in external circumstances, such as root-possibility, permission and obligation, were encoded also as possible event-oriented modality, even when the grammars do not provide explicit information about the possibility to use the gram with non-specific participants.
In the tables, event-oriented modality is presented between brackets, to indicate that this use is not always warranted by the information in the grammars. Furthermore, grams that express epistemic modality were encoded as both event-oriented and proposition-oriented if the examples or definitions in the grammar allowed both interpretations. Grams that express both root (π₁ or π₂) and epistemic modality (π₂ or π₃) were only encoded this way on the basis of explicit description in the grammars.

The first set of semantic functions often covered by a single expression are various modal categories with a basic sense of potentiality. Table 7-11 presents the combinations that occur and the frequencies thereof. Below the dotted line are the number of expressions that combine notions of potentiality with another semantic notion; these expressions are primarily assigned to another group of polysemous markers and return in one of the other tables.

**Table 7-11. Polysemous expressions with potentiality senses (N = 49)**

<table>
<thead>
<tr>
<th>N</th>
<th>Root-possibility / ability / permission (π₁)</th>
<th>Root-possibility / permission (π₂)</th>
<th>Epistemic possibility (π₃)</th>
<th>Uncertainty (π₄)</th>
<th>+ Other senses</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>X</td>
<td>X ᵘ ?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>X ᵘ ?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>X</td>
<td>X ᵘ ?</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>X</td>
<td>X ᵘ ?</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>X ᵘ ?</td>
<td>X</td>
<td></td>
<td>+ future</td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>X ᵘ ?</td>
<td>X</td>
<td></td>
<td>+ hypothetical</td>
</tr>
</tbody>
</table>

Basque is one of the seven languages that have a single marker for all types of potentiality, the suffix –ke. Consider (52), in which –ke has a root-possibility interpretation and (53), in which it has an epistemic possibility interpretation:

**Basque** (Saltarelli et al. 1988: 235-36)

(52) *hamarr-eta.*ko  *pelikula*-ra  *jioa*-n
   ten-LOC-REL. film-s.all go-PRF
g-a-ite-¹*-ke
   1.PABS-PRS-AUX²(SUBJ)-ABS.PL-POTN
‘we can go to the ten o’clock film’
The combinations of meanings in Table 7-11 are all in accordance with the developmental paths in Figure 7-5. Furthermore, it is evident that there are only combinations of semantic functions with adjacent scope levels. For example, there are no occurrences of markers that express ability (∏1) and uncertainty (∏3), without also expressing epistemic possibility (∏2).

Just as there are polysemous items with a basic sense of potentiality, there are also polysemous items that express different semantic functions with a basic sense of (weak) necessity. Table 7-12 presents which combinations of meanings were attested and their frequency of occurrence. Note that for one language, Cantonese, there is a marker *ying goi* that expresses obligation with all types of scope (Kwok 1971: 77). Recall that probability (∏2) is the epistemic function with the basic sense of weak necessity.

Table 7-12. Polysemous expressions with (weak) necessity senses (N = 40)

<table>
<thead>
<tr>
<th>N</th>
<th>Root-nec / obl (∏1)</th>
<th>Root-nec / obl (∏2)</th>
<th>Obligation (∏3)</th>
<th>Probability (∏2)</th>
<th>Certainty (∏3)</th>
<th>+ Other senses</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>X</td>
<td>(X?)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>(X?)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>(X?)</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>(X?)</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>X</td>
<td>(X?)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>(X?)</td>
<td></td>
<td></td>
<td></td>
<td>Prospective</td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>(X?)</td>
<td></td>
<td></td>
<td></td>
<td>Future /Volition</td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>(X?)</td>
<td></td>
<td></td>
<td></td>
<td>Irrealis</td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>(X?)</td>
<td></td>
<td>X</td>
<td></td>
<td>Future</td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>Hypothetical</td>
</tr>
</tbody>
</table>

*Notes.* nec = necessity; obl = obligation.
Most combinations in Table 7-12 are in line with the prediction. They only combine functions that have similar or adjacent scope. However, there are three markers that express participant-oriented necessity (π1) and certainty (π3) (see row 5) while it is uncertain whether these markers can also express event-oriented necessity or obligation (π2). The study of Bybee et al. (1994) did not reveal a universal diachronic path for the different semantic functions with the general sense of necessity. It is unsure how meanings of certainty (π3) exactly develop. I claimed in 4.2.3.3 that objective epistemic necessity is not marked in languages, since it is the default status of events. It may be the case that the subjective meaning of certainty develops directly from root-necessity or obligation, or, another possible development, by way of negative objective epistemic necessity: not necessary. It is thus not completely clear whether all polysemous items that express necessity meanings support the hypothesis. For most of the items, however, there is enough evidence that they do.

Other combinations of modal meanings occur less often. These are presented in Table 7-13.

Table 7-13. Polysemous expressions with sense of disposition (N = 2) and polysemous expressions with general basic senses (N = 4)

<table>
<thead>
<tr>
<th>N</th>
<th>Volition/desirability (π1)</th>
<th>Desirability (π2)</th>
<th>'General' modality (π1)</th>
<th>'General' modality (π2)</th>
<th>'General' modality (π3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>X</td>
<td>(X?)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>X</td>
<td>(X?)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There are two polysemous markers that express semantic functions with the basic sense of disposition, in Pangasinan and Meithei. The Pangasinan marker makaka- expresses disposition with narrow (54) and medial scope (55), as the translations in Benton (1971) show:

**Pangasinan** (Benton 1971: 133)

(54) *makakaaral*

'likes to learn, is inclined to learn'

(55) *Makakaarin ni met*

'it’s threatening rain.'
Furthermore, there are a few modal markers that do not express one of the four basic senses, such as defined in 4.2.1 but rather cover a larger part of the modal scale. In three languages there is a more general root-modal marker: In Kanuri \( \text{g} + \text{TONE} \) expresses volition/desirability and obligation (Lukas 1937: 35, 42); in Danish \( \text{måtte} \) expresses permission and strong obligation (Koefoed 1958: 185, 92); in Baluchi, V-\( \text{q}-\text{i} \) \( \text{bu}-\text{q} \) expresses senses from potentiality to weak entailment (Barker & Mengal 1969: 239). In Nakanai the particle \( \text{ge} \) expresses epistemic modality from potentiality to weak entailment (R. L. Johnston 1980: 63-65).

All the expressions in Table 7-13 comply with H3b: they only express semantic functions that have similar or adjacent scope levels.

A fourth group of polysemous expressions combines the categories of prospective or future tense with some category of root-modality. In Slave, for example, the particle \( \text{g} \) can express necessity and future tense. See (56)-(57):

**Slave** (Rice 1989: 416)

(56) \( \text{kee räyebdi g} \)  
shoe 1SG.buy NEC

(57) \( \text{wohé} g \)  
1SG.OPT:sing FUT

There are 13 polysemous expressions in the sample that combine these categories. Table 7-14 presents which meanings co-occur in single expressions.

**Table 7-14.** Polysemous expressions that combine participant-oriented modality and prospective or future \((N = 13)\)

<table>
<thead>
<tr>
<th>( N )</th>
<th>Obligation ((\pi_1))</th>
<th>Volition ((\pi_1))</th>
<th>Ability ((\pi_1))</th>
<th>Prospective ((\pi_1))</th>
<th>Future ((\pi_2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
With respect to the expressions in Table 7-14, the semantic paths in Figure 7-6 are relevant. Expressions for obligation (π₁), desire / volition (π₁) and root possibility / ability (π₁) may develop into expressions for intention (or prospective) and for future tense. The combinations of meanings for the polysemous items in the sample support the idea that the semantic functions expressed by a single item are adjacent on a developmental path. Furthermore, these markers are all in line with the hypothesis that only semantic functions with the same or adjacent scope can be expressed by single items.

A fifth combination of meanings that often occurs in polysemous expressions are epistemic modality, future tense and irrealis. There are 13 items that express two or more of these categories. The exact combinations and the frequencies thereof are presented in Table 7-15. The function of irrealis (10th column) is a general marker of irrealis, expressing both hypothetical and counterfactual.

Table 7-15. Polysemous expressions that combine epistemic modality, future tense and irrealis (N = 13)

<table>
<thead>
<tr>
<th>N</th>
<th>Ep.psb (π²)</th>
<th>Uncert (π³)</th>
<th>Prob (π²)</th>
<th>Wk cert (π₁/2)</th>
<th>Obl (π₁)</th>
<th>Cert (π³)</th>
<th>Fut (π²)</th>
<th>Hyp (π²)</th>
<th>Irr (π²)</th>
<th>Predc (π³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes. Ep.psb = Epistemic possibility; Uncert = Uncertainty; Prob = Epistemic probability; Wk cert = Weak certainty; Obl = Obligation; Cert = Certainty; Fut = Future; Hyp = Hypothetical; Irr = Irrealis; Predc = Prediction
The diachronic paths towards irrealis meanings are not described in Bybee et al. (1994). It is also not documented whether future tense and epistemic meanings, other than prediction, can develop into one another. Whether these meanings are adjacent on diachronic paths can therefore not be tested. However, with respect to the expectation that a single expression only covers functions with adjacent scope, the data in Table 7-15 are supportive.

A sixth group of meanings often expressed by polysemous items are quantification and continuative or imperfective aspect. There are 25 items that cover one or more of these categories. They are presented in Table 7-16.

Table 7-16. Polysemous expressions for quantification and aspect (N = 25)

<table>
<thead>
<tr>
<th>N</th>
<th>Iterative (π1)</th>
<th>Continuative (π1)</th>
<th>Imperf (π1)</th>
<th>Freq (π1)</th>
<th>Habitual (π2)</th>
<th>Repetitive (π2)</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Gugu-Yalanji has one marker that covers several aspectual and quantificational senses. Imperf = Imperfective; Freq = Frequentative

Most common combinations are an imperfective and a habitual (six times, in Margi, Kanakuru, Agau, Tucano, Nakanai and Shuswap) and an iterative and a habitual or a frequentative (ten times, in Alyawarra, Yessan-Mayo, Inuit, Margi, Chacobo, Worora, Krongo, Temne, Dakota, Rukai). Consider the examples (58) and (59) from Alyawarra, in which –iyn can express iterative and frequentative meaning:

Alyawarra (Yallop 1977: 61-62)

(58)  uu1u1-iyu-ika ayinga
search-ITR-PAST   I
I went around searching; I kept searching.
A single item can also express the continuative and frequentative. In Baluchi, for instance, the periphrastic construction \textit{V-an bux} is translated ‘keep (on) –ing’. It may combine with all kinds of verbs, like ‘weave’, ‘bake bread’, ‘sing’, ‘work’, ‘bring wood’, or ‘fall’ as in ‘snow will keep falling’. This may result in either a continuative or a frequentative reading. Finally, in Gugu-Yalanji (rightmost column) verbal reduplication ‘serves to indicate an ongoing, repeated or habitual action and/or a certain intensity in action’ (Patz 2002: 39-43). This is encoded as expressing general aspect (\(\pi_1\)), predicate- and event quantification (\(\pi_1\) and \(\pi_2\)).

Many of the combinations in Table 7-16 occur as adjacent meanings in the diachronic paths presented earlier in 7.2.3 in Figure 7-4: an iterative may develop into a continuative or a frequentative. However, the continuative and frequentative are on separate paths and the iterative and habitual or repetitive are not adjacent in Figure 7-4. However, the diachronic paths in 7.2.3 are not exhaustive: it is very well possible that a continuative may become a frequentative as well (consider for example English) and that an iterative may develop immediately into a habitual (see also Bybee et al. 1994: 159). The combinations in Table 7-16 are all in accordance with the predictions that follow from the Scope Hierarchy: they only combine functions that have similar or adjacent scope.

Finally, there are a few more, seemingly incidental combinations of semantic functions. First, there are four markers that combine aspect and tense: in Tigre (Leslau 1945: 8) and Karok (Bright 1957: 124) a single marker expresses prospective aspect and future tense; in Chepang, a single marker expresses perfective aspect and past tense and another marker expresses imperfective aspect and present tense, dependent on the order of expression (Caughley 1982: 49, 104-11). Second, the sample contains three expressions that combine irreals with past/perfect/perfective: Abkhaz has a marker that can express counterfactual (\(\pi_2\)) and past tense (\(\pi_2\)) and a marker that can express perfect aspect (\(\pi_1\)) and general irreals (\(\pi_2\)) (Hewitt 1979: 173, 75, 80-81); in Mano, the function of perfective (\(\pi_1\)) is combined with the function of irrealis (\(\pi_2\)) (Becker-Donner 1965: 38ff). Third, there are seven combinations of semantic functions that only occur in one language: Inuit has a marker \textit{sima} that expresses perfect aspect (\(\pi_1\)) and evidentiality (\(\pi_3\)) (Fortescue 1984: 265-66, 72-73); Baluchi combines remote past (\(\pi_2\)) and certainty (\(\pi_3\)) in one expression (Barker & Mengal 1969: 340-41); Udmurt has a single expression for future
tense (\(\pi_2\)) and frequentative (\(\pi_2\)) (Winkler 2001: 48); Tok Pisin combines habitual and ability (Mühlhäuser 1985: 339, 81); Nung has one marker for root possibility (\(\pi_1/2\)) and perfect aspect (\(\pi_1\)) (Barnard 1934: 20-22, 27-28); and Gugu-Yalanji for obligation (\(\pi_1/2\)) (Patz 2002: 37) and irrealis (\(\pi_2\)) (Hershberger 1964: 37). Of all these combinations of meanings, only the combination of perfect aspect and evidentiality in Inuit is in conflict with the hypothesis. It functions both as a \(\pi_1\)- and as a \(\pi_3\)-operator, but not as a \(\pi_2\)-operator. It is, however, in accordance with the idea that a single item can only express senses that are adjacent on a single diachronic path since a perfect can develop directly into an evidential marker, as was shown in Figure 7-2 in 7.2.3.

From the Scope Hierarchy it is predicted that different semantic functions of a single item can only have similar or adjacent scope. The combination that is thus impossible is an item that functions as a \(\pi_1\)-operator and as a \(\pi_3\)-operator, but not as a \(\pi_2\)-operator. Table 7-17 presents an overview of all the polysemous expressions in the sample according to their scope.

If the three necessity markers in Table 7-12, for which it is uncertain whether they can express \(\pi_2\)-modality, are ignored, there remains one counterexample to the predicted configurations. In Inuit, \textit{sim}a can express perfect (\(\pi_1\)) and evidentiality (\(\pi_3\)) (Fortescue 1984: 265-66, 72-73). As was shown in Figure 7-2, there is a universal diachronic path from perfect aspect to evidentiality. The meanings of \textit{sim}a do cover adjacent semantic notions in the diachronic path, but the diachronic path skips a scope level.

In sum, the polysemous markers in the GRAMCATS sample corroborate the adjacency hypothesis. For the semantic functions of which the diachronic paths are known, in most cases single items indeed express semantic functions that

| Table 7-17. Combinations of scopes in polysemous expressions (\(N = 159\)) |
|-----------------------------------------------|-------|-------|
| Predicted                                    | Frequency |
| \(\pi_1 + \pi_1\)                           | 8     | 5     |
| \(\pi_1 + \pi_2\)                           | 98    | 62    |
| \(\pi_2 + \pi_2\)                           | 4     | 2.5   |
| \(\pi_2 + \pi_3\)                           | 35    | 22    |
| \(\pi_1 + \pi_2 + \pi_3\)                   | 13    | 8     |
| Not predicted                                | \(\pi_1 + \pi_3\) | 1     | 0.5   |
are adjacent on a path. Of all the polysemous items, only one is not in accordance with the hypothesis that a single item can express only semantic notions with adjacent scope.

7.5.2.2 Portmanteau expressions

For portmanteau expressions, grams that at the same time express two or more different meanings, from different semantic domains, a comparable prediction holds as for polysemous items. There is no restriction expected on the specific semantic functions that may be expressed by one morpheme, but it is expected that operators only merge if they have the same or adjacent scope. What are the combinations of semantic functions in the portmanteau expressions of the sample?

There are 46 portmanteau expressions in the sample. Table 7-18 shows which combinations of semantic functions occur and how often.

Table 7-18. Frequency of combinations of meanings of portmanteau expressions (N = 46)

<table>
<thead>
<tr>
<th>N</th>
<th>Aspect (π1)</th>
<th>Tense (π2)</th>
<th>Quant. (π2)</th>
<th>Irrealis (π2)</th>
<th>Modality/Evid.</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Imperfective</td>
<td>Past</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Perfective</td>
<td>Past</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Perfect</td>
<td>Past</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Perfect</td>
<td>Present</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Perfect</td>
<td>Non-future</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Imperfective</td>
<td>Present</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Imperfective</td>
<td>Non-future</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Perfective</td>
<td>Future</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Imperfective</td>
<td>Future</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Continuative</td>
<td>Present</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Continuative</td>
<td>Past</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Past</td>
<td>Repetitive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Past</td>
<td>Inference</td>
<td>(π3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Past</td>
<td>Certainty</td>
<td>(π3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Counterf.</td>
<td>Obl. (π1/π2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>Weak obl.</td>
<td>(π1/π2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes.** Quant. = Quantification ; Evid. = evidentiality ; Counterf. = Counterfactual ; obl. = obligation
Most expressions combine notions of aspect and tense, in particular imperfective, perfective or perfect aspect often combine with tense ($\pi_2$). The most common combination is past tense and imperfective aspect, for example the suffix -šiŋ in Agau (Hetzron 1969: 28ff) or -iŋa in Alyawarra (Yallop 1977: 50-53). The complex portmanteau tense-aspect system in Guaymi (Kopesec 1975: 22) was already presented in Table 7-4. A second combination that occurs quite often is event quantification ($\pi_2$) with past tense. Kanakuru, Agau, Chacobo, Kui, Tem and Temne all have a habitual past marker; Temne also has a repetitive past (see Appendix G). Moreover, Slave has two expressions that combine modality and irrealis: there is an expression for counterfactual obligation and for counterfactual weak obligation (Rice 1989: 418-19). Finally, Uigur combines past tense with subjective epistemic modality: it makes an opposition between certain past and inferential past (Nadzhip 1971: 117-18).

As well as for polysemous grams, H3b predicts for portmanteau expressions that the fusion of two or more semantic functions into one single expression is restricted to operators that have equal or adjacent scope. In other words, the only combination that is not expected to occur in a portmanteau expression is the combination of a $\pi_1$- and a $\pi_3$-operator, since these scopes are not adjacent. The combinations of portmanteau expressions in Table 7-18 are presented again in Table 7-19, but now with respect to the scopes they combine. The table shows that most portmanteau expressions combine a $\pi_1$- and a $\pi_2$-operator. The only combination that is not predicted, $\pi_1 + \pi_3$, is in fact not attested. Two other combinations that were expected to be possible do not occur either, $\pi_1 + \pi_1$ and $\pi_1 + \pi_2 + \pi_3$. This makes sense if one realizes that portmanteau expressions can only evolve if they combine meanings that

<table>
<thead>
<tr>
<th>Configurations</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1 + \pi_1$</td>
<td>0</td>
</tr>
<tr>
<td>$\pi_1 + \pi_2$</td>
<td>37</td>
</tr>
<tr>
<td>$\pi_2 + \pi_2$</td>
<td>7</td>
</tr>
<tr>
<td>$\pi_2 + \pi_3$</td>
<td>2</td>
</tr>
<tr>
<td>$\pi_1 + \pi_2 + \pi_3$</td>
<td>0</td>
</tr>
<tr>
<td>$\pi_1 + \pi_3$</td>
<td>0</td>
</tr>
</tbody>
</table>
frequently co-occur. The fact that the combination of $\pi_1$, $\pi_2$ and $\pi_3$ does not occur may be because the frequency of a specific combination of three TMA expressions in one utterance is so low that it would never grammaticalize. The fact that the combination of $\pi_1 + \pi_1$ does not occur may lie in the function of $\pi_1$-operators. They specify the property or relation designated by the predicate, by aspect, modality or quantification. The absence of this combination suggests that the specific combinations do not occur often enough to grammaticalize: apparently specific modal markers do not combine frequently with specific aspect operators.

Table 7-19 clearly shows that all the portmanteau expressions in the sample adhere to the predicted combination of scopes: a single expression may only denote meanings with adjacent or similar scope. There are no counterexamples to the predicted co-occurrences.

**7.5.2.3 Conclusion**

The Scope Hierarchy correctly predicts the possible combinations of semantic functions within a single expression (H3b). Both polysemous and portmanteau expressions combine operators with similar or adjacent scope, with the exception of the polysemous marker in Inuit, that combines perfect aspect and evidentiality. These meanings are, however, also related diachronically (7.2.4). The data clearly show that there is a very strong tendency that the synchronic use of markers reflects the semantic relatedness of different meanings and the scope relations between these meanings. Scope plays an important role in the limits on variation.

**7.6 Expression Form**

This section will explore whether there is a correlation between scope and expression form of operators. Older grams are formally more grammaticalized, i.e. they are more fused with the main verb, more dependent on surrounding material and shorter in length. Although scope on the one hand and age of grams on the other hand are not identical, they are related and the hypothesis has been formulated as if they are alike:

**Q4:** What is the expression form of TMA expressions?

**H4:** Operators with wider scope show a higher or equal degree of formal grammaticalization than operators with narrower scope.

What is the degree of formal grammaticalization of specific expression forms? The degree of formal grammaticalization is higher if elements are phonologically more reduced, have fewer properties of their original lexical
category and have a fixed syntactic position or are even fused with other items. Abstracting away from language type, inflection shows the highest degree of formal grammaticalization, particles and auxiliaries show a lower degree of formal grammaticalization and periphrases are the least grammaticalized. Inflection is most reduced, has the most fixed position (is even fused), and shows the fewest lexical properties (see 5.4.4). However, the general language type, isolating, analytic, fusional or agglutinative, influences the highest possible degree of formal grammaticalization. Within isolating or analytic languages, most words are mono-morphemic and the highest degree of formal grammaticalization seems to be a particle, whereas in fusional or agglutinative languages, inflection is the most grammaticalized form.

### 7.6.1 Method

The TMA operators in the GRAMCATS sample are all coded very specifically for their expressions form. The different expression types are distributed as presented in Table 7-20. Almost half (49%) of the TMA expressions are inflectional (prefix, infix, suffix), whereas the other half consists for a substantial part of auxiliaries (22%) and particles (12%).

<table>
<thead>
<tr>
<th>Expression Form</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem change</td>
<td>0.8</td>
</tr>
<tr>
<td>Tone change</td>
<td>0.9</td>
</tr>
<tr>
<td>Reduplication</td>
<td>3.6</td>
</tr>
<tr>
<td>Zero</td>
<td>2.8</td>
</tr>
<tr>
<td>Suffix</td>
<td>42.2</td>
</tr>
<tr>
<td>Prefix</td>
<td>7.1</td>
</tr>
<tr>
<td>Infix</td>
<td>0.5</td>
</tr>
<tr>
<td>Periphrastic; bound + bound</td>
<td>1.8</td>
</tr>
<tr>
<td>Periphrastic; bound + unbound</td>
<td>5.4</td>
</tr>
<tr>
<td>Periphrastic; unbound + unbound</td>
<td>0.9</td>
</tr>
<tr>
<td>Auxiliary</td>
<td>21.6</td>
</tr>
<tr>
<td>Particle</td>
<td>11.8</td>
</tr>
<tr>
<td>Unclassified</td>
<td>0.4</td>
</tr>
</tbody>
</table>
In order to test H4, a general distinction is made between bound, non-bound and combined expression forms (periphrases). Bound expression forms include reduplication, stem- and tone change, zero expression, inflection (prefix, suffix, infix), or combinations thereof, the periphrastic constructions with two bound forms. They are considered to have the highest degree of formal grammaticalization. Non-bound expression forms are auxiliaries, and particles or combinations thereof. Finally, periphrastic constructions consisting of a bound and a non-bound element make up the third category. The few expression forms that are unclassified are excluded from further analysis.

In the next sections the relation between scope and expression form is examined. A complicating factor is that the potential degree of grammaticalization of expressions interferes with the morphological type of a language. Therefore, the relationship between morphological language type and the distribution of operators will be explored first.

### 7.6.2 Scope in different language types

If the morphological type of a language appears to be related to different distributions of operators, H4 cannot be tested in the complete sample; the proportion of certain language types might obscure the relation between scope and expression form. For this test, the languages are divided into three groups: (i) languages with only non-bound expression forms for operators, (ii) languages with only bound expression forms for operators, and (iii) languages with bound and non-bound expression forms for operators. In the sample, 23 languages have only inflectional or other bound expression forms for the TMA expressions, 7 languages have only non-bound expression forms and the other 46 languages have both bound and non-bound expression forms for TMA operators. For each language type the proportion of operators is presented in Table 7-21.

<table>
<thead>
<tr>
<th>Languages with expressions</th>
<th>No. of expressions</th>
<th>( \pi_1 )</th>
<th>( \pi_2 )</th>
<th>( \pi_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only non-bound forms ((N = 7))</td>
<td>70</td>
<td>59 %</td>
<td>35 %</td>
<td>6 %</td>
</tr>
<tr>
<td>Both forms ((N = 46))</td>
<td>630</td>
<td>49 %</td>
<td>42 %</td>
<td>9 %</td>
</tr>
<tr>
<td>Only bound forms ((N = 23))</td>
<td>256</td>
<td>39 %</td>
<td>49 %</td>
<td>12 %</td>
</tr>
</tbody>
</table>
The $\chi^2$-test is applied on the raw data for testing dependence of scope and language type ($\alpha = 0.5$). It appears that the distribution is significantly different from chance ($\chi^2 = 12.8; p = 0.012$). Morphological language type does correlate with the distribution of scope. The proportion of $\pi_1$-operators in languages with only non-bound forms is larger than the proportion of $\pi_1$-operators in languages with only bound forms, whereas the proportion of $\pi_2$- and $\pi_3$-operators is small in languages with only non-bound forms and large in languages with only bound forms. There is thus a clear relation between languages with non-bound expression forms and narrow scope operators and between languages with bound expression forms and medial and wide scope operators. To conclude, H4 can be tested only in languages that have different expression forms. The proportion of languages with only bound expression forms is much larger than that of languages with only non-bound expression forms, the former group would influence the results too strongly.

7.6.3 Scope in languages with different expression forms

In order to control for the influence of morphological language type on the results, H4 is tested in the restricted group of languages that allow bound, non-bound and periphrastic expression forms. For these 46 languages wider scope operators will be studied to see if they are more or less bound than narrow scope operators. Non-bound forms are considered formally less grammaticalized than bound forms and periphrastic constructions are considered the least grammaticalized. Table 7-22 presents an overview of the distribution of operator types over expression form.

In order to find out whether scope and expression form are independent variables, the $\chi^2$-test is applied. When all categories are compared, there is no reason to reject the independence hypothesis ($\chi^2 = 9.5; p = 0.05$). There does not seem to be a relation between scope and expression form. When, however, the distribution of $\pi_3$-operators over expression form is compared to the

<table>
<thead>
<tr>
<th>Operators</th>
<th>$N$</th>
<th>Bound</th>
<th>Periphrastic</th>
<th>Non-bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1$</td>
<td>312</td>
<td>51%</td>
<td>10%</td>
<td>39%</td>
</tr>
<tr>
<td>$\pi_2$</td>
<td>262</td>
<td>54%</td>
<td>7%</td>
<td>39%</td>
</tr>
<tr>
<td>$\pi_3$</td>
<td>56</td>
<td>38%</td>
<td>5%</td>
<td>57%</td>
</tr>
</tbody>
</table>
distribution of π₁- or π₂-operators, there appear to be significant differences. Operators with wide scope have a smaller proportion of bound expressions and a larger proportion of non-bound expressions compared to π₁-operators ($\chi^2 = 8.1; \ p = 0.018$) and compared to π₂-operators ($\chi^2 = 8.0; \ p = 0.018$). These differences are significant. The findings go against hypothesis 4. Wide scope seems to correlate with lower degrees of formal grammaticalization.

In sum, within languages that have different expression forms, **wide scope** correlates with non-bound expressions. However, when different language types are compared, **narrow scope** correlates with non-bound expressions and medial and wide scope with bound expressions.

### 7.6.4 Complicating factors

Since the results are very inconclusive, alternative factors will be explored. It is possible that semantic functions are blurring the picture, as was the case in 7.4. Certain semantic functions are more often expressed by bound or by non-bound forms than others and if they form a substantial part of a specific layer, they have a relative large influence on the figures. Therefore, a more detailed analysis is carried out, in which the relation between specific TMA domains and expression form is examined. Firstly, the distribution of TMA domains and expression form is examined within the group of languages with both bound and non-bound expression forms. See Table 7-23.

**Table 7-23.** Frequencies of expression forms for each semantic domain, in languages with both bound and non-bound expression forms ($N = 46$)

<table>
<thead>
<tr>
<th>Semantic domain</th>
<th>Expression forms in %</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bound</td>
<td>Periphrastic</td>
</tr>
<tr>
<td>Participant-or. modality (π₁)</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>Event-or. modality (π₂)</td>
<td>22</td>
<td>10</td>
</tr>
<tr>
<td>Proposition-or. modality (π₃)</td>
<td>37</td>
<td>7</td>
</tr>
<tr>
<td>Evidentiality (π₃)</td>
<td>38</td>
<td>0</td>
</tr>
<tr>
<td>Irrealis (π₂)</td>
<td>43</td>
<td>9</td>
</tr>
<tr>
<td>Aspect (π₁)</td>
<td>56</td>
<td>11</td>
</tr>
<tr>
<td>Tense (π₁)</td>
<td>65</td>
<td>5</td>
</tr>
<tr>
<td>Event quantification (π₂)</td>
<td>61.5</td>
<td>11.5</td>
</tr>
<tr>
<td>Property quantification (π₁)</td>
<td>96</td>
<td>4</td>
</tr>
</tbody>
</table>

*Note.* or. = oriented
Whether semantic domains and expression form are independent factors is again calculated on the raw numbers with the $\chi^2$-test. It appears that semantic categories are not equally distributed over expression form ($\chi^2 = 98.7$, $p < 0.001$). Domains that are relatively frequently expressed by non-bound forms are participant-, event-, and proposition-oriented modality. Domains that are relatively frequently expressed by bound forms are tense, aspect and quantification. The domains of evidentiality and irrealis do not correlate significantly with a specific expression form. As the domains of aspect and tense are in general much more frequently expressed than other domains (compare Table 7-8) the bound expression forms are proportionally overrepresented in the group of $\pi_1$- and $\pi_2$-operators. It might again be more insightful to compare only semantically related domains.

When the semantically related domains of predicate- and event quantification are compared, there is a significant difference in boundness between predicate and event quantification ($\chi^2 = 9.7$, $p < 0.01$) in that property quantification is more bound than event quantification. This goes against the hypothesis. However, when tense and aspect are compared, they show an equal level of boundness ($\chi^2 = 4.9$, $p = 0.08$), just like the different types of modality ($\chi^2 = 6.1$, $p = 0.19$). These results do not support the idea that scope and expression form are related. In fact, there seems to be a relation between semantic domain and expression form. Tense, aspect and property quantification are more often expressed by bound forms, whereas modality is more often expressed by non-bound forms.

This idea is further investigated in different language types. Do certain language types have a preference for certain semantic domains? Languages are again divided in three groups: languages (i) with only bound expression for TMA, (ii) with only non-bound expression and (iii) with both types of expression. Table 7-24 presents the proportion of different TMA domains for each language type. Whether language type and the presence of semantic domains are independent variables is calculated with the $\chi^2$-test. It appears that these variables are not independent ($\chi^2 = 37.6$, $p = 0.0017$). The proportion of expressions of participant-oriented modality is relatively large in languages with only non-bound expression forms and relatively small in languages with only inflection. The proportion of different aspect expressions is relatively small in languages with only inflection and large in languages with both expression forms. The proportion of different tense expressions is relatively large in languages with only bound expression and small in languages with only non-bound expression. Finally the proportion of evidential markers is relatively large in languages with only bound expression, compared to the other two language types.
Table 7-24. Proportions of semantic domains within different morphological types of languages (N = 76)

<table>
<thead>
<tr>
<th>TMA domain (in %)</th>
<th>Only bound forms N = 255</th>
<th>Both forms N = 630</th>
<th>Only non-bound N = 70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant-or. modality (π1)</td>
<td>9</td>
<td>11</td>
<td>27</td>
</tr>
<tr>
<td>Event-or. modality (π2)</td>
<td>9</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>Proposition-or. modality (π3)</td>
<td>7</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Evidentiality (π3)</td>
<td>5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Irrealis (π2)</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Aspect (π1)</td>
<td>26</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>Tense (π2)</td>
<td>29</td>
<td>23</td>
<td>16</td>
</tr>
<tr>
<td>Event quantification (π2)</td>
<td>8</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Property quantification (π1)</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

In sum, neither the distribution of semantic domains over language types nor the expression form for semantic domains is randomly distributed. As a result, languages with only bound forms have proportionally more operators with medial or wide scope, since the proportion of tense and evidential operators is relatively large and the proportion of aspect operators relatively small. Languages with only non-bound expression forms have proportionally more operators with narrow scope, because the proportion of participant-oriented modality markers is relatively large and the proportion of tense operators relatively small. Languages with different expression forms have a large proportion of aspect operators that are relatively often expressed by bound forms.

Notice that there is no one-to-one relation between the scope of the semantic categories and the expression forms. Aspect (π1) is often expressed by bound forms, whereas participant-oriented modality (π1) is often expressed by unbound forms.

The relation between semantic domain and expression form is probably even more complicated than the discussion shows, because the generalization over specific semantic functions may not be valid. Bybee & Dahl (1989) show that within broad semantic domains the specific semantic functions correlate with specific expression types. See Table 7-25. Within the sample of Dahl (1985),
progressive and perfect aspect are often periphrastically expressed, whereas perfective and imperfective are often expressed by bound forms. Past tense is most often expressed by bound forms, whereas future is in about half of the cases expressed periphrastically and in half of the cases by bound forms.

Table 7-25. Expression forms of major gram-types in sample of Dahl (1985, adapted from Bybee & Dahl 1989: 56)

<table>
<thead>
<tr>
<th>Periphrastic</th>
<th>Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>perfect (16/18)</td>
<td>past (33/45)</td>
</tr>
<tr>
<td>perfective (17/20)</td>
<td>perfective (17/20)</td>
</tr>
<tr>
<td>progressive (18/19)</td>
<td>imperfective (7/7)</td>
</tr>
<tr>
<td>future (27/50)</td>
<td>future (23/50)</td>
</tr>
</tbody>
</table>

7.6.5 Discussion and conclusion

In 6.6 it was shown that in English, expression form correlates with age of a gram, and only indirectly with scope. This section has shown that expression form correlates with semantic function, rather than with scope, and that semantic function correlates with morphological language type. Aspect, property quantification and tense are often expressed by bound expression in languages that have different expression forms. Since these categories are very frequent they constitute a large proportion of $\pi_1$- and $\pi_2$-operators; the high percentage of bound $\pi_1$- and $\pi_2$-operators in Table 7-22 may be solely the result of the semantics interfering. Language type also correlates with the presence of semantic functions. Participant-oriented modality markers are strongly represented in languages with only non-bound forms, whereas tense and evidential markers form relatively large groups within languages with only bound forms. The proportion of aspect operators is relatively large within languages with different expression forms.

An explanation for the correlations may lie in the grammaticalization paths. As modal expressions often have their sources in single lexical verbs, they may more easily grammaticalize in languages with only non-bound expression forms. As aspectual expressions often have their sources in periphrastic constructions, they may more easily grammaticalize in languages that allow bound expression forms. It would be a very interesting question for further research to find out what the influence is of the morphological type of languages on the ease with which specific constructions may grammaticalize.

Concerning the present research, it is evident that H4 should be rejected. Scope hardly seems to play a role in expression type, at most in an indirect way.
The Scope Hierarchy does not correctly describe the variation in expression forms.

### 7.7 ORDER OF EXPRESSION

The final area in which the Scope Hierarchy was expected to be reflected is the syntactic order of different TMA expressions. It is predicted that the order of TMA expressions reflects the Scope Hierarchy, because of the principle of iconicity.

**Q5:** What is the expression order of TMA expressions?

**H5:** The expression order of TMA expressions iconically reflects the scope relations as follows: \( \pi_3 \pi_2 \pi_1 \) Predicate \( \pi_1 \pi_2 \pi_3 \)

In order to find out whether the syntactic order of morphemes indeed reflects their semantic scope, the possible orders of different TMA expressions were documented. Unfortunately, for many languages it was impossible to find reliable information on the order of expression, since most grammars lack an explicit discussion of order phenomena. In several grammars there are incidental examples that contain different TMA expressions, however, if the order of different expressions is not explicitly discussed in the grammar, it is not clear whether the order found in specific examples is the only order allowed in the language. The clearest evidence was found in examples that present an order of TMA expressions that is different from the predicted one, the counterexamples to the hypothesis. Examples that show the predicted order of morphemes should, however, be treated with more reservation, since different orders of expression might also be possible. For some languages, a template is presented that indicates the positions of different markers with respect to the verb stem. Then, the order of expression is clear, although a problem remains in that the co-occurrence of morphemes is often not explicitly discussed.

In spite of the difficulties, as much information as possible was collected on the expression order of TMA operators in different languages. For establishing the order of expression the criteria of Bybee (1985: 34) were followed:

1. Portmanteau expressions have no relative order of two morphemes, because they are fused.
2. When two TMA markers are expressed on different sides of the predicate, their distance to the verb depends on the possibility to place another morpheme in between the marker in question and the verb. If this is
possible for one marker, but not for the other one, the first marker is considered farther away from the predicate.\textsuperscript{8}

3. When morphemes are mutually exclusive, there is no order between them.
4. Operators expressed by an affix are considered farther away from the predicate than operators expressed by reduplication or stem change.

First an overview is presented of the expression orders that are in accordance with the hypothesis, in Table 7-26. In this table, the examples are not repeated as in the original texts in the reference grammar: only the glosses are presented, thereby using my own analysis of the TMA expressions in question, which may be different from the analysis by the author of the grammar. A dash indicates that the gram is bound (inflection) and a space that the gram is unbound. A comma between two semantic functions indicates that these functions can be marked mutually exclusive by different expressions and a slash between two functions indicates that there are two possible interpretations of a single polysemous gram. In the rightmost column, the analysis with respect to scope is presented.

As an illustration, one of the examples will be discussed in detail. Consider (60), an example from Rukai in which different TMA expressions occur:

\begin{verbatim}
Rukai (P. J.-K. Li 1973: 160)
(60) wa-uda-udal-ŋa
PAST-IPFV-rain-COMPL
‘it has been raining.’
\end{verbatim}

In row 8, column 3 of Table 7-26, this example is represented using a general gloss only, according to my analysis: past–imperfectiveV–perfect. Furthermore it is indicated that this construction consists of a prefix for tense, a reduplicated verb stem for indicating imperfective and a suffix for perfect aspect. Column 4 presents the gloss in terms of scope: $\pi_2$–$\pi_1$Pred–$\pi_1$. It is considered an example that supports the hypothesis, since reduplication (imperfective) is considered closer to the stem than inflection (tense). The order of expression thus reflects the scope relation between past tense and imperfective aspect.

\textsuperscript{8} A slightly different, though comparable situation holds for Worora (Love 2000), in which tense is expressed on an auxiliary that comes in between the predicate and the tense inflection. In contrast, an iterative, frequentative and habitual are directly expressed on the verb without an intervening auxiliary. They are therefore considered closer to the predicate than tense.
Table 7-26. Examples in the sample in which expression order reflects scope relations

<table>
<thead>
<tr>
<th>Language</th>
<th>Reference</th>
<th>Gloss [according to analysis of AB]</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cheyenne</td>
<td>Leman (1980)</td>
<td>Uncertainty–tense–V</td>
<td>π3-π2-Pred</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V–progressive–epistemic.possibility/uncertainty</td>
<td>Pred-π1-π2/π3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V–iterative/frequentative/habitual</td>
<td>Pred-π1/π2/π2</td>
</tr>
<tr>
<td>Kui</td>
<td>Winfield (1928: 85-86)</td>
<td>V– (im)perfective AUX–tense</td>
<td>Pred-π1 AUX–π2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>progressive–V–future</td>
<td>π1-Pred-π2</td>
</tr>
<tr>
<td>Ono</td>
<td>Wacke (1930: 164-69)</td>
<td>V–progressive–tense</td>
<td>Pred-π1-π2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V–frequentative–tense</td>
<td>Pred-π2-π2</td>
</tr>
</tbody>
</table>
Table 7-26 continued. Examples in the sample in which expression order reflects scope relations

<table>
<thead>
<tr>
<th>Language</th>
<th>Reference</th>
<th>Gloss [according to analysis of AB]</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bongu</td>
<td>Hanke (1909: 53)</td>
<td>V AUX (perfect) – future</td>
<td>Pred π1-π2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“aspect” = completive, negative potentiality, continuable, iterative ; “tense” = present, near past, remote past, future</td>
<td></td>
</tr>
<tr>
<td>Temne</td>
<td>W.A.A. Wilson (1961: 26)</td>
<td>habitual ability V</td>
<td>π2 π1 Pred</td>
</tr>
<tr>
<td>Bari</td>
<td>Spagnolo (1933: 186)</td>
<td>future ability V</td>
<td>π2 π1 Pred</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V–progressive–past/present</td>
<td>Pred-π1-π2</td>
</tr>
<tr>
<td>Lao</td>
<td>Yates &amp; Sayasithsena (1970: 70)</td>
<td>non.past root.possibility V</td>
<td>π2 π2/π1 Pred</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V–perfective–non.past/irrealis/uncertainty</td>
<td>Pred-π1-π2/π2/π3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V–completive–irrealis/uncertainty</td>
<td>pred-π1-π2/π3</td>
</tr>
<tr>
<td>Cantonese</td>
<td>Kwok (1971: 77-78)</td>
<td>uncertainty/epistemic.possibility future/ability V</td>
<td>π3/π2 π2/π1 Pred</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V–past–probability</td>
<td>Pred-π2-π3</td>
</tr>
</tbody>
</table>
In a similar way, Table 7-27 presents the examples that were found in the corpus that do not support the hypothesis. One example from Chepang will be discussed as an illustration. Notice that Chepang also occurs in the first table, with examples that support the hypothesis. However, the example in (61) is a counterexample to the hypothesis.

**Chepang** (Caughley 1982: 106)

(61)  
\[ja? yom-\text{-te} i? yay\text{-ca?-na}\?\]

‘Tigers or bears may get you.’

This example is represented in row 8, column 3 of Table 7-27 by its gloss: V-irrealis/uncertainty-imperfective. In (61), -\text{ca}? is a marker of uncertainty (\(\pi_3\)), but it can also express irrealis (\(\pi_2\)). Column 4 presents the gloss in terms of its scope: Pred-\(\pi_2/\pi_3\)-\(\pi_1\). It shows an order that conflicts with the hypothesis, since operators with medial or wide scope are expressed closer to the predicate than an operator with narrow scope.

Palantla seems to be an exceptional case. The general template for the order of morphemes in this language is negation-aspect-tense-derivational/directive-V (Merrifield 1968: 21). The aspectual notions in this template are the imperfective and perfective, tense notions are past hodiernal and hesternal or remote, the derivational notions are not included in this research and directive morphemes include an habitual, prospective and completive marker. Although the template suggests Palantla to be a serious counterexample, as perfective and imperfective occur outside the scope of tense markers, the only co-occurrence of operators that in fact forms a counterexample is perfective–past–V. Other possible combinations in Palantla are all in line with the predicted order.
Table 7-27. Examples in the sample in which expression order does NOT reflect scope relations

<table>
<thead>
<tr>
<th>Language</th>
<th>Reference</th>
<th>Gloss [by AB]</th>
<th>Order</th>
</tr>
</thead>
</table>
| Chacobo  | Prost (1962: 116-18) | V-remote.past-perfective  
V-evidential-imperfective  
V-habitual.past-remote.past-perfective | Pred-π2-π1  
Pred-π3-π1  
Pred-π2-π2-π1 |
| Motu     | Lister-Turner & Clark  
(n.d.: 14) | perfect/progressive non.future V-continuative.present/continuative.past | π1/π1  
π2-π1 + π2/π1 + π2 |
“aspect” = habitual ; “tense” = perfect | Pred-π2-π1 |
| Kanuri   | Lukas (1937: 25) | perfective-past-V | π1-π2-Pred |
| Palantla | Merrifield (1968: 21) | perfective-past-V | π1-π2-Pred |
| Maidu    | Shipley (1964: 40) | [STEM-completive/ability/ingressive/continuative-evidential]-tense-(im)perfective/habitual  
[…] = Verb theme.  
NB: unclear which combinations are allowed | [Pred-π1/π1/π1-π3] – π2-π1/π2 |
| Chepang  | Caughley (1982: 106) | V-irrealis/uncertainty-imperfective | Pred-π2-π3-π1 |
| Labu     | Matisoff (1973: 215, 36) | obligation iterative/repetitive ingressive continuative V  
potentiality obligation continuative perfect resultative | π1/π2  
π1/2  
π1  
π1  
π1/Pred π1/π2  
π1/π2  
π1  
π1  
π1 |
For 22 languages of the 76 in the sample, examples or templates were found that support the hypothesis (Table 7-26), but it is not certain whether alternative orders are allowed in these languages. Although there is considerable support for the hypothesis, there is also a number of counterexamples, sometimes within the same languages (Tucano, Yessan-Mayo and Chepang). In nine languages, examples or templates were found that go against the hypothesis (Table 7-27).

It is clear that the hypothesis is formulated too absolutely. There seems to be a tendency for an iconic reflection of semantic scope in syntactic order, but it is a tendency, not an absolute principle. Apparently, even though the functional relations between operators are to a certain extent mirrored in the expression order, there are other factors that may distort this ‘ideal’ iconic picture.

Some additional evidence for H5 can be found in the study of Bybee (1985) that is based on a different sample of 50 languages. Bybee (1985: 34-35) examined the order of inflectional TMA expressions in her sample. In 18 languages both aspect and tense is expressed. For 10 of these languages, the ordering is not relevant, because the expressions are either mutually exclusive, expressed in a portmanteau expression or with equal distance to the verb. In the other 8 languages, aspect markers are expressed closer to the stem than tense, whereas the opposite order does not occur. In 23 languages, both aspect and epistemic modality are expressed. In 10 of these languages aspect is expressed closer to the stem than modality, whereas the opposite order does not occur. Finally, in 20 languages both tense and epistemic modality are expressed. In 8 of these languages, tense is expressed closer to the stem and in the other languages the order is not relevant or unknown. In one language, Ojibwa, epistemic modality (dubitative) is expressed closer to the stem than tense. In terms of scope, the results in Bybee (1985) show that $\pi_1$-operators of aspect are expressed closer to the predicate than $\pi_2$-operators of tense and than $\pi_3$-, or $\pi_3$-operators of epistemic modality in all relevant cases. Markers of tense ($\pi_2$) are expressed closer to the predicate than $\pi_2$- or $\pi_3$-operators of epistemic modality or mood in seven out of eight languages. These data clearly support the claim that there is indeed a strong tendency for scope to be iconically reflected in syntax.

To conclude, there is a rather strong tendency for iconic reflection of scope in syntactic order: examples in the GRAMCATS sample that support the hypothesis by far outnumber the counterexamples. Further support for H5 is found in the study of Bybee (1985). It would be interesting to find out why the scope relations are not always reflected. An explanation might be found in the continuous processes of grammaticalization. Old grams with medial or wide scope may be fused to the verb, whereas new grams with narrower scope are expressed by auxiliaries or periphrases. When these new grams also become
fused, they might merge with older grams. It is very improbable that they will be inserted in between the verb and the older gram, as this string is a stable construction in the language. Detailed diachronic surveys are needed to reveal how syntactic order and possible changes therein exactly arise.

7.8 DISCUSSION AND CONCLUSION

There is much evidence that the Scope Hierarchy is reflected in different linguistic features. However, the correlations between scope and certain linguistic domains are not always as clear as expected. With regard to the diachrony of operators, the layered structure makes the correct predictions (H1). Diachronic paths develop from narrow to wide scope and not the other way around. The only path that is not completely in line with the expectation is the development from a perfect aspect marker to an evidential expression, since this involves a large extension of scope. It also appears that frequency and variation generally decrease as scope increases (H2), although this decrease is less marked than expected: operators with narrow and medial scope are about equally represented, whereas operators with wide scope are clearly less frequent, both crosslinguistically and within languages. This suggests that narrow and medial scope operators do not differ fundamentally in markedness, whereas wide scope operators are undoubtedly more marked, that is, they are probably less relevant to communication.

The Scope Hierarchy appeared to be a very strong predictor of possible synchronic configurations of TMA expressions (H3a): there are (a few) languages that have only $\pi_1$-operators, languages that have $\pi_1$- and $\pi_2$-operators and languages that have $\pi_1$, $\pi_2$- and $\pi_3$-operators. Other configurations are not attested. It is however impossible to determine a dependency relation between different operator types. The difference in distribution can best be understood as a markedness cline, rather than a real implicational hierarchy. The Scope Hierarchy also makes correct predictions about single items that express multiple functions. Single items only express functions that have similar or adjacent scope, with the exception of a marker that expresses perfect aspect and evidentiality.

An area for which the Scope Hierarchy seems to be irrelevant is the expression form of operators. A correlation between scope and expression form (H4) could not be established. There were no indications of dependency between these variables. Expression form rather seems to correlate with semantic function. Finally, with respect to the expression order (H5), the predictions are too absolute, although there is a clear tendency for syntactic order to iconically reflect semantic scope. This is however not an absolute
principle. There are factors that compete with the Principle of Iconic Ordering (5.4.5).

All in all, the scope relations between operators are an important factor of influence on the possible variation in TMA systems. The results in this chapter that are based on crosslinguistic data, strongly support the findings in Chapter 6, that were based on only one language. The Scope Hierarchy cannot by itself explain all the facts about TMA systems. In particular, alternative explanations need to be found for the expression forms of TMA expressions. Furthermore, there are competing factors that influence the frequency and expression order of TMA operators. In spite of the shortcomings of the Scope Hierarchy, it does account for the limits on variation in adult languages to a large extent. Whether the Scope Hierarchy can also account for the limits on variation in stages of language acquisition will be explored in Part III.